EDITOR-IN-CHIEF

LEONARD R. JOHNSON
University of Tennessee College of Medicine, Memphis, Tennessee

ASSOCIATE EDITORS

DAVID H. ALPERS
Washington University School of Medicine, St. Louis, Missouri

KIM E. BARRETT
University of California, San Diego School of Medicine, San Diego, California

JOHN M. CARETHERS
University of California, San Diego School of Medicine and Veterans Affairs San Diego Healthcare System, San Diego, California

MARK FELDMAN
University of Texas Southwestern Medical School at Dallas and Presbyterian Hospital of Dallas, Dallas, Texas

GREGORY J. GORES
Mayo Clinic, Rochester, Minnesota

RICHARD J. GRAND
Harvard Medical School and Children's Hospital Boston, Boston, Massachusetts

MARTIN F. KAGNOFF
University of California, San Diego School of Medicine, San Diego, California

RODGER A. LIDDLE
Duke University Medical Center, Durham, North Carolina

SHELLY LU
Keck School of Medicine, University of Southern California, Los Angeles, California

JAMES LEE MADARA
University of Chicago School of Medicine, Chicago, Illinois

CHARLES M. MANSBACH II
University of Tennessee Health Science Center and Veterans Affairs Medical Center, Memphis, Tennessee

ALEXANDER R. MARGULIS
Weill Medical College, Cornell University, New York, New York

JOHN A. WILLIAMS
University of Michigan Medical School, Ann Arbor, Michigan

DOUG W. WILMORE
Harvard Medical School and Brigham and Women's Hospital, Boston, Massachusetts

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health, Columbus, Ohio

FOREWORD

People of my age often have mixed emotions about encyclopedias, probably because they conjure up memories of writing high-school term papers and reading from 26-volume encyclopedias in the library. Those individuals who came of age in the computer era may not fully understand to what I am referring. As I have grown older, I have rediscovered the usefulness of encyclopedias, especially those comprising only 1, 2, or 3 volumes. My favorite for general knowledge is the *Columbia Encyclopedia*, which was first published in 1935 and is now in its fifth edition. I have used this encyclopedia on many occasions. For example, while developing a Grand Rounds lecture on celiac disease, I discovered that the disease was described exceedingly well in the second century A.D. by Aretaeus the Cappadocian. Where is Cappadocia? Is it an ancient city of Greece? Knowing that someone would ask me these questions, I consulted the *Columbia Encyclopedia* and learned that Cappadocia was the ancient Hittite state located in what is now central Turkey. Recently, I was asked to be on the thesis committee of a young doctoral candidate in our Institute for Medical Humanities. His dissertation proposal referred to philosophies (e.g., hermeneutics and phenomenology) that I had never heard of. These topics were described in the *Columbia Encyclopedia* in a brief capsule format that did not overwhelm my untrained and uneducated mind. Therefore, I have become a fan of small-volume encyclopedias. They are quite useful.

Why, then, has a publisher not produced such encyclopedias for the fields of medicine and the biological sciences? In fact, indeed one has. Elsevier has published several such encyclopedias, including *The Encyclopedia of Cancer* (now in its second edition), *The Encyclopedia of Hormones*, and *The Encyclopedia of Toxicology*. Now, under the leadership of Editor-in-Chief Leonard R. Johnson, Ph.D., a respected gastrointestinal scientist, educator, editor, and author, Elsevier has published *The Encyclopedia of Gastroenterology*. Dr. Johnson has assembled a cadre of 15 Associate Editors, each a highly respected expert in one or more of the topics featured in the book. They have brought together over 700 authors to write 477 separate articles, divided into three volumes.

These articles go a long way toward covering the entire gamut of gastroenterology and hepatology and do so in an expert fashion. They cover gastrointestinal and hepatic diseases, as well as syndromes, diagnostic and treatment modalities, and physiological and pathological processes. Where necessary, discussions of some of the diseases are divided into separate articles describing the condition in pediatric patients and adult patients. The articles are easy to read, yet comprehensive, and usually encompass both the basic science and the clinical aspects of that disease or process. Each article begins with a glossary of terms that allows the uninitiated to read the article by filling in gaps in understanding; a brief abstract that gives a concise, but comprehensive, overview of the subject matter follows the glossary.

Modern gastroenterology and gastrointestinal science lend themselves well to an encyclopedia format. Gastroenterology is a broad clinical field comprising issues that concern human behavior and psychology, other disciplines that are useful for understanding functional gastrointestinal diseases, such as dyspepsia and irritable bowel syndrome, and very technical sciences that encompass endoscopy as a diagnostic and therapeutic tool and often surgery as definitive treatment. In the middle of the spectrum is classical internal medicine as it relates to the gastrointestinal tract and liver. Furthermore, the behavioral, medical, endoscopic, and surgical approaches may differ considerably for pediatric patients versus adult patients.

As regards basic gastrointestinal science, the gastrointestinal tract does more than simply process and assimilate nutrients and water through the action of its digestive enzymes and secretory and absorptive processes. Gastrointestinal science also encompasses diverse fields such as endocrinology, immunology, and the neurosciences. If all the endocrine cells of the gastrointestinal tract were combined into one organ, it would be the largest endocrine organ in the human body. Similarly, the mucosal immune cells and the gastrointestinal-associated lymphatic tissue together constitute perhaps the largest organ of the immune system in the body. Furthermore, this entire, complex epithelial, secretory, absorptive, endocrine,

and immunological organ is controlled by its own intrinsic brain and nervous system, the enteric nervous system.

Gastroenterology is made even more complicated by the frequency with which diseases of this system occur and by its close relationship to other disciplines. Gastrointestinal cancer is the second most common type of cancer, if men and women are considered together. The broad field of nutrition borders closely on and is intertwined with gastroenterology. Finally, the intestinal tract is colonized by commensal microbiota that are crucial for optimal health. Little is known about these microorganisms, but it is beginning to be understood that they may be a vehicle for the treatment or prevention of disease through probiotics.

This broad view of gastroenterology and gastrointestinal science must have made it difficult for the Editors to choose the individual articles that make up these three volumes. However, I find the list to be fairly complete and each article to be a good mixture of basic knowledge and clinical information. If certain subject areas are not represented as specific articles, they are reasonably well covered in other articles that are in the three volumes and can be located in the subject index at the end of the third volume.

Who will use *The Encyclopedia of Gastroenterology* and why? I believe that the range of potential users is wide. For example, medical students often have difficulty with medical textbooks. Either the textbooks are too advanced and the various words and terms are not explained, making it difficult for the student to comprehend the text, or else the content has been reduced to a "mini" textbook version that often lacks substance. I believe that many of the entries in this encyclopedia are geared perfectly for medical students new to clinical medicine. The glossary of terms at the beginning of each article and the abstracts should be extremely helpful for those who are medically naive. Furthermore, the articles are well-crafted combinations of basic science and clinical science and this is useful at the medical student level.

Physicians from other disciplines will undoubtedly find *The Encyclopedia of Gastroenterology* to be a valuable reference work. The explosion of medical knowledge has made it difficult, if not impossible, to keep up with advancements in other areas of medicine. This encyclopedia provides concise descriptions of the various gastrointestinal diseases that are easily readable, complete, and up-to-date. This should be quite useful for the primary care physician or a specialist in another discipline who needs to know about some specific gastrointestinal disease or process.

Basic scientists and nonphysician translational research scientists would certainly benefit from this encyclopedia also. For instance, the mixture of basic science and clinical science information in each article is precisely what the basic scientist needs as he or she writes the Introduction or Discussion sections of publications or the Background section of grant applications. In addition, the encyclopedia would prove quite valuable in rapidly bringing scientists up to date in a specific area of gastrointestinal disease. In this era of transgenic animals, it is not uncommon for the scientist who has been conducting research in a specific field, for example, immunology or rheumatology, to create a new knockout mouse that presents with a gastrointestinal phenotype rather than a rheumatological phenotype. Thus, the scientist who has spent his or her career gaining an understanding of rheumatologic disease will need to quickly acquire a basic understanding of Crohn's disease and ulcerative colitis.

Finally, the gastroenterologist or gastrointestinal scientist can certainly utilize this encyclopedia as well. It is impossible to stay abreast of all areas of gastroenterology and gastrointestinal science and yet the overlapping disciplines within gastroenterology may demand a more detailed knowledge of an otherwise distant field of expertise. Thus, I look forward to having these three volumes on my bookshelf. They will be helpful in my clinical practice of gastroenterology and also helpful to me as a scientist and to other research scientists in my laboratory.

In summary, the Editors should be proud of their contribution to the knowledge base of gastroenterology. They have found a niche in our field that has hitherto not been occupied. There will continue to be a need for more elementary dictionaries and for highly detailed, advanced textbooks and monographs. *The Encyclopedia of Gastroenterology* will play a role in the middle of this spectrum and should be extremely valuable to a wide range of users. I believe that the three volumes will find their way onto the bookshelves of most medical libraries, as well as those of individual practitioners of medicine and active gastrointestinal investigators. *The Encyclopedia of Gastroenterology* is an extremely useful addition to our field.

DON W. POWELL

Professor, Internal Medicine
Professor, Cellular Physiology and Molecular Biophysics
Associate Dean for Research
Interim Director, General Clinical Research Center

PREFACE

The *Encyclopedia of Gastroenterology* bridges basic science and clinical gastroenterology in a way that should appeal to the expert researching a topic outside his or her field of expertise as well as to students and the educated public. Although some articles on the basic medical sciences stand alone, most integrate these topics with areas of clinical medicine. These volumes appear at a time when general interest in, and knowledge of, the gastrointestinal tract is expanding at a rapid rate. Research has led to new approaches in the treatment of many gastrointestinal diseases and a plethora of new drugs have been added to the pharmaceutical armamentarium. Diagnostic procedures have advanced remarkably over the past few years, leading to an increased understanding of how diseases develop and an improved ability to detect them.

The reader will find articles related to all areas of gastroenterology. There are articles covering basic physiology, pharmacology, anatomy, immunology, and microbiology. Others relate these basic fields to specific diseases. Many of these articles are entitled with the name of a disease or pathological condition. When appropriate, nutritional aspects of clinical conditions are emphasized and several articles feature aspects of nutrition. Areas of parasitology of special importance in relation to the gastrointestinal tract are also covered. Separate articles treat topics relating primarily to pediatric gastroenterology and numerous entries are concerned with radiology, endoscopy, and surgery.

The articles are written and organized to serve as convenient, yet comprehensive sources of information. Each of the 477 entries begins with a glossary of cross-referenced terms followed by a brief abstract. Generous use of primary and secondary headings allows the reader to locate material rapidly. Tables and figures emphasize important points and concepts. Each article presents core knowledge, time-tested and generally accepted within the field. As a result, there are no specific references with the entries. Each contribution, however, concludes with a list of references for further reading. Many of these are review articles with comprehensive bibliographies.

This work began with the selection of a number of general areas of coverage. A group of 15 Associate Editors, each of whom is an expert in at least one of these areas, was subsequently recruited. The entire group, along with members of the Elsevier staff, then met for two days in San Diego. That meeting resulted in the refinement of the areas of coverage, selection of individual article topics within those areas, and identification of potential authors. The Associate Editors and I believe that we have covered the important topics of basic and clinical gastroenterology. Each article is written to stand alone as a complete subject, so there is, no doubt, a certain amount of overlap. This, however, should make it easier for the reader to locate a specific piece of information.

A product of this magnitude represents the knowledge and efforts of a large number of individuals. More than 600 authors contributed their expertise to the individual articles. I am especially grateful to those who produced articles on short notice, so that our deadline could be met. An outstanding group of 15 Associate Editors was the foundation for this project. Due to their great breadth of knowledge, they were able to propose topics for articles and recommend the authors to write them. They then recruited the authors and edited the completed articles.

Finally, I acknowledge the contributions of the staff at Elsevier. The *Encyclopedia of Gastroenterology* was initiated and supported by Jasna Markovac, Sr. vice President, Elsevier, Science and Technology. Nick Panissidi, Senior Developmental Editor, was indispensable as he contacted authors and kept up with all article submissions. Pat Gonzalez served as Production Manager, and Tari Paschall, Sr. Publishing Editor, and Judy Meyer, Associate Publishing Editor, provided overall management of the project.

LEONARD R. JOHNSON

A

Abdominal Aortic Aneurysm

GORAV AILAWADI AND GILBERT R. UPCHURCH, JR.
University of Michigan, Ann Arbor

abdominal aortic aneurysm Permanent dilation of the abdominal aorta at least 50% greater than the expected normal diameter.
arteriomegaly Diffuse arterial enlargement more than 50% above normal.
dissecting aneurysm Type of aneurysm that dissects; although aneurysms rarely dissect, a dissection may lead to aneurysmal changes over time.
ectasia Arterial dilation less than 50% of expected normal diameter.
false aneurysm (pseudoaneurysm) Type of aneurysm that involves a disruption of the arterial wall with containment by surrounding tissue or hematoma.
fusiform aneurysm Spindle-shaped aneurysm.
infrarenal aortic aneurysm In the anatomic classification scheme, limited to aorta below the renal arteries.
pararenal aortic aneurysm In the anatomic classification scheme, comprising a juxtarenal aneurysm (near, but not involving, the renal artery orifices) or a suprarenal aneurysm (involving the renal arteries but not the superior mesenteric artery).
saccular aneurysm Eccentrically shaped aneurysm.
thoracoabdominal aneurysm In the anatomic classification scheme, involving the suprarenal mesenteric vessels; may also involve the descending thoracic aorta in the chest.
true aneurysm Type of aneurysm that involves all three layers of arterial wall.

Abdominal aortic aneurysms are a disease primarily of elderly, Caucasian males; rupture risk correlates with aneurysm size. Elective repair is generally undertaken in patients with an abdominal aortic aneurysm of 5–5.5 cm. Unexplained abdominal or back pain should raise suspicion for aortic rupture. Aortic rupture carries a significant mortality.

NATURAL HISTORY

Abdominal aortic aneurysms (AAAs) are the 15th leading cause of death in the United States. AAAs have a 4 : 1 male : female predominance and are 3.5 times more common in Caucasians than in African-Americans. Nearly 90% of aortic aneurysms are infrarenal. A 5-cm aneurysm generally carries a 25–30% 5-year rupture risk, and increasing aortic size carries substantially higher risks of rupture. Ruptured AAAs result in an overall mortality of greater than 75%, with nearly one-half of these patients dying prior to reaching a hospital. Risk factors for rupture include smoking, hypertension, chronic obstructive pulmonary disease, and aneurysm size. Current recommendations are conservative management until an AAA reaches 5–5.5 cm, when the risk of rupture is greater than the risk of elective repair.

SIGNS AND SYMPTOMS

The majority of patients with AAAs are asymptomatic, and only approximately 50% of AAAs are detectable on physical exam. Many are discovered incidentally during workup for unrelated problems. The classic triad of a ruptured AAA includes a palpable pulsatile abdominal mass, hypotension, and abdominal or back pain.

SCREENING AND DIAGNOSIS

Ultrasound is an excellent screening test in an asymptomatic patient suspected of having an AAA. In preparation for AAA repair, patients often undergo an abdominal computer tomography (CT) scan with intravenous contrast to delineate the extent of the aneurysm. In addition, a CT scan with three-dimensional reconstructions may help to determine the feasibility of endovascular AAA repair (Fig. 1). Patients with significant azotemia may undergo magnetic resonance angiography to avoid the risk of contrast nephropathy. An aortogram is performed when concurrent renal or mesenteric disease is suspected.

AAA REPAIR

Charles DuBost performed the first aortic aneurysm repair in 1951 using an aortic homograft. With their contributions, DeBakey, Cooley, and Crawford improved outcomes of modern AAA repair. Despite improved survival with elective AAA repair, emergent

 Copyright 2004, Elsevier (USA). All rights reserved.

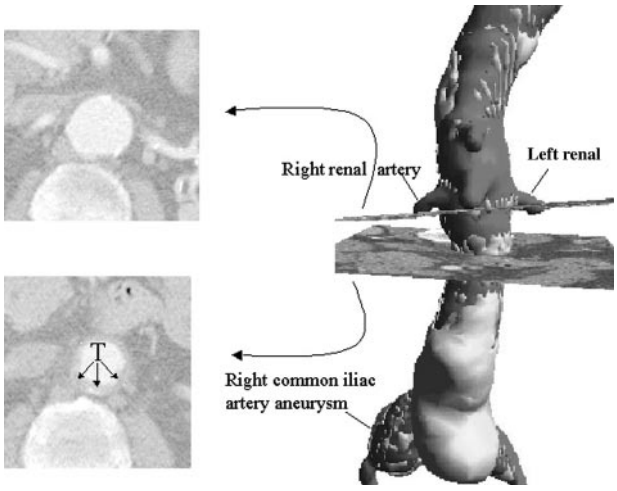


FIGURE 1 Three-dimensional CT scan of aorta in a patient being evaluated for endovascular AAA repair, showing the proximal aneurysm extent, involvement of the iliac arteries, and other anatomic landmarks important during repair of an AAA, such as excessive aortic thrombus (T) or calcification.

ruptured AAA repair is still associated with a high mortality rate. Endovascular treatment with aortic stent grafts, first performed in 1991, has become an option in patients who meet specific anatomic criteria.

POSTOPERATIVE COMPLICATIONS

Myocardial infarction and pulmonary failure may occur following AAA repair. Postoperative complications specific to the gastrointestinal system include ischemic colitis and aortoduodenal fistula. Early postoperative signs of abdominal pain, distension, or bloody stools should prompt immediate evaluation with flexible sigmoidoscopy. Aortoduodenal fistula is a late complication following AAA repair and is mainly due to erosion of the proximal aortic suture line into the duodenum. Hematemesis or hematochezia should prompt upper endoscopy in any patient with an AAA or a history of AAA repair.

See Also the Following Articles

Computed Tomography (CT) • Ultrasonography

Further Reading

Cronenwett, J. L., Murphy, T. F., Zelenock, G. B., Whitehouse, W. M., Jr., Lindenauer, S. M., Graham, L. M., Quint, L. E., Silver, T. M., and Stanley, J. C. (1985). Actuarial analysis of variable associated with rupture of small abdominal aortic aneurysms. *Surgery* **98**, 472–483.

Dimick, J. B., Stanley, J. C., Axelrod, D. A., Kazmers, A., Henke, P. K., Jacobs, L. A., Wakefield, T. W., Greenfield, L. J., and Upchurch, G. R., Jr. (2002). Variation in mortality after abdominal aortic aneurysmectomy in the United States: Impact of hospital volume, gender, and age. *Ann. Surg.* **235**, 579–585.

Katz, D. J., Stanley, J. C., and Zelenock, G. B. (1994). Operative mortality rates for intact and ruptured abdominal aortic aneurysms in Michigan: An eleven-year statewide experience. *J. Vasc. Surg.* **19**, 804–817.

Achalasia

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

achalasia Failure of the smooth muscle of the digestive tract to relax.
gastrointestinal sphincter A ring of circular muscle that contracts continuously and closes the lumen of the alimentary canal.
lower esophageal sphincter A ring of circular muscle that closes the orifice between the esophagus and the stomach.
sphincter of Oddi A ring of circular muscle that closes the entry of the bile duct into the small intestine.

The term achalasia is derived from the Greek word *chalasis*, which in English translates to relaxation. Achalasia is defined as failure of relaxation of the smooth muscle in any region of the digestive tract. It is most commonly applied to describe failure of relaxation in the various sphincters in the gut. Sphincters are regions where the circular muscle coat normally persists in a continuous state of contraction that produces a ring-like closure of the lumen. Sphincters function to prevent the back flux of lumenal contents from one digestive compartment to another. Prevention of the backward movement of the acidic contents of the stomach into the esophagus is an example of sphincteric function.

A nervous mechanism relaxes the sphincters with appropriate timing to open the lumenal orifice and permit passage from one compartment to the next. Relaxation is transient, with contraction and closure occurring after passage of the material through the sphincter. During a swallow, the lower esophageal sphincter relaxes to permit passage into the stomach. Emptying of the stomach occurs during transient relaxation of the sphincter located at the junction with the small intestine. Delivery of bile from the gallbladder to the small intestine occurs during relaxation of the sphincter of Oddi located at the opening of the bile duct into the small intestine. Passage of contents from the small intestine into the large intestine takes place during relaxation and opening of the sphincter that separates the two dissimilar compartments. The internal anal sphincter relaxes to permit passage of feces during defecation and then closes to prevent inopportune release of the contents of the large intestine. An obvious consequence of achalasia in any of these sphincters is obstruction of the forward passage of lumenal contents from one compartment to another.

PHYSIOLOGY

The specialized physiology of the sphincteric circular muscle coat accounts for the ability of the muscle to sustain contraction. The contractile apparatus of these smooth muscles consists of the two proteins actin and myosin. Contraction occurs during formation of cross-bridges between actin and myosin filaments. Contractile tension is maintained by a "catch" mechanism that latches the cross-bridges in place without expenditure of additional energy. Input from the nervous system leads to uncoupling of the cross-bridges, relaxation of contractile tension, and opening of the sphincter. The nervous input involves the release of chemical neurotransmitters at neuromuscular junctions.

Sphincteric muscles are innervated by motor neurons in the enteric nervous system. Most of the motor neurons to the sphincters are inhibitory motor neurons. Firing of nerve impulses by the inhibitory motor neurons releases inhibitory neurotransmitters at their junctions with the smooth muscle fibers of the sphincter. Two important inhibitory neurotransmitters are vasoactive intestinal polypeptide and nitric oxide. These neurotransmitters act to inhibit contraction of the smooth muscle and open the sphincter. Decisions as to when to open a sphincter are made by integrative neural networks located either in the central nervous system or in the gut itself. The neural networks control the firing of the inhibitory motor neurons. The inhibitory motor neurons are silent most of the time and the contractile behavior inherent in the muscle holds the sphincter in a closed state. Activation of the inhibitory motor neurons releases contractile tension in the muscle and the sphincter opens. Contractile tension redevelops and the sphincter closes coincident with cessation of motor neuronal firing.

 Copyright 2004, Elsevier (USA). All rights reserved.

PATHOGENESIS

Sphincteric physiology predicts that loss of the inhibitory motor innervation will result in achalasia. The inherent contractility of the sphincteric musculature keeps the sphincter closed and opening cannot occur when the inhibitory innervation is missing. The pathophysiologic result is an obstruction to passage of the lumenal contents through the sphincter. Material accumulates and dilates the digestive canal proximal to the sphincter because propulsive motility generally moves lumenal contents in the direction from mouth to anus. Achalasia in the sphincter between esophagus and stomach (i.e., the lower esophageal sphincter) can lead to a gross dilation of the esophagus described as a megaesophagus. Likewise, achalasia in the internal anal sphincter obstructs the passage of feces and can lead to a megacolon. Achalasia in the sphincter of Oddi obstructs the delivery of bile to the small intestine and can lead to overdistension of the biliary tree that is experienced as pain in the right upper abdominal quadrant.

Loss of the inhibitory innervation occurs in parallel with generalized dystrophy in the enteric nervous system that may be acquired or congenital. The most commonly acquired form reflects an autoimmune attack on the enteric nervous system that may be related to the presence of a tumor elsewhere in the body, may be related to an infectious agent, or may be idiopathic in nature. Autoimmune attack on the enteric nervous system occurs in association with small cell carcinoma of the lung. This is called paraneoplastic syndrome and occurs when the immune system attacks antigens on the surfaces of enteric neurons that are similar to antigens expressed by the tumor cells. Tests in patients with lower esophageal sphincter achalasia indicate that a significant proportion of these patients have circulating antibodies that react with enteric neurons. Achalasia in Chagas' disease is similar to paraneoplastic syndome in that antigenic epitopes expressed by the blood-borne parasite *Trypanosoma cruzi* are sufficiently similar to antigens on enteric neurons that an immune response to enteric neurons follows the response to the parasite.

Congenital absence of enteric neurons, including inhibitory motor neurons, occurs in Hirschsprung's disease. Mutations in Ret and endothelin genes prevent the fetal development of the enteric nervous system in a variable length of large intestine including the internal anal sphincter. The internal anal sphincter is achalastic and presents an obstruction to the passage of feces. A megacolon develops as feces accumulates proximal to the neurally deficient segment.

See Also the Following Articles

Anal Sphincter • Chagas' Disease • Paraneoplastic Syndrome

Further Reading

Castell, D., and Richter, J. (1999). "The Esophagus," 3rd Ed. Lippincott Williams & Wilkins, Philadelphia.

Verne, G., Sallustio, J., and Eaker, E. (1997). Anti-myenteric neuronal antibodies in patients with achalasia: A prospective study. *Digest. Dis. Sci.* **42**, 307–313.

Wood, J., Alpers, D., and Andrews, P. (1999). Fundamentals of neurogastroenterology. *Gut* **45**, 1–44.

Aging

MAKAU LEE
University of Mississippi Medical Center

aging Postmaturational changes occurring between middle age and old age, resulting in reduction of the functional capacity of the physiological systems and increase of the vulnerability of an organism to challenges and diseases.
digestive system Organ system responsible for digestion and absorption.

Aging, the postmaturational changes occurring between middle age and old age, reduces the functional capacity of the physiological systems and increases the vulnerability of an organism to challenges and diseases, thereby increasing the likelihood of death.

PHYSIOLOGY OF AGING

There are numerous theories of aging; and they can be classified into two main categories: the "programmed" theories, which view aging as the result of a predetermined genetic blueprint, and the "wear and tear" theories, which view aging as the consequence of continual stress and injuries. Although none of the theories of aging can singularly explain the aging process, available data indicate that the ultimate mechanism of aging is that of molecular changes: all of the changes involve both informational molecules (such as DNA and RNA) and structural molecules (such as lipids, carbohydrates, and proteins) in a process that is determined genetically and epigenetically.

Both clinical and basic investigations have demonstrated that aging is associated with specific physiological and structural changes in the digestive system. These age-related changes in the digestive system may contribute to the development of various digestive disorders that are more common among the elderly. In the clinical setting, when symptoms and signs occur in elderly patients, the physician must differentiate age-associated physiological changes from the consequences of diseases.

AGING AND OROPHARYNGEAL STRUCTURES

Age-related physiological changes in oropharyngeal structures include structural alterations (such as thinning of the tongue, involving both mucosa and muscles, and weakening of muscles of the mouth and pharynx) and functional changes (such as alterations in pharyngeal sensation, proprioception, and taste acuity; discoordination of masticatory muscles, resulting in prolonged swallowing; and altered peristaltic response after deglutition). Manometric studies have shown that upper esophageal pressure is decreased and upper esophageal sphincter relaxation after deglutition is delayed in the elderly. Salivary output and flow remain unchanged in the elderly.

Clinically, these age-related changes in oropharyngeal structures may affect the elderly patient's ability to swallow liquids and may lead to reduced food intake. Reduced food intake and malnutrition in the elderly can also occur for various reasons, including endogenous or medication-induced depression, anorexia, social isolation, physical handicaps that interfere with the elderly person's ability to prepare food, dental problems or ill-fitting dentures, and forgetfulness and failure to eat as the result of diseases of the central nervous system. Furthermore, structural lesions (such as Zenker's diverticulum) and upper esophageal sphincter dysfunction (such as cricopharyngeal achalasia) occur more frequently in the elderly.

AGING AND THE ESOPHAGUS

Manometric studies have demonstrated minor to mild changes in esophageal motility with aging. For instance, there are reductions in contractile velocity and amplitude, increases in synchronous contractions, polyphasic waves and tertiary contractions, upward displacement of lower esophageal sphincter, incomplete lower esophageal sphincter relaxation, and failure of contractions after deglutition in the distal esophagus. Radiographic evaluations frequently reveal dilatation of the esophagus and increases in synchronous and tertiary contractions in the elderly. Presbyesophagus (or corkscrew esophagus), which denotes the increase in tertiary contractions detected radiographically, has no known pathophysiologic consequence. Available data suggest

 Copyright 2004, Elsevier (USA). All rights reserved.

that most esophageal motility problems in the elderly are due to other concomitant medical problems and to the intake of numerous drugs. Clinically, gastroesophageal reflux disease, achalasia (a motility disorder), and esophageal cancer occur more frequently in the elderly.

AGING AND THE STOMACH

The majority of healthy elderly individuals have normal gastric acid secretion. Clinical studies have demonstrated that there are modest reductions in pepsin output and significant decreases in gastroduodenal mucosal prostaglandin biosynthesis in the elderly. Although gastric emptying of solids in healthy, elderly individuals remains unchanged, gastric emptying of liquids may be impaired.

Clinically, peptic ulcer disease and its complications are more common among the elderly. Potential explanations for the higher peptic ulcer incidence in the elderly include an age-related increase in the prevalence of *Helicobacter pylori* infection and the increasing use of nonsteroidal antiinflammatory drugs (NSAIDs) by older individuals. Moreover, both clinical and basic investigations have shown that age-related reductions in gastroduodenal mucosal protective factors (such as gastroduodenal prostaglandin synthesis, gastroduodenal bicarbonate secretion, and expression of mucosal protective growth factors) may predispose the elderly to the development of peptic ulcer disease.

Whereas at least 75% of healthy elderly persons have normal gastric acid production, up to 25% of elderly persons have acid hyposecretion (or hypochlorhydria) because of atrophic gastritis. Gastric hypochlorhydria leads to an increase in the luminal pH of the stomach and proximal intestine and may predispose these older individuals to various enteric infections (such as typhoid, *Salmonella*, cholera, and *Giardia*), because a low intragastric pH is a known defense mechanism against bacteria introduced into the upper digestive tract. Gastric hypochlorhydria may also contribute to the development of malabsorption of iron, folic acid, vitamin B_6 and vitamin B_{12}, calcium carbonate, and various trace minerals in the elderly. Finally, epidemiological studies have suggested that chronic gastritis and associated hypochlorhydria from *H. pylori* infection may lead to the progressive development of atrophic gastritis, gastric metaplasia, and adenocarcinomas of the stomach, which occur primarily in the elderly.

AGING AND THE INTESTINES

Due to the tremendous functional reserve that the intestine possesses, digestion and absorption are well preserved in the elderly. Data from animal and human studies have shown that with increasing age, there are modest changes in digestive enzyme secretion, subtle decreases in carbohydrate absorptive capacity, and altered calcium absorption. There are no significant structural differences in the proximal intestinal epithelium between the young and the old. Despite few changes in overall intestinal absorption with increasing age, clinical malabsorption syndrome occurs more frequently in the elderly. Common causes of malabsorption in patients older than 65 years of age include pancreatic insufficiency, celiac sprue, mesenteric ischemia, and bacterial overgrowth syndrome associated with small bowel diverticulosis, intestinal strictures, and postgastrectomy states.

Studies with radio-opaque markers have shown that bowel transit time does not differ between healthy young and old volunteers. However, delays in rectal evacuation and other anorectal dysfunctions (such as reduction in rectal wall elasticity and a blunting of rectal sensation) are frequently detected in elderly patients who complain of constipation. Clinically, aging is associated with an increased incidence of constipation, fecal impaction, fecal incontinence, megacolon, cecal volvulus, pseudoobstruction (including Ogilvie's syndrome), diverticular disease, ischemic bowel disease, and colorectal cancer.

AGING AND THE LIVER

Due to the enormous reserve capacity of the liver, age-related changes in liver functions are not clinically significant. Both clinical and basic investigations have shown that there are modest reductions in liver size, blood flow and perfusion, and dynamic liver functions with increasing age. Animal studies have also shown that although the rate and time course of hepatic regeneration after partial hepatectomy are slightly delayed with aging, regeneration of the liver in old animals is as complete as in young animals. These observations have led to the important decision to raise the age limit for potential liver donors, resulting in an increased use of older donor livers for hepatic transplantation. Although the clinical course of liver diseases in the elderly does not differ from that in the young, autoimmune liver diseases (such as primary biliary cirrhosis), covert alcoholism, and hepatocellular carcinoma occur more frequently in the elderly.

AGING AND THE GALLBLADDER AND BILIARY TRACT

Age-related changes in the biliary tract include gradual narrowing of the distal common bile duct and increased incidence of juxtapapillary duodenal diverticula. Aging is also associated with an increased incidence of gallstones, reflecting the formation of more lithogenic bile (due to elevated biliary cholesterol secretion, reduced bile acid synthesis, and resultant supersaturation of bile with cholesterol) and less forceful gallbladder contractions (the gallbladder becomes more distensible with age). Consequently, gallstone disease and its complications (such as cholecystitis, choledocholithiasis, and cholangitis), biliary tract diseases, cholangiocarcinoma, and cancer of the gallbladder occur more frequently in the elderly.

AGING AND THE PANCREAS

Age-related changes in the pancreas include modest reduction in pancreatic size and weight, minor atrophic or fibrotic changes, symmetrical dilatation of the interlobular and intralobular ducts, and an increased incidence of pancreatic duct stones. Basal and stimulated exocrine pancreatic secretion changes little with increasing age. Because only 10–20% of the maximal pancreatic exocrine output is needed for normal digestive function, minor declines in pancreatic enzyme output with aging are not clinically significant. Clinically, pancreatic malignancies occur primarily in older individuals.

AGING, GASTROINTESTINAL PROLIFERATION, AND CARCINOGENESIS

The gastrointestinal tract is an epithelial organ with constant cell turnover and rapid mucosal proliferation. Both animal and human studies have shown that aging is associated with increases in gastrointestinal mucosal proliferation, impaired proliferative response to mucosal injury or to feeding, reduced capacity to repair mucosal injury and oxidative damage to key cellular molecules (such as DNA, proteins, and lipids), and altered expression of various growth factors in response to injury. In experimental models, dietary (or calorie) restriction, which is a proved antiaging intervention, retards the development of age-related hyperproliferative changes in the gastrointestinal tract and reduces carcinogen-induced gastrointestinal tumors. These observations suggest that age-associated increases in gastrointestinal proliferation may contribute to a higher risk of tumor formation, because various malignancies of the digestive tract occur primarily in the elderly.

Acknowledgment

Adapted with permission from "Gastrointestinal Function" in *The Encyclopedia of Aging*, 3rd ed. (Maddox, *et al.*) pp. 410–412. Copyright 2001 Springer Publishing Company, Inc., New York.

See Also the Following Articles

Achalasia • Atrophic Gastritis • Cholelithiasis, Complications of • Esophageal Cancer • *Helicobacter pylori* • Nutrition in Aging • Stomach Adenomas and Carcinomas of the • Zenker's Diverticulum

Further Reading

Armbrecht, H. J. (2001). The biology of aging. *J. Lab. Clin. Med.* **138**, 220–225.

Cohen, S. (1996). Pancreatic disease in the elderly. *In* "Clinical Gastroenterology in the Elderly" (A. M. Gelb, ed.), pp. 213–226. Marcel Dekker, New York.

Holt, P. R. (1995). Approach to gastrointestinal problems in the elderly. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 2nd Ed., pp. 968–988. J. B. Lippincott, Philadelphia.

James, O. F. W. (1997). Parenchymal liver disease in the elderly. *Gut* **41**, 430–432.

Lee, M. (2001). Gastrointestinal function. *In* "The Encyclopedia of Aging" (G. L. Maddox, ed.), 3rd Ed., pp. 410–412. Springer Publ., New York.

Lee, M., and Feldman, M. (1997). The aging stomach: Implications for NSAID gastropathy. *Gut* **41**, 425–426.

Siegel, J. H., and Kasmin, F. E. (1997). Biliary tract diseases in the elderly: Management and outcomes. *Gut* **41**, 433–435.

Tack, J., and Vantrappen, G. (1997). The aging oesophagus. *Gut* **41**, 422–424.

AIDS, Biliary Manifestations of

TOMMY T. OEI
Keck School of Medicine, University of Southern California

cholangiopathy Any disease process involving the biliary system.
cholestasis Stoppage or suppression of the flow of bile due to either intrahepatic or extrahepatic causes.
endoscopic retrograde cholangiopancreatogram A procedure performed by cannulation of the common bile duct and pancreatic duct by means of a flexible endoscope and retrograde injection of radiopaque contrast media in order to demonstrate all portions of the biliary tree and pancreatic ducts.
magnetic resonance cholangiopancreatogram A noninvasive, radiological procedure performed by magnetic resonance technology that provides images of the biliary tree and pancreatic ducts.
papillary stenosis Narrowing or stricture of the papilla of Vater, the common exit site of fluid passing through the common bile and pancreatic ducts.
sclerosing cholangitis Inflammation of the bile ducts leading to fibrotic/stenotic changes.
sphincterotomy Division of a sphincter; in the case of AIDS cholangiopathy, refers to endoscopic cutting of the papilla of Vater in order to relieve biliary obstruction.

AIDS (acquired immune deficiency syndrome) cholangiopathy is a clinical condition characterized by right upper quadrant pain, fever, marked elevation in serum alkaline phosphatase levels, nausea, and vomiting. Typically, these patients do not have jaundice. The underlying pathology includes acalculous cholecystitis, papillary stenosis, and sclerosing cholangitis, conditions that often occur concomitantly. This type of biliary disease takes place only in the presence of significant immunosuppression as the typical CD4 count in these patients is less than 200/mm^3. One must keep in mind, however, that although patients with AIDS are uniquely susceptible to this type of cholangiopathy, more common conditions such as gallstone disease should still be considered in the differential diagnosis.

INTRODUCTION

Diseases of the biliary tract and gallbladder in patients with AIDS (acquired immune deficiency syndrome) have been well described. Prior to 1981, when AIDS was first recognized, there were no reports of cholangiopathy associated with opportunistic pathogens such as cryptosporidia, cytomegalovirus (CMV), or microsporidia. In 1983, the first cases of biliary tract disease associated with AIDS were published as human cryptosporidial infections of the bile ducts associated with biliary tract obstruction. These first patients, as well as those that followed, tended to be homosexual men with low CD4 counts and a history of other opportunistic infections. Investigation by endoscopic retrograde cholangiopancreatogram (ERCP) frequently demonstrated intra- and/or extrahepatic biliary changes characteristic of sclerosing cholangitis, often associated with papillary stenosis. Culture of the bile and biopsy of the immediately surrounding duodenal tissue often demonstrated the previously mentioned opportunistic organisms. Antimicrobial and endoscopic treatments were attempted, but subsequent studies demonstrated that whereas the latter led to substantial pain relief, the former did little to alter the patients' clinical course.

The prevalence of AIDS cholangiopathy is unknown. Most studies focusing on this disease have been either case reports or series. One study suggested that, at least prior to the highly active antiretroviral therapy (HAART) era, as many as 30% of patients with AIDS-associated refractory diarrhea may have had AIDS cholangiopathy. It is unknown what effect HAART has had on this disease, though one would postulate that it has been beneficial due to the resultant increase in the CD4 count.

ETIOLOGY

The exact cause of AIDS cholangiopathy is unclear. Many investigators favor an infectious etiology as different opportunistic organisms, most commonly *Cryptosporidium parvum*, microsporidia species, and CMV (see Table I), have been cultured from the bile and surrounding duodenal mucosa. Antimicrobials aimed specifically at eradicating these pathogens, however, have neither reversed the underlying disease

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Pathogens That Are Associated with Cholangiopathy and Cholecystitis in Patients with AIDS

Cryptosporidium parvum
Microsporidia
Enterocytozoon bieneusi
Encephalitozoon intestinalis
Encephalitozoon cuniculi
Cytomegalovirus
Mycobacterium avium complex
Cyclospora cayetanesis
Isospora belli
Salmonella enteritidis
Salmonella typhimurium
Enterobacter cloacae
Campylobacter fetus
Candida albicans

process nor prevented its progression. It is unclear whether these pathogens serve as inciting factors or are merely incidentally found in an environment void of the typical immunologic defenses.

Regardless of whether it is these opportunistic organisms, human immunodeficiency virus, or another causative factor that initiates the process leading to AIDS cholangiopathy, what develops is an inflammatory reaction of the bile duct epithelium leading to the hallmark cholangiographic changes. Microscopic examination of the bile duct epithelium demonstrates that the inflammatory changes are different from those seen in primary sclerosing cholangitis (PSC). In PSC, the diseased bile ducts are surrounded by T4 lymphocytes, the cell population that is depleted in AIDS patients. Instead, others have noted the presence of squamous metaplasia of the bile ducts and pancreatic ducts associated with cryptosporidial infection. Though detection of a single organism in the biliary tree has been most commonly documented, multiple organisms in individual patients have been noted also. Earlier studies reported a significant number of patients with no identifiable organism, but these investigations were conducted prior to the discovery of the presence of microsporidia in the bile ducts.

Although infection by opportunistic organisms is associated with most cases, bile duct infiltration with neoplasms such as Kaposi's sarcoma (KS) and Burkitt's lymphoma of the bile ducts may present in a similar fashion. In addition, pancreatic disorders resulting in disease of the distal common bile duct (CBD) may mimic an infectious or neoplastic process. Differentiating among these possible causes of cholangiopathic changes is important as there are potentially beneficial treatments available for some, but not others.

CLINICAL PRESENTATION

The clinical presentation of patients with AIDS cholangiopathy is variable, ranging from asymptomatic elevation in serum alkaline phosphatase to cholangitis with right upper quadrant (RUQ) pain, fever, and chills (see Table II). Patients in case series tended to be homosexual, middle-aged men who had AIDS for over 1 year. It is unknown whether an individual's sexual practice increases the likelihood of developing cholangiopathy or whether this finding is due to homosexual men, as a group, being affected and seeking medical attention earlier in the AIDS epidemic. Due to their low CD4 count, these patients often have had other opportunistic infections prior to presenting with cholangiopathy. Diarrhea commonly occurs in these patients as the organisms that have been associated with this cholangiopathy are also known to cause diarrhea by infecting the small bowel. Jaundice is uncommon and its presence should initiate the search for other coexisting disease processes.

DIAGNOSIS

Laboratory Tests

Abnormal liver tests are commonly seen in AIDS patients. The etiology of these abnormalities is varied and often multifactorial (e.g., medications, infections). Patients with AIDS cholangiopathy most commonly present with anicteric cholestasis (i.e., markedly elevated serum alkaline phosphatase levels with normal bilirubin levels). Significantly elevated bilirubin levels are uncommon (< 5%). Transaminase levels may be mildly elevated, but are usually not markedly so unless there is a coexisting disease process. Other cholestatic conditions that should be considered include granulomatous hepatitis, drugs, viral hepatitis, and intrahepatic lymphoma.

TABLE II Cardinal Features of AIDS Cholangiopathy

RUQ pain
Marked elevation in serum alkaline phosphatase levels
No or minimal elevations in transaminase and bilirubin levels
Low CD4 count (<200/mm^3)
History of opportunistic infections
Diarrhea
Nausea
Vomiting
No jaundice
Fever

Imaging

Ultrasound

Abdominal ultrasound (US) and ERCP should be considered complementary tests for the diagnosis of this condition. The most effective initial imaging test for AIDS patients with RUQ pain and abnormal liver tests remains the abdominal ultrasound. The typical findings in patients with AIDS cholangiopathy are dilated intra- and/or extrahepatic bile ducts with focal strictures. Biliary ductal wall thickening also supports the presence of disease. Although ERCP is considered the gold standard, ultrasound has a sensitivity of 75–87%. US, however, is superior to ERCP in demonstrating ductal wall thickening, which suggests pericholangitis. In one series, a normal ultrasound was found in 25% of patients with ERCP-confirmed AIDS cholangiopathy. Therefore, ERCP should still be performed in patients with a strong clinical suspicion for cholangiopathy despite a normal ultrasound.

US findings of thickened gallbladder walls and pericholecystic fluid suggest the presence of cholecystitis. In patients without gallstones, a hydroxyiminodiazetic acid (HIDA) scan should be performed to diagnose acalculous cholecystitis. In these cases, HIDA scan will demonstrate an absence of gallbladder filling despite imaging of the CBD.

Computed Tomography Scan

CT (computed tomography) scans do not usually contribute much additional information beyond that gained by a good-quality abdominal US in patients who have only AIDS cholangiopathy. However, CT scans may provide additional insight for patients in whom an US was technically difficult (e.g., due to obesity) or who are jaundiced and are suspected of having other pathology, such as mass lesions.

ERCP

ERCP is the gold-standard diagnostic test for AIDS cholangiopathy. With the improving technological advances of magnetic resonance cholangiopancreatogram, however, it is conceivable that this noninvasive test will replace ERCP for diagnosis. ERCP should be performed only on symptomatic patients with RUQ pain, since only then would a sphincterotomy benefit the patient.

Typically, ERCP demonstrates distal CBD tapering with dilation of the larger intrahepatic ducts, proximal CBD, and common hepatic ducts. Focal intraductal debris is present as is beading of the mucosa to suggest intramural submucosal infiltration and edema. These sclerotic changes occur more commonly in the left intrahepatic system. These findings initially suggested a condition similar to PSC; however, there are some differences (see Table III). In PSC, the entire CBD and the larger intrahepatic ducts are ordinarily sclerotic with minimal focal dilations. Extrahepatic strictures rarely exceed 4–5 mm and saccular deformities involve the intra- and extrahepatic ducts. Isolated strictures of the CBD are rarely due to AIDS cholangiopathy and may be due to primary CBD lymphoma, external compression of the CBD by nodes enlarged by lymphoma, or pancreatic disease caused by chronic pancreatitis, infections, or neoplasms.

Four different cholangiographic patterns of disease have been described. These include sclerosing cholangitis with papillary stenosis, papillary stenosis alone, sclerosing cholangitis without papillary stenosis, and long bile duct strictures (see Table IV). It is unknown whether these patterns represent a progression of disease, but the detection of papillary stenosis is meaningful as sphincterotomy in these patients can lead to substantial symptom relief.

TABLE III Comparative ERCP Findings of AIDS Cholangiopathy versus Primary Sclerosing Cholangitis

AIDS cholangiopathy	Primary sclerosing cholangitis
Distal CBD tapering	Extrahepatic strictures rarely exceeding 4–5 mm
Larger intrahepatic bile ducts, proximal CBD, and common hepatic ducts are, in general, dilated	Sclerotic extrahepatic biliary system and larger intrahepatic ducts with minimal focal dilation
Irregular dilation of the intra- and extrahepatic biliary system	Entire CBD involved with irregular strictures
Beading of the mucosa	No beading of the mucosa
Focal intraductal debris	No intraductal debris
Disproportionately distorted left intrahepatic ducts	Equal distribution of deformities between right and left systems
Intrahepatic irregular focal sacculations and dilations	Saccular deformities involving the intra- and extrahepatic ducts
Pruning of the smaller intrahepatic bile ducts	Paucity of intrahepatic ducts

TABLE IV Cholangiographic Patterns of AIDS Cholangiopathies

Type	Description	Percentage of total cases
Papillary stenosis (PS)	CBD diameter greater than 8 mm with tapering of the distal 2–4 mm with marked retention of contrast beyond 30 min	7% (0–15)[a]
Sclerosing cholangitis (SC) without PS	Focal strictures and dilations of the intra- and extrahepatic bile ducts	31% (20–88)
PS and intra- and extrahepatic SC	As above	54% (13–73)
Long, extrahepatic bile duct strictures	Strictures with lengths in excess of 1–2 cm in patients without prior CBD exploration or documented chronic pancreatitis	8% (0–15)

[a] Numbers in parentheses are the percentage ranges seen in different studies.

TREATMENT

Endoscopic

Endoscopic treatment is the most effective intervention for symptomatic patients who are found to have papillary stenosis on ERCP. Sphincterotomy results in prompt and sustained pain relief. However, this procedure does not alter the course of the disease. Patients' alkaline phosphatase levels continue to increase, suggesting progression of the underlying disease process. Studies evaluating repeat ERCP have demonstrated progressive intrahepatic sclerosing cholangitis in some patients despite symptom relief after endoscopic treatment. Sphincterotomy should be viewed as an effective method of improving these patients' quality of life. Due to the good response to sphincterotomy, the pain has been postulated to be due to sphincter of Oddi dysfunction incited by infection and/or inflammation rather than to biliary obstruction.

Medical

Effective treatment is lacking for patients who are found not to have papillary stenosis. Medical treatments aimed at eradicating the underlying infection have been attempted. Studies utilizing paromomycin for cryptosporidiosis and gancyclovir for CMV infections have improved neither symptoms nor disease progression. Other studies have tried ursodeoxycholic acid, but there are no large prospective trials demonstrating efficacy. In addition, there have been no published reports describing the effect of HAART. It is plausible that HAART may improve symptoms or at least hinder disease onset or progression, but this has not been proven. In patients discovered to have KS or lymphoma, chemotherapy should be considered as there was a case report of cholangiopathy secondary to disseminated non-Hodgkin's lymphoma that resolved after treatment with chemotherapy.

Surgery

Surgical intervention has no role in AIDS cholangiopathy limited to ductal disease. However, in patients with cholecystitis, whether due to stones or infection, surgery can be both beneficial and curative.

PROGNOSIS

The prognosis in patients diagnosed with AIDS cholangiopathy is poor. This condition is rarely fatal by itself, however. Other opportunistic infections and progression of the underlying immunodeficiency are usually the more immediate causes of death. Different studies have reported a 1-year survival rate of 14–40% (median survival time being 10 months). The likelihood of survival is affected by neither the level of the liver tests nor the degree of ERCP abnormalities. Rather, the patient's CD4 count provides the best prognostic value. It is unknown whether HAART improves prognosis in patients with AIDS cholangiopathy as most studies were conducted prior to the introduction of this treatment.

See Also the Following Articles

AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • *Campylobacter* • Candidiasis • Cholangitis, Sclerosing • Computed Tomography • *Cryptosporidium* • Cytomegalovirus • *Salmonella* • Sphincterotomy • Ultrasonography

Further Reading

Cello, J. P. (1992). Human immunodeficiency virus-associated biliary tract disease. *Semin. Liver Dis.* **12**, 213–218.

Cello, J. P. (1998). AIDS-related biliary tract disease. *Gastrointest. Endosc. Clin. North Am.* **8**, 963–973.
Cello, J. P., and Chan, M. F. (1995). Long-term follow-up of endoscopic retrograde cholangiopancreatography sphincterotomy for patients with acquired immune deficiency syndrome papillary stenosis. *Am. J. Med.* **99**, 600–603.
Ducreux, M., Buffet, C., Lamy, P., *et al.* (1995). Diagnosis and prognosis of AIDS-related cholangitis. *AIDS* **9**, 875–880.
Wilcox, C. M., and Monkemuller, K. E. (1998). Hepatobiliary diseases in patients with AIDS: Focus on AIDS cholangiopathy and gallbladder disease. *Digest. Dis.* **16**, 205–213.

AIDS, Gastrointestinal Manifestations of

PETER V. CHIN-HONG AND ROBERT L. OWEN
University of California, San Francisco

highly active antiretroviral therapy Combination antiretroviral therapy used to treat HIV. Initial therapy usually includes one or two protease inhibitors together with two nucleoside analogues, or a nonnucleoside reverse transcriptase inhibitor (NNRTI) with two nucleoside analogues. Since the introduction of highly active antiretroviral therapy, there has been a marked reduction in mortality, incidence of opportunistic infections, and hospitalizations in patients on therapy.

immune reconstitution Immune-specific responses that can be transiently associated with unusual manifestations of opportunistic infections, such as *Mycobacterium avium* complex lymphadenitis. This occurs several weeks to months after highly active antiretroviral therapy is initiated, as CD4 lymphocytes expand. Supportive therapy is administered and highly active antiretroviral therapy is not usually discontinued.

nonnucleoside reverse transcriptase inhibitor One of the classes of drugs used in combination as highly active antiretroviral therapy to treat HIV disease. Examples in this class include efavirenz and nevirapine, which are both highly potent medications.

nucleoside reverse transcriptase inhibitor One of the first classes of drugs used to treat HIV. Examples in this class include zidovudine and stavudine.

protease inhibitor Introduced in 1995, drugs in this class allowed for the first time the use of effective combination therapy against HIV. Examples include indinavir and nelfinavir. Most recently, there has been widespread use of ritonavir-boosted regimens (lower dose ritonavir in conjunction with another protease inhibitor), which exploit the fact that ritonavir is a potent inhibitor of the P450 enzyme pathway.

Gastrointestinal manifestations are common in the human immunodeficiency virus (HIV)-positive population, occurring in up to 93% of patients in studies prior to the widespread use of highly active antiretroviral therapy (HAART). Despite the advances in the therapeutic arsenal, gastrointestinal complaints are still a major source of morbidity and mortality in this population. The use of HAART has changed the spectrum of gastrointestinal involvement in HIV disease, treating or preventing the occurrence of opportunistic infections, but at the same time, contributing to a significant cause of drug-related toxicity. Furthermore, as the life expectancy of individuals living longer with autoimmune deficiency syndrome (AIDS) is increased because of HAART, other gastrointestinal-specific causes of morbidity, such as chronic hepatitis C virus, have become increasingly important.

APPROACH TO THE HIV-POSITIVE PATIENT WITH A GASTROINTESTINAL COMPLAINT

An approach to an HIV-positive patient with gastrointestinal (GI) symptoms classically depends on the degree of immunosuppression. Symptoms are varied and include dysphagia and odynophagia, nausea, vomiting, abdominal pain, GI bleeding, diarrhea, and anorectal tenderness. Signs such as jaundice and hepatomegaly are common. Because symptoms can be protean and nonspecific, the decision to pursue a

 Copyright 2004, Elsevier (USA). All rights reserved.

Cello, J. P. (1998). AIDS-related biliary tract disease. *Gastrointest. Endosc. Clin. North Am.* **8**, 963–973.
Cello, J. P., and Chan, M. F. (1995). Long-term follow-up of endoscopic retrograde cholangiopancreatography sphincterotomy for patients with acquired immune deficiency syndrome papillary stenosis. *Am. J. Med.* **99**, 600–603.
Ducreux, M., Buffet, C., Lamy, P., *et al.* (1995). Diagnosis and prognosis of AIDS-related cholangitis. *AIDS* **9**, 875–880.
Wilcox, C. M., and Monkemuller, K. E. (1998). Hepatobiliary diseases in patients with AIDS: Focus on AIDS cholangiopathy and gallbladder disease. *Digest. Dis.* **16**, 205–213.

AIDS, Gastrointestinal Manifestations of

PETER V. CHIN-HONG AND ROBERT L. OWEN
University of California, San Francisco

highly active antiretroviral therapy Combination antiretroviral therapy used to treat HIV. Initial therapy usually includes one or two protease inhibitors together with two nucleoside analogues, or a nonnucleoside reverse transcriptase inhibitor (NNRTI) with two nucleoside analogues. Since the introduction of highly active antiretroviral therapy, there has been a marked reduction in mortality, incidence of opportunistic infections, and hospitalizations in patients on therapy.

immune reconstitution Immune-specific responses that can be transiently associated with unusual manifestations of opportunistic infections, such as *Mycobacterium avium* complex lymphadenitis. This occurs several weeks to months after highly active antiretroviral therapy is initiated, as CD4 lymphocytes expand. Supportive therapy is administered and highly active antiretroviral therapy is not usually discontinued.

nonnucleoside reverse transcriptase inhibitor One of the classes of drugs used in combination as highly active antiretroviral therapy to treat HIV disease. Examples in this class include efavirenz and nevirapine, which are both highly potent medications.

nucleoside reverse transcriptase inhibitor One of the first classes of drugs used to treat HIV. Examples in this class include zidovudine and stavudine.

protease inhibitor Introduced in 1995, drugs in this class allowed for the first time the use of effective combination therapy against HIV. Examples include indinavir and nelfinavir. Most recently, there has been widespread use of ritonavir-boosted regimens (lower dose ritonavir in conjunction with another protease inhibitor), which exploit the fact that ritonavir is a potent inhibitor of the P450 enzyme pathway.

Gastrointestinal manifestations are common in the human immunodeficiency virus (HIV)-positive population, occurring in up to 93% of patients in studies prior to the widespread use of highly active antiretroviral therapy (HAART). Despite the advances in the therapeutic arsenal, gastrointestinal complaints are still a major source of morbidity and mortality in this population. The use of HAART has changed the spectrum of gastrointestinal involvement in HIV disease, treating or preventing the occurrence of opportunistic infections, but at the same time, contributing to a significant cause of drug-related toxicity. Furthermore, as the life expectancy of individuals living longer with autoimmune deficiency syndrome (AIDS) is increased because of HAART, other gastrointestinal-specific causes of morbidity, such as chronic hepatitis C virus, have become increasingly important.

APPROACH TO THE HIV-POSITIVE PATIENT WITH A GASTROINTESTINAL COMPLAINT

An approach to an HIV-positive patient with gastrointestinal (GI) symptoms classically depends on the degree of immunosuppression. Symptoms are varied and include dysphagia and odynophagia, nausea, vomiting, abdominal pain, GI bleeding, diarrhea, and anorectal tenderness. Signs such as jaundice and hepatomegaly are common. Because symptoms can be protean and nonspecific, the decision to pursue a

 Copyright 2004, Elsevier (USA). All rights reserved.

definitive etiology must be carefully assessed by both provider and patient, based on the impact of the symptoms on quality of life and on the invasiveness of the recommended procedure. If a patient is on HAART, it is important to carefully obtain a drug history. Many GI symptoms in individuals on HAART can be ascribed to the medicines being taken.

ESOPHAGEAL DISEASE

Although a wide range of infectious and noninfectious etiologies are possible, *Candida albicans*, found in two-thirds of patients with AIDS, causes the most disease. Patients with esophageal candidiasis usually complain of dysphagia and odynophagia. This is often a clinical diagnosis, particularly if there is oral thrush present and CD4+ cell count is $<200/\mu l$. Endoscopy provides the definitive diagnosis and is usually pursued only if symptoms persist despite at least a week of empiric therapy. Multiple yellow–white, friable plaques are visualized. Treatment is fluconazole (100–400 mg/day). Fluconazole can be used as maintenance therapy, but only if there are frequent or severe recurrences. Resistance may develop, however.

Cytomegalovirus (CMV), herpes simplex virus (HSV), and aphthous ulcers can also cause esophageal symptoms. Compared to esophageal candidiasis, patients with CMV, HSV, or aphthous ulcers typically have more odynophagia and less dysphagia. Fever is common with CMV. Extensive, deep ulcerations on endoscopy suggest these diagnoses, but biopsies are required. CMV is treated with ganciclovir (5 mg/kg, intravenously twice/day), foscarnet, or cidofovir for 2–3 weeks. Relapses may necessitate retreatment followed by maintenance therapy. Esophageal ulcers due to HSV are treated with acyclovir for 2–3 weeks and the more common aphthous ulcers can be treated with prednisone over 4 weeks or with thalidomide, observing careful precautions to avoid use in patients who are or might become pregnant.

Other less common infectious etiologies include *Mycobacterium avium*, tuberculosis, cryptosporidia, *Pneumocystis carinii*, histoplasmosis, and primary HIV infection. Neoplasms such as Kaposi's sarcoma and lymphoma may present with esophageal manifestations. Drug-induced dysphagia may be due to use of zidovudine (AZT) or didanosine (ddI).

GASTRIC DISEASE AND ABDOMINAL PAIN

Symptomatic gastric infections are rare in patients with AIDS and usually present as part of systemic and disseminated infections. CMV is the most common opportunistic infection in the stomach and is sometimes associated with gastrointestinal bleeding. Abdominal pain is a common complaint in HIV-positive patients, and can involve a variety of organs. A similar approach can be taken as for a patient without AIDS. Many of these specific etiologies are covered elsewhere in this article.

INTESTINAL DISEASE AND DIARRHEA

Diarrhea was reported to occur in as many as 90% of patients with AIDS prior to the introduction of HAART. With effective HIV medicines, diarrhea is still common but the etiologic agents responsible now include a high proportion of adverse drug effects and infections that are not HIV specific. The differential diagnosis is large but can be narrowed based on acuity, clinical presentation, and CD4+ lymphocyte count. As with HIV-negative patients, a careful history must first be obtained to exclude lactose or other food intolerance. Many of the causes of acute diarrhea mimic those seen in nonimmunocompromised patients. These include *Salmonella*, *Shigella*, *Campylobacter jejuni*, and *Clostridium difficile*. However, in patients with AIDS, enteric bacteria such as *Salmonella* are often associated with bacteremia and are more virulent. In up to 30% of HIV-infected individuals, enteric viruses (adenoviruses, picobirnaviruses, astroviruses, and caliciviruses) may predominate. Although most of these will resolve with only supportive therapy, up to one-third of patients will progress to chronic diarrhea.

The spectrum of organisms that predominate in chronic diarrhea is different. Protozoa such as cryptosporidia, microsporidia, *Cyclospora*, and *Isospora* are commonly identified, because many are unresponsive to medical therapy and cause chronic diarrhea, particularly in patients with CD4+ cell counts $<200/\mu l$. Cryptosporidia can cause clinically significant disease and associated weight loss. Large-volume, watery diarrhea occurs. Diagnosis is made by demonstrating oocysts similar in size to erythrocytes in an acid-fast stain of the stool. Therapeutic options only have marginal efficacy and include paromomycin and octreotide. The best response has been seen with HAART. Microsporidia are common in some series of AIDS patients with chronic diarrhea and have a similar presentation to cryptosporidia. These organisms are extremely small and electron microscopy is often needed for diagnosis. Albendazole can be effective for *Enterocytozoon intestinalis* but not *Enterocytozoon bieneusi*, and HAART is probably the best current therapeutic option. *Cyclospora* and *Isospora belli* are less common protozoa,

probably because both are responsive to trimethoprim–sulfamethoxazole, which is used widely for *Pneumocystis carinii* prophylaxis in the same at-risk populations. The protozoa *Giardia lamblia* and *Entamoeba histolytica* can also cause diarrhea in nonimmunocompromised hosts, but probably do not occur more frequently in patients with AIDS.

CMV is the most common viral cause of chronic diarrhea in patients with AIDS, particularly in those with CD4+ cell counts $<50/\mu l$. The most common presentation is chronic watery diarrhea in association with abdominal cramps. Diagnosis is typically made by the identification of intranuclear inclusion bodies with associated inflammation on biopsy. Treatment can be effective in about 75% of cases and options include ganciclovir or foscarnet intravenously.

Mycobacteria such as *M. avium* have been demonstrated to be a common cause of chronic diarrhea in individuals with CD4+ cell counts $<50/\mu l$. Patients can present with diffuse abdominal pain, watery diarrhea, and wasting. Diagnosis can be made by blood cultures or by acid-fast organisms in biopsy specimens. Treatment is chronic with clarithromycin and ethambutol. Rifampin is sometimes added. HAART can increase the CD4+ cell count and allows the discontinuation of therapy if immune function is regained. *Mycobacterium tuberculosis* presents less commonly with chronic diarrhea, although other extrapulmonary manifestations are common in AIDS patients.

In patients with no identifiable pathogen (up to 30% in some series), enteric HIV infection is thought to cause small bowel malabsorption via mucosal atrophy. Symptomatic treatments with bulking and antidiarrheal agents have had variable success, and HAART has been associated with some improvement. In patients with persistent large-volume diarrhea without an identifiable pathogen, Kaposi's sarcoma and lymphoma must be ruled out.

HEPATOBILIARY DISEASE

Patients with HIV disease commonly have liver disease that can be specific to AIDS or associated with therapy used to treat it. The clinical spectrum of disease is large, from isolated liver test abnormalities to liver failure, and varies by degree of immunosuppression. For organizational purposes, liver parenchymal disease and biliary disease can be considered separately.

Drug-induced liver disease, infection, and neoplasm are the principal etiologies of hepatic parenchymal disease. With the use of multiple potent antiretrovirals and antibiotics, medication-related hepatotoxicity has become quite common. Usual implicated agents include sulfonamides, zidovudine, nevirapine, and many in the protease inhibitor class, including ritonavir. HAART-related immune reconstitution is sometimes associated with an exacerbation of underlying hepatitis. Hepatic steatosis in conjunction with lactic acidosis has been observed in association with nucleoside reverse transcriptase inhibitors such as stavudine and zidovudine. Pancreatitis, myopathy, and peripheral neuropathy may coexist. Providers must be vigilant and discontinue medicines if necessary, because this syndrome has been linked to several deaths.

Hepatic infections continue to be a significant cause of morbidity and mortality in AIDS patients. Hepatitis B, D, and C viruses interact with HIV in distinct ways. Coinfection with hepatitis B virus and HIV has been associated with increased hepatitis B DNA polymerase, reappearance of hepatitis B surface antigen (HBsAg), loss of anti-HBs, and increased prevalence of hepatitis B e antigen expression. Despite an increase in the chronic carrier state, there is an attenuation of histologic and chemical hepatitis in most of these coinfected patients with more advanced immunosuppression. The institution of HAART in hepatitis B virus chronic carriers may be disastrous; immune activation can lead to hepatitis flares and even fulminant liver failure. Hepatitis D virus and hepatitis B virus act similarly in the HIV-positive patient. Unlike hepatitis B and D viruses, infection with hepatitis C virus appears to worsen with increased immune suppression in HIV-positive patients. The prevalence of cirrhosis and hepatic failure is higher, and there is a more accelerated course. The effect of HAART on hepatitis C virus disease is variable, but because patients live longer, there may be an increased risk of complications such as hepatocellular carcinoma. The role of hepatitis C virus combination therapy with interferon α and ribavirin in HIV-coinfected patients is being carefully evaluated in randomized controlled trials, and initial reports look promising. *Mycobacterium avium*-related hepatic disease is a common diagnosis in individuals with CD4+ cell counts $<50/\mu l$. Extrapulmonary tuberculosis is common in HIV-positive patients and hepatic tuberculosis can occur at all CD4+ cell counts. Although CMV inclusion bodies in the liver are commonly found in autopsy studies, CMV hepatitis is rarely diagnosed clinically. *Bartonella* spp., *P. carinii*, histoplasmosis, *Cryptococcus*, and *Candida* are rare causes of hepatic disease in patients with AIDS. Neoplasms such as Kaposi's sarcoma and non-Hodgkin's lymphoma (NHL) can have significant hepatic involvement in patients infected with HIV.

Prognosis of NHL is related to the degree of immunosuppression.

Infectious cholangitis and neoplasm are AIDS-specific causes of biliary disease. AIDS cholangiopathy is pathologically similar to sclerosing cholangitis with papillary stenosis, and is clinically indistinguishable, with right upper quadrant abdominal pain and an alkaline phosphatase elevated out of proportion to bilirubin, aspartate transaminase (AST), and alanine transaminase (ALT). This entity is thought to have an infectious etiology because *Cryptosporidium*, *Microsporidium*, and CMV have been isolated in biliary epithelium and bile. Ultrasound or CT may be diagnostic, and sphincterotomy may provide symptomatic relief. Kaposi's sarcoma, primary bile duct lymphoma, and lymph node obstruction may also manifest as biliary disease in patients with AIDS.

RECTAL AND ANAL DISEASE

Patients with AIDS have a high prevalence of anorectal disease, including anal fistulas, perirectal abscesses, and idiopathic ulcers. Proctitis can be caused by a host of infectious agents, including *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, and herpes simplex. CMV can cause ulcers. Human papillomavirus is a common sexually transmitted infection that causes a high prevalence of anal intraepithelial neoplasia in HIV-positive men and women. There is a corresponding high prevalence of anal cancer in certain populations at risk, such as HIV-positive men who have sex with men. Other neoplasms such as lymphoma and Kaposi's sarcoma have anorectal manifestations. Although HAART has resulted in a decrease in lymphoma and Kaposi's sarcoma, there has been no reduction in the incidence of anal cancer or its precursors in HIV-positive men and women.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Hepatic Manifestations of • Bacterial Toxins • *Campylobacter* • Candidiasis • *Cryptosporidium* • Cytomegalovirus • Diarrhea • Dysphagia • Lymphomas • Mycobacterial Infection • *Salmonella* • *Shigella*

Further Reading

Chin-Hong, P. V., and Palefsky, J. M. (2002). Natural history and clinical management of anal human papillomavirus disease in men and women infected with human immunodeficiency virus. *Clin. Infect. Dis.* **35**, 1127–1134.

HIV InSite (2003). AIDS Knowledge Base sponsored by the University of California, San Francisco. Web site retrieved at http://hivinsite.ucsf.edu.

Koch, J., and Owen, R. L. (1998). Small intestine pathogens in AIDS: Conventional and opportunistic. *Gastrointest. Endosc. Clin. North Am.* **8**, 869–888.

Mokrzycki, M. H., Harris, C., May, H., *et al.* (2000). Lactic acidosis associated with stavudine administration. A report of five cases. *Clin. Infect. Dis.* **30**, 198.

Smith, P. D., and Janoff, E. N. (2002). Gastrointestinal infections in HIV-1 disease. *In* "Infections of the Gastrointestinal Tract" (M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant, eds.), 2nd Ed., pp. 415–443. Lippincott Williams & Wilkins, Philadelphia.

Sulkowski, M. S., Thomas, D. L., Chaisson, R. E., and Moore, R. O. (2000). Hepatotoxicity associated with antiretroviral therapy in adults infected with human immunodeficiency virus and the role of hepatitis C or B virus infection. *JAMA* **283**, 74.

Wilcox, C. M. (2002). Gastrointestinal consequences of infection with human immunodeficiency virus. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., pp. 487–506. W. B. Saunders, Philadelphia.

AIDS, Hepatic Manifestations of

JOSH LEVITSKY*, VANI J. KONDA*, SHANIKA P. SAMARASINGHE*, AND THOMAS W. FAUST†
*University of Chicago Hospitals and †University of Pennsylvania Health System

B symptoms Fever, weight loss, and night sweats; associated with worse prognosis in patients with lymphoma.
highly active anti-retroviral therapy Combination of different anti-human immunodeficiency virus (HIV) drugs that have markedly increased the prognosis for patients with HIV.
silver stain Special stain used to detect fungal organisms not seen by regular microscopy.
YMDD mutation A mutation in the hepatitis B virus polymerase gene in which methionine is replaced by serine; associated with lamivudine therapy and drug resistance.

Liver disease in patients with human immunodeficiency virus and acquired immunodeficiency syndrome (AIDS) is a common, often complex, problem. The clinical presentations of liver disease in this population are highly variable, ranging from asymptomatic liver enzyme abnormalities to fulminant hepatic failure. Infections, medication toxicity, and malignancies are the most common causes of liver disease in patients with AIDS and will be the main focus of discussion in this article.

INTRODUCTION

The hepatic manifestations of acquired immunodeficiency syndrome (AIDS) often correlate with the degree of immunosuppression, success of medical therapy, and socioeconomic status of the human immunodeficiency virus (HIV)-infected individual. Patients with normal or near-normal immune function are typically susceptible to liver diseases that affect the immunocompetent population, with the exception of hepatic toxicity from highly active anti-retroviral therapy. Hepatitis B virus (HBV)- or hepatitis C virus (HCV)-related liver diseases are leading causes of morbidity and mortality in HIV patients, even in well-controlled HIV. In patients with AIDS and severe immunosuppression, the liver is often affected by opportunistic infections and neoplasms, such as Kaposi's sarcoma and lymphoma. This article will examine the general approach to liver disease in an HIV-positive patient and then specifically discuss the most common causes of liver disease in this population.

APPROACH TO THE PATIENT

When liver disease is suspected in an HIV-positive patient, a rational and focused diagnostic approach should be utilized to narrow the broad differential diagnosis (Table I). The most important initial step is a complete history and physical to search for possible risk factors or evidence of hepatic disease. The clinician should inquire about intravenous drug abuse and blood transfusions, use of hepatotoxic medications, recent travel, exposure to opportunistic infections or sick acquaintances, sexual practices, and alcohol use. Symptoms that may reveal clues to the presence or cause of liver disease include fever, weight loss, night sweats, jaundice, right upper quadrant abdominal pain, bilirubinuria, acholic stools, and skin lesions. Physical examination should focus on searching for signs of liver disease, such as jaundice, hepatomegaly, spider angiomata, gynecomastia, testicular atrophy, and proximal muscle wasting.

If liver disease is suspected, a complete hepatic function panel, complete blood count with platelets, coagulation profile, and CD4 count should be obtained. The CD4 count is probably the most important test in formulating a differential diagnosis. A patient with abnormal liver tests and a normal CD4 count is unlikely to have an opportunistic infection of the liver. However, the clinician should initially focus on ruling out opportunistic infections of the liver in patients with CD4 counts of less than 200. Other chronic liver diseases that affect the healthy population should also be considered.

The pattern of liver test abnormalities is often helpful in determining the cause of liver disease in HIV-positive patients. Elevation in transaminases out of proportion to alkaline phosphatase suggests hepatocellular injury. Chronic viral hepatitis, alcohol, medications, immunologic, and metabolic diseases should be considered. Elevations in alkaline phosphatase, direct bilirubin, and γ-glutamyl transpeptidase suggest cholestatic liver disease, commonly caused by granulomatous infiltration or biliary obstruction. If a diagnosis cannot be made with laboratory testing, the next step should be

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Etiologies of Hepatic Disease in HIV Infection

Mycobacterial infections
Mycobacterium avium complex
Mycobacterium tuberculosis
Fungal infections
Cryptococcus neoformans
Histoplasma capsulatum
Coccidioides immitis
Candida albicans
Pneumocystis carinii
Aspergillus fumigatus
Protozoal infections
Cryptosporidia
Leishmania donovani
Encephalitozon cuniculi
Schistosoma ssp.
Toxoplasma gondii
Bacillary peliosis hepatitis
Viral infections
Cytomegalovirus
Herpes simplex virus
Epstein–Barr virus
Adenovirus
Varicella–zoster virus
Hepatitis A–E
AIDS cholangiopathy
Neoplasms
Kaposi's sarcoma
Non-Hodgkin's lymphoma
Hodgkin's lymphoma
Hepatocellular carcinoma
Leiomyoma
Leiomyosarcoma
Cholangiocarcinoma
Hepatotoxic drugs
Anti-retrovirals
Antimicrobials
Anticonvulsants
Anti-emetics
Ethanol
Herbal medications
Miscellaneous
HIV infection
Malnutrition
Total parenteral nutrition
Amyloid

an ultrasound and/or a computerized tomography (CT) scan of the abdomen. CT is superior to ultrasound for delineation of hepatic lesions, lymphadenopathy, cirrhosis, and splenomegaly, whereas ultrasound is more sensitive for gallbladder and biliary tract disease. Since neither test is highly specific or sensitive, liver biopsy may be required to determine the etiology or evaluate the extent and severity of liver disease. Although a specific diagnosis can be made in the majority of patients, liver biopsy should be performed only if other diagnostic measures, including cultures and less invasive procedures such as skin or bone marrow biopsy, have failed to determine an etiology.

ETIOLOGIES AND TREATMENT

Opportunistic Infections

Mycobacterium avium complex (MAC) is the most common opportunistic pathogen of the liver and can be seen in 20–55% of autopsies of patients with AIDS. Patients usually present with fever, weight loss, and diarrhea. The CD4 count is almost always <100 in these patients. Initial workup often reveals an elevated alkaline phosphatase and liver biopsy may show poorly formed noncaseating granulomas; however, with advanced immunosuppression, tissue reaction and granulomas are often absent. Cultures or special stains are required for diagnosis in patients with suspected MAC and inconclusive histopathology. Treatment requires a 6- to 12-month course of a multiple drug regimen, including either clarithromycin or azithromycin and either two or three of the following: rifampin, ethambutol, or ciprofloxicin.

Mycobacterium tuberculosis (MTB) infection presents with fever, weight loss, and cough. Unlike MAC, patients may develop pulmonary or extrapulmonary MTB with any CD4 count. Liver biopsy may show granulomas and occasional hepatic abscesses. Similar to MAC, patients with low CD4 counts may not have classic histopathology on liver biopsy. Only culture can differentiate between MAC and MTB. Patients are placed on a multidrug regimen of isoniazid, rifampin, ethambutol, and pyrazinamide for 9 to 12 months, unless there is evidence of drug resistance.

Fungal infections such as *Cryptococcus neoformans*, *Histoplasma capsulatum*, and *Pneumocystis carinii* (PCP) may affect the liver. Cryptococcal liver involvement is usually asymptomatic. Organisms may be seen on biopsy with silver stain as small round glassy structures. The treatment of *Cryptococcus* is intravenous (iv) amphotericin B $\pm$ oral flucytosine. Disseminated histoplasmosis may cause severe constitutional symptoms and systemic disease even though hepatic involvement is usually asymptomatic. Detection of the polysaccharide antigen in the blood is used for diagnosis, and liver biopsy with silver stain reveals round organisms with occasional budding, mild inflammation, and rarely granulomas. The treatment for histoplasmosis is iv amphotericin or itraconazole.

Disseminated PCP is associated with moderate elevations in transaminases and alkaline phosphatase.

Foamy periportal or diffuse nodules containing cysts on silver stain are typical of hepatic involvement and are usually detected only at autopsy. Disseminated PCP is found primarily in patients who receive aerosolized pentamidine. Treatment is usually with trimethoprim-sulfamethoxazole ± prednisone. Other less common hepatic fungal infections include *Coccidioides immitis*, *Candida albicans*, and *Aspergillus fumigatus*. Protozoal infections are also rare in the liver and include schistosomiasis, *Leishmania donovani*, *Encephalitozon cuniculi*, and *Toxoplasma gondii*.

Bacillary peliosis hepatitis has been associated with *Bartonella quintana* or *Bartonella henselae*. Peliosis lesions are blood-filled cavities distributed throughout the liver parenchyma. Gross inspection reveals gray or hemorrhagic nodules, whereas granulation tissue intermixed with neutrophils and eosinophilic material is seen on histopathology. Any macrolide antibiotic or doxycyline should be given for at least 4 months.

Cytomegalovirus (CMV) of the liver is present in 33–44% of AIDS autopsy cases, usually in patients with widespread organ involvement. Most patients who have hepatic CMV infection are asymptomatic and rarely present with fever and hepatomegaly. Transaminases are often only mildly elevated. CMV infection can be detected via polymerase chain reaction (PCR), antigen assays, and blood cultures; definitive CMV involvement of the liver is determined via biopsy. Microabscesses and CMV inclusions in hepatocytes, Kupffer cells, bile duct epithelium, and endothelial cells are characteristic but not always present. Immunohistochemical techniques and culture of tissue specimens increase the sensitivity of diagnosis. CMV can also cause mass lesions, granulomatous disease, or biliary tract disease and obstruction. Treatment options include ganciclovir, foscarnet, valganciclovir, and cidofovir. Other opportunistic viral infections of the liver in AIDS include herpes simplex virus, Epstein–Barr virus, adenovirus, and varicella-zoster virus.

CMV and other infections, such as cryptosporidia, may cause inflammation and fibrosis of the biliary tract known as AIDS cholangiopathy. Patients present with fever, jaundice, and right upper quadrant pain. The alkaline phosphatase is typically twofold higher than transaminases. Significant obstruction is associated with hyperbilirubinemia. Papillary stenosis, long extrahepatic strictures, or intrahepatic and extrahepatic sclerosing lesions resembling primary sclerosing cholangitis are seen with cholangiography. Endoscopic sphincterotomy is the treatment of choice for papillary stenosis and stenting or balloon dilation has been used for dominant strictures.

Other Hepatotrophic Viruses

With advances in antiretroviral therapy and improved life expectancy in HIV patients, chronic HCV infection has become a leading cause of morbidity and death in this population. Because of similar risk factors for transmission, it is estimated that 30–50% of patients with HIV are co-infected with HCV. Several studies suggest that HIV increases HCV replication and accelerates the progression to cirrhosis. All HIV patients should be screened for HCV with antibody testing and confirmed with either the recombinant immunoblot assay or HCV-RNA PCR. HCV in patients with advanced immunosuppression can often be diagnosed only by measurement of RNA levels. HCV infection is usually asymptomatic and liver enzymes are often normal early in the course of disease. Liver biopsy establishes the degree of necroinflammatory activity and fibrosis and should be performed if treatment is considered. In general, treatment of HIV with anti-retroviral therapy is initiated first, especially if the CD4 count is low. For patients with stable HIV disease, combination therapy with pegylated interferon and ribavirin can be administered, although safety and efficacy data in this population have been reported only in small series.

Co-infection with HIV and HBV increases the rate of HBV replication and risk of a chronic carrier state; however, the inflammatory response in the liver is generally milder in HIV–HBV co-infected patients than in patients infected with HBV alone. Reactivation of quiescent HBV may occur with progression of HIV and can lead to fulminant liver failure in patients with co-infection. Indications for treatment include evidence of ongoing viral replication, elevated liver enzymes, and histological confirmation of active or progressive disease. Lamivudine is often used as a first-line agent as it has activity against both HIV and HBV. HBV resistance in the form of the YMDD mutation occurs in patients treated with lamivudine and may respond to another nucleoside analogue, such as entecavir or adefovir.

Patients with HIV exposed to hepatitis A virus do not have more severe disease than immunocompetent patients. Hepatitis D virus requires concurrent infection with HBV and can cause severe liver disease in immunosuppressed patients. Hepatitis E virus, a fecal–orally transmitted virus found almost exclusively in poor countries, usually causes an acute self-limited disease.

Neoplasms

Kaposi's sarcoma (KS) and non-Hodgkin's lymphoma (NHL) are the most common hepatic neoplasms in HIV patients. KS is a multicentric neoplasm derived from lymphatic endothelial cells and is predominantly

seen in homosexual men. Hepatic KS is found in 8.6% of patients with AIDS but is rarely diagnosed prior to the development of cutaneous disease. Patients are usually asymptomatic, but may occasionally present with hepatomegaly and an elevated alkaline phosphatase. KS lesions are often too small to delineate by ultrasonography, but may enhance with iv contrast on CT imaging. Gross pathology reveals multifocal purple-brown soft nodules composed of vascular endothelial cells forming slit-like spaces, extravasated red blood cells, and central necrosis. Since hepatic KS does not appear to shorten life expectancy, only symptomatic patients require treatment with chemotherapy (liposomal anthracyclines, paclitaxel, or vinorelbine).

NHL is a known complication of HIV infection. The majority are high- or intermediate-grade B-cell tumors and extranodal presentation in common. Hepatic masses are often asymptomatic, but may cause hepatomegaly and jaundice due to hepatic infiltration or biliary tract compression. B symptoms are nonspecific in AIDS patients. CT typically shows hypodense masses with rim enhancement surrounded by normal hepatic parenchyma. Biopsy may reveal lymphomatous infiltration of the portal tracts or lobules. Diffuse portal infiltrates may be seen in infiltrating lymphomas. The median survival for patients with AIDS and lymphoma is under 6 months despite chemotherapy. Other hepatic neoplasms found in HIV patients include Hodgkin's lymphoma, hepatocellular carcinoma, smooth muscle tumors, and cholangiocarcinoma. Hodgkin's lymphoma involving the liver is uncommon in HIV patients and is almost exclusively of the mixed cellularity or nodular sclerosis subtype. Hepatocellular carcinoma in HBV or HCV co-infected individuals occurs at a younger age and is more aggressive than in patients without HIV infection. Cholangiocarcinoma and hepatic leiomyoma or leiomyosarcoma are less commonly described.

Drugs

Patients with AIDS may develop a spectrum of liver problems with HIV medications, from mild liver test abnormalities to fulminant hepatic failure. Nucleoside reverse transcriptase inhibitors, such as zidovudine, have been implicated in a rare but potentially fatal mitochondrial hepatopathy. Patients usually present with nonspecific symptoms, but have severe microvesicular steatosis on liver biopsy. Mortality due to refractory lactic acidosis or hepatic failure can approach 60%. Early discontinuation of nucleoside reverse transcriptase inhibitors may improve clinical outcome, but there are reported cases of progressive acidosis despite their discontinuation. Nutritional therapy is the mainstay of treatment, although prognosis is poor. Nevirapine, a common nonnucleoside reverse transcriptase inhibitor, can cause a hypersensitivity hepatitis. Ritonavir, a protease inhibitor, is associated with toxicity via inhibition of the cytochrome P450 system. Pentamindine and sulfa drugs are often associated with hepatic granulomas and elevations of alkaline phophatase. Ethanol, antifungals, antivirals, herbal medications, and nonprescription drugs may also contribute to hepatic disease.

Miscellaneous

Nonspecific histopathologic findings are described in nearly one-third of biopsy specimens in patients with HIV. Possible causes include malnutrition with or without use of total parenteral nutrition, amyloid deposition, and HIV infection itself. HIV has been found within hepatocytes and Kupffer cells; however, the direct relationship between HIV and liver disease has not been clearly delineated.

CONCLUSION

Patients with HIV and/or AIDS can have a spectrum of liver diseases, from isolated abnormal liver enzymes to fulminant hepatic failure. Instead of considering all of the possible causes, the clinician should narrow the differential diagnosis based on the history, physical exam, pattern of abnormal liver tests, and degree of immunosuppression. The least invasive approach to diagnosis is usually the most appropriate initially. The most common causes of liver disease in HIV patients are opportunistic infections, neoplasms, and medications. Treatment should be based upon the most likely diagnosis.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • Cytomegalovirus • Fungal Infections • Hepatitis A • Hepatitis B • Hepatitis C • Hepatitis D • Hepatitis E • Hepatocellular Carcinoma • Kaposi's Sarcoma • Lymphomas • Mycobacterial Infection

Further Reading

Bonacini, M. (1992). Hepatobiliary complications in patients with human immunodeficiency virus infection. *Am. J. Med.* **92**, 404–411.

Bruno, R., Sacchi, P., Puoti, M., *et al.* (2002). HCV chronic hepatitis in patients with HIV: Clinical management issues. *Am. J. Gastroenterol.* **97**, 1598–1606.

Dodig, M., and Tavil, A. S. (2001). Hepatitis C and human immunodeficiency virus coinfections. *J. Clin. Gastroenterol.* **33**, 367–374.

Herndier, B. G., and Friedman, S. L. (1992). Neoplasms of the gastrointestinal tract and hepatobiliary system in acquired immunodeficiency syndrome. *Semin. Liver Dis.* **12**, 128–141.

Poles, M. A., Lew, E. A., and Dieterich, D. T. (1997). HIV infection and the gastrointestinal tract. *Gastroenterol. Clin.* **26**, 291–321.

Stenzel, M. S. (2000). The management of the clinical complications of antiretroviral therapy. *Infect. Dis. Clin. North Am.* **14**, 851–878.

Wilcox, C. M. (1996). AGA technical review: Malnutrition and cachexia, chronic diarrhea, and hepatobiliary disease in patients with human immunodeficiency virus infection. *Gastroenterology* **111**, 1724–1752.

Alagille Syndrome

BINITA M. KAMATH AND DAVID A. PICCOLI
The Children's Hospital of Philadelphia

Alagille syndrome A disease caused by mutations in the Jagged1 gene, resulting in variable manifestations in the liver, heart, spine, eyes, kidneys, face, and other systems.

cholestasis Decrease in bile flow, typically associated with increased bilirubin, increased bile salts, or both. Elevations typically result in the visible signs of jaundice.

embryotoxon, posterior A finding in the anterior chamber of the eye, sometimes visible as a discrete line, due to a prominent Schwalbe's ring, seen in Alagille syndrome, in some other genetic disorders, and in approximately 10% of normal individuals.

hepatitis, neonatal Term used to define a large and diverse group of structural and metabolic disorders unique to the neonatal period, resulting in hepatic disease and cholestasis.

hepatomegaly Increase in the size of the liver. In children, normal size varies with age.

hyperbilirubinemia Increase in the level of bilirubin (unconjugated, indirect hyperbilirubinemia) or conjugates of bilirubin (conjugated, direct hyperbilirubinemia).

Jagged1 mutation A defect in a gene encoding a primary ligand for the NOTCH signaling pathway, resulting in Alagille syndrome.

jaundice Eevation of bilirubin, typically conjugated bilirubin, to the point that it is visible in the eyes or skin as a yellowish pigmentation.

paucity, bile duct Decrease in the number of bile ducts present in a portal tract, defined as a ratio of bile ducts to portal tracts averaging less than 0.9.

splenomegaly Increase in size of the spleen. In children, normal size varies with age.

xanthoma, Alagille syndrome Nodular or plaque-like lipid deposits in the skin that occur commonly in Alagille syndrome when the cholesterol level exceeds 500 mg/dl.

Alagille syndrome (AGS) is a complex dominantly inherited multisystem disorder involving predominantly the liver, heart, eyes, face, and skeleton. The main clinical manifestations of AGS are cholestasis, characterized by bile duct paucity on liver biopsy; congenital cardiac defects, primarily involving the pulmonary arteries; posterior embryotoxon in the eye; typical facial dysmorphism; and butterfly vertebrae. Renal and central nervous abnormalities have also been described. A classic feature of the syndrome is highly variable expressivity of the clinical features, even within families. *Jagged1* (*JAG1*), a ligand in the evolutionary conserved Notch signaling pathway, has been identified as the AGS gene.

DIAGNOSIS OF ALAGILLE SYNDROME

Alagille originally defined the syndrome by bile duct paucity in association with at least three of five major criteria: cholestasis, characteristic facies, vertebral abnormalities, ocular anomalies, and a heart murmur. Since that time, a wide range of manifestations in many organ systems have been associated with Alagille syndrome (AGS). Renal disease, pancreatic disease, and intracranial vascular events are recognized to be significant manifestations of AGS.

Bile duct paucity on liver biopsy has been considered to the most important and constant feature of AGS. This paucity, however, is not present in infancy in many patients ultimately shown to have AGS. Overall, paucity is present in approximately 89% of patients. Several studies of serial liver biopsies have demonstrated that

 Copyright 2004, Elsevier (USA). All rights reserved.

Herndier, B. G., and Friedman, S. L. (1992). Neoplasms of the gastrointestinal tract and hepatobiliary system in acquired immunodeficiency syndrome. *Semin. Liver Dis.* **12**, 128–141.

Poles, M. A., Lew, E. A., and Dieterich, D. T. (1997). HIV infection and the gastrointestinal tract. *Gastroenterol. Clin.* **26**, 291–321.

Stenzel, M. S. (2000). The management of the clinical complications of antiretroviral therapy. *Infect. Dis. Clin. North Am.* **14**, 851–878.

Wilcox, C. M. (1996). AGA technical review: Malnutrition and cachexia, chronic diarrhea, and hepatobiliary disease in patients with human immunodeficiency virus infection. *Gastroenterology* **111**, 1724–1752.

Alagille Syndrome

BINITA M. KAMATH AND DAVID A. PICCOLI
The Children's Hospital of Philadelphia

Alagille syndrome A disease caused by mutations in the Jagged1 gene, resulting in variable manifestations in the liver, heart, spine, eyes, kidneys, face, and other systems.

cholestasis Decrease in bile flow, typically associated with increased bilirubin, increased bile salts, or both. Elevations typically result in the visible signs of jaundice.

embryotoxon, posterior A finding in the anterior chamber of the eye, sometimes visible as a discrete line, due to a prominent Schwalbe's ring, seen in Alagille syndrome, in some other genetic disorders, and in approximately 10% of normal individuals.

hepatitis, neonatal Term used to define a large and diverse group of structural and metabolic disorders unique to the neonatal period, resulting in hepatic disease and cholestasis.

hepatomegaly Increase in the size of the liver. In children, normal size varies with age.

hyperbilirubinemia Increase in the level of bilirubin (unconjugated, indirect hyperbilirubinemia) or conjugates of bilirubin (conjugated, direct hyperbilirubinemia).

Jagged1 mutation A defect in a gene encoding a primary ligand for the NOTCH signaling pathway, resulting in Alagille syndrome.

jaundice Eevation of bilirubin, typically conjugated bilirubin, to the point that it is visible in the eyes or skin as a yellowish pigmentation.

paucity, bile duct Decrease in the number of bile ducts present in a portal tract, defined as a ratio of bile ducts to portal tracts averaging less than 0.9.

splenomegaly Increase in size of the spleen. In children, normal size varies with age.

xanthoma, Alagille syndrome Nodular or plaque-like lipid deposits in the skin that occur commonly in Alagille syndrome when the cholesterol level exceeds 500 mg/dl.

Alagille syndrome (AGS) is a complex dominantly inherited multisystem disorder involving predominantly the liver, heart, eyes, face, and skeleton. The main clinical manifestations of AGS are cholestasis, characterized by bile duct paucity on liver biopsy; congenital cardiac defects, primarily involving the pulmonary arteries; posterior embryotoxon in the eye; typical facial dysmorphism; and butterfly vertebrae. Renal and central nervous abnormalities have also been described. A classic feature of the syndrome is highly variable expressivity of the clinical features, even within families. *Jagged1* (*JAG1*), a ligand in the evolutionary conserved Notch signaling pathway, has been identified as the AGS gene.

DIAGNOSIS OF ALAGILLE SYNDROME

Alagille originally defined the syndrome by bile duct paucity in association with at least three of five major criteria: cholestasis, characteristic facies, vertebral abnormalities, ocular anomalies, and a heart murmur. Since that time, a wide range of manifestations in many organ systems have been associated with Alagille syndrome (AGS). Renal disease, pancreatic disease, and intracranial vascular events are recognized to be significant manifestations of AGS.

Bile duct paucity on liver biopsy has been considered to the most important and constant feature of AGS. This paucity, however, is not present in infancy in many patients ultimately shown to have AGS. Overall, paucity is present in approximately 89% of patients. Several studies of serial liver biopsies have demonstrated that

 Copyright 2004, Elsevier (USA). All rights reserved.

paucity is more common later in infancy and childhood. Ductular proliferation is present in a small number of infants with AGS, leading to potential diagnostic confusion and misdiagnoses of biliary atresia.

Diagnosis of the syndrome is also hindered by highly variable expressivity of the clinical manifestations. Large series of patients reported by different groups have demonstrated differing frequencies of the manifestations of AGS. Diagnosis can also be difficult because several of the characteristic features are seen in normal individuals.

The identification of the disease gene in AGS and the availability of molecular testing have assisted in the confirmation of the diagnosis in affected individuals. However, it has also provided a tool for identifying individuals with few or only one feature of AGS who carry a mutation in *Jagged1* (*JAG1*). It remains to be clarified whether these individuals who carry mutations in *JAG1* but are without overt clinical disease should be classified as having AGS or merely as carriers of a gene defect with minimal expression. Since the transmission risk for progeny is high, the designation of a diagnosis of AGS seems appropriate.

Based on these considerations, the diagnostic criteria for AGS can be modified significantly. A revised list of diagnostic criteria is proposed in Table I.

CLINICAL FEATURES AND COMPLICATIONS OF ALAGILLE SYNDROME

Hepatic Manifestations

The hepatic manifestations of AGS vary from mild to severe cholestasis. The majority of symptomatic patients present in the first year of life. Hepatitis is present in many infants, but is generally less important than the cholestasis. Synthetic liver failure is extremely uncommon under the age of 1 year. Hepatomegaly is recognized in 93–100% of AGS patients and is common in infancy. Splenomegaly is unusual early in the course of the disease, but is eventually found in up to 70% of patients. Jaundice is present in the majority of symptomatic patients and typically presents as a conjugated hyperbilirubinemia in the neonatal period. In half of these infants, it is persistent, resolving only in later childhood. The magnitude of the hyperbilirubinemia is typically less than the degree of cholestasis and pruritus. The pruritus is among the most severe of any chronic liver disease. It rarely is present before 3 to 5 months of age, but is seen in most children by the third year of life, even in some who are anicteric.

Multiple xanthomas are common sequelae of severe cholestasis. The formation of xanthomas correlates with serum cholesterol greater than 500 mg/dl. They typically form on the extensor surfaces of the fingers, the palmar creases, nape of the neck, the ears, the popliteal fossa, the buttocks, and around the inguinal creases. These xanthomas increase in number over the first few years of life and then may disappear subsequently as cholestasis improves.

The most striking laboratory abnormalities are in the measures of cholestasis and bile duct damage. Elevations of bilirubin up to 30 times normal and bile salt elevations of 100 times normal are not uncommon. Bile salt elevations are common, even when the bilirubin is normal. Markers of bile duct damage, including γ-glutamyl transferase and alkaline phosphatase, are usually markedly elevated. Likewise, other substances typically excreted in bile are increased in blood. Cholesterol levels may exceed 1–2000 mg/dl. The aminotransferases are typically elevated 3- to 10-fold, but may be normal in some patients with cholestasis. Hepatic synthetic function is usually well preserved.

Liver transplantation is eventually necessary in 21–50% of patients. Indications for transplantation

TABLE I Revised Diagnostic Criteria for the Diagnosis of Alagille Syndrome

AGS family history	Paucity	*JAG1* defect	Criteria needed
None (proband)	Present	Not identified	3 or more features[a]
None (proband)	Absent	Not identified	4 or more features
None (proband)	Absent	Identified	1 or more features
Present	Present	Not identified	1 or more features
Present	Unknown	Not identified	1 or more features
Present	Absent	Identified	Any or no features

Note. Major clinical criteria include consistent (1) cardiac, (2) ocular disease, (3) butterfly vertebrae, (4) characteristic "Alagille" facies, or (5) renal disease.

[a] A number of index cases with two criteria or even 1 criterion will ultimately be shown to have AGS by molecular testing, but two criteria should be considered insufficient to establish the diagnosis in a proband.

include synthetic dysfunction, intractable portal hypertension, bone fractures, pruritus, xanthomata, and growth failure. The survival of AGS patients undergoing liver transplantation has been comparable to other pediatric patients undergoing liver transplantation; however, survival may be limited in patients with significant cardiac or renal complications of AGS.

Cardiac Manifestations

The presence of a heart murmur is the most common manifestation of AGS. The majority of these murmurs are due to stenosis at some level in the pulmonary outflow tract or peripheral pulmonic vessels. Peripheral pulmonic stenosis may occur in isolation or in combination with structural intracardiac disease. The most common congenital defect is Tetralogy of Fallot, which occurs in 7–11% of patients. Other cardiovascular lesions include truncus arteriosus, ventricular septal defect complex, atrial septal defect, ventricular septal defect, and isolated pulmonic stenosis.

Cardiac disease accounts for nearly all the early mortality in AGS. Patients with intracardiac disease have approximately 40% survival to 6 years of life compared to 95% survival in AGS patients without intracardiac lesions. The operative mortality for cardiac surgery in AGS patients is greater than for those without AGS. Cardiovascular disease has also been implicated in the increased posttransplantation mortality seen in some series.

Ocular Anomalies

A large and varied number of ocular abnormalities have been described in AGS. A few of the abnormalities are secondary to chronic vitamin deficiencies. Of the primary ocular abnormalities, posterior embryotoxon is the most important diagnostically. Posterior embryotoxon is a prominent, centrally positioned Schwalbe's ring (or line), at the point where the corneal endothelium and the uveal trabecular meshwork join. Posterior embryotoxon occurs in 56–95% of patients with AGS, but also occurs in 8–15% of normal eyes. Axenfeld anomaly, seen in 13% of AGS patients, is a prominent Schwalbe's ring with attached iris strands and is associated with glaucoma. Numerous other ocular anomalies have also been described, most of which have no functional significance.

Facial Features

The characteristic facial features of AGS include a prominent forehead, deep-set eyes with moderate hypertelorism, a pointed chin, and a saddle or straight nose with a bulbous tip. The combination of these features gives the face a triangular appearance (see Fig. 1). Other facial characteristics include a flat appearance of the face in profile and prominent ears. Additional abnormalities include large ears, recurrent sinusitis, recurrent otitis, and a high-pitched voice. The facies of an adult with AGS do not resemble the childhood features; the forehead becomes less prominent and the chin is more protuberant, making the face less triangular in appearance.

It has been suggested that the facies are due to the effects of chronic cholestasis, but they are not seen in other diseases with severe cholestasis. It is likely that

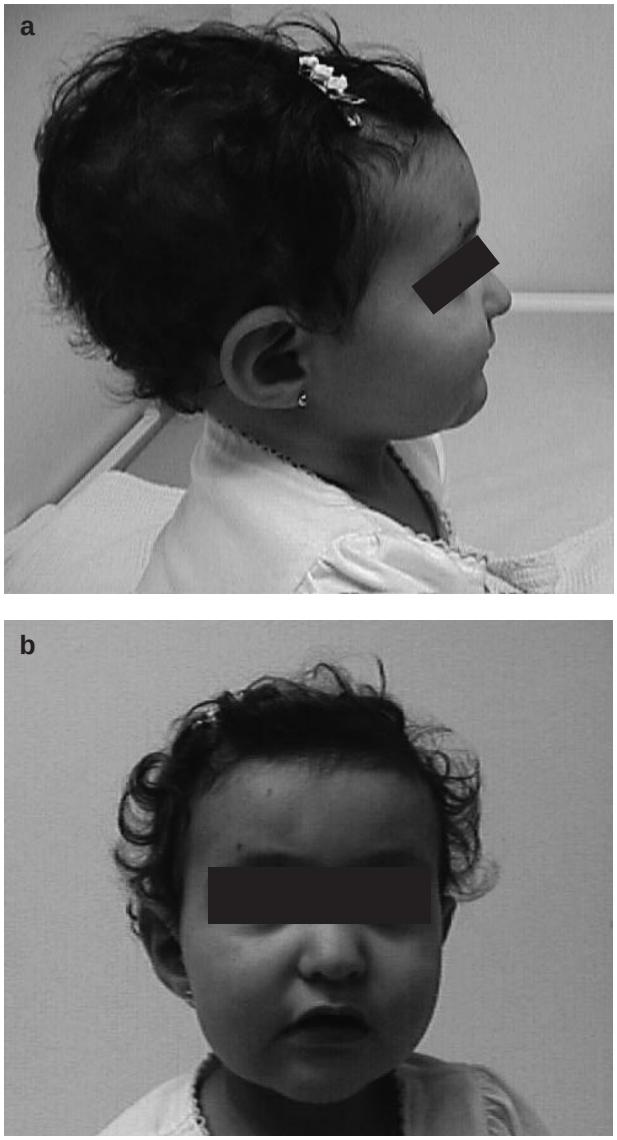


FIGURE 1 (a and b) Typical facies seen in a 4-year-old girl with Alagille syndrome.

the facial bone formation is directly affected by mutations in *JAG1*.

Skeletal Manifestations

The most characteristic skeletal finding in AGS is the sagittal cleft or butterfly vertebrae, which is found in 33–87% of patients (see Fig. 2). This anomaly may occur in normal individuals. The affected vertebral bodies are split sagittally into paired hemivertebrae, due to a failure of the fusion of the anterior arches of the vertebrae. Generally, these are asymptomatic and of no structural significance.

Other associated skeletal abnormalities include, among others, an abnormal narrowing of the adjusted interpedicular space in the lumbar spine, fusion of the adjacent vertebrae, hemivertebrae, and the presence of a bony connection between ribs. The fingers may seem short, with broad thumbs. Supernumerary digital flexion creases are also seen.

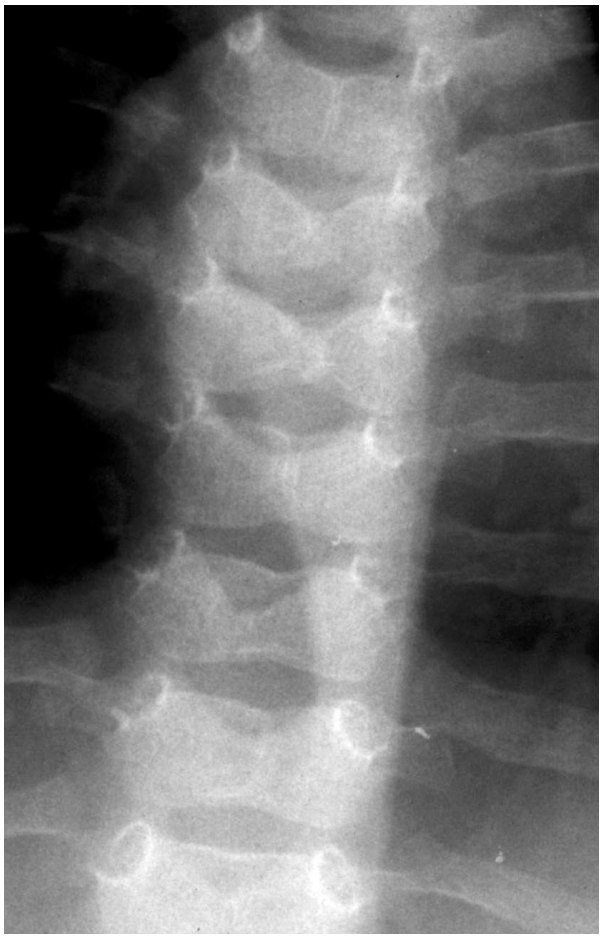


FIGURE 2 Butterfly vertebrae with variable degrees of clefting in the thoracic spine of an infant with Alagille syndrome.

Severe metabolic bone disease with osteoporosis and pathologic fractures is common in AGS. A number of factors may contribute to osteopenia and fractures, including severe chronic malnutrition, vitamin D and K deficiency, chronic hepatic and renal disease, magnesium deficiency, and pancreatic insufficiency.

Central Nervous System Manifestations

Intracranial bleeding is the most important neurologic complication of AGS and a significant cause of morbidity and mortality. It occurs in approximately 15% of patients and in 30–50% of these events the hemorrhage is fatal. The intracranial bleeding varies significantly in location and severity. The majority of this bleeding occurs in the absence of significant coagulopathy. Head trauma, typically of a minor degree, has been associated with the bleeding in a number of cases. The majority of cases of bleeding are spontaneous, however, with no clear risk factors. Structural vascular abnormalities have been identified in only some of the patients with bleeding episodes.

Renal Manifestations

Renal disease is an important feature of AGS. Structural and functional renal abnormalities occur in 40–50% of patients with AGS. These include solitary kidney, ectopic kidney, bifid pelvis, reduplicated ureters, and multicystic and dysplastic kidneys. Renal artery stenosis is a cause of systemic hypertension in AGS. Tubulointerstitial nephropathy, renal tubular acidosis, a characteristic "lipidosis" of the glomeruli, and adult-onset renal insufficiency may also occur.

Additional Manifestations

Severe growth retardation is common in patients with AGS, particularly in the first 4 years of life. Malnutrition due to malabsorption is a major factor in this failure to thrive. There appear to be limitations in linear growth even when protein-calorie malnutrition is not evident.

Intrinsic pancreatic disease may also occur in AGS. Clinically, the identification of pancreatic insufficiency is important, as therapy with enzyme supplementation is available.

MORBIDITY, MORTALITY, AND OUTCOME IN ALAGILLE SYNDROME

Diseases of the cardiac, hepatic, and central nervous systems account for the majority of morbidity and mortality in AGS. The presence of complex intracardiac

disease at diagnosis is the only predictor of excessive early mortality. Hepatic complications account for most of the later mortality, although recent series document significant mortality from intracranial bleeding. The 20-year survival has been estimated to be 75% overall.

THERAPY FOR ALAGILLE SYNDROME

Patients with AGS present significant management challenges. Cholestasis is commonly profound. Bile flow may be stimulated with the choleretic ursodeoxycholic acid, but in many patients the pruritus continues unabated. Care should be taken to keep the skin hydrated with emollients and fingernails should be trimmed. Therapy with antihistamines may provide some relief, but many patients require additional therapy with agents such as rifampin or naltrexone. Biliary diversion has been successful in a limited number of patients, but intractable pruritus continues to be an indication for transplantation in refractory patients.

Malnutrition and growth failure should be treated with aggressive nutritional therapy. There will be significant malabsorption of long-chain fat and therefore formulas supplemented with medium-chain triglycerides have some nutritional advantage. Fat-soluble vitamin deficiency is present to a variable degree in most patients. Multivitamin preparations may not provide the correct ratio of fat-soluble vitamins and thus vitamins are best administered as individual supplements.

GENETICS OF ALAGILLE SYNDROME

AGS is inherited in autosomal dominant fashion with highly variable expressivity ranging from subclinical features to severely affected individuals. It is one of the most common genetic causes of cholestasis in infancy, with an estimated frequency of 1 in 70,000 live births.

The site of the gene responsible for AGS was first suggested by the identification of cytogenetically visible deletions on the long arm of chromosome 20 in multiple AGS patients. In 1997, mutations in *JAG1* were shown to be the cause of AGS. Molecular analysis has demonstrated *JAG1* mutations in approximately 70% of patients.

JAG1 is a cell surface protein that functions as a ligand for the Notch transmembrane receptors, which are part of the evolutionarily conserved Notch signaling pathway. The Notch pathway regulates cell fate determination in many different cell types throughout development. Approximately 50% of AGS patients have protein truncating (frameshift or nonsense) mutations. The mutations are distributed widely over the entire coding region. Fifty to 70% of mutations are *de novo*. Mutations in *JAG1* are thought to cause disease by haploinsufficiency whereby there is a decrease in the amount of the normal protein.

Although the AGS phenotype is highly variable, there is no apparent genotype–phenotype correlation. No phenotypic differences have been identified based on type or location of mutation. Furthermore, the extreme variability of the AGS phenotype within families suggests that other genetic or environmental factors contribute significantly to the clinical manifestations of the disease.

Genetic evaluation for *JAG1* mutations is currently available on a research basis. Current methods cannot identify a mutation in approximately 30% of clearly affected probands. Prenatal testing is available, but only if the mutation is identified in the proband. Molecular testing has also aided in the diagnosis of AGS for patients with minor or atypical manifestations and has expanded the spectrum of AGS manifestations.

The identification of *JAG1* as the cause of AGS has opened a window on human development, which will surely lead to an enhanced understanding of embryogenesis and multisystem diseases of humans.

See Also the Following Articles

Biliary Tract, Developmental Anomalies of the • Liver Transplantation • Malabsorption • Malnutrition • Neonatal Cholestasis and Biliary Atresia • Neonatal Hyperbilirubinemia

Further Reading

Emerick, K. M., Rand, E. B., Goldmuntz, E., *et al.* (1999). Features of Alagille syndrome in 92 patients: Frequency and relation to prognosis. *Hepatology* **29**, 822–829.

Li, L., Krantz, I. D., Deng, Y., *et al.* (1997). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat. Genet.* **16**, 243–251.

Piccoli, D. A. (2001). Alagille syndrome. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), 2nd Ed., pp. 327–342. Lippincott, Williams & Wilkins, Philadelphia, PA.

are typical primary physical signs of cirrhosis. Secondary phenomena include portal hypertension with splenomegaly, edema, and ascites; encephalopathy; and gastrointestinal hemorrhage from bursting esophageal or gastric varices, and bleeding tendencies due to clotting factor deficiencies resulting from hepatocyte malfunction. Tertiary complications include spontaneous peritonitis caused by anaerobic bacteria.

Typically, blood tests reveal a decrease in serum albumin with increased and abnormal globulins. Serum transaminases and bilirubin are usually only mildly elevated, with more striking increases in the end stages. As previously discussed, the outcome is often dismal, with a 1-year mortality of 50% or more in patients over the age of 60 years.

See Also the Following Articles

Alcohol Metabolism • Ascites • Cirrhosis • Hepatitis C • Hepatocellular Carcinoma • Hepatocytes

Further Reading

Chedid, A., Mendenhall, C. L., Gartside, P., French, S. W., Chen, T., Rabin, L., and the VA Cooperative Group (1991). Prognostic factors in alcoholic liver disease. *Am. J. Gastroenterol.* **82**, 210–216.

Lieber, C. S. (ed.) (1992). "Medical and Nutritional Complications of Alcoholism: Mechanisms and Management." Plenum Press, New York.

Lieber, C. S. (2001). Alcohol and hepatitis C. *Alcohol Res. Health* **25**, 245–254.

Worner, T. M., and Lieber, C. S. (1985). Perivenular fibrosis as precursor lesion of cirrhosis. *JAMA* **254**, 627–630.

Alcohol Metabolism

CHARLES S. LIEBER

Bronx Veterans Affairs Medical Center and Mount Sinai School of Medicine, New York

cytosol The fluid component of cytoplasm, excluding organelles.

endoplasmic reticulum The membrane network in cytoplasm that is composed of tubules or cisternae. Some membranes carry ribosomes on their surfaces (rough endoplasmic reticulum) and others are smooth.

free radicals (1) An uncharged atom or group of atoms having at least one unpaired electron, which makes it highly reactive. (2) An organic compound having some unpaired valence electrons; a normal by-product of oxidation reactions in metabolism.

microsomes Small particles in the cytoplasm of a cell, typically consisting of fragmented endoplasmic reticulum to which ribosomes are attached.

mitochondrion A spherical or elongated organelle in the cytoplasm of the cell, containing genetic material and many enzymes important for cell metabolism, including those responsible for the conversion of food to usable energy. It consists of two membranes: an outer smooth membrane and an inner membrane arranged to form cristae.

peroxisomes Cell organelles containing enzymes, such as catalase and oxidase, that catalyze the production and breakdown of hydrogen peroxide.

Ethanol (also called alcohol or ethyl alcohol) is the most commonly abused psychoactive drug, yet it is legal in most countries. Its metabolism is the cause of a vast array of pathologic manifestations.

INTRODUCTION

Ethanol is readily absorbed from the gastrointestinal tract but only 2 to 10% is eliminated through the kidneys and lungs; the rest is metabolized in the body. Many tissues contain enzymes capable of ethanol oxidation or nonoxidative metabolism, but significant activity occurs only in the liver (Fig. 1) and, to a lesser extent, in the stomach. Accordingly, medical consequences are

Copyright 2004, Elsevier (USA). All rights reserved.

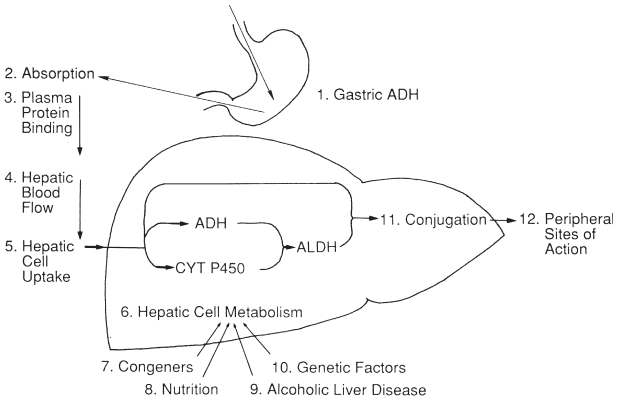


FIGURE 1 Schematic illustration of the main pathways of ethanol disposition in stomach and liver. Metabolic and drug interactions may affect conjugation, microsomal cytochrome P450-dependent pathways (CYT P450), alcohol dehydrogenase (ADH), and acetaldehyde dehydrogenase (ALDH). Reprinted from Lieber (1997), © Lippincott Williams & Wilkins, with permission.

predominant in these organs and thus these two sites of ethanol metabolism are the focus of this article.

HEPATIC METABOLISM OF ETHANOL AND ITS CONSEQUENCES

The hepatocyte contains three main pathways for ethanol metabolism, each located in a different subcellular compartment: (1) the alcohol dehydrogenase (ADH) pathway of the cytosol (the soluble fraction of the cell); (2) the microsomal ethanol-oxidizing system, located in the endoplasmic reticulum; and (3) catalase, located in the peroxisomes. Each of these pathways produces specific metabolic or toxic disturbances and all three generate acetaldehyde, a highly toxic metabolite.

The ADH Pathway

This is the main pathway of alcohol metabolism.

Multiple Forms of ADH

Human ADH is a zinc metalloenzyme with five classes of molecular forms that arise from the association of eight different types of subunits (α, $\beta 1$, $\beta 2$, $\beta 3$, $\gamma 1$, $\gamma 2$, π, and χ) into active dimeric molecules. A genetic model accounts for this multiplicity as products of seven gene loci, ADH1 through ADH7.

Metabolic Disorders Associated with the ADH Pathway

In ADH-mediated oxidation of alcohol, hydrogen is transferred from the substrate to the cofactor nicotinamide adenine dinucleotide (NAD), converting it to its reduced form (NADH), and acetaldehyde is produced (Fig. 1). Thus, the first step in this oxidation of alcohol generates an excess of reducing equivalents in the cytosol, primarily as NADH, with a marked shift in the redox potential of the cytosol. The altered redox state, in turn, is responsible for a variety of metabolic abnormalities, including hyperlactacidemia, which contributes to the acidosis and also reduces the capacity of the kidneys to excrete uric acid, leading to secondary hyperuricemia. Hyperuricemia explains the common clinical observation that excessive consumption of alcoholic beverages frequently aggravates or precipitates gouty attacks.

Short-term alcohol intoxication occasionally causes severe hypoglycemia, which can result in sudden death. Hypoglycemia is due, in part, to the block of hepatic gluconeogenesis by ethanol, again as a consequence of the increased NADH/NAD ratio in subjects whose glycogen stores are already depleted by starvation or who have preexisting abnormalities in carbohydrate metabolism.

The reducing equivalents can also be transferred to NADPH, and the increased NADPH can be utilized for synthetic pathways in the cytosol and in the microsomes. Some of the hydrogen equivalents are transferred from the cytosol into mitochondria. The mitochondrial membrane is impermeable to NADH and the reducing equivalents are thought to enter the mitochondria via shuttle mechanisms such as the malate cycle (quantitatively, probably the most important), the fatty acid elongation cycle, and the α-glycerophosphate cycle. Normally, fatty acids are oxidized via β-oxidation in the citric acid cycle of the mitochondria, which serves as a "hydrogen donor" for the mitochondrial electron transport chain. When alcohol is oxidized, the generated hydrogen equivalents, which are shuttled into mitochondria, supplant the citric acid cycle as a source of hydrogen. The activity of the citric acid cycle is actually depressed, partly because of a slowing of the reactions of the cycle that depend on the NAD/NADH ratio. Consequently, the mitochondria will use the hydrogen equivalents originating from ethanol rather than those derived from the oxidation of fatty acids that normally serve as the main energy source of the liver. This substitution of alcohol for fat favors hepatic fat accumulation.

Microsomal Ethanol-Oxidizing System (MEOS)

Characterization of the MEOS and Its Role in Ethanol Metabolism

Although recognized only three decades ago, this new pathway is still the subject of extensive research,

reviewed in detail elsewhere. The first indication of an interaction of ethanol with the microsomal fraction of the hepatocyte was provided by the morphologic observation that alcohol feeding results in a proliferation of the smooth endoplasmic reticulum (SER), both in animals and in human. This increase in SER resembles that seen after the administration of a wide variety of hepatotoxins, therapeutic agents, and some food additives. Since most of the substances that induce a proliferation of the SER are metabolized, at least in part, by the cytochrome P450 enzyme system that is located on the SER, the possibility that alcohol may also be metabolized by similar enzymes was raised. Such a system was indeed demonstrated in liver microsomes *in vitro* and found to be inducible by chronic alcohol feeding *in vivo* and was named the microsomal ethanol-oxidizing system. Its distinct nature was shown by (1) isolation of a P450-containing fraction from liver microsomes that, although devoid of any ADH or catalase activity, could still oxidize ethanol as well as higher aliphatic alcohols (e.g., butanol, which is not a substrate for catalase) and (2) reconstitution of ethanol-oxidizing activity using NADPH−cytochrome P450 reductase, phospholipid, and either partially purified or highly purified microsomal P450. The purified human protein (now called CYP2E1) was obtained in a catalytically active form, with a high turnover rate of ethanol and other specific substrates. MEOS has a relatively high K_m for ethanol (8−10 m*M*, compared to 0.2−2 m*M* for hepatic ADH) but, contrasting with hepatic ADH, which is not inducible in primates as well as most other animal species, enhanced levels of both hepatic CYP2E1 protein and its mRNA were found in actively drinking patients.

The presence of CYP2E1 was also shown in extrahepatic tissues and in nonparenchymal cells of the liver, including Kupffer, but not stellate, cells. In rats, ethanol treatment caused a sevenfold increase in CYP2E1 content of the Kupffer cells.

Increased Xenobiotic Toxicity and Carcinogenicity; Oxidative Stress

Much of the medical significance of the MEOS (and its ethanol-inducible CYP2E1) results not only from the oxidation of ethanol but also from the unusual and unique capacity of CYP2E1 to generate reactive oxygen intermediates, such as superoxide radicals. Indeed, there is increased evidence that ethanol toxicity may be associated with an increased production of reactive oxygen intermediates. Increased generation of oxygen- and ethanol-derived free radicals occurs at the microsomal level, especially through the intervention of the ethanol-inducible CYP2E1. This induction is accompanied by increased oxidation of NADPH with resulting H_2O_2 generation. There is also increased superoxide radical production. This oxidative stress contributes to the lipid peroxidation associated with alcoholic liver injury. Lipid peroxidation correlates with the amount of CYP2E1 in liver microsomal preparations and it can be inhibited by antibodies against CYP2E1.

CYP2E1 also activates many xenobiotic compounds to their toxic metabolites, often free radicals. This pertains, for instance, to carbon tetrachloride and other industrial solvents such as bromobenzene and vinylidene chloride as well as anesthetics such as enflurane and halothane. Ethanol abuse also markedly increases the activity of microsomal benzene-metabolizing enzymes and aggravates the hematopoietic toxicity of benzene. In addition, enhanced metabolism (and toxicity) pertains to a variety of prescribed drugs, including isoniazid and phenylbutazone and some over-the-counter medications such as acetaminophen (paracetamol or Tylenol), all of which are substrates for, or inducers of, CYP2E1. Therapeutic amounts of acetaminophen (2.5 to 4 g per day) can cause hepatic injury in alcoholics. In animals given ethanol for long periods, hepatotoxic effects peaked after withdrawal when ethanol was no longer competing for the microsomal pathway but levels of the toxic metabolites were at their highest. Thus, alcoholics are most vulnerable to the toxic effects of acetaminophen shortly after cessation of chronic drinking. In fact, such patients hospitalized with acetaminophen toxicity related to accidental misuse had higher rates of morbidity and mortality than those who attempted suicide, even though the latter had taken more acetaminophen.

There is an association between alcohol misuse and an increased incidence of upper alimentary and respiratory tract cancers. Many factors have been incriminated, one of which is the effect of ethanol on enzyme systems involving activation of carcinogens by CYP2E1.

Alcoholics are also commonly heavy smokers, and the synergistic effects of alcohol consumption and smoking on cancer development are striking, with a 44-fold increase of esophageal cancer in human.

Most importantly, induction of the MEOS results in interaction with many nutrients, as reviewed elsewhere, and in enhanced acetaldehyde production (Fig. 2), which, in turn, aggravates the oxidative stress directly as well as indirectly by impairing defense systems against it. Acetaldehyde is a major cause of complications (Fig. 2) that contribute to the development of liver injury.

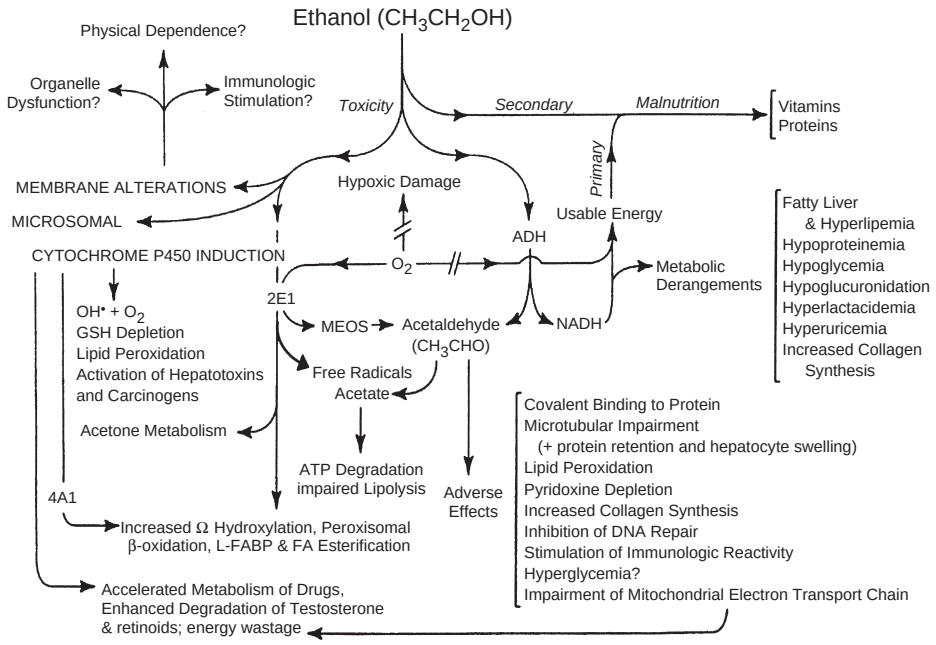


FIGURE 2 Hepatic, nutritional and metabolic abnormalities after ethanol abuse. Malnutrition, whether primary or secondary, has been differentiated from direct toxicity of ethanol. The latter results from redox changes, effects secondary to microsomal induction, the generation of acetaldehyde, direct membrane alterations, and/or hypoxia.

The Catalase Pathway

Catalase is capable of oxidizing alcohol *in vitro* in the presence of an H_2O_2-generating system and its interaction with H_2O_2 in the intact liver was demonstrated. However, its role is limited by the small amount of H_2O_2 generated and, under physiological conditions, catalase appears to play no major role in ethanol oxidation.

The catalase contribution might be enhanced if significant amounts of H_2O_2 become available through β-oxidation of fatty acids in peroxisomes. Indeed, long-term ethanol consumption is associated with increases in the content of a specific cytochrome (CYP4A1) that promotes microsomal ω-hydroxylation of fatty acids, which may compensate, at least in part, for the deficit in fatty acid oxidation due to the ethanol-induced injury of the mitochondria. Products of ω-oxidation also increase liver cytosolic fatty acid-binding protein and peroxisomal β-oxidation, an alternate but modest pathway for fatty acid disposition.

OXIDATION OF ETHANOL IN THE STOMACH: GENDER AND ETHNIC DIFFERENCES

Alcohol was known to disappear from the stomach and this was considered to be part of its "absorption" from the gastrointestinal tract (Fig. 1). It is now apparent that some of this absorbed ethanol is in fact metabolized in the gastric wall. As a result, when alcohol is taken orally, blood levels achieved are generally lower than those obtained after administration of the same dose intravenously, so-called first-pass metabolism (FPM). Indeed, the gastric mucosa contains the same diversity in isozymes of ADH (except for π) as the liver (*vide supra*). However, in addition, it has a class IV ADH (called σ-ADH) that is not present in the liver. This enzyme has now been purified, its full-length cDNA has been obtained, and the complete amino acid sequence has been deduced. Furthermore, the gene (ADH7) was cloned and localized to chromosome 4 by Yokoyama *et al*. in 1996. The upstream structure of the human ADH7 gene and the organ distribution of its expression were also defined. σ-ADH was found to have a high capacity for ethanol oxidation, greater than that of the other isozymes. Its affinity for ethanol is relatively low, with a K_m of approximately 30 m*M*, but this is not a drawback in the stomach, where ethanol is commonly present at much higher concentrations.

In vitro, gastric ADH was found to be responsible for a large fraction of the ethanol metabolism observed in cultured rat and human gastric cells. Thus, FPM reflects, at least in part, gastric metabolism.

The concept of a significant ethanol metabolism in the stomach was also supported indirectly by the observation that commonly used drugs, such as aspirin, and some histamine-2 blockers, which decrease the activity of gastric ADH and/or accelerate gastric emptying, also increase blood alcohol levels *in vivo*. This was particularly apparent after the repeated intake of low alcohol doses, mimicking social drinking. Although questioned at first, such increases in blood levels have now been widely confirmed. The blood level achieved by each single administration of a low dose is small, but social drinking is usually characterized by repetitive consumption of small doses. Under those conditions, the effect of the drug is cumulative, and the increase in blood alcohol becomes sufficient to reach levels known to impair cognitive and fine motor functions.

Some ethnic differences also support the concept of the role of gastric ADH in FPM of ethanol. Indeed, σ-ADH is absent or markedly decreased in activity in a large percentage of Japanese subjects. Their FPM is correspondingly reduced.

Gender differences have also been described: In Caucasians, gastric ADH activity is lower in women than in men, at least below the age of 50 years, a difference mainly due to lower χ-ADH activity in women. This is associated with higher blood alcohol levels, an effect more striking in alcoholic women than in nonalcoholic women because FPM is partly lost in the alcoholic, together with decreased gastric ADH activity. Furthermore, in women, the alcohol consumed is distributed in a 12% smaller water space because of a difference in body composition (more fat and less water). The larger proportion of ethanol that enters the systemic circulation in women than in men may contribute to their greater vulnerability to alcohol, not only in terms of central nervous system manifestations but also for liver disease, especially since women have less effective defense mechanisms against alcoholic liver injury than men.

The magnitude of FPM also depends on the concentration of the alcoholic beverages used. Indeed, gastric ADH isozymes require a relatively high ethanol concentration for optimal activity. Therefore, the concentration of alcoholic beverages affects the amount metabolized, with lesser gastric FPM (and hence more alcohol reaching the blood) after weak beverages (such as beer) than for equivalent amounts of ethanol consumed in strong beverages (such as whiskey). Fasting also strikingly decreases FPM, most likely because of accelerated gastric emptying, resulting in shortened exposure of ethanol to gastric ADH and its more rapid intestinal absorption.

In summary, gastric FPM of ethanol represents a useful barrier against excess penetration of ethanol into the body.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Cirrhosis • Cytochrome P450 • Hepatocytes

Further Reading

Baraona, E., Abittan, C. S., Dohmen, K., Moretti, M., Pozzato, G., Chayes, Z. W., Schaefer, C., and Lieber, C. S. (2001). Gender differences in pharmacokinetics of alcohol. *Alcohol Clin. Exp. Res.* **25**, 502–507.

Baraona, E., Yokoyama, A., Ishii, H., Hernandez-Munoz, R., Takagi, T., Tsuchiya, M., and Lieber, C. S. (1991). Lack of alcohol dehydrogenase isoenzyme activities in the stomach of Japanese subjects. *Life Sci.* **49**, 1929–1934.

Lieber, C. S. (1997). Alcoholic liver disease. *In* "Gastrointestinal Pharmacology and Therapeutics" (G. Friedman, E. D. Jacobson, and R. W. McCallum, eds.), Chap. 36, pp. 465–487, Lippincott–Raven Press, Philadelphia.

Lieber, C. S. (1999). Microsomal ethanol-oxidizing system (MEOS), the first 30 years (1968–1998)—A review. *Alcohol Clin. Exp. Res.* **23**, 991–1007.

Lieber, C. S. (2002). Alcohol: Its metabolism and interaction with nutrients. *In* "Handbook of Nutrition and Food" (C. Berdanier, ed.), Chap. 45, pp. 915–940. CRC Press, Boca Raton, FL.

Lieber, C. S., and DeCarli, L. M. (1968). Ethanol oxidation by hepatic microsomes: Adaptive increase after ethanol feeding. *Science* **162**, 917–918.

Yokoyama, H., Baraona, E., and Lieber, C. S. (1995). Upstream structure of human ADH7 gene and the organ distribution of its expression. *Biochem. Biophys. Res. Commun.* **216**, 216–222.

Yokoyama, H., Baraona, E., and Lieber, C. S. (1996). Molecular cloning and chromosomal localization of ADH7 gene encoding human class IV ADH. *Genomics* **31**, 243–245.

Alcoholic Liver Injury, Hepatic Manifestations of

CHARLES S. LIEBER

Bronx Veterans Affairs Medical Center and Mount Sinai School of Medicine, New York

ascites Accumulation of fluid in the peritoneal cavity.
cholestasis Arrest in the flow of bile.
collagen Fibrous protein constituent of bone, cartilage, tendon, and scar tissue.
edema Accumulation of excessive watery fluid in tissues, or serous cavities.
encephalopathy Any of various diseases of the brain.
endoplasmic reticulum Membrane network in cytoplasm composed of tubules or cisternae. Some membranes carry ribosomes on their surfaces (rough endoplasmic reticulum), whereas others are smooth.
fenestration Opening in the surface of a structure, as in a membrane.
hemochromatosis Hereditary disorder of iron metabolism characterized by excessive accumulation of iron in tissues.
hepatocellular carcinoma Cancer derived from parenchymal cells in the liver.
hepatocyte Main liver cell.
necrosis Death of cells or tissues through injury or disease.
peritonitis Inflammation of the lining of the abdominal cavity.
polymorphonuclear leukocyte White blood cell, usually neutrophilic, having a nucleus that is divided into lobes.
steatosis Accumulation of fat within the cells of an organ, such as the liver, resulting in diminished functioning.
ultrastructural Visible under the electron microscope.

Liver disease in the alcoholic patient has long been believed to be due exclusively to malnutrition, but at present, the additional role of ethanol hepatotoxicity is well established. In the absence of dietary deficiencies, and even in the presence of protein-, vitamin-, and mineral-enriched diets, ethanol produces a fatty liver with striking ultrastructural lesions, both in rats and in human volunteers, and scarring or fibrosis with cirrhosis in nonhuman primates. In humans, there is a clear link between the amount of alcohol consumed and the incidence of cirrhosis, associated with deadly complications.

INTRODUCTION

The clinical course and ultimate outcome of alcoholic liver disease is dismal. In a prospective survey of 280 patients with alcoholic liver injury, Chedid and coworkers found that, within 48 months of followup, 30% of those with a fatty liver, more than half of those with cirrhosis, and two-thirds of those with cirrhosis plus alcoholic hepatitis died. This outcome is more severe than that of many cancers, yet it is attracting much less concern, both among the public and the medical profession. This may be due, at least in part, to a prevailing and pervasive perception that not much can be done about this major public health issue. Pathogenic concepts of alcoholic liver disease, however, are evolving, and elucidation of the biochemical effects of ethanol allows for a more optimistic outlook in terms of diagnosis and treatment.

NATURAL COURSE OF ALCOHOLIC LIVER DISEASE

Fatty Liver

Fat accumulation in the liver cells is the earliest and most common response to alcohol metabolism in the liver. The normal liver weighs about 1.5 kg, whereas the alcoholic fatty liver weighs 2.0–2.5 kg. Macroscopically, on hematoxylin- and eosin-stained sections, parenchymal accumulation of lipid is seen as clear intracytoplasmic vacuoles, or as black or red globules with various Sudan stains. In massive steatosis (Fig. 1), the hepatocytes are uniformly filled by larger fat droplets. The cell nucleus may be eccentrically placed, and when the cell membranes between adjacent hepatocytes rupture, fatty cysts are formed. Increased hepatic lipid accumulation (involving mainly a 3- to 10-fold rise in triglycerides) can be demonstrated by biochemical measurements before it becomes histologically apparent. Evidence of necrosis is usually sparse but sometimes the hepatocytes are surrounded by a mild accumulation of inflammatory cells. When pronounced, these formations are called "lipogranulomas."

Ultrastructural changes reveal enlarged and distorted mitochondria with shortened cristae, sometimes

 Copyright 2004, Elsevier (USA). All rights reserved.

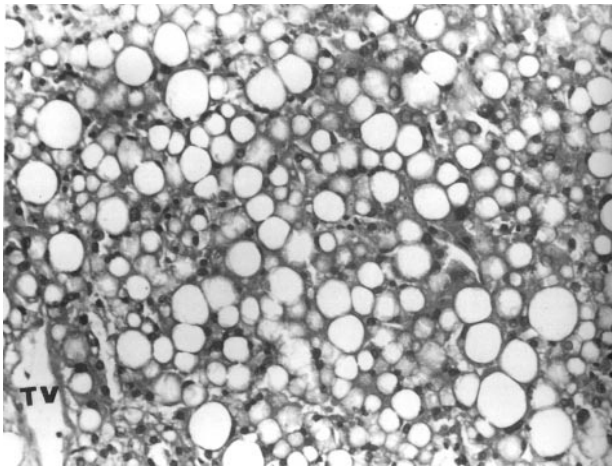


FIGURE 1 Severe fatty liver of an alcoholic. The hepatocytes are uniformly filled by large fat droplets and nuclei are eccentrically placed. TV, Terminal venule. Hematoxylin and eosin staining, ×250. Reproduced with permission from Lieber (1992).

containing crystalline inclusions. The endoplasmic reticulum shows vacuolar dilatation and proliferation. Despite adequate nutrition, alcohol was found to induce these alterations in rats and baboons (Fig. 2) and in alcoholic human volunteers.

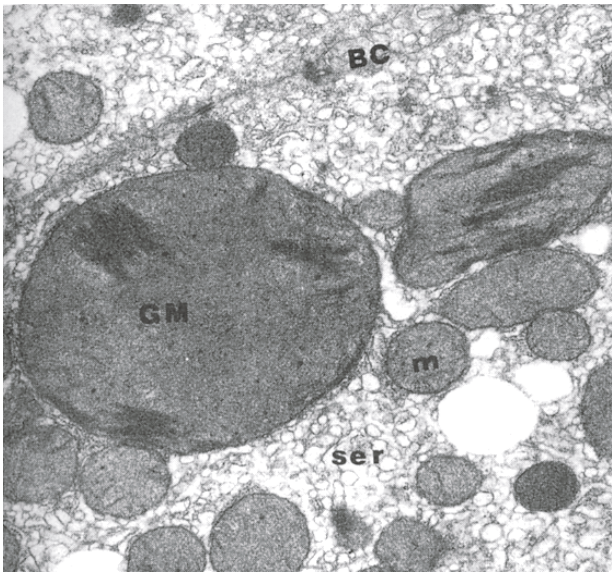


FIGURE 2 Ultrastructural changes in a baboon liver following prolonged ethanol ingestion. Note the prominence of the smooth endoplasmic reticulum (ser). Mitochondrial changes include swelling, shortening of the cristae, and some giant forms (GM). Some normal-sized mitochondria (m) are still present. Disruptions of mitochondrial membranes are common. BC, Bile canaliculus. Uranyl acetate and lead staining, ×9500. Reproduced with permission from Lieber (1992).

The clinical spectrum of alcoholic fatty liver may extend from silent, nonsymptomatic hepatomegaly to severe hepatocellular failure with cholestasis and impaired blood flow through the liver. Most of the patients with simple fatty liver are virtually asymptomatic. Hepatomegaly is the commonest clinical sign, present on physical examination in 75% of patients. In more advanced cases, hepatic tenderness, anorexia, nausea, emesis, jaundice, and even fluid accumulation in the abdomen (ascites) are present. Severe forms of fatty liver may present a clinical picture mimicking extrahepatic obstructive jaundice, especially if associated with dark urine and light-colored stools. Typical abnormalities in laboratory tests are slightly or moderately elevated γ-glutamyl transferase (GGT) and serum (alanine and aspartate) transaminases (AST and ALT). In contrast to more advanced alcoholic liver injury, all the abnormalities of the laboratory tests tend to return to normal within the first days of abstinence.

Perivenular Fibrosis

Although it can occur anywhere in the liver, the earliest deposition of fibrous tissue is generally seen around the central veins and venules—now called terminal hepatic venules. Fibrosis around these veins has been described in alcoholic hepatitis and is often associated with a necrotizing process called sclerosing hyaline necrosis. When intensive, it may obliterate the terminal hepatic veins and lead to postsinusoidal portal hypertension with ascites prior to the development of cirrhosis. It is important to note that perivenular fibrosis can be seen in the absence of, or prior to, widespread inflammation and necrosis, in association with what pathologists would label as "simple" fatty liver.

Once perivenular fibrosis has developed, it indicates that the patient has already entered the fibrotic process and that, on continuation of drinking, he or she will rapidly develop more severe stages, including cirrhosis. Thus, this lesion can be considered as a marker of vulnerability to the development of subsequent cirrhosis and therefore can be used as an indication for active intervention.

Alcoholic Hepatitis

This stage is characterized by the appearance of necrosis with an inflammatory reaction, including polymorphonuclear cells. Although the incidence of fatty liver in alcoholics is very common, alcoholic hepatitis develops in only a fraction of heavy drinkers. Histological characteristics of alcoholic hepatitis are ballooning and a great disarray of hepatocytes (predominantly in perivenular areas) with disseminated infiltration by

polymorphonuclear inflammatory cells and varying degrees of steatosis, necrosis, fibrosis, and cholestasis. Mallory's alcoholic hyaline (irregular cytoplasmic bodies) can be considered as a diagnostic hallmark, but it is not always present. Ultrastructural changes in alcoholic hepatitis are similar but are more severe than the ones seen in fatty liver. Collagen fibers may surround the liver cells and interfere with their exposure to the blood.

The clinical spectrum of alcoholic hepatitis may thus range from mild hepatomegaly to fatal disease with jaundice, ascites, liver coma, gastrointestinal hemorrhage, and with all the complications of severe hepatic insufficiency. In general, there is a correlation between the severity of the histological features of alcoholic hepatitis (hepatocellular necrosis, leukocytic infiltration, frequency of alcoholic hyaline) and the degree of hepatomegaly, leukocytosis, and elevations of serum transaminases and the intensity of symptoms.

Among the laboratory abnormalities, an AST/ALT ratio greater than 2.0 is considered indicative for alcoholic liver injury. However, this ratio may be less helpful in the presence of cirrhosis. The depression of serum albumin and the prolongation of the prothrombin time have been shown to correlate with the severity of the histological lesions. Other common laboratory findings include serum bilirubin elevations (in 90% of patients). It should, however, be remembered that serum bilirubin may also be increased in the absence of alcoholic hepatitis (for instance, because of red cell destruction or pancreatitis, an inflammation of the pancreas).

Cirrhosis

Macroscopically, the early cirrhotic liver is golden yellow with fine uniform nodules (from 1 to 5 mm) on the surface. Traditionally, this type of cirrhosis is called micronodular, or Laennec's, cirrhosis. The size of the liver varies, depending on the degree of fibrosis, inflammation, and steatosis, from a small, shrunken, and hard liver to a large organ weighing up to 4 kg.

Microscopically, scar tissue distorts the normal architecture of the liver by forming bands of connective tissue joining portal and central zones (Fig. 3). At first, nodules are regular in size and shape. In more advanced cirrhosis, some nodules become larger and irregular.

One of the most characteristic features of cirrhosis is the change in the hepatic blood circulation. Inflammation and perivenular fibrosis (*vide supra*) may impair blood flow and the regenerative nodules may also compress the hepatic veins, resulting in localized increased blood pressure. With the progression of the fibrosis, the blood flow is further depressed. In addition, the total blood flow through the cirrhotic liver decreases. This results from extrahepatic shunts bypassing the scarred liver. These shunts further isolate the hepatocytes from their blood supply. As a result, marked deterioration in hepatic functions occurs.

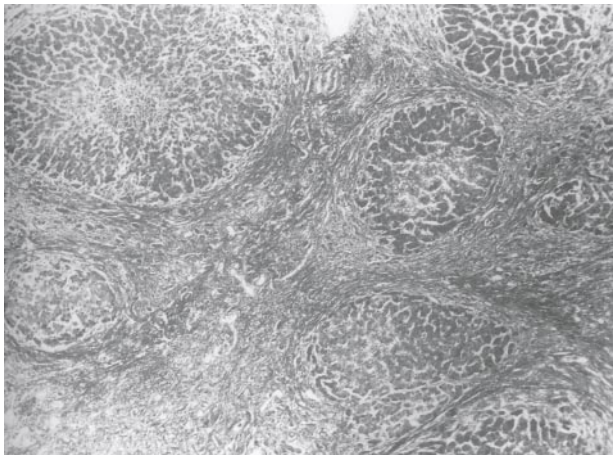


FIGURE 3 Section of liver from an alcoholic patient, showing cirrhosis. The normal liver parenchyma is replaced by nodules surrounded by abundant bands of fibrotic tissue. Hematoxylin and eosin staining, ×50. Reproduced with permission from Lieber (1992).

In addition to fibrosis, fat, inflammatory reactions, cholestasis, and histologically stainable iron deposits are common in cirrhotic livers, especially in alcoholic subjects. But total iron stores are very seldom increased and never exceed 179 mmol (10 g). The differentiation from primary hemochromatosis is possible by measuring the total hepatic iron content from liver biopsies.

Hepatocellular carcinoma (HCC) has become a relatively frequent finding in liver cirrhosis, presumably because of improved management and prolonged survival. Pathogenetically, it has been traditionally related to cirrhosis, but possible carcinogenic properties of alcohol must also be considered, because primary hepatocellular carcinoma can complicate noncirrhotic alcoholic liver disease. Furthermore, at least in some patients, there is a significant interaction between alcohol and viral hepatitis, especially hepatitis C, in hepatocarcinogenesis.

Most patients are diagnosed when they seek treatment for complications. Autopsy studies show that cirrhosis may remain unrecognized antemortem in 40% of the patients, and it is discovered fortuitously in about 20% of patients on routine examination or during the evaluation of some other, unrelated, disease.

Low-grade and continuous fever is also common in decompensated cirrhosis. Jaundice and hepatomegaly

are typical primary physical signs of cirrhosis. Secondary phenomena include portal hypertension with splenomegaly, edema, and ascites; encephalopathy; and gastrointestinal hemorrhage from bursting esophageal or gastric varices, and bleeding tendencies due to clotting factor deficiencies resulting from hepatocyte malfunction. Tertiary complications include spontaneous peritonitis caused by anaerobic bacteria.

Typically, blood tests reveal a decrease in serum albumin with increased and abnormal globulins. Serum transaminases and bilirubin are usually only mildly elevated, with more striking increases in the end stages. As previously discussed, the outcome is often dismal, with a 1-year mortality of 50% or more in patients over the age of 60 years.

See Also the Following Articles

Alcohol Metabolism • Ascites • Cirrhosis • Hepatitis C • Hepatocellular Carcinoma • Hepatocytes

Further Reading

Chedid, A., Mendenhall, C. L., Gartside, P., French, S. W., Chen, T., Rabin, L., and the VA Cooperative Group (1991). Prognostic factors in alcoholic liver disease. *Am. J. Gastroenterol.* **82**, 210–216.

Lieber, C. S. (ed.) (1992). "Medical and Nutritional Complications of Alcoholism: Mechanisms and Management." Plenum Press, New York.

Lieber, C. S. (2001). Alcohol and hepatitis C. *Alcohol Res. Health* **25**, 245–254.

Worner, T. M., and Lieber, C. S. (1985). Perivenular fibrosis as precursor lesion of cirrhosis. *JAMA* **254**, 627–630.

Alcohol Metabolism

CHARLES S. LIEBER

Bronx Veterans Affairs Medical Center and Mount Sinai School of Medicine, New York

cytosol The fluid component of cytoplasm, excluding organelles.

endoplasmic reticulum The membrane network in cytoplasm that is composed of tubules or cisternae. Some membranes carry ribosomes on their surfaces (rough endoplasmic reticulum) and others are smooth.

free radicals (1) An uncharged atom or group of atoms having at least one unpaired electron, which makes it highly reactive. (2) An organic compound having some unpaired valence electrons; a normal by-product of oxidation reactions in metabolism.

microsomes Small particles in the cytoplasm of a cell, typically consisting of fragmented endoplasmic reticulum to which ribosomes are attached.

mitochondrion A spherical or elongated organelle in the cytoplasm of the cell, containing genetic material and many enzymes important for cell metabolism, including those responsible for the conversion of food to usable energy. It consists of two membranes: an outer smooth membrane and an inner membrane arranged to form cristae.

peroxisomes Cell organelles containing enzymes, such as catalase and oxidase, that catalyze the production and breakdown of hydrogen peroxide.

Ethanol (also called alcohol or ethyl alcohol) is the most commonly abused psychoactive drug, yet it is legal in most countries. Its metabolism is the cause of a vast array of pathologic manifestations.

INTRODUCTION

Ethanol is readily absorbed from the gastrointestinal tract but only 2 to 10% is eliminated through the kidneys and lungs; the rest is metabolized in the body. Many tissues contain enzymes capable of ethanol oxidation or nonoxidative metabolism, but significant activity occurs only in the liver (Fig. 1) and, to a lesser extent, in the stomach. Accordingly, medical consequences are

Copyright 2004, Elsevier (USA). All rights reserved.

Alimentary Tract, MRI of the

SUSAN M. ASCHER
Georgetown University Hospital

gadolinium chelate Magnetic resonance intravenous contrast agent that produces T1 shortening of adjacent water molecules, resulting in tissue brightening.

inflammatory bowel disease Group of idiopathic bowel disorders characterized by inflammation. The most common types of inflammatory bowel disease are Crohn's disease and ulcerative colitis.

magnetic resonance imaging Technique that is based on the interaction between an external magnetic field and a nucleus that possesses spin. The patient is exposed to energy at a specific, correct frequency; this energy is absorbed and a short time later is released, at which time it can be detected and processed to yield images.

T1-weighted images Precontrast, provide information on abnormally increased fluid content or fibrous tissue content (e.g., low signal intensity) and information on the presence of subacute blood or concentrated protein (e.g., high signal intensity). The addition of fat suppression allows blood to be distinguished from fat. Post-Gd images increase the conspicuity of disease processes.

T2-weighted images Provide information about the presence of increased fluid in diseased tissue (e.g., high signal intensity) and the presence of chronic fibrotic tissue or iron deposition (e.g., low signal intensity).

Alimentary tract magnetic resonance imaging is an important and well-established tool for colon cancer and inflammatory bowel disease evaluation. Many more applications are emerging as higher gradients and faster imaging techniques allow for noninvasive, real-time evaluation of the gastrointestinal system.

INTRODUCTION

Hardware and software advancements have expanded the role of magnetic resonance imaging (MRI) of the alimentary tract. Breath-hold imaging (T1-weighted fat-suppressed spoiled gradient echo and T2-weighted single-shot fast spin) combined with intravenous gadolinium (Gd) chelates will result in reproducible, high-quality images that arrest bowel motion, decrease susceptibility artifacts, remove the competing high signal intensity of intraabdominal fat, expand the dynamic range of signal intensities, and facilitate distinction of bowel wall from intraluminal contents. The role of oral contrast agents remains controversial and their use is not routine. Established and evolving applications of alimentary tract MRIs include colon carcinoma evaluation (screening, staging, and distinguishing recurrent tumors from radiation therapy changes) and inflammatory bowel disease evaluation (type and severity).

TECHNIQUE

To limit peristalsis, patients should fast 4–6 hours prior to the examination; alternatively, intravenous antiperistaltics (1 mg glucagon) may be given prior to imaging. Imaging is optimized with the use of a torso-phased array coil. A standard protocol involves orthogonal T1-weighted spoiled gradient echo (T1-W SGE) sequences (with and without fat suppression), orthogonal T2-W single-shot fast spin echo sequences (T2-W SSFSE) (Fig. 1), and orthogonal post-Gd (0.1 mmol/kg) fat-suppressed T1-W SGE sequences. High-resolution (512 matrix) T2-W fast spin echo (T2-W FSE) sequences are reserved for imaging the rectum. This is because the rectum is relatively fixed in position and therefore refractory to blurring artifacts associated with bowel motion. The rectum may also be imaged with an endoluminal coil. This coil optimizes spatial resolution and improves conspicuity of the different rectal wall layers, sphincter complex, and disease states. Magnetic resonance (MR) colonography has been gaining momentum and involves distending the large bowel with a 2-liter solution of water spiked (20 ml) with gadolinium chelate (0.5 mol/liter) and obtaining coronal three-dimensional (3D) spoiled gradient echo volumetric slabs and two-dimensional (2D) T2-W SSFSE images. The 3D data sets are then postprocessed using commercially available software (Fig. 2).

COLON CARCINOMA

Screening

Colon cancers predilect the rectosigmoid colon. Tumors may be polypoid, circumferential, or plaquelike. There is ongoing controversy concerning the most appropriate screening program (hemoccult stool testing, conventional colonoscopy, barium enema, and, most

 Copyright 2004, Elsevier (USA). All rights reserved.

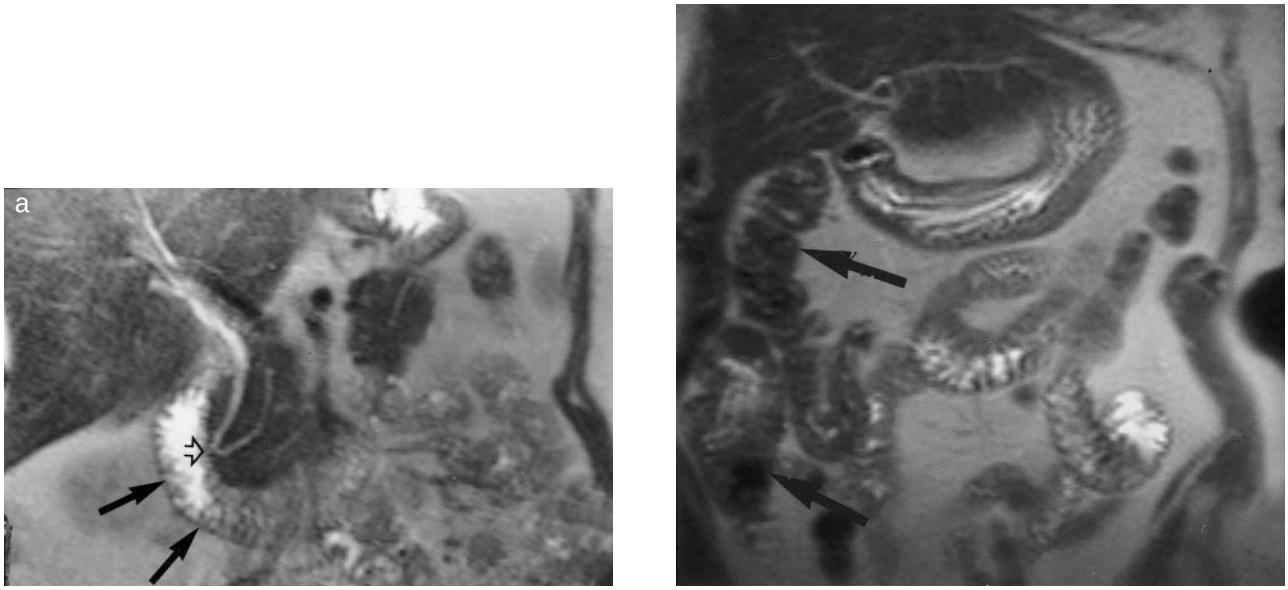


FIGURE 1 Normal bowel. T2-W SSFSE of normal bowel. Coronal (a, b) images highlight normal small (arrows, a) and large (arrow, b) bowel anatomy. The junction of the common bile duct and pancreatic duct at the major papilla (arrowhead, a) is easily visible, as are the valvulae conniventes in the fluid-filled small intestine. From Semelka *et al*. (2002), Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

recently, virtual colonography/colonoscopy). The diagnostic performance of MR colonoscopy is directly related to lesion size. For example, the sensitivity, specificity, positive predictive value, and negative predictive value for detecting polyps 10 mm or greater are 93, 99, 93, and 99%, respectively (Fig. 3). This falls to 75, 92, 72, and 93%, respectively, when using a lesion cutoff value of 7 mm. The performance of MR colonography is expected to improve with optimization of spatial resolution associated with advances in receiver coil technology,

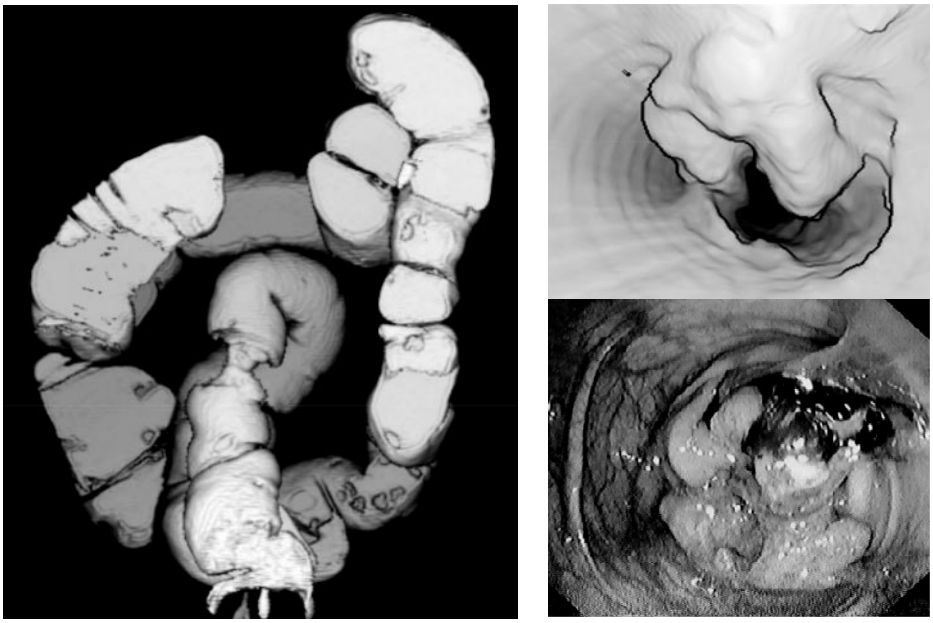


FIGURE 2 MR colonoscopy (sigmoid cancer). The MR colonoscopy images and corresponding endoscopic image show an apple-core lesion in the sigmoid colon. The MR colonoscopy was performed by distending the colon with Gd-doped water. Images courtesy of Dr. Jorg Debatin, Germany.

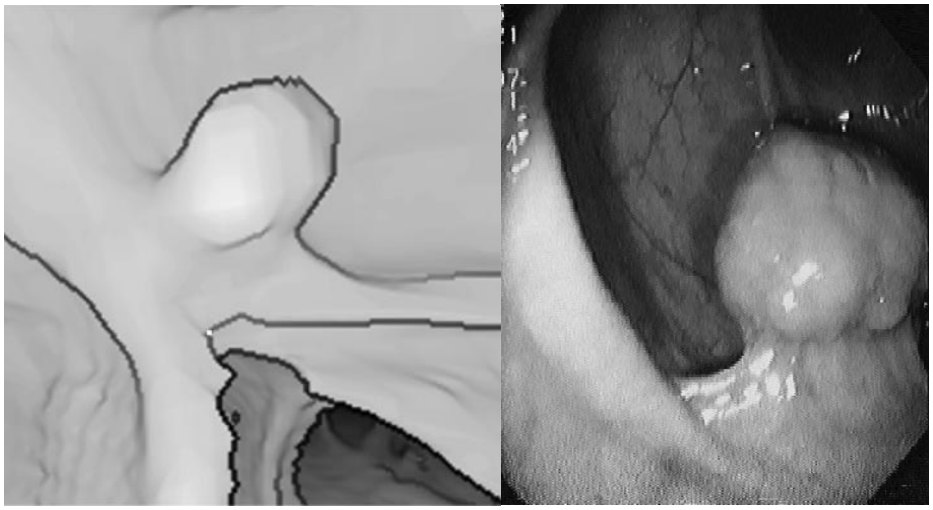


FIGURE 3 MR colonoscopy (12-mm pedunculated polyp). The MR colonoscopy and corresponding endoscopic image demonstrate a polyp. The sensitivity, specificity, and positive and negative predictive values for MR colonoscopy are maximal for polyps equal to or greater than 10 mm. Images courtesy of Dr. Jorg Debatin, Germany.

faster gradients systems, and more efficient pulse sequences.

Preoperative Staging

Once colon cancer is diagnosed, MRI is useful in staging, especially when the tumor involves the rectum. Specifically, for locoregional spread, there is good correlation between MRI techniques and surgical specimens for tumor size, wall involvement, peritumoral invasion, and lymph node involvement. For distant metastases (e.g., liver metastases), a dynamic contrast-enhanced scan of the abdomen is often done at the time of pelvic imaging. Colon cancer is staged according to the TNM classification (T, tumor size assessment; N, degree lymph node involvement; M, degree of metastasis).

Primary colon tumors are often heterogeneous in signal intensity on T2-W images. Also, though many tumors have moderately high signal intensity components, these may be difficult to appreciate on T2-W SSFSE and FSE sequences, where the surrounding fat is also high in signal intensity—though this phenomenon is usually less problematic on standard T2-W FSE sequences. Tumors with a shaggy outer margin and loss of the normal low signal intensity outer bowel muscularis propria suggest lymphovascular invasion. Colon cancers enhance following the administration of contrast. Similarly, peritumoral extension, lymph node metastases, and peritoneal seeding all enhance on fat-suppressed Gd T1-W SGE imaging (Fig. 4). Lymph node metastases are common with alimentary tract malignancies and may be present in "nonpathologically" enlarged nodes; their presence may be inferred, however, when noting greater than five lymph nodes measuring < 10 mm in short axis in a regional distribution related to the primary tumor. Liver metastases are well depicted on MRIs and tend to have irregular borders.

There have been no large randomized clinical trials to compare the staging accuracy of state-of-the-art MR and multidetector computer tomography (CT) for patients with colon cancer. Summarizing older literature, the overall staging accuracy of these two modalities is equivalent, approximately 70%. And though the report from the Radiology Diagnostic Oncology Group (RDOG) II found that CT was more accurate in detection of tumor penetration of the muscularis propria; current MR techniques (e.g., breath-hold imaging and contrast-enhanced fat-suppressed imaging) render the RDOG findings obsolete.

Recurrent Tumor versus Radiation Fibrosis

Rectosigmoid cancers recur at a rate of 8–50%. This range reflects differences in the stage of primary tumor at presentation. Local recurrences are most common and, if diagnosed early, are amenable to curative resection. Recurrent rectal carcinomas are best imaged in the sagittal plane. One study reported 93.3% accuracy in detecting recurrent rectal cancers using T1-W, T2-W, and Gd T1-W sequences. Others have shown MRI to be superior to conventional CT and transrectal ultrasound (US) for identifying recurrent disease. Data on multidetector CT are at present, unknown.

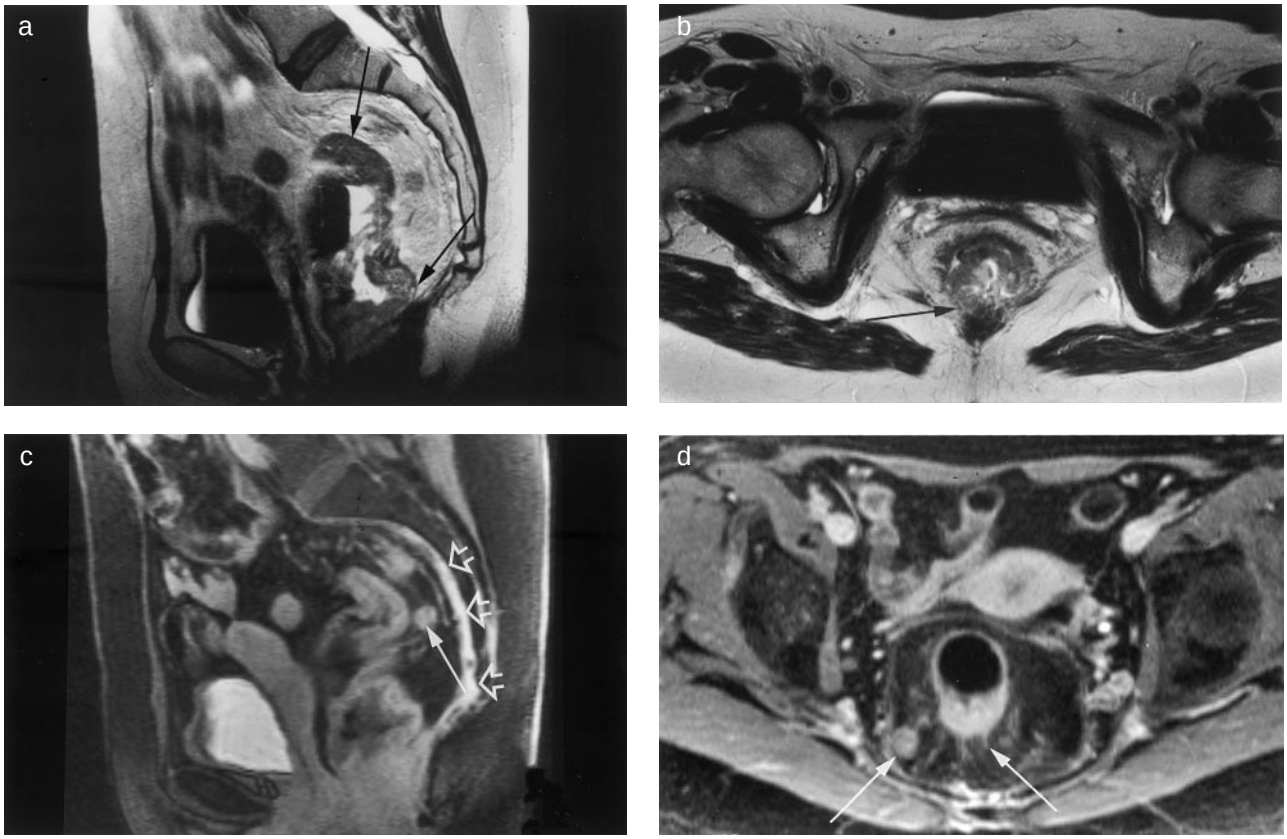


FIGURE 4 Rectal cancer. Sagittal (a) and axial (b) T2-W FSE and sagittal (c) and axial (d) fat-suppressed Gd T1-W SGE images in a patient with advanced rectal cancer. The sagittal image demonstrates the craniocaudad extent of the tumor (arrows, a). Extension to the anal verge (arrow, b) is confirmed on the axial images. On the T2-W sequences, the tumor's shaggy border suggests lymphovascular involvement that is confirmed on the contrast-enhanced images (arrows, c, d). Similarly, presacral spread of tumor is more conspicuous following gadolinium (open arrows, c). From Semelka *et al*. (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

A recurrent tumor is often low in signal intensity on T1-W techniques and enhances less than primary tumors following the administration of gadolinium chelates. On T2-W images, recurrences are often moderately high in signal intensity, but, as with primary tumors, this may not be conspicuous on T2-W SSFSE sequences. Recurrent tumors tend to be more nodular and "masslike" compared to postradiation fibrosis; though the two entities may coexist.

Following 1 year of treatment, radiation fibrosis is often low in signal intensity on both T1-W and T2-W sequences and has negligible enhancement (Fig. 5). Unfortunately, associated granulation tissue may also demonstrate high signal intensity for up to 3 years posttreatment. This phenomenon is most pronounced with patients who have received >45 mGy. Similarly, overlap in signal intensity between recurrent disease and fibrosis on T2-W SSFSE can occur when normal high-signal-intensity fat is admixed with low-signal-intensity fibrosis (Fig. 6). And, though radiation fibrosis is morphologically a plaquelike process, recurrent tumor with desmoplastic features may mimic these radiation-induced changes. Because recurrent carcinoma and radiation changes share many imaging features, interpretation should be performed with appropriate clinical correlation [e.g., rising carcinogenic embryonic antigen (CEA) levels].

INFLAMMATORY BOWEL DISEASE

MRI is robust for evaluating patients with inflammatory bowel disease (IBD). It correlates well with clinical indices, endoscopy, and histology. Furthermore, MRI may aid in diagnosing the type of IBD (Crohn's disease vs. ulcerative colitis), establishing a baseline of involvement and monitoring treatment response—all without exposing a patient to ionizing radiation. It is even useful for imaging complications associated with IBD

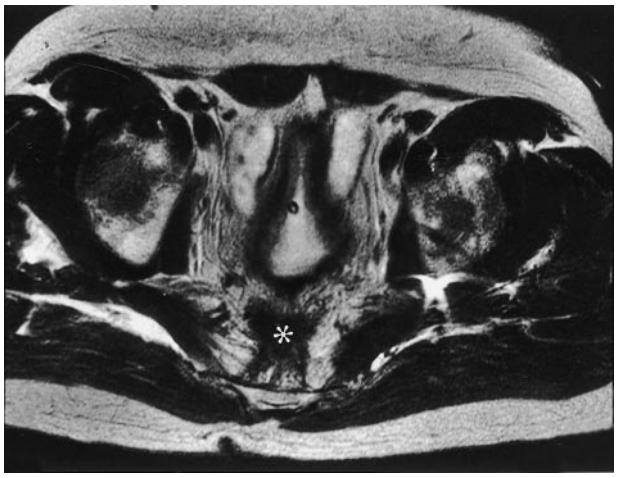


FIGURE 5 Posttreatment fibrosis. Axial high-resolution T2-W FSE image in a patient after surgery. The low signal intensity in the surgical bed (asterisk) is consistent with fibrosis. Its plaquelike morphology favors postradiation change, whereas recurrent tumor tends to be more nodular. From Semelka *et al.* (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

(e.g., abscess, fistulae, strictures/obstruction, and toxic megacolon).

Crohn's Disease

For patients with Crohn's disease, T2-W SSFSE and fat-suppressed Gd T1-W SGE sequences demonstrate classic findings: transmural involvement, skip lesions, and mesenteric inflammation. The terminal ileum, alone or in conjunction with other regions of the alimentary tract, is involved in approximately 70% of cases; whereas isolated Crohn's colitis occurs in about 25% of cases. MRI correlates of clinical disease activity have been worked out. Severity of disease (mild, moderate, or severe) is based on wall thickness, length of diseased segment, and percent bowel wall enhancement (Table I) (Fig. 7). MR assessment is performed on the nondependent bowel surface. For mural enhancement, qualitative assessment is performed by comparing bowel wall enhancement to renal parenchyma. Specifically, bowel should never enhance to the same degree as renal cortex on either the initial capillary phase or the interstitial phase images. For quantitative analysis, the abnormal segment is measured 2.5 minutes after injection. Imaging at this time (interstitial phase) reflects capillary leakage and impaired venous egress in the inflamed segment. Acute-on-chronic disease may deviate from expected MR findings: there is often marked enhancement of the mucosa of the involved loop of thickened bowel, with minimal enhancement of its outer layer.

Fistulae are common complications of Crohn's disease and MR images them well. Depending on their contents, they may have high signal intensity on T2-W images (e.g., fluid filled); alternatively, they may be signal void (e.g., gas filled). Multiplanar imaging combined with fat suppression increases the

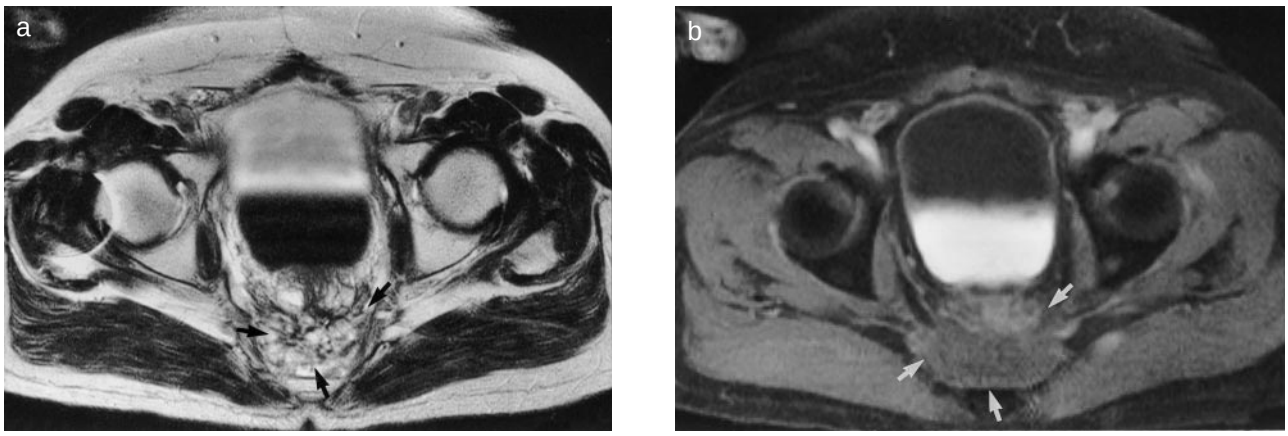


FIGURE 6 Recurrence simulating posttreatment fibrosis. Axial high-resolution T2-W FSE (a) and fat-suppressed T1-W SGE (b) images in a patient 1.5 years after treatment for rectal cancer. Heterogeneous and bulky high-signal-intensity soft tissue (arrrows, a) occupies the surgical bed, a worrisome indication for recurrent disease. Granulation tissue from radiation, inflammation, or infection may image similarly. The heterogeneity is misleading and results from high-signal-intensity fat—a phenomenon associated with fast spin echo techniques—admixed with low-signal-intensity fibrotic tissue. Negligible enhancement favors the diagnosis of fibrosis (arrows, b). From Semelka *et al.* (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

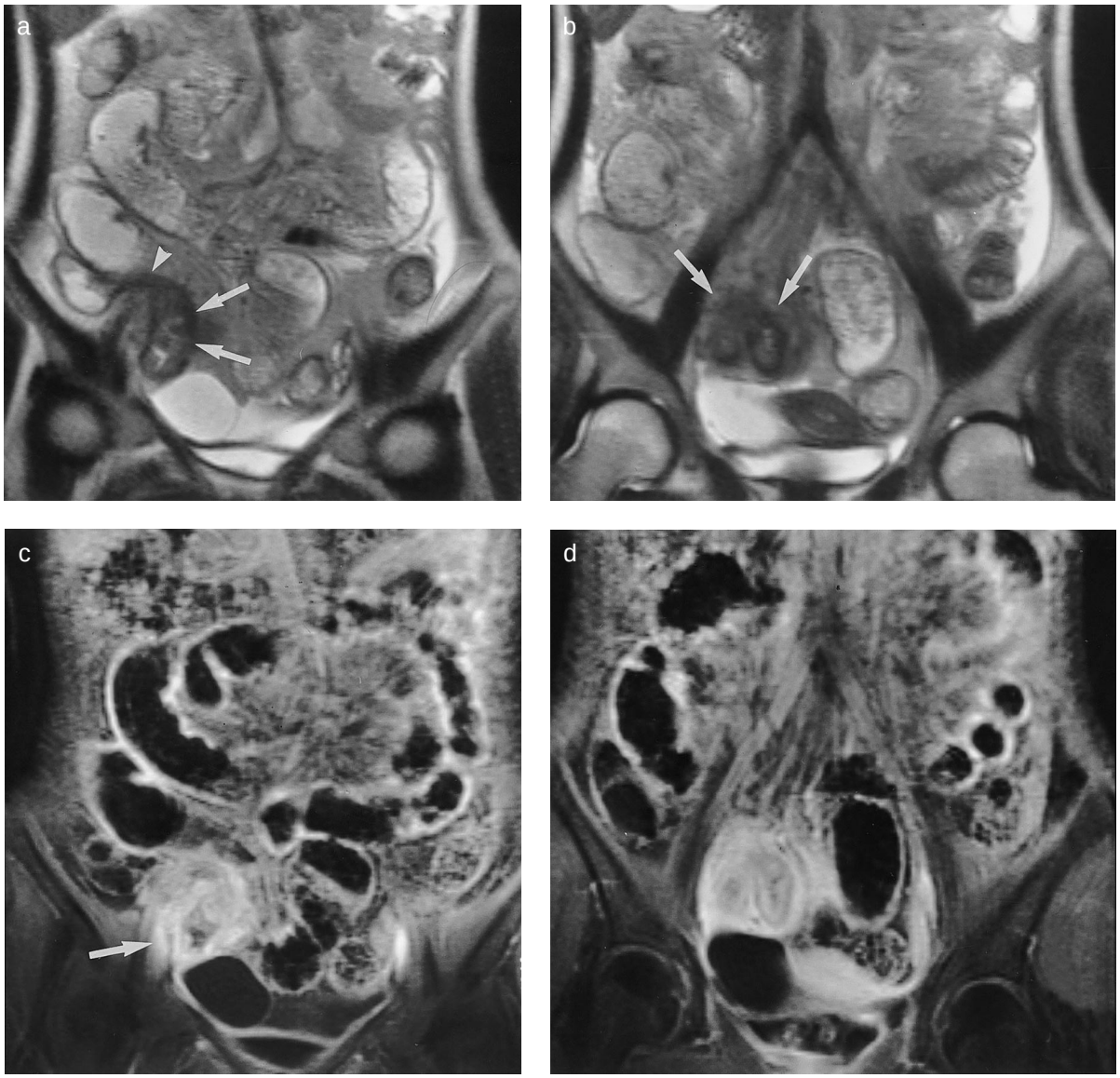


FIGURE 7 Severe Crohn's disease. Coronal T2-W SSFSE (a, b) and fat-suppressed Gd T1-W SGE (c, d) images in a patient with inflammatory bowel disease. The coronal T2-W SSFSE images demonstrate thickened loops of small bowel (arrows, a, b) to include the diseased terminal ileum entering the cecum (arrowhead, a). The corresponding enhanced images show marked transmural enhancement of the affected segments and surrounding tissues. In addition, a fistula associated with the ileo-cecal valve is highlighted (arrow, c). The wall thickness, length of involved bowel, and percent enhancement are consistent with severe Crohn's disease. From Semelka *et al*. (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

conspicuity of the enhancing fistulous tracts on post-contrast images (Fig. 7c). Focal loss of integrity of the involved organ at the site of fistula penetration is diagnostic. In the case of strictures, T2-W SSFSE demonstrates dilated obstructed bowel proximal to the site of narrowing, as well as the offending stricture.

Ulcerative Colitis

Ulcerative colitis (UC) is a chronic mucosal disease that begins in the rectum and extends in a proximal fashion to involve all or part of the colon. Though the small bowel is not directly affected in UC, secondary

TABLE I Crohn's Disease Severity Criteria

Severity	Contrast enhancement (%)	Wall thickness (mm)	Length of diseased segment (cm)
Mild[a]	<50	<5	<5
Moderate	50–100	5–20	Variable
Severe	>100	>10	>5

[a] Wall thickening must be at least 4 mm, and one of the other of the two categories must be satisfied.

inflammation of the terminal ileum, or "backwash ileitis" (reflux of colon contents into the small bowel), does occur with pancolonic disease. The major complications of UC include increased risk of developing colon carcinoma and toxic megacolon.

The MRI appearance is predictable: rectal involvement with variable continuous retrograde progression and submucosal sparing. Fat-suppressed Gd T1-W imaging in the interstitial phase highlights the enhancing high-signal-intensity mucosa surrounded by the negligibly enhancing low-signal-intensity submucosa (Fig. 8). The vasa rectae also enhance prominently. In long-standing disease, the submucosal sparing is accentuated. This is a function of submucosal edema and lymphangectasia. Interestingly, the imaging features of toxic megacolon differ from that of uncomplicated UC. Toxic megacolon is a transmural process—full-thickness enhancement usually accompanies the involved segment, though regions with submucosal sparing may coexist (Fig. 9).

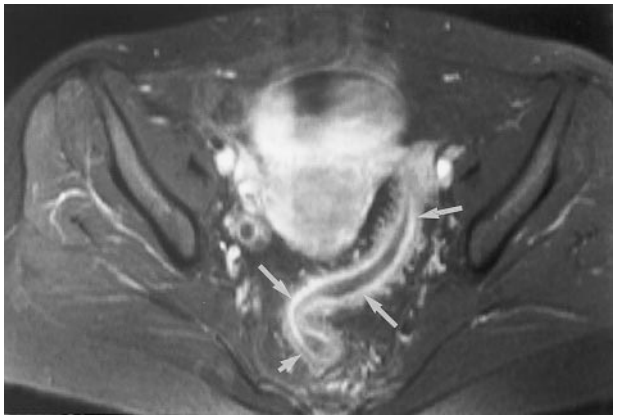


FIGURE 8 Ulcerative colitis. Axial interstitial fat-suppressed Gd T1-W spin echo image in a patient with ulcerative colitis. There is marked mucosal enhancement (long arrows) with prominent vasa rectae and submucosal sparing (short arrow). From Semelka *et al.* (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

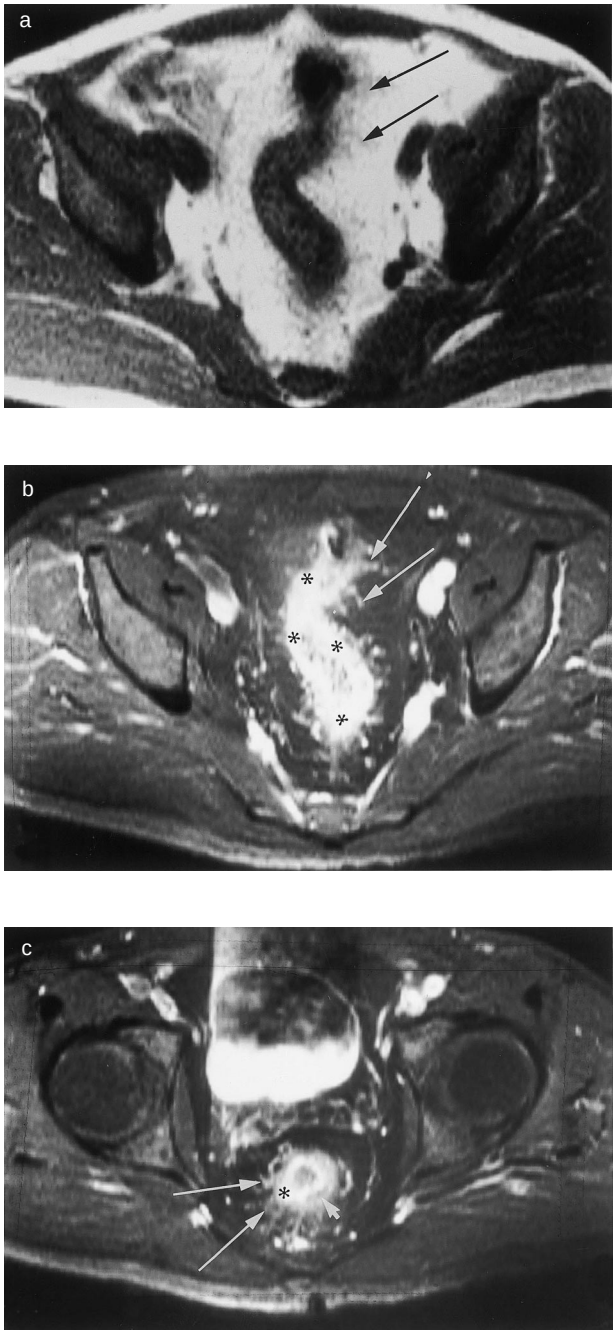


FIGURE 9 Toxic megacolon. Axial T1-W SGE (a) and fat-suppressed Gd T1-W spin echo (b, c) images in a patient with ulcerative colitis complicated by toxic megacolon. Precontrast, there are low-signal-intensity strands in the pericolonic fat (arrows, a). After contrast, there is intense enhancement of the sigmoid colon (asterisks, b, c) and vasa rectae (long arrows, b, c). Though a feature of ulcerative colitis is submucosal sparing (short arrow, c), cases complicated by toxic megacolon will also demonstrate transmural abnormalities. From Semelka *et al.* (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI." Copyright 2002 John Wiley. This material is used by permission of John Wiley & Sons, Inc.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Computed Tomography • Crohn's Disease • Magnetic Resonance Imaging

Further Reading

Beets-Tan, R. G. H., Betts, G. L., Borstlap, A. C. W., Oei, T. K., Teune, T. M., and von Meyenfeldt, M. F. (2000). Preoperative assessment of local tumor extent in advanced rectal cancer: CT or high-resolution MRI? *Abdom. Imaging* **25**, 533–541.

Berlin, J. W., Gore, R. M., Yaghmai, V., Newmark, G. M., and Miller, F. H. (2000). Staging of colorectal cancer. *Semin. Roentgenol.* **35**, 370–384.

Kinkel, K., Tardivon, A. A., Soyer, P., Spatz, A., Lasser, P., Rougier, P., and Vanel, D. (1996). Dynamic contrast-enhanced subtraction versus T2-weighted spin-echo MR imaging in the follow-up of colorectal neoplasm: A prospective study of 41 patients. *Radiology* **200**, 453–458.

Knollman, F. D., Dietrich, T., Bleckmann, B. S., Bock, J., Maurer, J., Radtke, C., and Felix, R. (2002). Magnetic resonance imaging of inflammatory bowel disease: Evaluation in a rabbit model. *J. Magn. Reson. Imaging* **15**, 65–173.

Lauenstein, T., Holtmann, G., Schoenfelder, D., Bosk, S., Ruehm, S. G., and Debatin, J. F. (2001). MR colonography without colonic cleansing: A new strategy to improve patient acceptance. *Am. J. Roentgenol.* **177**, 823–827.

Low, R. N., and Francis, I. R. (1997). MR imaging of the gastrointestinal tract with IV gadolinium and diluted barium oral contrast medica compared with unenhanced MR imaging and CT. *Am. J. Roentgenol.* **169**, 1051–1059.

Luboldt, W., Bauerfeind, R., Wildermuth, S., Marincek, B., Fried, M., and Debatin, J. F. (2000). Colonic masses: Detection with MR colonography. *Radiology* **216**, 383–388.

Marcos, H. B., and Semelka, R. C. (2000). Evaluation of Crohn's disease using half-fourier RARE and gadolinium-enhanced SGE sequences: Initial results. *Magn. Reson. Imaging* **18**, 263–268.

Markus, J., Morrissey, B., deGara, C., and Tarulli, G. (1997). MRI of recurrent rectosigmoid carcinoma. *Abdom. Imaging* **22**, 338–342.

Müller-Schimpfle, M., Brix, G., Layer, G., *et al.* (1993). Recurrent rectal cancer: Diagnosis with dynamic MR imaging. *Radiology* **189**, 881–889.

Schnall, M. D., Furth, E. E., Rosato, E. F., and Kressel, H. Y. (1994). Rectal tumor stage: Correlation of endorectal MR imaging and pathologic findings. *Radiology* **190**, 709–714.

Schoenut, J. P., Semelka, R. C., Magro, C. M., Silverman, R., Yaffe, C. S., and Mickflier, A. B. (1994). Comparison of magnetic resonance imaging and endoscopy in distinguishing the type and severity of inflammatory bowel diseases. *J. Clin. Gastroenterol.* **19**, 31–35.

Semelka, R. C., Hricak, H., Kim, B., *et al.* (1997). Pelvic fistulas: Appearances on MR images. *Abdom. Imaging* **22**, 91–95.

Semelka, R. C., Pedro, M. S., Armao, D., Marcos, H. B., and Ascher, S. M. (2002). Gastrointestinal tract. *In* "Abdominal-Pelvic MRI" (R.C. Smelka, ed.), 1st Ed., pp. 527–649. Wiley-Liss Inc., New York.

Semelka, R. C., (2002). "Abdominal-Pelvic MRI," pp. 527–649. 1st Ed., Wiley-Liss Inc., New York.

Stevenson, G. W. (2000). Colorectal cancer imaging: A challenge for radiologists. *Radiology* **214**, 615–621.

Thoeni, R. F. (1997). Colorectal cancer. Radiologic staging. *Radiol. Clin. North Am.* **35**, 457–485.

Zerhouni, E. A., Rutter, C. R., Hamilton, S. R., *et al.* (1996). CT and MR imaging in the staging of colorectal carcinoma: Report of the Radiology Diagnostic Oncology Group II. *Radiology* **200**, 443–451.

Alpha-1-Antitrypsin (α1AT) Deficiency

MICHAEL K. PORAYKO
Vanderbilt University Medical Center

conformational diseases A group of disorders resulting from gene mutations affecting the proper folding and three-dimensional structure of intracellular proteins, leading to changes in protein interactions, abnormal intracellular accumulations, and altered function.

D-PAS-positive globules Diastase-resistant, periodic acid-Schiff-positive globules of α1-antitrypsin within the liver.

isoelectric focusing An electrophoretic technique on polyacrylamide gel allowing better discrimination of α1-antitrypsin glycoprotein variants.

loop-sheet polymerization The process whereby abnormally folded α1-antitrypsin molecules link together to form polymers and inclusions.

molecular chaperones Intracellular helper proteins that assist in the proper folding of polypeptides into their functional structures.

PI*QO Denotes the null genotype, characterized by the total absence of serum α1-antitrypsin.

PI*Z Denotes the α1AT genotype.

PI Z Denotes the α1AT phenotype.

serpin Serine proteinase inhibitors constitute a family of circulating molecules that share structural similarities yet carry out diverse physiological functions in the body.

targeted homologous recombination A genetic technique whereby a targeted or defective gene can be cleaved from DNA and replaced by a normal form of the gene.

Among inherited diseases, α1-antitrypsin deficiency (α1ATD) is a relatively common disorder affecting the cellular secretion of one of the body's protective proteinase inhibitors. Specific mutations of the α1-antitrypsin (α1AT) gene lead to changes in the three-dimensional structure of its translated protein, which becomes trapped in hepatocytes where it is formed. The resultant circulating deficiencies of α1AT allow destructive proteases to cause tissue injury, particularly in the lung, and accumulations of mutant α1AT in liver cells disrupt normal function, sometimes leading to cirrhosis. α1-Antitrypsin deficiency is only one member of a family of conformational diseases requiring more detailed understanding in order to counteract common pathophysiological abnormalities, namely, the improper folding and processing of critical proteins, as a means of restoring normal biological function. This article focuses primarily on the current understanding of liver-related injury secondary to α1ATD.

INTRODUCTION

In the United States alone, more than 100,000 individuals have α1-antitrypsin deficiency (α1ATD). However, the disease goes undetected in many people because the disorder is underappreciated and misinterpreted. Given the protean functions of α1-antitrypsin (α1AT), patients with inherited deficiency mutations can present in a variety of ways depending on the tissue involved (e.g., lungs, liver, blood vessels, kidneys) and the extent of injury. Clearly, multiple inherited and environmental factors are involved in the full expression of disease in patients with α1ATD. Disease is not inevitable, even in patients with the most severe forms of deficiency. Alternatively, there are individuals with heterozygous genotypes and partial deficiencies of α1AT who develop disease with all the manifestations of a homozygote carrying two mutant alleles. Cigarette smoking is one well-recognized example of an environmental factor capable of accelerating pulmonary disease in patients with α1ATD but modulating variables contributing to liver injury in these patients remain speculative and are under investigation. Nonetheless, earlier identification of patients at risk is important because manifestations of disease can be moderated in some instances by increasing awareness of potential aggravating conditions and applying preventative strategies. If liver failure develops, liver transplantation can be curative but the success of surgery is greatest when performed at an optimal time in the course of the disease. Proper monitoring of the patient plays an important role in making this determination. Genetic counseling should be available for family members but physicians must be cognizant of the potential implications of discovering an underlying genetic disorder, especially in those yet unaffected by disease.

BIOLOGICAL AND MOLECULAR CHARACTERISTICS

α1-Antitrypsin is one of a family of molecules called serine protease inhibitors that share structural homology but are functionally diverse, regulating a variety of proteolytic processes essential to life (Table I). The

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Serine Proteinase Inhibitors (Serpin Superfamily)

α1-Antitrypsin
α1-Antichymotrypsin
α2-Antiplasmin
Plasminogen activator I
Antithrombin III
Angiotensinogen
Kallistatin
Leukocyte inhibitor
Thyroxine-binding globulin
Cortisol-binding globulin
Heparin cofactor II
Protease nexin I
Protein C inhibitor

activities of these molecules are highly dependent on amino acid sequence and posttranslational intracellular processing in order to evolve into a characteristic structure and shape. With proper conformational changes, these molecules expose specific reactive centers critical to their function. Misfolding of these polypeptides can change the usual interactions with target molecules and other proteins.

α1-Antitrypsin is a relatively small glycoprotein made up of 394 amino acids and three carbohydrate side chains. Its primary role appears to be inhibition of the proteolytic activities of neutrophil elastase, cathepsin G, and protease 3. The advantage of having adequate levels of circulating α1AT is most evident when considering the destructive actions of neutrophil elastase in the lungs. Inhalations of pulmonary pollutants, particularly cigarette smoke, attract and stimulate neutrophils, releasing destructive enzymes such as elastase. Controlled release of these enzymes is intended to break down components of the extracellular matrix, bacterial cell walls, and other proteins to facilitate the movement of inflammatory cells into affected areas of the lung. The presence of α1AT, with its ability to bind irrevocably with neutrophil elastase and other destructive enzymes, limits tissue injury, whereas severe deficiency of α1AT has been clearly associated with premature development of emphysema.

Synthesis of the α1AT protein complex, which occurs almost exclusively in the liver, involves a series of steps within hepatocytes before the complex is secreted into the circulation. The α1AT molecule is folded into a highly ordered three-dimensional structure composed of three β-sheets and nine α-helices as it traverses the cell. Final assembly and folding of α1AT occurs in the endoplasmic reticulum and Golgi apparatus with the help of one or more molecular chaperones (e.g., calnexin, heat shock proteins) facilitating the process (Fig. 1). When properly folded, a unique reactive site centered around two amino acids, Met^{358}–Ser^{359}, is exposed on the surface of the α1AT molecule. The interaction of this site with its target protease results in permanent inactivation of the coupled molecules. Small amounts of flawed or improperly folded proteins appear, ordinarily, to enter a separate intracellular pathway for degradation and disposal. Slight differences in the amino acid sequence of α1AT can lead to significant changes in the final shape of the molecule, altering its interactive surface. In the case of α1AT, single point mutations can alter the protein complex in such a way that the mutant α1AT molecules begin to interconnect in a process called "loop-sheet" polymerization. One polypeptide hooks into another and so on, eventually forming conglomerations of polymerized molecules that aggregate in the cell's endoplasmic reticulum, forming globules of varying size that are visible by light microscopy. These globules can be degraded over time, at different rates depending on genetically determined processes, sometimes releasing small quantities of viable α1AT molecules from the cell. Individuals with more efficient mechanisms of aberrant α1AT disposal may be less susceptible to the consequences of intracellular accumulation.

NOMENCLATURE—PROTEASE INHIBITOR PHENOTYPING

A gene located on the long arm of chromosome 14 regulates the production of α1AT. This gene locus is extremely pleiomorphic, with more than 100 different allelic variants (protease inhibitor or PI types) identified to date. Each inherited parental allele is co-dominantly expressed. Therefore, the quantity of circulating α1AT represents the combined product of the independent expression of each allele (Table II). In order to separate these variant glycoproteins into clinically recognizable "phenotypes," a classification system has been devised, making use of distinctive patterns of bands on polyacrylamide gel electrophoresis using a special technique called isoelectric focusing. Phenotypic variants are assigned letters from the alphabet based on the locations of these bands on gel slabs. During the process, faster moving molecules travel further along the gel strip and are assigned letters at the beginning of the alphabet, whereas slower moving variants are assigned letters near the end of the alphabet. It is important to note that the electrophoretic mobility does not indicate specific function or the capacity to be secreted from the hepatocyte. A majority of individuals (>95%) produce a normal M variety of α1AT, whereas several of

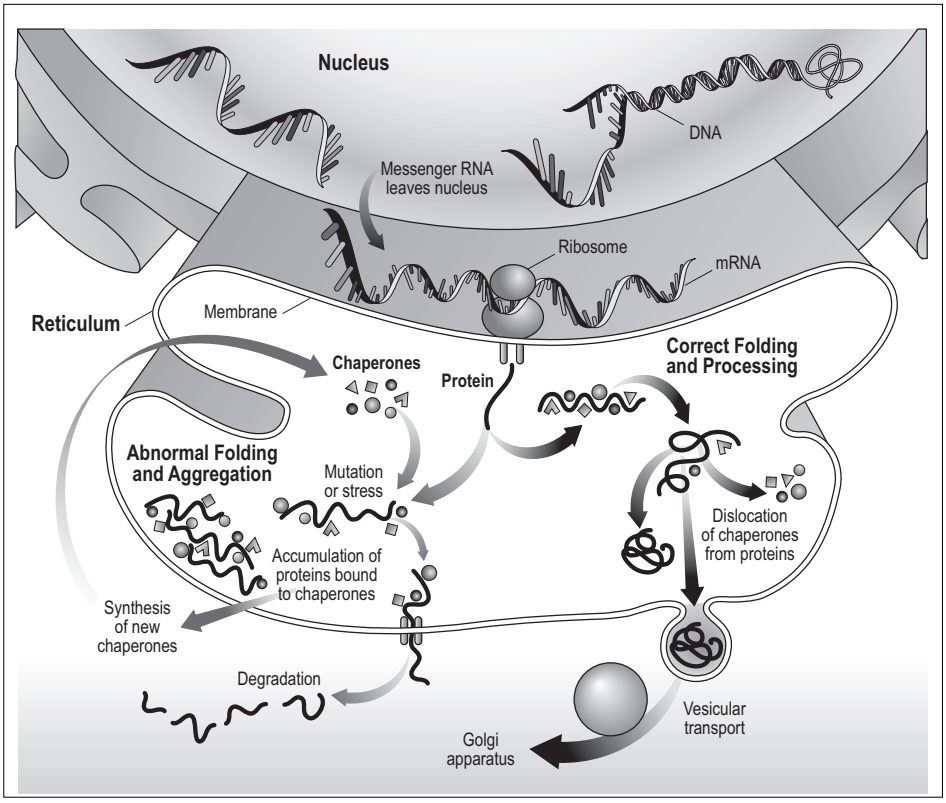


FIGURE 1 α1-Antitrypsin protein folding in the endoplasmic reticulum.

the more common "deficiency mutants" have been assigned letters such as S and Z. Homozygotes for the Z allele make up 95% of all deficiency phenotypes in patients with α1ATD and they are readily identifiable using isoelectric focusing. At times, variant molecules associated with α1ATD have patterns that are not easily detected by isoelectric focusing techniques but may be suspected by physicians because of characteristic clinical presentation, recognition of afflicted family members, detection of low circulating levels of α1AT, or detection of typical globules of mutant proteins on liver biopsy. In these instances, the specific mutation can be isolated by DNA sequencing. When determining an individual's phenotype, the laboratory will provide a letter for each of the two proteins representing the products of each allele (e.g., MM, MZ, ZZ).

TABLE II Serum Concentrations of α1AT Depending on Phenotype

Phenotype	Concentration (% of normal)
MM	100
MS	80
SS	60
MZ	55–60
SZ	35–40
ZZ	10–15
M_{duarte} (homozygote)	<15
M_{malton} (homozygote)	<15
QOQO (null)	0

The multiple different variants of α1AT generally can be placed into four groups based on circulating quantity and function: normal, deficient, null, and dysfunctional. The availability of α1AT in the blood can be normal (e.g., PI M), reduced (e.g., the deficiency alleles S and Z), or absent (e.g., extremely rare mutations called null alleles). Even though there are reduced amounts of circulating α1AT associated with many of these deficiency variants, the molecules that are excreted are usually fully functional. However, there are exceptionally rare mutations that result in a protein that escapes the hepatocyte but is folded is such a way that it changes function. The Pittsburgh variant is one example of a dysfunctional variant in which α1AT acts as a thrombin inhibitor, resulting in a severe bleeding disorder.

EPIDEMIOLOGY

Allelic frequencies of α1AT gene mutations vary considerably around the world and between study

populations. Utilizing global population studies to map the putative α1AT genes, investigators have suggested that the Z mutation originated in Scandinavia approximately 2000 years ago, eventually spreading from Northern Europe when people migrated from the area to other parts of the world. Therefore, the highest frequencies of the mutant gene are seen in Northern Europeans or in individuals of European ancestry. In Sweden, the frequency of the Z allele (PI*Z) is estimated to be 0.026 with 4−5% of the population carrying the mutation and 1 : 1600 live births being homozygous for this deficiency. The PI*Z gene frequency is somewhat lower in the United States among individuals of European descent, estimated to be between 0.01 and 0.02 with a carrier rate of approximately 3%. Some 80,000 to 100,000 people in the United States are PI ZZ homozygotes with considerable potential to develop disease. The Z allele is confined predominantly to Caucasians and is uncommon among individuals of Asian or African ancestry unless there is ethnic intermixing in the population. PI*S is a more common deficiency mutation than PI*Z, having a gene frequency between 0.02 and 0.03 in U. S. Caucasians. However, this genotypic variant is not associated with as severe a reduction in circulating levels of α1AT as the PI*Z allele and hence does not leave the individual susceptible to disease unless it is co-inherited with a Z allele. Of greater concern is the risk of disease in the 1 : 1000 to 1 : 1500 Caucasian individuals expressing a PI SZ phenotype. Although the matter is controversial, anyone carrying a single Z allele may be at some, as yet undefined, risk of developing liver disease regardless of the circulating levels of α1AT because of intracellular accumulation of the abnormally folded protein.

CLINICAL MANIFESTATIONS

The risk of developing emphysema in someone with α1ATD is closely related to the amount of circulating α1AT, especially when serum levels fall consistently below a "protective" level of 80 mg/dl (11 μ*M*). Even in patients with severe deficiencies of α1AT (< 50 mg/dl, < 7 μ*M*), a majority of patients presenting with pulmonary symptoms also report a history of smoking cigarettes. Interestingly, liver disease has not been a commonly reported problem in patients diagnosed with chronic lung disease related to α1ATD. The development of liver disease has been more closely linked to those genotypic variants associated with intracellular accumulation of polymerized polypeptides within the endoplasmic reticulum. Aside from the well-recognized Z variant, other rare mutations have been associated with intracellular globules identified in the liver (Table III). Accumulation of α1AT deposits can begin *in utero*, sometimes affecting neonates at an early stage in development. PI Z infants frequently present with low birth weight and coagulation abnormalities; as many as 70% have biochemical evidence of liver injury as indicated by elevations of liver enzymes. Prolonged cholestatic hepatitis, lasting up to 8 months, is clinically manifest in approximately 10% of these newborns. Although α1AT globules are not always easily identified on liver biopsy at this early stage, hepatocellular necrosis associated with acute and chronic portal inflammatory infiltrates, bile duct proliferation, bile plugs, and even ductopenia (intrahepatic bile duct loss) have been described in deficient individuals. Males appear to be affected more frequently than females (2 : 1) but a majority of children do not have long-term consequences at least as followed into their late teens. Approximately 3% have been shown to progress to hepatic fibrosis or cirrhosis and 12−15% have persistent mild elevations of liver enzymes.

Adults with the most prevalent form of α1ATD (PI ZZ) typically present with pulmonary symptoms (i.e., dyspnea) rather than with indicators of liver disease. It has been estimated that as many as 85% of patients with severe deficiency of α1AT have evidence of chronic obstructive pulmonary disease, often identified in the third and fourth decades of life. The prevalence of liver disease in these patients has not been well investigated even though it has been stated that few patients have liver test abnormalities when presenting with pulmonary manifestations. However, liver disease in adults with α1ATD progresses insidiously, often escaping detection until patients are beyond the age of 50 years. Commonly, a diagnosis is made when signs and symptoms of more advanced liver disease and portal hypertension (e.g., weakness, muscle wasting, ascites, splenomegaly, variceal hemorrhage, encephalopathy) appear. Therefore, a true incidence of liver disease in these patients

TABLE III α1ATD Phenotypic Variants with Potential Risk of Liver Disease

Associated with liver cirrhosis or dysfunction
Homozygous or heterozygous PI Z
ZZ, MZ, SZ, FZ, PZ
Rare PI M variants
M_{malton}, M_{duarte}
Other rare variants
$W_{salerno}$
Rare PI variants showing α1AT hepatic globules without clinical evidence of liver disease
$M_{cagliari}$, $M_{nichinan}$, $M_{elemberg}$, M_{cobalt}, M_{leuven}, $QO_{hong\ kong}$, S_{iiyama}

may be underappreciated, especially when the full expression of liver disease is attenuated by premature mortality from emphysema in younger age groups. This may be a reason that carefully conducted autopsy studies in populations with a high prevalence of severe α1ATD have estimated an incidence of cirrhosis as high as 40% over the lifetime of affected individuals. Most investigators feel that the numbers who develop cirrhosis may be somewhere between 10 and 15%, acknowledging that the risk increases with age, particularly in males. A North American study conducted in Canada found the cumulative risk of cirrhosis in susceptible men with α1ATD to be 23%, whereas the same risk in women amounted to only 5%. Of additional interest, these investigators observed that death from liver failure frequently occurred within 4 years of onset of clinical symptoms of liver disease, indicating that detection is often delayed.

Several population surveys and autopsy series find an increased risk of hepatocellular cancer in patients with severe deficiencies of α1AT and cirrhosis. In Sweden, liver cancer has been detected in approximately 15% of afflicted patients but once again the finding is predominantly in males.

A subject of ongoing interest and controversy is whether patients carrying a single Z allele (heterozygotes) are at some, as yet undefined, risk of developing cirrhosis and hepatic cancer. Studies to date have been unable to address this issue confidently because of the inherent difficulty in identifying and following subjects with the inherited trait longitudinally, along with proper controls. Nevertheless, it has been well recognized that heterozygotes for the Z allele often develop α1AT globules in the liver and can present with cirrhosis without any obvious alternative etiology for liver disease. Cohorts of patients with "cryptogenic" cirrhosis have shown a higher prevalence of α1ATD heterozygotes (e.g., PI MZ), raising suspicion about the single gene's role in the pathogenesis of liver disease. The relative importance of other variables such as viral hepatitis infection, alcohol, iron excess, generalized inflammatory conditions creating biological stress, or genetic factors in the natural history and progression of disease in these patients remains unclear. The coexistence of one or more of these factors may accelerate the accumulation or reduce the processing and breakdown of α1AT inclusions, resulting in enhanced hepatotoxicity. Despite this potential increased risk of developing significant liver disease in individuals with heterozygous α1ATD, the risk is estimated to be exceptionally small and should not alarm patients. Undue concern by the physician can have psychological consequences for the individual carrying the genetic trait. Prudent preventative measures and follow-up are all that may be recommended.

DIAGNOSIS

Laboratory tests used to screen for α1ATD should be considered in any patient, particularly Caucasians of European decent, with clinical, biochemical, radiological, or histological findings suggesting an underlying liver disorder. Additional signs or symptoms of pulmonary disease or history of an affected family member should heighten interest in pursuing this diagnosis.

Screening is best accomplished by obtaining both a quantitative level of α1AT and phenotype analysis. Reliance on semiquantitative analyses such as agarose gel electrophoresis, seeking a diminished or absent α1-globulin band, or determining quantitative levels of α1AT (e.g., immunoturbidimetry, nephelometry) alone can be misleading. Transient increases in α1AT levels have been observed in patients with putative deficiency variants, presumably as part of a general acute-phase reaction. Alternatively, several phenotypic variants, such as PI SS and null subtypes, have reduced circulating levels of α1AT but do not have the attendant risk of liver disease because of a lack of intracellular accumulation of glycoprotein. The combination of an α1AT level and phenotype helps to more directly identify patients with potential disease-causing variants.

PI phenotyping requires significant skill and experience to ensure accurate readings. Commonly, samples must be sent to a reference laboratory for interpretation. One method of phenotype identification uses a technique called isoelectric focusing (IEF) within a narrow pH gradient in polyacrylamide gel strips. Based on the isoelectric point, the α1AT molecules migrate along the gel to different positions, forming one or more bands that correspond to a particular phenotype. Some of the more unusual variants are not always easily visualized by routine IEF and can be misinterpreted as another more common pattern; for example, PI M_{malton} and PI M_{duarte} have electrophoretic patterns similar to that of PI M especially when heterozygous with a normal M allele. Null alleles do not generate any circulating product; therefore, IEF bands are absent. When the null allele is present in the heterozygous state, it may not be appreciated given the presence of bands from another allelic protein. However, a null allele should be suspected when the quantitative level of α1AT is significantly reduced in the setting of what appears to be a normal phenotype. At times, more unusual variants can be more accurately discriminated by gene analysis and other DNA-based techniques but many of these procedures are time-consuming. However, polymerase

chain reaction-based methods have been utilized as tools for screening populations when narrowing the focus of investigation to specific alleles.

After an initial screening, if an individual tests positive for a higher risk α1AT phenotype and has any indication of liver involvement, such as clinical signs, elevated liver enzymes, or hepatic architectural changes by radiological imaging, a liver biopsy should be considered. Evaluating liver histology helps to support the diagnosis of α1ATD, is useful in evaluating for alternative diagnoses, and can stage liver damage for purposes of making decisions about future management. Histological features can vary depending on the age of the individual (Table IV). A characteristic finding on biopsy is the presence of diastase-resistant, periodic acid-Schiff (D-PAS)-positive globules predominantly found in periportal hepatocytes and adjacent to fibrotic bands when present. These globules have a distinctive magenta color, making them highly visible with D-PAS staining, but they can be missed using light microscopy with standard hematoxylin and eosin staining.

Rarely, D-PAS-positive globules have been found in patients without α1ATD. D-PAS-positive globules have been found in liver biopsy specimens from patients who are elderly or terminally ill but a clue is the atypical location of the inclusions within the lobule (e.g., centrilobular region). These aberrant globules may represent lysosomes filled with proteins requiring degradation resulting from sinusoidal congestion and hypoxia. Occasionally, patients with alcoholic cirrhosis and oral contraceptive-associated hepatic tumors have D-PAS-positive globules within the liver despite normal α1AT phenotypes. If necessary, further characterization of intracellular inclusions can be performed using immunohistochemical techniques with mono-specific antiserum against α1AT.

TABLE IV α1-Antitrypsin—Hepatic Histopathology

Neonates and children
Hepatocytes vary in size with slight acinar formation
Occasional giant cell formation
Bile ducts may be normal, proliferated (sometimes confused with biliary atresia), or hypoplastic with duct loss
Low-grade inflammation, piecemeal necrosis, and fibrosis initially found in portal areas as disease progresses
D-PAS globules found in periportal hepatocytes predominantly but may be rare or inconspicuous during the first 2 months of life[a]
Adults
Low-grade lymphocytic infiltration in portal tracts, often inconspicuous
Intralobular bile ducts may be reduced in number
D-PAS-positive globules distributed in areas of inflammation and along bands of fibrosis[a]

[a] The presence, number, or size of globules does not predict occurrence of liver disease.

POTENTIAL THERAPIES

At present, there are no practical therapeutic modalities capable of halting progression of liver disease in patients with inflammation and fibrosis on liver biopsy. The usual preventative measures (i.e., vaccinations for hepatitis viruses, avoidance of aggravating factors such as alcohol), supportive care in case of complications, and monitoring to determine the optimal time for liver transplantation if it becomes necessary are useful in patients with advancing liver disease; reassurance is all that is necessary for patients with benign forms of the trait. Given that an estimated 50 to 75% of patients with liver disease secondary to α1ATD have some degree of underlying pulmonary dysfunction as well, pulmonary function testing along with a discussion of preventative behaviors (e.g., avoiding tobacco) and potential therapies should be part of the management plan. Genetic counseling and additional screening of family members, particularly first-degree relatives, can be considered in order to identify individuals at risk of developing lung or liver disease. However, a physician or genetic counselor must be aware of the potential social, psychological, ethical, and legal consequences of discovering a genetic disorder, especially in persons without obvious signs or symptoms of disease.

Patients who have progressed to end-stage liver disease should be considered for liver transplantation. Multiple transplant centers report excellent patient survival rates, ranging between 70 and 92%, in patients with liver failure secondary to α1ATD. Moreover, transplantation of a donor allograft with a normal genotype cures the disease by normalizing production of α1AT.

Therapies designed to replace or augment production of α1AT, sometimes utilized in the management of patients with progressive pulmonary disease, are not useful and may be detrimental in patients with hepatic disease arising from α1ATD. Increasing cellular production of α1AT by medications such as tamoxifen or danazol act only to increase accumulation within the endoplasmic reticulum, presumably the reason for toxicity. Intravenous replacement with purified preparations of α1AT may also result in an increased production of the mutant form of the molecule within the cell through a feedback mechanism stimulated by α1AT−protease complexes. Currently, experimental exploration of pharmaceuticals capable of augmenting the breakdown of intracellular α1AT polymers or acting

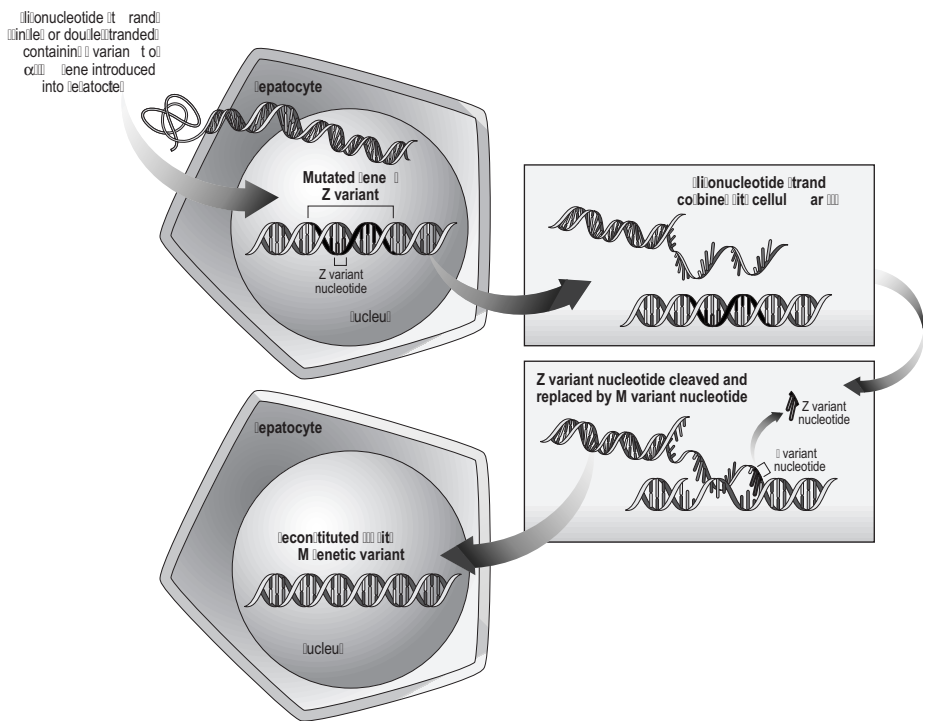


FIGURE 2 Mechanisms of gene manipulation and repair.

as chemical chaperones reducing protein misfolding and polymerization may offer some promise.

Ultimately, gene therapy for both α1ATD lung and liver disease appears to offer the best solution for cure, conceptually, but many practical problems must be solved before these therapies can be implemented clinically. Conformational diseases such as α1ATD offer unique challenges for gene therapy because it is not only important to introduce new fully functional genes to appropriate target cells but native mutant genes or their products must be removed or inactivated in order to have the desired effect—increasing production of a normal gene product without causing accumulations of abnormal proteins in the cell. Several techniques of site-directed gene repair already have proven successful in the laboratory including targeted homologous recombination and nucleotide exchange using chimeric RNA–DNA oligonucleotides (Fig. 2). Other investigators have utilized specialized ribozymes to inactivate mutant α1AT mRNA produced by cells in a human hepatoma cell line followed by transfecting the cells with a modified wild-type gene capable of producing normal α1AT. These preliminary experimental approaches are exciting but many questions must be answered before they can be used safely in human subjects.

See Also the Following Articles

Alpha-1-Antitrypsin (α1AT) Deficiency, Pediatric • Cirrhosis • Hepatocellular Carcinoma • Hepatocytes • Smoking, Implications of

Further Reading

Blank, C. A., and Brantly, M. (1994). Clinical features and molecular characteristics of α1-antitrypsin deficiency. *Ann. Allergy* **72**, 105–120.

Carrell, R. W., and Lomas, D. A. (2002). α1-Antitrypsin deficiency—A model for conformational diseases. *N. Engl. J. Med.* **346**, 45–53.

Coakley, R. J., Taggart, C., O'Neill, S., and McElvaney, N. G. (2001). α1-Antitrypsin deficiency: Biological answers to clinical questions. *Am. J. Med. Sci.* **321**, 33–41.

Cox, D. W. (1995). α1-Antitrypsin deficiency. *In* "The Molecular and Metabolic Basis of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), pp. 4125–4158. McGraw-Hill, New York.

Crystal, R. G. (1990). α1-Antitrypsin deficiency, emphysema, and liver disease: Genetic basis and strategies for therapy. *J. Clin. Invest.* **85**, 1343–1352.

Eriksson, S. (1999). α1-Antitrypsin deficiency. *J. Hepatol.* **30**, 34–39.

Norman, M. R., Mowat, A. P., and Hutchison, D. C. S. (1997). Molecular basis, clinical consequences and diagnosis of α1-antitrypsin deficiency. *Ann. Clin. Biochem.* **34**, 230–246.

Perlmutter, D. H. (1998). α1-Antitrypsin deficiency. *Semin. Liver Dis.* **18**, 217–225.

Qu, D., Teckman, J. H., and Permutter, D. H. (1997). α1-Antitrypsin deficiency associated liver disease. *J. Gastroenterol. Hepatol.* **12**, 404–416.

Richardson, P. D., Kren, B. T., and Steer, C. J. (2002). Gene repair in the new age of gene therapy. *Hepatology* **35**, 512–518.

World Health Organization (1997). α1-Antitrypsin deficiency: Memorandum from a WHO meeting. *Bull. World Health Org.* **75**, 397–415.

Wulfsberg, E. A., Hoffmann, D. E., and Cohen, M. M. (1994). α1-Antitrypsin deficiency. *J. Am. Med. Assoc.* **271**, 217–222.

Alpha-1-Antitrypsin (α1AT) Deficiency, Pediatric

JEFFREY H. TECKMAN

Washington University and St. Louis Children's Hospital, Missouri

M Nomenclature signifying the normal allele of the Alpha-1-antitrypsin (α1AT) gene.

PAS-positive, diastase-resistant globule Accumulations of mutant α1-antitrypsin protein within hepatocytes in PIZZ α1AT deficiency that stain red by periodic acid–Schiff and are resistant to diastase digestion.

PI or PI-type Diagnostic phenotype analysis of α1-antitrypsin protein from patient serum by isoelectric-focusing gel electrophoresis.

S Nomenclature signifying a mutant allele of the α1-antitrypsin gene that yields an intermediate level of α1AT deficiency.

Z Nomenclature signifying the most common mutant allele of the α1-antitrypsin gene that yields profound α1AT deficiency.

ZZ or PIZZ Homozygosity for the mutant Z allele of the α1-antitrypsin gene and the definition of classical α1AT deficiency.

Alpha-1-Antitrypsin (α1AT) is a glycoprotein synthesized primarily in the liver and secreted into the blood where its function is to inhibit neutrophil protease-induced host tissue damage. Homozygosity for the Z mutant allele of α1AT (PIZZ) causes the classical form of α1AT deficiency. The protein product of the mutant Z gene accumulates within hepatocytes rather than being efficiently secreted. PIZZ homozygous adults have a markedly increased risk of developing emphysema by a loss-of-function mechanism, i.e., reduced levels of circulating α1AT in the lung to inhibit connective tissue breakdown by neutrophil proteases. PIZZ homozygous children and adults may also develop liver disease and hepatocellular carcinoma by a gain-of-function mechanism; i.e., the accumulation of mutant α1ATZ protein within hepatocytes is toxic to liver cells.

CLASSICAL PIZZ α1-ANTITRYPSIN DEFICIENCY

Pathophysiology and Genetics

Classical, PIZZ α1-antitrypsin deficiency is caused by homozygosity for a point mutation in the α1AT gene encoding substitution of lysine for glutamate at position 342. The 12.2 kb α1AT gene is located on chromosome 14q and encodes a 55 kDa glycoprotein of 395 amino acids. PIZZ homozygotes occur at a frequency of 1 in 1500–3500 in North American and European populations. It is the most common genetic cause of liver disease in children and the most common genetic disease leading to liver transplantation in children and it can also cause chronic liver disease, hepatocellular carcinoma, and premature pulmonary emphysema in adults.

The Z mutation confers polymerogenic properties on the mutant α1ATZ protein molecule, which is primarily synthesized in the liver but is also found in leukocytes and other tissues. The mutant α1ATZ protein is retained in the endoplasmic reticulum (ER) of hepatocytes rather than being secreted into the blood and body fluids where its normal function is to inhibit neutrophil proteases and thereby protect host tissues from inflammation-induced damage. PIZZ homozygous individuals have a markedly increased risk of developing emphysema as adults by a loss-of-function mechanism, i.e.,

 Copyright 2004, Elsevier (USA). All rights reserved.

Qu, D., Teckman, J. H., and Permutter, D. H. (1997). α1-Antitrypsin deficiency associated liver disease. *J. Gastroenterol. Hepatol.* **12**, 404–416.

Richardson, P. D., Kren, B. T., and Steer, C. J. (2002). Gene repair in the new age of gene therapy. *Hepatology* **35**, 512–518.

World Health Organization (1997). α1-Antitrypsin deficiency: Memorandum from a WHO meeting. *Bull. World Health Org.* **75**, 397–415.

Wulfsberg, E. A., Hoffmann, D. E., and Cohen, M. M. (1994). α1-Antitrypsin deficiency. *J. Am. Med. Assoc.* **271**, 217–222.

Alpha-1-Antitrypsin (α1AT) Deficiency, Pediatric

JEFFREY H. TECKMAN

Washington University and St. Louis Children's Hospital, Missouri

M Nomenclature signifying the normal allele of the Alpha-1-antitrypsin (α1AT) gene.

PAS-positive, diastase-resistant globule Accumulations of mutant α1-antitrypsin protein within hepatocytes in PIZZ α1AT deficiency that stain red by periodic acid–Schiff and are resistant to diastase digestion.

PI or PI-type Diagnostic phenotype analysis of α1-antitrypsin protein from patient serum by isoelectric-focusing gel electrophoresis.

S Nomenclature signifying a mutant allele of the α1-antitrypsin gene that yields an intermediate level of α1AT deficiency.

Z Nomenclature signifying the most common mutant allele of the α1-antitrypsin gene that yields profound α1AT deficiency.

ZZ or PIZZ Homozygosity for the mutant Z allele of the α1-antitrypsin gene and the definition of classical α1AT deficiency.

Alpha-1-Antitrypsin (α1AT) is a glycoprotein synthesized primarily in the liver and secreted into the blood where its function is to inhibit neutrophil protease-induced host tissue damage. Homozygosity for the Z mutant allele of α1AT (PIZZ) causes the classical form of α1AT deficiency. The protein product of the mutant Z gene accumulates within hepatocytes rather than being efficiently secreted. PIZZ homozygous adults have a markedly increased risk of developing emphysema by a loss-of-function mechanism, i.e., reduced levels of circulating α1AT in the lung to inhibit connective tissue breakdown by neutrophil proteases. PIZZ homozygous children and adults may also develop liver disease and hepatocellular carcinoma by a gain-of-function mechanism; i.e., the accumulation of mutant α1ATZ protein within hepatocytes is toxic to liver cells.

CLASSICAL PIZZ α1-ANTITRYPSIN DEFICIENCY

Pathophysiology and Genetics

Classical, PIZZ α1-antitrypsin deficiency is caused by homozygosity for a point mutation in the α1AT gene encoding substitution of lysine for glutamate at position 342. The 12.2 kb α1AT gene is located on chromosome 14q and encodes a 55 kDa glycoprotein of 395 amino acids. PIZZ homozygotes occur at a frequency of 1 in 1500–3500 in North American and European populations. It is the most common genetic cause of liver disease in children and the most common genetic disease leading to liver transplantation in children and it can also cause chronic liver disease, hepatocellular carcinoma, and premature pulmonary emphysema in adults.

The Z mutation confers polymerogenic properties on the mutant α1ATZ protein molecule, which is primarily synthesized in the liver but is also found in leukocytes and other tissues. The mutant α1ATZ protein is retained in the endoplasmic reticulum (ER) of hepatocytes rather than being secreted into the blood and body fluids where its normal function is to inhibit neutrophil proteases and thereby protect host tissues from inflammation-induced damage. PIZZ homozygous individuals have a markedly increased risk of developing emphysema as adults by a loss-of-function mechanism, i.e.,

 Copyright 2004, Elsevier (USA). All rights reserved.

reduced levels of circulating α1AT in the lung to inhibit connective tissue breakdown by neutrophil proteases. A subgroup of PIZZ homozygous children or adults may develop liver disease and hepatocellular carcinoma by a gain-of-function mechanism; i.e., the accumulation of polymerized, mutant α1ATZ within the ER is toxic to liver cells.

The key step in the pathophysiology of α1AT deficiency is retention of the newly synthesized mutant Z protein molecule in the ER of hepatocytes. During biogenesis, the nascent mutant Z polypeptide chain is appropriately assembled on the ribosome and translocated into the ER lumen. However, in the ER the mutant Z protein molecule folds slowly and inefficiently into its final, secretion-competent conformation and has a tendency to form unique polymers. A system of proteins within the ER, termed the "quality control" apparatus, recognizes these mutant Z molecules as abnormal and directs them to a series of proteolytic pathways rather than allowing progression down the secretory pathway. The result is a significantly deficient, approximately 15% of normal, serum level of α1AT. Accumulation of the retained mutant Z protein molecules within hepatocytes appears to cause liver injury. A small proportion of the retained molecules may remain in a polymerized conformation and accumulate as aggregations within dilated areas of ER. These accumulations may become so large that they are visible by light microscopy as eosinophilic hepatocellular inclusions by hematoxylin and eosin or as the periodic acid—Schiff (PAS)-positive, diastase-resistant globules classically described within hepatocytes in this disease.

The exact role of protein polymerization in the pathophysiology of α1AT deficiency is still unclear. The presence of the Z mutation allows a protruding surface loop of one molecule to insert into a groove in a neighboring Z molecule. A conformational change in the molecules then occurs, holding them together in the absence of covalent bonds. Long chains of Z protein polymers can form in this way and physical—chemical analysis suggests that this conformation is extremely stable and long-lived. However, it is still unclear whether polymerization, the quality control apparatus, both, or neither is responsible for retention of the mutant Z protein within hepatocytes. The exact mechanism of cellular injury resulting from α1ATZ intracellular retention and/or polymerization is also unclear.

Presentation and Natural History

The presentation of patients with PIZZ α1AT deficiency can be highly varied, ranging from chronic liver disease to fulminant hepatic failure to adult emphysema. In infancy, the typical presentation is one of neonatal cholestasis (also called the neonatal hepatitis syndrome) and may include symptoms and signs of jaundice, abdominal distension, pruritis, poor feeding, poor weight gain, hepatomegaly, and splenomegaly. Laboratory evaluation may reveal elevated total and direct bilirubin, elevated serum aspartate aminotransferase (AST) and alanine aminotransferase (ALT), hypoalbuminemia, or coagulopathy due to vitamin K deficiency or to liver synthetic dysfunction. Liver biopsy findings may be highly variable in infants including giant cell transformation, lobular hepatitis, fibrosis, hepatocellular necrosis, bile duct paucity, or bile duct proliferation. The PAS-positive, diastase-resistant hepatocellular globules characteristic of this disease may occasionally be absent in very young infants. Although life-threatening liver disease can occur in the first few years of life, prospective, population-based studies indicate that 80% of PIZZ patients with neonatal cholestasis are healthy and free of chronic liver disease by the age of 18 years. These data suggest that the overall risk of life-threatening liver disease in childhood may be as low as 3%, but that the risk of varying degrees of liver dysfunction in children may range from 15 to 60%.

In toddlers and older children, PIZZ α1AT deficiency may present as failure to thrive, possibly with poor feeding or hepatomegaly. Some children come to medical attention when asymptomatic hepatomegaly or splenomegaly is detected during routine medical checkups. Many children appear to be completely healthy, without evidence of liver injury, except for mild and usually clinically insignificant elevations in serum AST or ALT. Occasionally, children with previously unrecognized chronic liver disease and cirrhosis present with ascites, gastrointestinal bleeding, or hepatic failure. However, many PIZZ children with cirrhosis may remain stable and grow and develop normally for a decade or more before beginning a process of decompensation.

Liver disease in adults may present as chronic hepatitis, with or without cirrhosis, and the risk of clinically significant disease may increase with advancing age. The findings may be similar to those of adult alcoholic liver disease, which may lead to diagnostic confusion. There appears to be an increased risk of hepatocellular carcinoma in adults with PIZZ α1AT deficiency, although the magnitude of the risk is unclear. Liver biopsy findings in older children and adults may include lobular inflammation, variable hepatocellular necrosis, fibrosis, cirrhosis, and PAS-positive, diastase-resistant globules in some, but not all, hepatocytes in nearly every patient.

Children with PIZZ α1AT deficiency generally do not develop clinically detectable emphysema, although

they may be at increased risk for severe childhood asthma. Asthma or emphysema may develop during adulthood and chronic pulmonary disease is the most common life-threatening complication of PIZZ α1AT deficiency in adults. Smoking significantly increases the risk of progressive, irreversible lung disease as do occupational lung exposures and environmental atmospheric pollutants, all of which are thought to increase the uninhibited proteolytic attack on lung connective tissue. Some studies suggest that exposure to secondhand smoke and environmental air pollutants in childhood is also a significant risk for the development of adult lung disease.

Diagnosis

The gold standard for the diagnosis of α1AT deficiency is the analysis of the phenotype ("PI" or "PI-type") of α1AT protein in a sample of patient serum by isoelectric-focusing gel electrophoresis. It is then inferred from the PIZZ phenotype results that the patient is carrying two copies of the mutant Z α1AT gene. The phenotype gel analysis is technically demanding and is therefore best performed in an experienced reference laboratory. Since the presentations of α1AT deficiency are quite variable and the serum testing is of relatively low expense and risk, α1AT serum phenotyping is applied in a wide variety of clinical situations (Table I). Measurement of the level of α1AT in peripheral blood can be used as a complementary test to compare the phenotype result against what would be an appropriate predicted level (Table II) and to assist in the elucidation of unusual alleles whose protein products yield confusing phenotype results. Liver biopsy is not required for the diagnosis of α1AT deficiency, although it may be useful in selected cases to evaluate disease progression or to investigate the contribution of co-morbid states.

Management

There is no specific treatment for the liver disease associated with α1AT deficiency. Management focuses on preventing the complications of chronic liver disease, such as bleeding, ascites, pruritis, malnutrition, fat-soluble vitamin deficiency, infection, and growth disturbances, or attenuating the systemic repercussions if they do occur. Some authorities advocate aggressive treatment of any systemic infectious or inflammatory episodes, as there is concern that these may be important in increasing end-organ injury. However, many patients have normal health and can be monitored conservatively with infrequent visits to a physician knowledgeable in liver disease. Some patients with significant degrees of liver injury, and even cirrhosis, often remain stable for many years with very little intervention. If life-threatening liver disease does develop, then liver transplantation has been employed with excellent published success rates.

All pediatric and adult patients with α1AT deficiency should be urgently cautioned against personal smoking, secondhand smoke, and environmental lung exposures. Prospective studies indicate that identification of α1AT-deficient patients as children dramatically reduces their incidence of smoking as adults and therefore decreases morbidity and mortality from lung disease. Exogenous α1AT protein replacement is available as a treatment for the adult emphysema associated with α1AT deficiency; however, the efficacy of the therapy is controversial. Exogenous α1AT replacement has no effect on the development of liver disease since liver injury is related to the accumulation of the α1AT mutant Z protein within hepatocytes, not a lack of circulating anti-protease activity.

TABLE I Clinical Presentations and Indications for Testing for α1AT Deficiency

Infant	Cholestatic jaundice
Child	Unexplained failure to thrive or poor feeding
Any age	Unexplained, asymptomatic hepatomegaly, or elevated AST/ALT
Any age	Unexplained liver disease, cirrhosis, or hepatocellular carcinoma
Adult	Severe asthma, any emphysema <50 years old or any age in a nonsmoker

TABLE II α1AT Phenotypes and Corresponding Typical α1AT Serum Levels

Phenotype	Level (μM)[a]
PIMM	20–48
PIMZ	12–35
PISS	15–33
PISZ	8–19
PIZZ	2.5–7.0
Null–Null	0.0

[a] Convert micromolar to milligrams per deciliter by multiplying by conversion factor of 5.2.

HETEROZYGOSITY AND OTHER ALLELES

Individuals who are heterozygous for α1AT, carrying one normal M allele and one mutant Z allele ("PIMZ" or "MZ") are generally considered asymptomatic and healthy. Data from referral center studies suggest that

some rare PIMZ adults may develop liver disease, although the possible genetic or environmental influences on the development of this injury remain unclear. Limited but unselected population-based studies have thus far failed to confirm this increased risk. PIMZ children appear to be completely healthy, and even in adults a PIMZ phenotype result is not readily accepted as the cause of otherwise unexplained liver disease.

Individuals who are heterozygous PISZ and who have developed liver disease identical to PIZZ patients, including PAS-positive, diastase-resistant globules, have been clearly described. However, the risk of liver disease to PISZ individuals, though of unclear absolute magnitude, appears to be less than the risk to PIZZ individuals. PISZ individuals may also be at risk for adult emphysema. In contrast, PISS individuals are generally accepted as normally healthy.

A large number of other mutations in the α1AT gene have been described. Some of these gene products yield a normal M result on the phenotype test but when present in the heterozygous state with a Z allele can accumulate within the liver and have been associated with liver disease. Such patients are usually recognized by a profoundly low α1AT level in peripheral blood, which is inappropriate when compared to an apparently PIMZ phenotype result. Null mutations of the α1AT gene that produce no protein secreted into the peripheral blood, resulting in unusually low blood levels and confounding phenotype results, have also been described. Several pilot newborn screening programs for α1-antitrypsin deficiency have been tested in the United States and in Europe. However, universal, population-based screening is still not recommended because there is not yet a proven, immediate health benefit or treatment available to asymptomatic newborns that would outweigh the risk of psychological trauma or genetic discrimination.

See Also the Following Articles

Alpha-1-Antitrypsin (α1AT) Deficiency • Cirrhosis • Hepatocellular Carcinoma • Neonatal Cholestasis and Biliary Atresia

Further Reading

Carrell, R. W., and Lomas, D. A. (2002). Alpha-1-Antitrypsin deficiency—Model for conformational diseases. *N. Engl. J. Med.* **346**, 45–53.

Cox, D. W. (1995). Alpha-1-Antitrypsin deficiency. *In* "The Molecular and Metabolic Basis of Inherited Disease" (C. R. Scriver and A. L. Beaudet, eds.), pp. 4125–4158. McGraw-Hill, New York.

Eriksson, S., Propst, A., Sveger, T., and Teckman, J. H. Liver and other diseases in Alpha-1-antitrypsin deficiency. *In* "Standards for the Diagnosis and Management of Individuals with Alpha-1-Antitrypsin (AAT) Deficiency: A Report of the Alpha-1-Antitrypsin Deficiency Task Force." [*Am. J. Resp. Crit. Care Med.*] In press.

Teckman, J. H., and Perlmutter, D. H. (1999). Metabolic disorders of the liver. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (R. Wyllie and J. S. Hyams, eds.), pp. 579–599. W. B. Saunders, Philadelphia, PA.

Teckman, J. H., and Perlmutter, D. H. (1996). Molecular pathogenesis of liver disease in Alpha-1-antitrypsin deficiency. *Hepatology* **24**, 1504–1516.

World Health Organization (1997). Alpha-1-Antitrypsin deficiency: Memorandum from a WHO meeting. *Bull. World Health Organ.* **75**, 397–415.

Amebiasis

SAMUEL L. STANLEY, JR.
Washington University, St. Louis, Missouri

abscess A localized collection of pus in a part of the body, surrounded by inflamed tissue.
amebic colitis Intestinal disease caused by *Entamoeba histolytica*.
amebic liver abscess The major extraintestinal manifestation of amebiasis, which is fatal if not recognized and treated appropriately.
Entamoeba dispar Morphologically identical commensal that is not associated with disease, but may lead to incorrect diagnoses of amebiasis.
Entamoeba histolytica The intestinal protozoan parasite that causes amebiasis.
toxic megacolon Dilation of the colon during the course of fulminant colonic inflammation.

Amebiasis remains a major cause of morbidity and mortality worldwide and is responsible for as many as 100,000 deaths yearly. The details of the relationship between the causative agent, *Entamoeba histolytica*, and its human host remain the subject of much study and new insights into the role of the host response to infection are emerging rapidly. The recognition that what was previously called *E. histolytica* based on morphology is really two distinct species, *E. histolytica*, the pathogen, and *Entamoeba dispar*, a harmless commensal, has forced the rethinking of approaches to the diagnosis and treatment of infected individuals. Fortunately, the problem of drug resistance that complicates the treatment of many pathogenic microorganisms has not emerged for *E. histolytica* and the nitroimidazoles remain very effective therapy for amebiasis.

INTRODUCTION

Amebiasis was known to the Greeks, and Hippocrates wrote, "Dysenteries, when they set in with fever, alvine discharges of a mixed character, or with inflammation of the liver . . . are bad." However, it was not until 1875, when the St. Petersburg physician Fedor Aleksandrovich Lösch described amebic trophozoites in the stool and colonic ulcerations of a farmer with dysentery, that ameba were established as a cause of colitis. Since that initial description, much has been learned about the biology of the causative agent, *Entamoeba histolytica*, and effective therapies for this potentially deadly disease have been developed.

THE ORGANISM

E. histolytica is a single-celled eukaryote that lacks mitochondria (probably through secondary loss) and derives energy by the anaerobic conversion of glucose or pyruvate to ethanol. *E. histolytica* has a simple life cycle, existing as either the quadranucleate cyst form (Fig. 1) or the active, vegetative trophozoite form (Fig. 2). The cyst is the infectious form and leads to colonization and disease when it is ingested in food or water contaminated with human feces. The cysts survive the gastric acidity of the stomach and, under yet to be determined stimuli, excyst, forming the trophozoite stage. It is the highly motile trophozoites that are capable of colonizing the colon, invading into the colonic mucosa, and causing disease. The life cycle continues when the trophozoites encyst and the infectious cysts are passed out of the human host in feces.

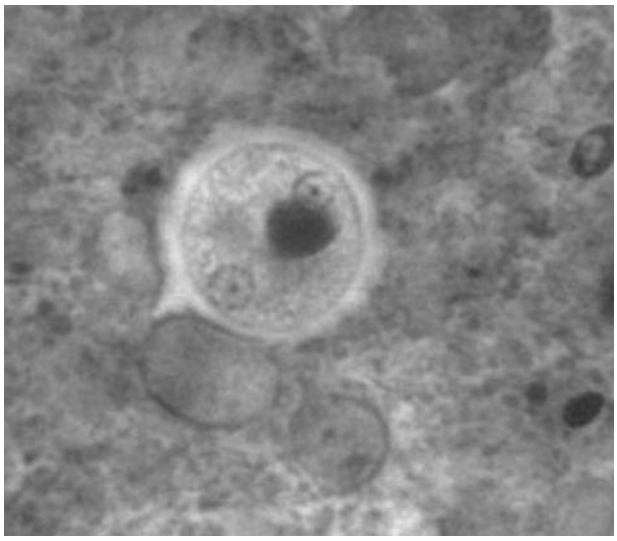


FIGURE 1 *E. histolytica* cyst in stool; chlorazol black stain. Two of the four nuclei, each with the characteristic central karyosome, are clearly visible. Courtesy of Dr. George Healy, Centers for Disease Control.

 Copyright 2004, Elsevier (USA). All rights reserved.

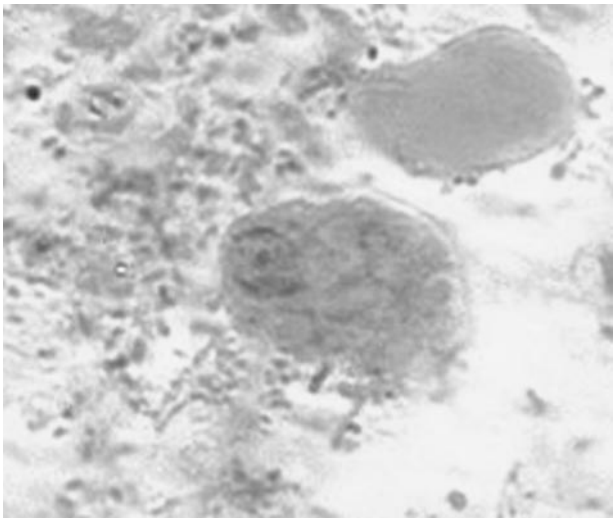


FIGURE 2 *E. histolytica* trophozoite in stool; Wheatley's trichrome stain. Note the round nucleus and central karyosome. Courtesy of Centers for Disease Control.

EPIDEMIOLOGY

E. histolytica is cosmopolitan in distribution and is found wherever there are poor boundaries between human feces and food and drinking water. *E. histolytica* infects only humans and perhaps some nonhuman primates. Data on the worldwide prevalence of *Entamoeba histolytica* infection are difficult to come by, complicated by the recent recognition that what was previously considered to be *E. histolytica*, based on microscopy, is a mixture of two species, *E. histolytica*, a pathogen, and *E. dispar*, a harmless commensal. A reasonable estimate, based on newer studies using techniques that can differentiate between *E. histolytica* and *E. dispar*, is that worldwide approximately 450,000,000 individuals are infected with *E. dispar*, whereas 50,000,000 individuals are infected with *E. histolytica*. Many of the individuals infected with *E. histolytica* are asymptomatic, but have an approximately 10% chance yearly of developing disease. Looking at disease, rather than infection rates, provides a more meaningful look at the impact of amebiasis worldwide. As many as 100,000 people die yearly from amebiasis, making it the second leading cause of death from parasitic diseases. In Mexico, more than 1.3 million cases of intestinal amebiasis were reported in a single year, and in one region of Vietnam, a population center of 1 million people experienced 1500 cases of amebic liver abscess over a 5-year period. In the United States, amebiasis is primarily seen among immigrants, with most cases seen in states that border Mexico. Amebic colitis can be seen in all ages and both sexes, but amebic liver abscess is primarily a disease of men between the ages of 18 and 50, with rates that are 3 to 20 times higher than other populations. The explanation for the increased susceptibility of young and middle-aged men is unknown, but hormonal influences (postmenopausal women also have a higher rate) seem most likely.

PATHOGENESIS

Amebic colitis begins when amebic trophozoites adhere to epithelial cells in the colonic mucosa. Adhesion is primarily mediated by a surface lectin that recognizes N-terminal galactose and *N*-acetyllactosamine residues. *E. histolytica* trophozoites can lyse human cells on contact through the action of amoebapores, small peptides that assemble to form pores in the surface membranes of eukaryotic and bacterial cells. *E. histolytica* trophozoites also secrete abundant quantities of cysteine proteinases, which can lyse extracellular matrix proteins and facilitate amebic invasion into the submucosal tissue. *E. histolytica* trophozoites invade laterally through the submucosal spaces, creating the classic flask-shaped ulcer (Fig. 3) of amebiasis. There can be a marked inflammatory response to the parasite and amebic colitis may be difficult to distinguish from inflammatory bowel disease. Colonic findings can range from diffuse mucosal thickening, multiple discrete ulcers, diffusely inflamed and edematous mucosa, to necrosis and perforation of the intestinal wall (Fig. 4). In a human intestinal xenograft model of amebic colitis, the host inflammatory response is dependent on the activation of the transcription factor nuclear factor κB in intestinal epithelial cells, and the actions of interleukin-8 (IL-8), IL-1, tumor necrosis factor α, and cyclooxygenase 2. Although the inflammatory response to *E. histolytica* may actually exacerbate tissue damage, innate immunity is probably key to containing the infection, and

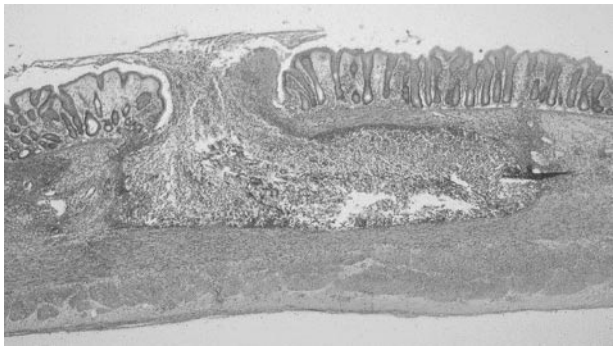


FIGURE 3 Classic flask-shaped ulcer of amebiasis showing mucosal ulceration with widespread submucosal invasion. Courtesy of Dr. Mae Melvin, Centers for Disease Control.

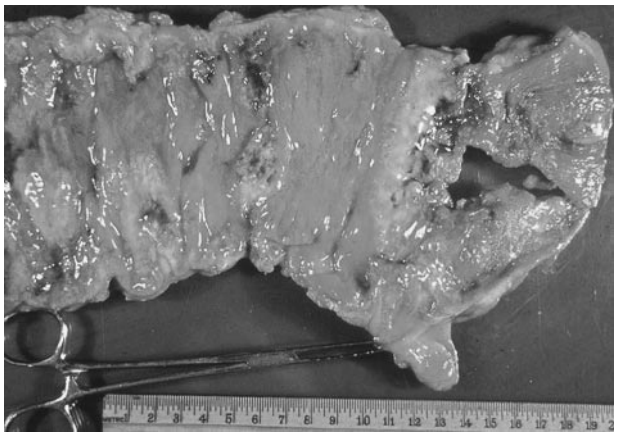


FIGURE 4 Gross pathology of amoebic colitis showing multiple ulcer formation. Courtesy of Dr. Mae Melvin, Centers for Disease Control.

individuals with amebic colitis who were mistakenly given corticosteroids for what was thought to be inflammatory bowel disease had worse outcomes with an increased frequency of fulminant amebic colitis and amebic liver abscess. In approximately 10% of individuals with amebic colitis, *E. histolytica* trophozoites reach the portal circulation and cause amebic liver abscesses, well-circumscribed areas of hepatocyte death and liquefied debris (Fig. 5).

CLINICAL FEATURES

Individuals with amebic colitis usually present with the gradual onset of bloody diarrhea, abdominal pain, and abdominal tenderness. Some individuals will have multiple small-volume mucoid stools and others will have watery diarrhea, but, because *E. histolytica* is invasive, stools almost always contain blood. Fever is seen in less than 40% of patients; some individuals will report weight loss and anorexia. Fulminant amebic colitis, characterized by profuse bloody diarrhea, fever, marked leukocytosis, peritoneal signs, and extensive colonic involvement, occurs rarely, but carries a high mortality. Pregnant women, individuals treated with corticosteroids, malnourished people, and immunocompromised individuals are clearly at higher risk for this complication. Toxic megacolon has been reported with amebiasis, and amebomas, localized inflammatory masses that may mimic carcinomas by causing mass obstructing lesions, may also complicate amebic colitis.

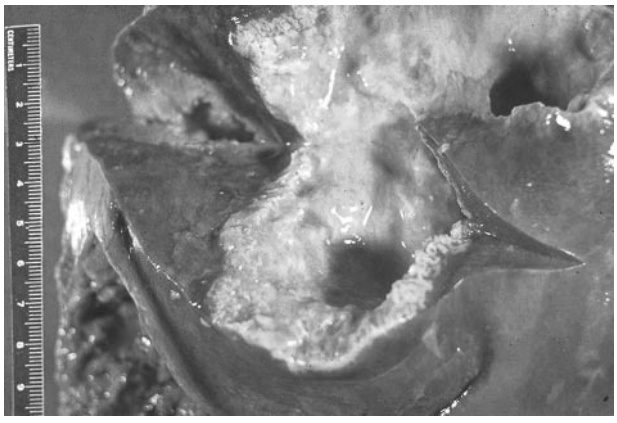


FIGURE 5 Gross pathology of amoebic liver abscess. Section showing necrotic material within the abscess and the surrounding fibrinous border. The adjacent liver parenchyma is usually normal. Courtesy of Dr. Mae Melvin, Centers for Disease Control and Dr. E. West of Mobile, Alabama.

Amebic liver abscesses arise from the hematogenous spread of amebic trophozoites from the colon to the liver and are the most common extraintestinal manifestation of amebiasis outside the intestinal tract. Individuals present with the classic triad of fever, right upper quadrant abdominal pain, and hepatic tenderness. Symptoms are usually acute (onset within the past 10 days), but some individuals present with a more chronic disease with associated weight loss and anorexia. Individuals with amebic liver abscess may present years after travel or residency in an endemic area so a careful travel history may be key to making the diagnosis. Most patients with amebic liver abscess do not have evidence for concurrent intestinal infection based on the microscopic examination of stool, but more sensitive diagnostic methods [e.g., polymerase chain reaction (PCR)] suggest that concurrent intestinal infection may be relatively common. Cough may be present and dullness and rales in the right lung base may be seen. Jaundice is unusual, as is eosinophilia, but leukocytosis and elevated alkaline phosphatase are relatively common.

The most common complication of amebic liver abscess is rupture into the pleural space, with the formation of an amebic empyema, a hepatobronchial fistula (where individuals cough up the contents of their amebic liver abscess), and/or an amebic lung abscess. Pulmonary extension of the amebic liver abscess may give rise to a clinical syndrome (productive cough, fever, chest pain) that may be confused with pneumonia. Less common, but more dangerous is amebic liver abscess rupture into the peritoneum, with associated peritonitis and shock symptoms, and amebic liver abscess rupture into the pericardium, with tamponade and/or pericarditis. Mortality associated with pericardial involvement is very high (30%).

Other extraintestinal manifestations of amebiasis are very rare, with amebic brain abscesses (seen in less than 0.1% of individuals with amebic liver

abscesses), urinary tract involvement, genital disease, perianal disease, and cutaneous lesions all reported.

DIAGNOSIS

For years the diagnosis of amebic colitis was based on the microscopic demonstration of amebic trophozoites in the stool of an individual with diarrhea. Microscopy is still used in much of the world, but the recognition that microscopy cannot distinguish between infection with *E. histolytica* and the more common commensal *E. dispar* has changed the approach to diagnosis and emphasized the need for more specific and more sensitive ways to identify *E. histolytica* intestinal infection. In the proper clinical setting, an individual with bloody diarrhea, appropriate exposure history, and a microscopic exam that shows *E. histolytica* trophozoites that have ingested red blood cells, stool microscopy may still be diagnostic for amebic colitis. However, given the high prevalence of *E. dispar* in many areas, relying solely on microscopy may lead to the false diagnosis of amebiasis in cases of *Shigella*, *Campylobacter jejuni*, or other forms of dysentery. One solution to this problem has been the development of *E. histolytica*-antigen detection enzyme-linked immunosorbent assay (ELISA) tests that use antibodies to recognize specific *E. histolytica* antigens in stool and can distinguish between *E. histolytica* and *E. dispar*. These tests performed well in initial studies and may become the diagnostic method of choice, but recent reports that some of the ELISA tests were relatively insensitive in field studies (compared with PCR or stool culture) provide a cautionary note. Molecular diagnostics using PCR to amplify *E. histolytica*-specific sequences from stool may become more widespread—these offer the advantage of allowing genotyping for strain identification. The major barrier for PCR-based tests is the requirement for specialized equipment and reagents that will be unavailable in many of the countries where the tests are needed the most.

The diagnosis of amebic liver abscess is based on the demonstration of a space-occupying lesion in the liver and a positive amebic serology. Both ultrasound and computed tomography scanning (Fig. 6) are effective in detecting amebic liver abscesses. Abscesses can be solitary or multiple spherical lesions and are most commonly found in the right lobe of the liver. Amebic serology is very sensitive (>90%) and very specific (>98%) and the only caveat is that serology may be negative very early in disease (within the first week) and should be repeated if the index of suspicion for amebic liver abscess is high.

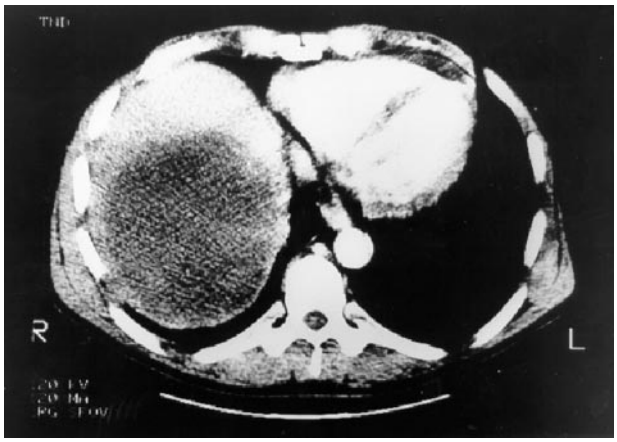


FIGURE 6 CT scan from a patient with amebic liver abscess. A large solitary lesion in the right lobe of the liver is present.

TREATMENT

The mainstay of therapy for amebiasis remains the nitroimidazole compounds. Since only metronidazole is available in the United States, the recommendations listed in Table I are focused on metronidazole. Amebic colitis is treated with metronidazole, followed by a lumenal agent to eliminate colonization. Asymptomatic individuals infected with *E. histolytica* should be treated with a lumenal agent to eliminate colonization. This is based on the fact that they are at risk to develop invasive disease and because they pose a risk of spreading *E. histolytica* infection to others. If an individual is found to be colonized with *E. dispar*, no therapy is necessary, but the physician should be alerted to the fact that the patient has ingested fecally contaminated food or water.

Amebic liver abscess is also treated with metronidazole, and, remarkably, given the size of the abscess, single-dose therapy can be used. A lumenal agent should be administered as well to eradicate intestinal colonization, even if stool microscopy is negative. Unlike most abscesses, amebic abscesses can resolve without drainage and percutaneous aspiration and drainage should be reserved for diagnostic purposes (if a bacterial abscess or suprainfection is suspected), for large abscesses in the left lobe of the liver (because of the risk of rupture into the pericardium), when individuals are not responding to therapy (continued pain, fever after 72 h of treatment), and when rupture seems imminent (large abscess, accelerated clinical course with increasing pain). Some authorities advocate the addition of dehydroemetine in complicated cases of amebic colitis or amebic liver abscess because of its rapid amebicidal activity, but there are no controlled

TABLE I Drugs of Choice for the Treatment of Amebic Colitis and Amebic Liver Abscess

Drug	Adult dosage	Adverse effects	Comments
Metronidazole	750 mg po or iv tid for 5 to 10 days; for uncomplicated liver abscess (limited experience) 2.4 g po daily for 2 days	Metallic aftertaste, nausea, vomiting, diarrhea; rarely—sensory neuropathies, central nervous system toxicity with ataxia, vertigo, seizures, and encephalopathy	Drug of choice for amebic colitis and amebic liver abscess
Dehydroemetine	1−1.5 mg/kg/day im for up to 5 days	Cardiotoxicity, diarrhea, nausea, vomiting, muscle weakness	No indication for use in standard therapy; may offer some benefit in fulminant colitis or patients with ruptured amebic liver abscess when administered in combination with metronidazole, but controlled trials are lacking
Lumenal agents			
Paromomycin	30 mg/kg/day po in three divided doses for 5 to 10 days	Nausea, vomiting, cramps, diarrhea	Drug of choice for treatment of lumenal *E. histolytica* infection; should be administered to all individuals following completion of metronidazole therapy
Diloxanide furoate	500 mg po tid for 10 days	Flatulence	Excellent alternative to paromomycin for treatment of lumenal *E. histolytica* infection; should be administered to all individuals following completion of metronidazole therapy; not readily available in the United States
Iodoquinol	650 mg po tid for 20 days	Headache, nausea, vomiting; optic nerve damage and peripheral neuropathy reported in individuals exceeding recommended dosage	Alternative to paromomycin or diloxanide furoate for treatment of lumenal *E. histolytica* infection; should be administered to all individuals following completion of metronidazole therapy

studies indicating that it offers specific advantages over metronidazole therapy alone.

See Also the Following Articles

Liver Abscess • Parasitic Diseases, Overview

Further Reading

Haque, R., Ali, I. M., Sack, R. B., Farr, B. M., Ramakrishnan, G., and Petri, W. A., Jr. (2001). Amebiasis and mucosal IgA antibody against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. *J. Infect. Dis.* **183**, 1787–1793.

Ravdin, J. I. (1988). "Amebiasis: Human Infection by *Entamoeba histolytica*." Wiley, New York.

Ravdin, J. I. (2000). "Amebiasis." Imperial College Press, London.

Reeves, R. E. (1984). Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv. Parasitol.* **23**, 105–142.

Seydel, K. B., Li, E., Zhang, Z., and Stanley, S. L., Jr. (1998). Epithelial cell-initiated inflammation plays a crucial role in early tissue damage in amebic infection of human intestine. *Gastroenterology* **115**, 1446–1453.

Seydel, K. B., and Stanley, S. L., Jr. (1998). *Entamoeba histolytica* induces host cell death in amebic liver abscess by a non-Fas-dependent, non-tumor necrosis factor α-dependent pathway of apoptosis. *Infect. Immun.* **66**, 2980–2983.

Stanley, S. L., Jr. (2003). Amoebiasis. *Lancet* **361**, 1025–1034.

Stanley, S. L., Jr., and Reed, S. L. (2001). VI. Entamoeba histolytica: Parasite–host interactions. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G1049–G1054.

Amylase

MARK E. LOWE

Washington University School of Medicine, St. Louis, Missouri

amylopectin Branched-chain glucose polymer; the major component of dietary starch.

amylose Straight-chain glucose polymer present in starch.

isoenzymes Multiple forms of the same enzyme; have subtle differences in amino acid sequence or posttranslational modifications.

posttranslational modification Alteration of a protein after it has been synthesized. Examples include the addition of sugar chains, of phosphate, or of sulfate.

Amylase is a digestive enzyme secreted primarily by the pancreas and some salivary glands. Because of the almost exclusive production of amylase by the pancreas, serum amylase levels are of diagnostic importance in assessing acute pancreatitis.

INTRODUCTION

The digestive enzyme amylase secreted by the pancreas and some salivary glands is responsible for the initial process of digestion of dietary starch. In the Western world, adults consume about 400 g of carbohydrates each day, with starches and sucrose providing the largest sources. Starch is the storage form of carbohydrate in plants and can account for 10–80% of the plant volume. All starches are glucose polymers with molecular masses ranging from 10^5 to more than 10^6 Da. The two major starches are amylose, a straight-chain α-1,4-linked glucose polymer, and amylopectin, a branched starch with a backbone of α-1,4-linked glucose and α-1,6-linked glucose branches about every 20–25 residues. Amylose and amylopectin account for 20 and 80%, respectively, of dietary starch. Because the intestinal epithelium absorbs only monosaccharides, dietary starch, to serve as a nutrient and energy source, must first be hydrolyzed into glucose, a process facilitated by α-amylase.

PHYSIOLOGY

In humans and other primates and in rodents and lagomorphs (rabbits, hares, and pikas), pancreatic acini and certain salivary glands, primarily the parotids, secrete amylase. No other tissues express significant levels of amylase, although amylases are present in the fallopian

 Copyright 2004, Elsevier (USA). All rights reserved.

studies indicating that it offers specific advantages over metronidazole therapy alone.

See Also the Following Articles

Liver Abscess • Parasitic Diseases, Overview

Further Reading

Haque, R., Ali, I. M., Sack, R. B., Farr, B. M., Ramakrishnan, G., and Petri, W. A., Jr. (2001). Amebiasis and mucosal IgA antibody against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. *J. Infect. Dis.* **183**, 1787–1793.

Ravdin, J. I. (1988). "Amebiasis: Human Infection by *Entamoeba histolytica*." Wiley, New York.

Ravdin, J. I. (2000). "Amebiasis." Imperial College Press, London.

Reeves, R. E. (1984). Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv. Parasitol.* **23**, 105–142.

Seydel, K. B., Li, E., Zhang, Z., and Stanley, S. L., Jr. (1998). Epithelial cell-initiated inflammation plays a crucial role in early tissue damage in amebic infection of human intestine. *Gastroenterology* **115**, 1446–1453.

Seydel, K. B., and Stanley, S. L., Jr. (1998). *Entamoeba histolytica* induces host cell death in amebic liver abscess by a non-Fas-dependent, non-tumor necrosis factor α-dependent pathway of apoptosis. *Infect. Immun.* **66**, 2980–2983.

Stanley, S. L., Jr. (2003). Amoebiasis. *Lancet* **361**, 1025–1034.

Stanley, S. L., Jr., and Reed, S. L. (2001). VI. Entamoeba histolytica: Parasite–host interactions. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G1049–G1054.

Amylase

MARK E. LOWE

Washington University School of Medicine, St. Louis, Missouri

amylopectin Branched-chain glucose polymer; the major component of dietary starch.

amylose Straight-chain glucose polymer present in starch.

isoenzymes Multiple forms of the same enzyme; have subtle differences in amino acid sequence or posttranslational modifications.

posttranslational modification Alteration of a protein after it has been synthesized. Examples include the addition of sugar chains, of phosphate, or of sulfate.

Amylase is a digestive enzyme secreted primarily by the pancreas and some salivary glands. Because of the almost exclusive production of amylase by the pancreas, serum amylase levels are of diagnostic importance in assessing acute pancreatitis.

INTRODUCTION

The digestive enzyme amylase secreted by the pancreas and some salivary glands is responsible for the initial process of digestion of dietary starch. In the Western world, adults consume about 400 g of carbohydrates each day, with starches and sucrose providing the largest sources. Starch is the storage form of carbohydrate in plants and can account for 10–80% of the plant volume. All starches are glucose polymers with molecular masses ranging from 10^5 to more than 10^6 Da. The two major starches are amylose, a straight-chain α-1,4-linked glucose polymer, and amylopectin, a branched starch with a backbone of α-1,4-linked glucose and α-1,6-linked glucose branches about every 20–25 residues. Amylose and amylopectin account for 20 and 80%, respectively, of dietary starch. Because the intestinal epithelium absorbs only monosaccharides, dietary starch, to serve as a nutrient and energy source, must first be hydrolyzed into glucose, a process facilitated by α-amylase.

PHYSIOLOGY

In humans and other primates and in rodents and lagomorphs (rabbits, hares, and pikas), pancreatic acini and certain salivary glands, primarily the parotids, secrete amylase. No other tissues express significant levels of amylase, although amylases are present in the fallopian

 Copyright 2004, Elsevier (USA). All rights reserved.

tube, lungs, tears, sweat, and human milk. Amylase represents about 5–6% of the total protein in pancreatic secretions and, along with mucins, amylase is a major secretory protein of the parotid gland. Closely related, but distinct, genes encode salivary and pancreatic amylase. In the human genome, there are three salivary genes and two pancreatic genes encoding amylase. Expression of the genes encoding human pancreatic amylase is developmentally regulated. Human newborns have little to no pancreatic amylase at birth and the levels remain low (< 1.0% of adult levels) throughout the first months of life and may not reach adult levels until the second or third year of life.

ENZYMOLOGY

α-Amylase, an endoenzyme, preferentially cleaves interior α-1,4 linkages and has very low activity against the bonds of terminal glucose units. Additionally, it cannot hydrolyze the α-1,6 linkages in amylopectin. The resulting products of amylase acting on starch, referred to as dextrins, are α-1,4-linked glucose dimers (maltose), α-1,4-linked glucose trimers (maltotriose), and branched oligosaccharides of 6 to 8 glucose units that contain both α-1,6 and α-1,4 linkages (limit dextrins). Starch digestion can begin in the mouth and in a swallowed bolus of food, but primarily occurs in the lumen of the upper small intestine. Digestion of starch is completed in the intestine by the brush border enzymes, maltase and isomaltase.

The active site of α-amylase contains multiple subsites, each of which is capable of binding one glucose residue of the substrate. The porcine and human enzymes appear to have five subsites, and subsite three is probably the catalytic site. Substrates can bind with the first glucose residue in subsite one or two so that cleavage can occur between the first and second or second and third residues. During a single enzyme–substrate encounter, multiple glucose bonds are cleaved. Three acidic residues, one glutamic acid and two aspartic acids, are thought to be the catalytic residues. The glutamic acid is believed to be the proton donor and one of the aspartic acids acts as a nucleophile. α-Amylase has an absolute requirement for calcium ions and is activated by anions such as chloride, bromide, iodide, or fluoride. Heavy metals inhibit the enzyme. The importance of serum amylase levels in the diagnosis of acute pancreatitis has generated widespread interest in its assay. Amylase is most commonly measured by absorbance or fluorescence assays in which a labeled substrate is cleaved.

PROTEIN STRUCTURE

Both the proteins and the cDNAs encoding amylase have been isolated from the pancreas and the salivary glands. α-Amylase has a molecular mass of about 57 kDa and contains a single carbohydrate chain, although an unglycosylated form is made in the parotid gland. Chromatographic and electrophoretic methods have demonstrated multiple isoenzymes of both pancreatic and salivary amylase. The presence of multiple genes for both salivary and pancreatic amylase accounts for some of the isoenzymes, but additional isoenzymes are formed by posttranslational changes. The amylases are multidomain proteins consisting of three domains. The substrate-binding site lies in a cleft between domains A and B. Residues in both of these domains contribute to calcium binding near the active site. A chloride ion binds on domain A near the active site cleft.

See Also the Following Articles

Carbohydrate Digestion and Absorption • Pancreatic Digestive Enzymes • Salivary Glands, Physiology

Further Reading

Alpers, D. H. (1994). Digestion and absorption of carbohydrates and proteins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1723–1750. Raven Press, New York.

Lowe, M. (1994). The structure and function of pancreatic enzymes. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1531–1542. Raven Press, New York.

Amyloidosis

MARVIN J. STONE AND MICHAEL J. GUIRL
Baylor University Medical Center, Dallas

amyloid A hyaline eosinophilic substance, as viewed by light microscopy, that is deposited extracellularly in blood vessels and other tissues in a wide variety of disorders. At least 18 distinct proteins that can form amyloid have been identified thus far. All amyloids are fibrillar and look the same under light, polarization, and electron microscopy.

amyloid, cerebral Amyloid occurring in the brain in patients with Alzheimer's disease, in aged Down's syndrome patients, and in individuals with spontaneous cerebral hemorrhage consisting of deposits of β-protein (Aβ) or cystatin C. Prion diseases are due to another protein (PrP^{SC}) that deposits in Creutzfeldt-Jacob disease, bovine spongiform encephalopathy, and other human and animal neurodegenerative disorders.

amyloid, dialysis-related Amyloid occurring in a minority of patients on long-term (7–10 years) hemodialysis due to deposition of β2 microglobulin.

amyloid, hereditary (familial) Familial amyloid polyneuropathy occurs in patients with late-onset (midlife) polyneuropathy and is usually due to deposition of variant transthyretin. Some patients have cardiomyopathy or nephropathy. The disorders are inherited as autosomal dominant conditions. Transthyretin is also the protein found in patients with senile systemic (cardiac) amyloidosis. In some kindreds, other proteins form the amyloid (apolipoprotein A-I, gelsolin, fibrinogen Aα, lysozyme).

amyloid, light chain origin Also known as primary systemic or immunocytic amyloid; this amyloid occurs *de novo*, i.e., without coexisting or preexisting chronic disease. The protein deposited is derived from immunoglobulin light chains produced and secreted by monoclonal plasma cells (plasma cell dyscrasia).

amyloidogenesis The process by which a precursor protein undergoes proteolysis, yielding fragments that assume a β-pleated sheet conformation and deposit extracellularly in blood vessels and various tissues.

amyloid P component A nonfibrillar component of all amyloids that is derived from a normal precursor, serum amyloid P.

amyloid, reactive Also known as secondary systemic amyloid. This type, amyloid A, occurs in patients with chronic infectious or inflammatory disease and in the autosomal recessively inherited disorder, familial Mediterranean fever. It is also the type present in experimentally induced amyloidosis in mice.

Bence Jones protein Free monoclonal immunoglobulin light chains produced and secreted by a single clone of plasma cells. Fifteen to 20% of Bence Jones proteins appear to be amyloidogenic—more frequently in the λ class than in the κclass. Such amyloid is of the amyloid light chain origin type.

β-pleated sheet conformation (β-conformation) The protein conformation common to all amyloids that is responsible for Congo red staining and the fibrillar structure. Proteolysis can convert serum precursor proteins into twisted β-pleated sheet fibrils.

polarization birefringence A common physical property of all amyloid fibrils related to their β-pleated sheet conformation and associated with an apple green color after Congo red staining of involved tissue.

The amyloidoses are diverse disorders characterized by extracellular deposits of various fibrillar proteins in tissues. They constitute one group of an expanding class of conditions referred to as diseases of protein misfolding. Virtually any site in the body can be affected. Many controversies arose about the cause and composition of amyloid for a century after it was named by Rudolph Virchow in 1853. The first major breakthrough occurred in 1959, when electron microscopy showed that this apparently structureless material was actually fibrillar. The ability to solubilize the fibrils enabled subsequent characterization of their major protein constituents. At least 18 distinct proteins have been identified thus far as amyloid precursors in human diseases and it is likely that more will be described in the future. All amyloids share the same physical properties under light, polarization, and electron microscopy.

INTRODUCTION

Historically, amyloidosis was classified according to whether it occurred *de novo* ("primary") or was "secondary" to a recognizable preexisting or coexisting chronic infectious or inflammatory disease. During the past 70 years, rare hereditary amyloid syndromes have been well documented. Most primary and secondary amyloid syndromes are systemic; i.e., they involve more than one organ system. Many patients with hereditary amyloidosis have systemic disease as well. Localized

 Copyright 2004, Elsevier (USA). All rights reserved.

or tumor-like collections of amyloid in various organs also have been described. Recently, new amyloid disorders have been recognized.

Several amyloid syndromes involve the gastrointestinal (GI) tract. The initial presentation and/or the dominant manifestation may be due to GI amyloid deposition. Thus, the amyloid syndromes are of interest to gastroenterologists and other physicians who see patients with GI disease.

DEFINITION AND DIAGNOSIS

Amyloid is an eosinophilic substance that, under the light microscope, has a hyaline appearance and is deposited extracellularly in the walls of small blood vessels and various organs. These deposits, when extensive, interfere with normal function. Multiple proteins can form amyloid, but all share the common physical properties of polarization birefringence after Congo red staining, linear nonbranching fibrils with a diameter of 7.5 to 10 nm by electron microscopy (EM), and a twisted β-pleated sheet conformation by X-ray diffraction. All amyloid deposits also contain a nonfibrillar glycoprotein moiety, the P component. This amyloid P (AP component) is derived from a normal serum precursor [serum amyloid P (SAP)] structurally related to an acute-phase reactant, C-reactive protein. Apolipoprotein E and other proteoglycans are additional nonfibrillar components of amyloid.

The diagnosis of amyloidosis is based on biopsy of involved tissue. Apple green birefringence under polarized light after Congo red staining and the typical fibrillar structure evident by EM constitute the most reliable methods (Fig. 1).

Because the GI tract is an easily accessible biopsy site, tissue sampling from the GI tract is often employed. The sensitivity of endoscopic biopsy is dependent on the presence of amyloid involvement, the site sampled, and adequate tissue procurement. The biopsy should be obtained with standard endoscopic forceps to include submucosa since amyloid is best identified in the walls of small blood vessels. Abdominal fat pad aspirate has a 60–85% sensitivity in systemic amyloidosis and is the initial diagnostic procedure of choice since there is a low risk of complications. Bone marrow biopsies may also reveal the presence of amyloid. Liver and rectal biopsies are associated with a low but definite risk of severe hemorrhage. Whatever the biopsy site, it is important to emphasize that Congo red stains must be performed and examined under polarized light if amyloidosis is suspected.

Once amyloid is recognized on tissue biopsy by either apple green birefringence after Congo red staining or the characteristic fibrillar appearance by EM, it is important to identify the protein deposited since the

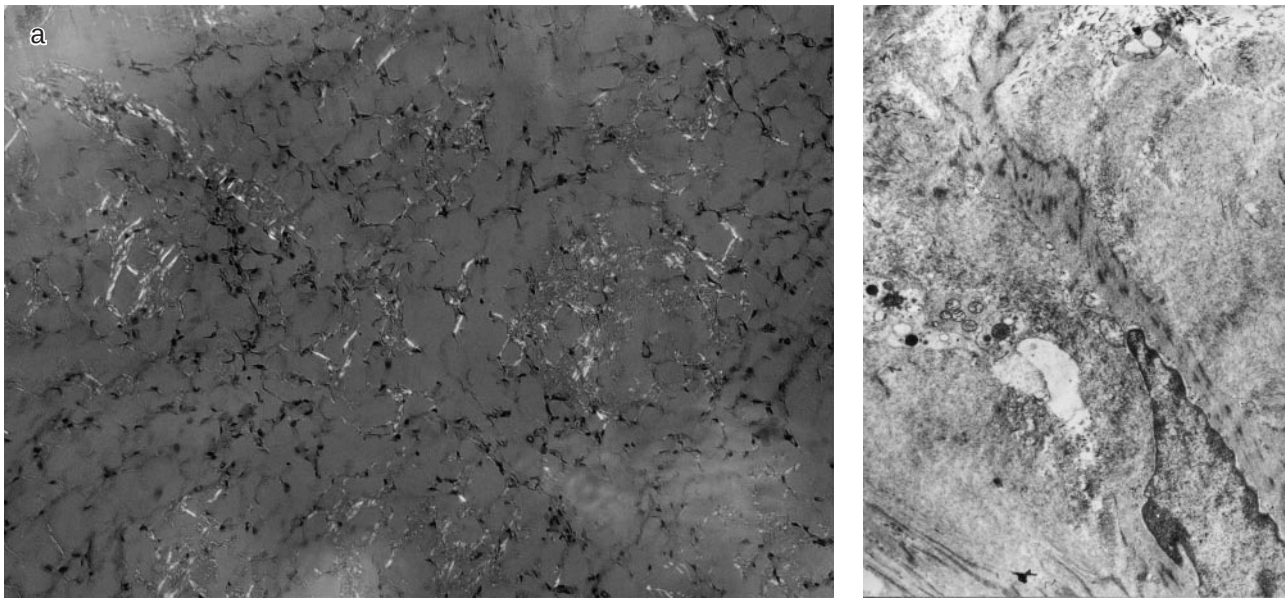


FIGURE 1 (a) Polarization birefringence after staining in subcutaneous fat. The patient had familial amyloid polyneuropathy and chronic diarrhea due to transthyretin amyloidosis. Magnification, ×200. (b) Electron micrograph shows fibrillar structure of amyloid (AL) in a rectal biopsy. Magnification, ×12,000. The majority of the picture shows gray fibrillar material (amyloid).

natural history and therapy vary among the different disease entities. Immunohistochemistry using antibodies to amyloid precursor proteins, such as immunoglobulin light (L) chains, serum amyloid A, transthyretin, and β2 microglobulin, is helpful, though not always definitive, in delineating the type of amyloid present in the fibrils of a particular patient. Immune EM and/or chemical analysis of extracted fibrillar protein may be necessary for unequivocal diagnosis. Radiolabeled SAP component scintigraphy has been shown to be useful in staging systemic AL (amyloid of light chain origin) and AA (amyloid A) amyloidoses, but is not generally available.

In most circumstances, a circulating precursor protein results from overproduction of either intact or aberrant molecules (plasma cell dyscrasias), reduced degradation or excretion (secondary amyloid syndromes and patients on long-term hemodialysis), or genetic abnormalities associated with variant proteins (familial autosomal dominant polyneuropathies). Amyloidogenesis is characterized by proteolysis of a larger protein precursor molecule with production of low-molecular-weight fragments that polymerize and assume a β-conformation as extracellular tissue deposits (Fig. 2). Thus, amyloidosis is a generic term referring to a final common pathophysiologic pathway for tissue protein deposition in a wide variety of diseases.

Except for their similar morphologic and physical properties, the various amyloid diseases are disparate and occur in diverse clinical settings. The classification of amyloidosis based on the major protein subunits present in the fibrils is shown in Table I. As noted, the terms primary and secondary refer to the absence or presence of a preexisting or coexisting chronic inflammatory or infectious disease. Because of significant overlap, primary and secondary designations should not be based on anatomic sites of involvement in individual patients.

PRINCIPAL AMYLOID PROTEINS AND DISEASES

Primary (Immunocytic) Systemic Amyloid

Primary (immunocytic) systemic amyloid of light chain origin is a disorder closely related to multiple myeloma in which a monoclonal immunoglobulin component, most often a free monoclonal L chain [Bence Jones protein (BJP)], is produced by a single family (clone) of plasma cells. Fragments of BJP polymerize and form β-pleated sheets that deposit as fibrillar material in tissues. Primary amyloidosis is the most common nonhereditary systemic amyloid type in the United States. Serum and urine immunofixation electrophoresis will demonstrate a monoclonal immunoglobulin component (M component) in serum, urine, or both, in 80% of cases. Amyloid deposition in these individuals tends to be distributed in the heart, tongue, gastrointestinal tract, skin, ligaments, and peripheral nerves. Involvement of liver, kidneys, spleen, and adrenals, a distribution more characteristic of secondary systemic amyloid, also may occur because of overlap in amyloid deposition in the immunocytic and reactive syndromes. Approximately 15 to 20% of Bence Jones proteins appear to be "amyloidogenic" in that they have the property of precipitating as fibrillar material resembling amyloid after *in vitro* proteolytic digestion. This amyloidogenic property is associated with the variable region of the molecule and is more commonly observed with λ than with κ monoclonal L-chains, a finding in accord with the L-chain distribution noted

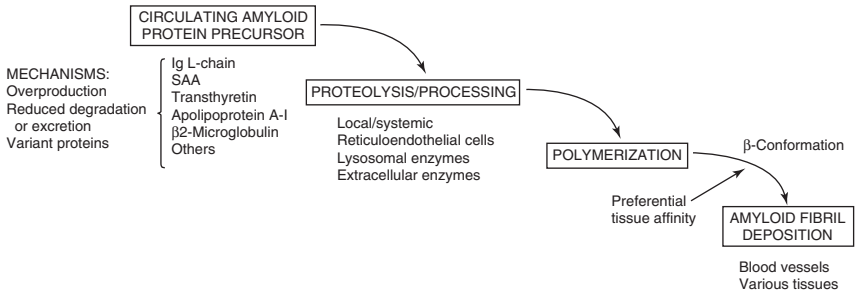


FIGURE 2 Pathogenesis of amyloidosis: Sequence of events by which a number of circulating precursor proteins are cleaved by proteolytic enzymes and deposited in tissues as amyloid fibrils. A similar sequence occurs in locally overproduced proteins, as in the case of amyloid associated with endocrine tumors. Modified from Stone, M. J. (1990). Amyloidosis: A final common pathway for protein deposition in tissues. *Blood* 75, 531–545. Copyright American Society of Hematology, used by permission.

TABLE I Classification of Amyloid Diseases According to Major Protein Constituent in Fibrils

Clinical type	Protein component
Primary systemic or localized (immunocytic)	AL immunoglobulin light chain (Bence Jones protein); rarely heavy chain (AH)
Myeloma-associated (immunocytic)	AL (rarely AH)
Secondary systemic (reactive)	AA
Familial: autosomal recessive—familial Mediterranean fever	AA
Reactive/induced in animals	AA
Familial: autosomal dominant polyneuropathies	Transthyretin, apolipoprotein A-I, gelsolin
Senile cardiac	Transthyretin
Cerebral	
Alzheimer's disease	β-Protein (Aβ)
Down's syndrome	β-Protein (Aβ)
Hereditary cerebral amyloid angiopathy (Dutch)	β-protein (Aβ)
Hereditary cerebral amyloid angiopathy (Icelandic)	Cystatin C
Prion diseases (spongiform encephalopathies)	PrP^{SC}
Dialysis-related	β2 Microglobulin
Endocrine-associated	
Medullary thyroid carcinoma	Procalcitonin
Type II diabetes mellitus, Islet cell tumors	Islet amyloid polypeptide
Others	Atrial natriuretic peptide, fibrinogen α-chain, prolactin, insulin, lysozyme, ? keratin

Note. Modified from Stone. M. J. (1990) Amyloidosis: A final common pathway for protein deposition in tissues. *Blood* **75**: 531–545. Copyright American Society of Hematology, used by permission.

in AL patients. The data are consistent with the hypothesis that patients with plasma cell dyscrasias who secrete Bence Jones proteins that possess amyloidogenic properties develop a clinical picture dominated by the features of primary systemic amyloidosis with fewer monoclonal plasma cells compared with the usual findings and cell burden observed in multiple myeloma. The resulting clinical illness therefore is more dependent on the molecular structure of the individual BJP synthesized than on any intrinsic difference between primary systemic amyloidosis and multiple myeloma. Such a hypothesis does not dictate that every patient producing amyloidogenic L-chains necessarily will develop clinical amyloidosis; some clearly do not, suggesting that additional factors play an important role in tissue deposition of amyloid fibrils.

Secondary (Reactive) Systemic Amyloidosis

Secondary (reactive) systemic amyloidosis is due to deposition of a nonimmunoglobulin protein (amyloid A) derived from a circulating protein precursor, serum amyloid A (SAA), that acts as an acute-phase reactant. This type of amyloidosis occurs in association with chronic inflammatory or infectious diseases. It occurs rarely in patients with tumors such as Hodgkin's disease, renal cell carcinoma, or other neoplasms. Chronic inflammatory diseases, especially rheumatoid arthritis and juvenile rheumatoid arthritis, are associated with amyloid of this type. Patients with other connective tissue disorders and Crohn's disease rarely develop AA amyloidosis. The same AA protein forms amyloid in patients with the autosomally recessively inherited disorder, familial Mediterranean fever (FMF). Chronic infections such as in patients with tuberculosis, leprosy, chronic osteomyelitis, bronchiectasis, decubitus ulcers, paraplegia, chronically infected burns, chronic skin infections associated with parenteral drug abuse, hypogammaglobulinemia, and Whipple's disease are associated with AA amyloidosis. Amyloid A is also the type occurring in mice with experimentally induced amyloidosis. An "amyloid-enhancing factor" that accelerates amyloidogenesis in animal models may consist of the AA fibril itself.

SAA, which has a molecular weight of approximately 12,500 Da, is a heterogeneous minor component of

normal plasma and is transported in association with high-density lipoprotein. Three forms of SAA have been described. Amyloid A is a single polypeptide chain consisting of 76 amino acids and having a molecular weight of approximately 7500 Da. Larger and smaller sizes of the protein have been identified as well.

Familial Amyloid Polyneuropathy

Familial amyloid polyneuropathy is an autosomal dominantly inherited late-onset syndrome usually due to deposits of transthyretin (TTR). TTR deposition also occurs in "senile" cardiac amyloid. TTR, previously known as thyroxin-binding prealbumin, is a transport protein that also binds retinal-binding protein. It is synthesized by the liver as a single polypeptide chain of 127 amino acids. The TTR that deposits in various tissues is a structurally abnormal protein. Over 80 amino acid substitutions at more than 50 different sites in the TTR molecule have been described. TTR is a negative acute-phase reactant; that is, its concentration decreases with inflammation; serum levels of the protein are also low in many patients with TTR amyloidosis. Not all patients with familial amyloid polyneuropathy have associated TTR deposits. In the Iowa type characterized by lower limb neuropathy, peptic ulcers, and nephrotic syndrome, a variant form of apolipoprotein A-I has been demonstrated in the amyloid fibrils. Apolipoproteins occur in other types of amyloid. Mutations of gelsolin, fibrinogen α-chain, and lysozyme rarely have been reported to cause familial amyloidosis.

Cerebral Amyloids

Cerebral amyloids are amyloid deposits in the brain occurring in Alzheimer's disease patients, in aged Down's syndrome patients, and in some individuals with hereditary cerebral amyloid angiopathy consist of a component called amyloid β-protein (Aβ), having a molecular weight of approximately 4200 Da. As with other chemical types of amyloid, the β-protein originates from a larger precursor (AβPP) and is found in blood vessels, plaques, and neurofibrillary tangles in the brain. The Icelandic form of cerebral amyloid angiopathy is characterized by a protein closely related to the cysteine protease inhibitor, cystatin C, which deposits in blood vessels. A different form of brain amyloid is present in the prion diseases, which include Creutzfeldt-Jakob disease and bovine spongiform encephalopathy ("mad cow" disease). In these disorders, a normal or variant protein appears to acquire a greater degree of β-pleated sheet conformation associated with its deposition in the brain. These amyloid diseases are generally restricted to the central nervous system.

Dialysis-Related Amyloid

Some patients on long-term (7–10 years) hemodialysis develop carpal tunnel syndrome, usually bilateral, due to amyloid deposition. It may be accompanied by cystic bone lesions or pathologic fractures. Occasionally deposits are found in other organs including those in the GI tract. The fibrillar protein in this circumstance is composed of intact β2 microglobulin (β 2M), a protein of 11,800 Da. The serum β 2M concentration in chronic hemodialysis patients is approximately 50 times higher than normal, as β 2M is too large to pass through dialysis membranes. However, elevated levels alone do not correlate well with the risk of developing amyloid. Thus, other systemic or local tissue factors may be important in determining whether amyloid deposition occurs.

CLINICAL FINDINGS

Clinical manifestations of amyloidosis vary widely depending on the organ system predominantly involved. A number of hematologic findings may occur in patients with amyloidosis, especially "scratch purpura" and spontaneous periorbital purpura after a Valsalva maneuver (Fig. 3a). Acquired factor X deficiency is an unusual but well-documented complication of amyloidosis. The coagulation factor appears to be rapidly cleared from the circulation and bound by amyloid deposits. Splenectomy may alleviate the bleeding diathesis resulting from factor X deficiency. Occasionally, factors IX and V are reduced. Other abnormalities in coagulation or fibrinolytic pathways are rare. Lytic bone lesions of the type characteristic of myeloma rarely occur in patients with amyloidosis. Uncommon sites of amyloid involvement include endocrine tumors, serosal membranes, lymph nodes, breast tissue, and thyroid.

Patients with primary AL amyloidosis most commonly present with nephrotic range proteinuria, refractory congestive heart failure, unexplained hepatomegaly, or peripheral neuropathy. Patients with AA amyloidosis present most often (90%) with proteinuria/renal insufficiency, but 20% have GI findings (diarrhea, pseudo-obstruction, constipation, or malabsorption). Patients with familial autosomal dominant amyloidosis usually present with late-onset polyneuropathy, and some have significant GI involvement.

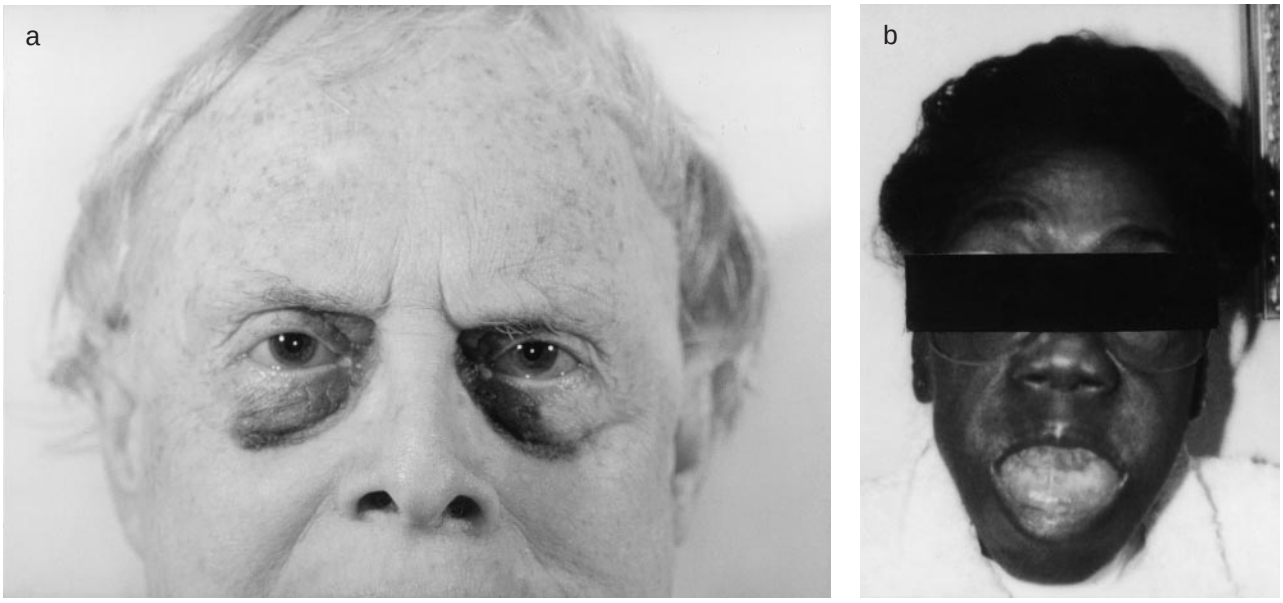


FIGURE 3 (a) Bilateral periorbital purpura in a patient with AL amyloidosis. (b) Nodular macroglossia in a patient with AL amyloidosis.

GASTROINTESTINAL AMYLOIDOSIS

Many patients with systemic amyloidosis have involvement of the GI tract with amyloid deposition occurring anywhere from the tongue to the rectum, as well as the liver, spleen, and pancreas. The GI manifestations may be the mode of presentation and often are accompanied by evidence of organ involvement elsewhere, namely, the heart, kidney, and autonomic nervous system. The symptoms, physical signs, and clinical manifestations of GI amyloidosis are listed in Table II. Endoscopic appearance and radiographic appearance are also described below.

Endoscopic Appearance

There is no endoscopic appearance specific for amyloidosis. The most common endoscopic findings include a fine granular appearance, erosions, ulcerations, and mucosal friability. These findings are most often seen in the duodenum, followed by the stomach, colorectum, and esophagus.

Radiographic Appearance

Barium radiography is of limited diagnostic utility since the changes that may be seen are not specific for amyloidosis. Nevertheless, esophageal abnormalities include reflux and esophageal dysmotility and dilation. The stomach may show rigid gastric rugal folds, dilation with barium retention, and narrowing or obstruction in the antral region. Small intestinal radiographs are often the initial test suggesting GI amyloid involvement. Commonly reported findings include sharply demarcated thickening of the valvular conniventes, altered intestinal transit, bowel dilation, and multiple nodular lesions. Colonic contrast studies most often show involvement of the rectosigmoid characterized by altered colonic transit with dilation or multiple filling defects with narrowing and rigidity secondary to ischemia. Abdominal computerized tomography imaging may show wall thickening or bowel dilation and rarely mesenteric thickening or lymphadenopathy.

Specific Organ Involvement

Mouth

Oral amyloidosis may affect the tongue, buccal mucosa, and gingiva. Bleeding from the latter two sites is common. Tongue involvement causes macroglossia in 20–50% of primary (AL) amyloid patients (Fig. 3b). Macroglossia may cause difficulties with speech, mastication, swallowing, and breathing due to reduced tongue mobility. Tracheostomy may be necessary in severe cases. Amyloid infiltration of the salivary

TABLE II Signs, Symptoms, and Clinical Manifestations of GI Amyloidosis

Location	Symptoms	Physical signs	Clinical manifestations
Mouth	Enlarged tongue	Macroglossia	Dysphonia
	Bleeding	Reduced tongue mobility and induration	Difficulty with mastication and deglutition
	Toothache	Nodular lesions of tongue and buccal mucosa	Sicca syndrome
	Paresthesias	Oral hemorrhagic bullae	Jaw claudication
	Dry mouth		Upper airway obstruction/sleep apnea
Esophagus	Heartburn		Esophageal dysmotility
	Waterbrash		Esophagitis
	Dysphagia		
Stomach	Nausea	Succussion splash	Erosions and ulcerations
	Vomiting	Cachexia	Gastric amyloid nodules
	Epigastric pain		Gastroparesis
	Anorexia		Gastric outlet obstruction
	Abdominal fullness		
	Bleeding		
	Weight loss		
Small intestine	Diarrhea	Cachexia	Intestinal ischemia/bleeding
	Constipation		Pseudo-obstruction
	Abdominal pain		Malabsorption (steatorrhea, protein-losing enteropathy)
	Bleeding		Obstruction due to amyloidoma
	Weight loss		Intestinal infarction/perforation
Colon	Diarrhea	Cachexia	Colonic ischemia/bleeding
	Constipation		Pseudo-obstruction/megacolon
	Abdominal pain		Fecal incontinence
	Bleeding		Volvulus
			Intestinal infarction/perforation
Liver	Jaundice	Hepatomegaly	Well-preserved hepatic synthetic function
	Abdominal pain	No evidence of cirrhosis	Elevated alkaline phosphatase
		Occasional evidence of portal hypertension	Focal intrahepatic masses
			Spontaneous hepatic rupture
Spleen	Abdominal pain	Splenomegaly	Functional hyposplenism
			Spontaneous splenic rupture
Pancreas	Diarrhea		Exocrine pancreatic insufficiency
	Abdominal pain		Pancreatitis

glands can cause symptoms of dry mouth associated with the sicca syndrome.

Esophagus

Esophageal involvement usually presents with symptoms of dysphagia and gastroesophageal reflux. Histological studies show that amyloidosis can infiltrate both the striated and smooth muscle portions of the esophagus, vagus nerve, myenteric plexus, and vasa nervosa. Manometric studies may show a nonspecific esophageal dysmotility and abnormal lower esophageal sphincter relaxation. Treatment with a proton pump inhibitor is usually recommended.

Stomach

Patients with gastric amyloid may present with symptoms of nausea, vomiting, epigastric pain, abdominal bloating, anorexia, and weight loss. Some patients develop mucosal or gastric amyloid tumor ulcerations with signs of upper GI bleeding. Others present with gastric motility disturbances. In familial amyloid polyneuropathy with autonomic nervous system involvement, gastroparesis with delayed gastric emptying and hypotonia is common. Treatment with prokinetic agents has limited effectiveness. Total parenteral nutrition as a means of providing temporary nutrition or drainage procedures such as gastrojejunostomies have shown some benefit.

Small Intestine

In systemic amyloidosis, small intestinal deposits are present histologically in more than 70% of cases. Amyloid may be seen in the intrinsic and extrinsic nervous system, the mucosa, submucosa, and muscle wall of the small intestine. The variable location of these deposits accounts for the myriad of symptoms and clinical manifestations associated with small bowel involvement including diarrhea, constipation, malabsorption, obstruction, pseudo-obstruction, bleeding, and vascular insufficiency. Chronic intestinal ischemia may occur secondary to progressive occlusion of the vessels of the submucosa. This will lead to sloughing of the intestinal lining and hemorrhage. Rarely, protein-losing enteropathy or progression to intestinal infarction and perforation is seen. Small bowel obstruction due to an amyloid tumor also has been described.

The clinical spectrum of symptoms, course, and prognosis associated with pseudo-obstruction may depend on the type of amyloid. Patients with AL and Aβ 2M amyloidosis typically present with irreversible chronic, intermittent obstructive symptoms and have evidence of extensive infiltration of the smooth muscle of the bowel wall. On the other hand, AA amyloidosis patients may present with reversible, acute obstructive symptoms and have evidence of myenteric plexus involvement. Surgery is generally not beneficial in the treatment of amyloid-induced pseudo-obstruction. Promotility agents have not been proven effective and some patients may require long-term TPN.

Diarrhea is often severe and uncontrollable in systemic amyloidoses. Malabsorption as evidenced by steatorrhea and protein-losing enteropathy is seen in less than 5% of patients but has been reported in AL, AA, and TTR amyloidoses. Diarrhea is particularly common in familial amyloid polyneuropathy due to TTR amyloid and often leads to cachexia and early mortality. The pathogenesis of diarrhea in patients with amyloidosis is unknown. A variety of mechanisms have been advanced as possible explanations for the disruption of gastrointestinal function. These include (1) malabsorption of fat, protein, and/or carbohydrate by multiple processes including infiltration of the mucosa producing a mechanical barrier, mucosal atrophy secondary to amyloid-induced vascular ischemia, and pancreatic amyloidosis with concomitant pancreatic exocrine insufficiency; (2) bile salt malabsorption; and (3) altered intestinal transit. Slow intestinal motility may result from amyloid deposition within the muscular layer of the intestinal wall causing a myopathy or from amyloid deposition in the GI autonomic nervous system causing a neuropathy. Reduced motility permits bacterial overgrowth, bile acid deconjugation, and consequent diarrhea and steatorrhea. Treatment for amyloid-induced diarrhea includes opioids, antibiotics, cholestyramine, and octreotide.

Colon

Colonic amyloid is associated with symptoms similar to small intestinal involvement including diarrhea, constipation, abdominal pain, intestinal ischemia, and bleeding. Volvulus, megacolon, and fecal incontinence due to neuropathy or amyloid deposition in the anal sphincter have been reported in patients with colonic amyloidosis.

Liver

Hepatic amyloidosis is manifested by hepatomegaly with relatively well-preserved hepatic synthetic function. Modest elevations of serum alkaline phosphatase are the most common laboratory abnormalities. Hypoalbuminemia may be present and due to nephrotic syndrome, decreased hepatic synthetic function, or rarely, protein-losing enteropathy. Elevated serum aminotransferase levels are less commonly seen and elevated serum bilirubin levels are rare. Cholestatic jaundice portends a poor prognosis as it may lead to liver failure and is usually due to advanced AL amyloidosis. Although many patients with amyloidosis have hepatic deposits, there is poor correlation between the degree of liver dysfunction and the extent of amyloid deposition. Rarely, patients present with focal intrahepatic mass lesions and hepatic rupture, the latter usually being a fatal event.

Clinical features of chronic liver disease (e.g., spider angiomas, palmar erythema) are uncommon as most patients die of extrahepatic amyloid deposition. Portal hypertension is seen occasionally. Its complications, including ascites and esophageal variceal hemorrhage, are seen to variable degrees. Ascites may develop from portal hypertension with or without hypoalbuminemia. Treatment of hepatic amyloidosis is directed at management of these complications and the underlying mechanism of the amyloid deposition.

Hepatic amyloidosis is diagnosed by liver biopsy. Increased risk of hemorrhage with liver biopsy is thought to be secondary to amyloid infiltration of blood vessels with consequent increased fragility and inability of the blood vessels to contract. Several studies have reported cases of fracture of the liver with hemorrhage, capsular rupture, and death following liver biopsy in patients with suspected amyloid liver involvement. One series showed that all amyloidosis patients with hemorrhagic complications related to a

diagnostic procedure had a prior history of a bleeding episode. Therefore, liver biopsy should be approached with caution in patients with established or suspected amyloidosis.

Spleen

Splenomegaly is initially present in approximately 10% of patients with amyloidosis and splenic rupture has been reported rarely. Hypersplenism does not occur with splenomegaly, probably because of massive replacement of splenic tissue by amyloid. Approximately 20 to 25% of patients with systemic amyloidosis and 60% with hepatic amyloidosis develop functional hyposplenism characterized by a normal-sized or large spleen on imaging studies and a peripheral blood smear showing Howell-Jolly bodies (nuclear remnants), target cells, and large platelets. The presence of Howell-Jolly bodies on a blood smear is a highly specific, but not sensitive, indicator of splenic amyloidosis. Functional hyposplenism is a valuable clue to the diagnosis of systemic amyloidosis in the patient presenting with nephrotic syndrome, hepatomegaly, or refractory heart failure of unknown etiology. The absence of hyposplenism is not a predictable sign of the absence of splenic involvement. In rare cases, amyloidosis-related functional hyposplenism is reversible if the patient responds to treatment.

Pancreas

Amyloid deposition can involve both the exocrine and endocrine portions of the pancreas. Exocrine pancreatic involvement can cause acinar atrophy and destruction. Amyloidosis has been reported in patients with cystic fibrosis and exocrine pancreatic insufficiency. These patients respond to enzyme replacement therapy. Rarely, pancreatic involvement results in pancreatitis secondary to ductal obstruction by amyloid deposits. Deposits in the pancreatic islets occur in type II diabetes mellitus and islet cell tumors.

Crohn's Disease-Associated Amyloidosis

AA amyloidosis is an uncommon yet important complication of Crohn's disease. In the largest series of patients with inflammatory bowel disease collected over a 50-year span, Greenstein *et al*. found a 0.9% incidence of amyloidosis in Crohn's disease. The prevalence may be higher as amyloid deposits are often documented only at autopsy. The mean time from the clinical onset of Crohn's disease to development of amyloidosis was 15 years. Amyloidosis in Crohn's disease was associated with more extensive intestinal disease, suppurative complications (fistulas, abscesses), extraintestinal manifestations (arthritis, pyoderma gangrenosum, aphthous stomatitis), and male gender.

Crohn's disease-associated amyloidosis most often develops in the kidney, but generalized amyloidosis with involvement of the GI tract, heart, thyroid, liver, and spleen has been described. Patients initially present with proteinuria or renal insufficiency, which may progress to nephrotic syndrome and renal failure.

TREATMENT

Therapy of amyloidosis is unsatisfactory. Evaluation of various approaches has been hindered by the lack of ability to accurately determine the extent of involvement in individual patients and by the widely disparate etiology of the various disorders that lead to amyloid deposits. Despite its apparently inert nature and inaccessible extracellular location, amyloidosis occasionally is reversible and radiolabeled SAP scintigraphy has demonstrated that amyloidogenesis is a dynamic process. Potential therapeutic approaches consist of those directed at prevention of amyloid precursor protein synthesis, prevention of amyloid fibril deposition, and removal or dissolution of amyloid deposits from tissues.

AL Amyloidosis

For AL amyloidosis, a myeloma type of chemotherapy regimen (melphalan and prednisone) is generally employed in an attempt to reduce production of the circulating monoclonal L-chain precursor. Colchicine may have some activity though less than cytotoxic chemotherapy. Iodo-doxorubicin has been reported to have anti-amyloid activity, but its role in therapy is currently undefined. High-dose chemotherapy with stem cell autografting has been employed by several groups and preliminary reports seem encouraging. However, follow-up is limited and the heterogeneity of individual amyloid patients makes subgrouping necessary and analysis difficult. A high incidence of GI bleeding after stem cell autografting has been reported. Preclinical studies have indicated that AL amyloid resolution can be induced by passive administration of an amyloid-reactive antibody. Such an approach may be effective in other amyloid types as the antibody appears to be directed to a fibrillar epitope. This observation has potentially important implications. Clinical trials in patients with AL and other amyloid diseases are eagerly awaited.

AA Amyloidosis

AA amyloidosis sometimes improves with therapy designed to prevent initial amyloid deposition or disease progression. Effective treatment of chronic infection

may result in amyloid stabilization or reversal. Colchicine prevents amyloid deposition and further deterioration of renal and cardiac function in AA amyloidosis associated with FMF and Crohn's disease. Colchicine may induce transient diarrhea and abdominal pain that must be distinguished from FMF, Crohn's disease, or GI tract amyloidosis. Even low-dose colchicine may cause diarrhea, especially in patients who have GI tract amyloid involvement.

In selected AA amyloidosis patients, renal and cardiac transplantation has been performed. Renal transplantation appears to have a role in FMF patients. In patients with Crohn's disease-associated amyloidosis, progression of the amyloid-related disease appears to cease following effective therapy directed at the underlying inflammatory bowel disease. Resection of the inflamed bowel is controversial with a possible increased risk of postoperative morbidity and mortality.

Familial Amyloid Polyneuropathic Syndromes Due to TTR Amyloidosis

In hereditary amyloidosis due to TTR, more than 95% of the mutant transthyretin is produced by the liver. A liver transplant should replace mutated TTR with the normal (wild-type) molecule and thereby stop amyloid formation. More than 400 liver transplants have been performed for TTR amyloidosis. Techniques have included both orthotopic and living–related donor liver transplantation. Combined liver/heart or liver/kidney transplants also have been performed. The goal of liver transplantation is to prevent further disease progression and onset of new complications. Liver transplantation has been reported to halt amyloid progression in familial amyloid polyneuropathy patients. Modest improvement in peripheral and autonomic neuropathy has been claimed in some patients. Improvement in gastrointestinal symptoms and nutrition has been noted posttransplant. Factors associated with a favorable outcome after liver transplantation include the presence of the TTR Val30Met mutation, symptomatic disease duration of less than 7 years, and good nutritional status with lack of severe autonomic impairment. Use of DNA testing to identify persons at risk for development of TTR amyloidosis while still in the preclinical phase of their disease will be helpful in assessing the efficacy of various new therapeutic regimens.

TTR amyloidosis rarely involves the liver and has not been associated with liver failure. In addition, its onset occurs in the third decade of life or later and carriers of the trait may never develop the disease. For these reasons, liver explants from patients with familial amyloid polyneuropathy have been used for sequential ("domino") liver transplantation. To date, no amyloid disease has been reported in recipients of these livers.

Effective therapy for other types of amyloidosis listed in Table I awaits elucidation of further insights into the origin and pathogenesis of fibrillar deposition in patients with these disorders.

Acknowledgments

The authors thank Dr. John S. Fordtran for helpful discussions and review of the manuscript. Ms. Linda Miller provided expert assistance with manuscript preparation. This work was supported in part by the Edward and Ruth Wilkof Foundation, the Robert Schanbaum Memorial Fund, and the James E. Nauss Cancer Research Fund.

See Also the Following Article

Crohn's Disease

Further Reading

Araoz, P. A., Batts, K. P., and MacCarty, R. L. (2000). Amyloidosis of the alimentary canal: Radiologic–pathologic correlation of CT findings. *Abdom. Imaging* **25**, 38–44.

Buxbaum, J. N., and Tagoe, C. E. (2000). The genetics of the amyloidoses. *Annu. Rev. Med.* **51**, 543–569.

Comenzo, R. L. (2000). Hematopoietic cell transplantation for primary systemic amyloidosis: What have we learned. *Leukocyte Lymph.* **37**, 245–258.

Falk, R. H., Comenzo, R. L., and Skinner, M. (1997). The systemic amyloidoses. *N. Engl. J. Med.* **337**, 898–909.

Friedman, S., and Janowitz, H. D. (1998). Systemic amyloidosis and the gastrointestinal tract. *Gastroenterol. Clin. North Am.* **27**(3), 595–614.

Gertz, M. A., and Kyle, R. A. (1997). Hepatic amyloidosis: Clinical appraisal in 77 patients. *Hepatology* **25**, 118–121.

Gertz, M. A., Lacy, M. Q., and Dispenzieri, A. (1999). Amyloidosis. *Hematol. Oncol. Clin. North Am.* **13**, 1211–1233.

Gillmore, J. D., Hawkins, P. N., and Pepys, M. B. (1997). Amyloidosis: A review of recent diagnostic and therapeutic developments. *Br. J. Haematol.* **99**, 245–256.

Greenstein, A. J., Sacher, D. B., Panday, A. K., *et al.* (1992). Amyloidosis and inflammatory bowel disease: A 50 year experience with 25 patients. *Medicine* **71**, 261–270.

Hrncic, R., Wall, J., Wolfenbarger, D. A., *et al.* (2000). Antibody-mediated resolution of light chain-associated amyloid deposits. *Am. J. Pathol.* **157**, 1239–1246.

Kyle, R. A., Gertz, M. A., Greipp, P. R., *et al.* (1997). A trial of three regimens for primary amyloidosis: Colchicine alone, melphalan and prednisone, and melphalan, prednisone, and colchicine. *N. Engl. J. Med.* **336**, 1202–1207.

Lachmann, H. J., Booth, D. R., Booth, S. E., *et al.* (2002). Misdiagnosis of hereditary amyloidosis as AL (primary) amyloidosis. *N. Engl. J. Med.* **346**, 1786–1791.

Murphy, C. L., Eulitz, M., Hrncic, R., *et al.* (2001). Chemical typing of amyloid protein contained in formalin-fixed

paraffin-embedded biopsy specimens. *Am. J. Clin. Pathol.* **116**, 135–142.
Röcken, C., Schwotzer, E. B., Linke, R. P., *et al.* (1996). The classification of amyloid deposits in clinicopathological practice. *Histopathology* **29**, 325–335.
Stone, M. J. (1990). Amyloidosis: A final common pathway for protein deposition in tissues. *Blood* **75**, 531–545.
Stone, M. J., and Hirsch, V. J. (1986). Splenic function in amyloidosis. *In* "Amyloidosis" (G. Glenner, E. Osserman, E. Benditt, E. Calkins, A. Cohen, and D. Zucker-Franklin, eds.), pp. 583–590. Plenum Press, New York.
Suhr, O. B., Herlenius, G., Friman, S., *et al.* (2000). Liver transplantation for hereditary transthyretin amyloidosis. *Liver Transplant* **6**, 263–276.

Anal Canal

AHMED SHAFIK
Cairo University, Egypt

rectal neck Embryologically, the anal canal is the lower part of the hindgut from which the rectum develops. The hindgut extends down and fixes itself to the perineal skin. Therefore, the anal canal is actually the lower narrow part of the rectum and it is referred to as the rectal neck.

The rectal neck (anal canal) is the gateway to the gut; it is surrounded by sphincters and muscles that regulate the passage of the rectal contents to the exterior. These muscles are responsible for fecal continence and defecation.

EXTERNAL ANAL SPHINCTER

The external anal sphincter (EAS) is a triple-loop system consisting of top, intermediate, and base loops (Fig. 1). Each loop is separated from the others by a fascial septum (Fig. 2) and has its individual attachment, direction of muscle bundles, and innervation. The top loop comprises the deep part of the EAS and puborectalis, which are intimately fused together. Its muscle bundles loop around the upper part of the rectal neck (RN) and are attached to the symphysis pubis. It forms a downward extension, which descends along the RN and contributes to the formation of the longitudinal muscle (Fig. 2). It is innervated by the inferior rectal nerve. The intermediate loop embraces the midportion of the RN and is innervated by the perineal branch of the fourth sacral nerve. The base loop encloses the lower RN and is innervated by the inferior rectal nerve. It consists of only loop fibers in its upper part and of inner circular and outer loop fibers in its lower part.

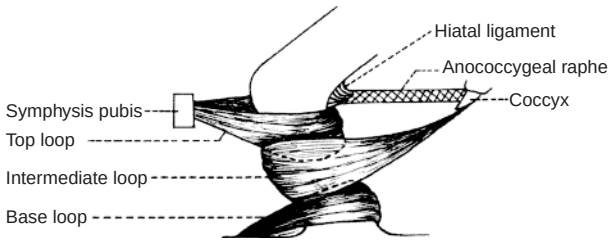


FIGURE 1 Diagram illustrating the triple-loop system of the external anal sphincter. Reprinted from Shafik (1981), with permission from Excerpta Medica.

Mechanism of Action

The EAS induces voluntary continence by a double-fold action: (1) prevention of internal sphincter relaxation on detrusor contraction, which is termed "voluntary anorectal inhibition reflex," and (2) direct compression of the RN, or the "mechanical action."

Voluntary Anorectal Inhibition Reflex

As stools enter the rectum, the rectal detrusor contracts and the internal sphincter relaxes reflexly to open the RN (Fig. 3). The latter does not open unless the EAS relaxes voluntarily. However, if there is no desire to defecate, the EAS contracts, mechanically preventing relaxation of the internal sphincter. Failure of the latter to relax reflexly inhibits contraction of the rectal detrusor, which relaxes and dilates to accommodate the new contents (Fig. 4). Voluntary EAS contraction to inhibit

Copyright 2004, Elsevier (USA). All rights reserved.

paraffin-embedded biopsy specimens. *Am. J. Clin. Pathol.* **116**, 135–142.
Röcken, C., Schwotzer, E. B., Linke, R. P., *et al.* (1996). The classification of amyloid deposits in clinicopathological practice. *Histopathology* **29**, 325–335.
Stone, M. J. (1990). Amyloidosis: A final common pathway for protein deposition in tissues. *Blood* **75**, 531–545.
Stone, M. J., and Hirsch, V. J. (1986). Splenic function in amyloidosis. *In* "Amyloidosis" (G. Glenner, E. Osserman, E. Benditt, E. Calkins, A. Cohen, and D. Zucker-Franklin, eds.), pp. 583–590. Plenum Press, New York.
Suhr, O. B., Herlenius, G., Friman, S., *et al.* (2000). Liver transplantation for hereditary transthyretin amyloidosis. *Liver Transplant* **6**, 263–276.

Anal Canal

AHMED SHAFIK
Cairo University, Egypt

rectal neck Embryologically, the anal canal is the lower part of the hindgut from which the rectum develops. The hindgut extends down and fixes itself to the perineal skin. Therefore, the anal canal is actually the lower narrow part of the rectum and it is referred to as the rectal neck.

The rectal neck (anal canal) is the gateway to the gut; it is surrounded by sphincters and muscles that regulate the passage of the rectal contents to the exterior. These muscles are responsible for fecal continence and defecation.

EXTERNAL ANAL SPHINCTER

The external anal sphincter (EAS) is a triple-loop system consisting of top, intermediate, and base loops (Fig. 1). Each loop is separated from the others by a fascial septum (Fig. 2) and has its individual attachment, direction of muscle bundles, and innervation. The top loop comprises the deep part of the EAS and puborectalis, which are intimately fused together. Its muscle bundles loop around the upper part of the rectal neck (RN) and are attached to the symphysis pubis. It forms a downward extension, which descends along the RN and contributes to the formation of the longitudinal muscle (Fig. 2). It is innervated by the inferior rectal nerve. The intermediate loop embraces the midportion of the RN and is innervated by the perineal branch of the fourth sacral nerve. The base loop encloses the lower RN and is innervated by the inferior rectal nerve. It consists of only loop fibers in its upper part and of inner circular and outer loop fibers in its lower part.

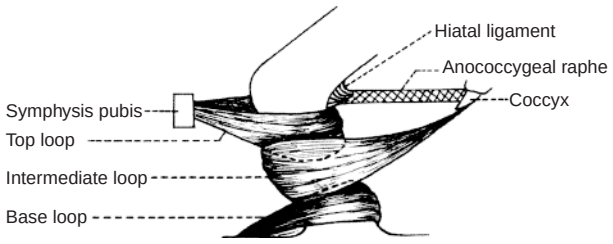


FIGURE 1 Diagram illustrating the triple-loop system of the external anal sphincter. Reprinted from Shafik (1981), with permission from Excerpta Medica.

Mechanism of Action

The EAS induces voluntary continence by a double-fold action: (1) prevention of internal sphincter relaxation on detrusor contraction, which is termed "voluntary anorectal inhibition reflex," and (2) direct compression of the RN, or the "mechanical action."

Voluntary Anorectal Inhibition Reflex

As stools enter the rectum, the rectal detrusor contracts and the internal sphincter relaxes reflexly to open the RN (Fig. 3). The latter does not open unless the EAS relaxes voluntarily. However, if there is no desire to defecate, the EAS contracts, mechanically preventing relaxation of the internal sphincter. Failure of the latter to relax reflexly inhibits contraction of the rectal detrusor, which relaxes and dilates to accommodate the new contents (Fig. 4). Voluntary EAS contraction to inhibit

Copyright 2004, Elsevier (USA). All rights reserved.

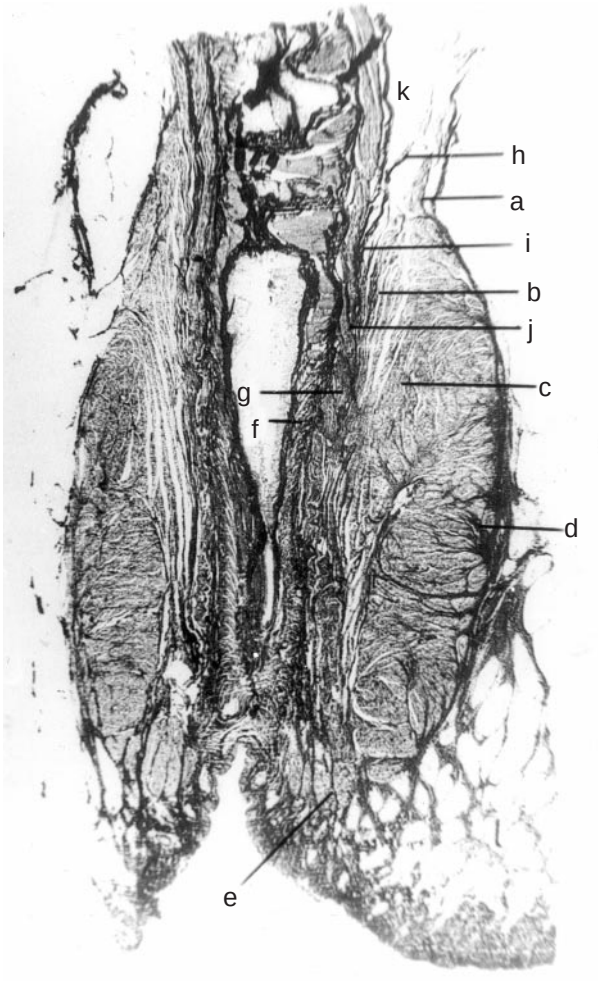


FIGURE 2 Coronal section at the level of the midanal orifice shows the three loops of the external anal sphincter. It also shows the external anal fascia investing the whole sphincter and sending inward extensions between its loops. Verhoeff-van Gieson. Magnification, ×7. (a) Levator plate; (b) suspensory sling; (c) top loop (fused puborectalis and deep external anal sphincter); (d) intermediate loop of external anal sphincter; (e) base loop of external anal sphincter; (f) internal anal sphincter; (g) longitudinal anal muscle; (h) fascia on pelvic surface of levator plate; (i) hiatal ligament; (j) tunnel septum; (k) pelvirectal space. Reprinted from Shafik (1981), with permission.

reflex internal sphincter relaxation is the "voluntary anorectal inhibition reflex." This is the main action responsible for voluntary continence. The internal sphincter integrity is thus necessary not only for involuntary continence, but also for voluntary continence, because the internal sphincter mediates the voluntary inhibition reflex. For this reason, internal sphincter reconstruction should be considered an essential step in rectal incontinence repair.

Voluntary Mechanical Action

In addition to the voluntary inhibition reflex, EAS contraction firmly seals the RN by mechanical compression. Because it is a striated muscle, the EAS cannot contract for a long period to maintain continence mechanically. The mechanical compression action is thus momentary (40–60 s) and serves to occlude the RN by the time the detrusor relaxes as a result of the voluntary inhibition reflex.

Stress Defecation

Under conditions of internal sphincter damage, voluntary continence is induced only by the mechanical action of the EAS. The voluntary inhibition reflex is lost. Because it is a striated muscle, the EAS cannot contract long enough to withstand the noninhibited prolonged contraction of the loaded rectal detrusor. Detrusor contraction continues despite EAS contraction till the latter fatigues and relaxes and the detrusor evacuates itself. Hence, in cases of internal sphincter damage, once the desire to defecate is initiated, evacuation should occur. This condition, which is termed "stress defecation," is observed in patients after internal sphincterotomy for anal fissure. It could also explain the impaired control of feces and flatus after internal sphincterotomy.

Single-Loop Continence

As a result of the separate arrangement of the three EAS loops and because each loop has its own separate and bilateral innervation, any single loop can function as a sphincter. The EAS continence action can be achieved by a single-loop contraction and not necessarily by the three loops. This constitutes the basis of "single-loop continence." On contraction, a single loop induces continence by both the voluntary

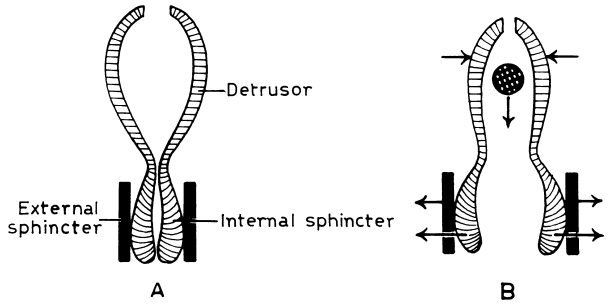


FIGURE 3 External and internal sphincters at rest and during defecation. (A) At rest: detrusor is relaxed and internal sphincter is involuntarily contracted. (B) During defecation: detrusor is contracted and external and internal sphincters are relaxed. Reprinted from Shafik and El-Sibai (2001), with permission. Copyright 2001 Lippincott Williams & Wilkins.

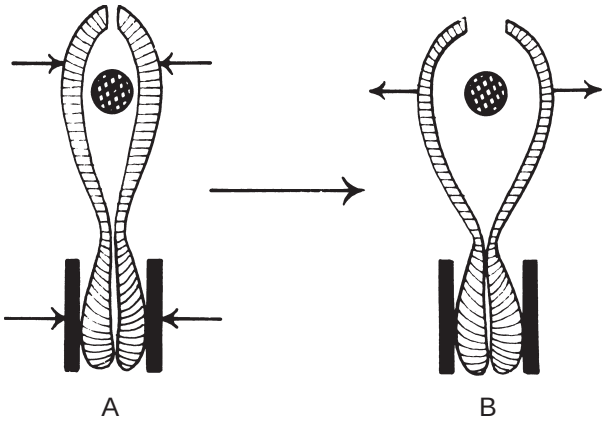


FIGURE 4 Mechanism of voluntary inhibition reflex to oppose a call to defecate. (A) Detrusor contraction with failure of internal sphincter relaxation due to voluntary external sphincter contraction. (B) Reflex detrusor relaxation due to failure of internal sphincter relaxation, the voluntary inhibition reflex. Reprinted from Shafik and El Sibai (2001), with permission. Copyright 2001 Lippincott Williams & Wilkins.

inhibition reflex and mechanical occlusion. The latter action creates significantly tight loop contraction, being effected not only by direct compression but also by RN kinking.

THE ANOGENITAL MUSCLE

A recent study has demonstrated that the base loop of the EAS extends uninterrupted across the perineum to the bulb of the penis where it becomes continuous with the bulbocavernosus muscle. Lying over the bulb, the muscle bundles are arranged into three groups: the median fibers and two lateral bundles of fibers. The median fibers form the "retractor penis muscle," which is inserted into the corpora cavernosa, and the lateral fibers, which form the "compressor bulbae muscle," are inserted into the perineal membrane. Upon glans stimulation, both the EAS and the bulbocavernosus muscle contract synchronously with similar latency and action potentials. The bulbocavernosus muscle is an integral part of the EAS, and the muscle in its entirety is appropriately named the "anogenital muscle." The muscle plays a dual and synchronous role in fecal control and sexual response. It is suggested that EAS disorders lead to sexual dysfunction and vice versa.

LONGITUDINAL MUSCLE

The longitudinal muscle consists of three layers: medial, intermediate, and lateral (Figs. 2 and 5). The medial longitudinal muscle is a continuation of the longitudinal rectal muscle coat. The intermediate muscle is the suspensory sling of the levator ani, whereas the lateral muscle is the longitudinal extension of the top loop of the EAS. The fleshy longitudinal muscle ends at the level of the lower border of the internal sphincter by giving rise to a fascial condensation called the "central tendon." The latter splits into multiple fibrous septa. The medial septum attaches to the RN lining, whereas the lateral muscle passes into the ischiorectal fossa. The intermediate septa penetrate the EAS base loop, decussate to form the corrugator cutis, and insert in the perineal skin (Figs. 2 and 5). The longitudinal muscle plays an important role in the mechanism of defecation. On contraction at stool, it shortens and widens the RN. Furthermore, it helps to fix the RN during straining at defecation, thus preventing rectal prolapse. Subluxation of the longitudinal muscle shares in rectal prolapse genesis.

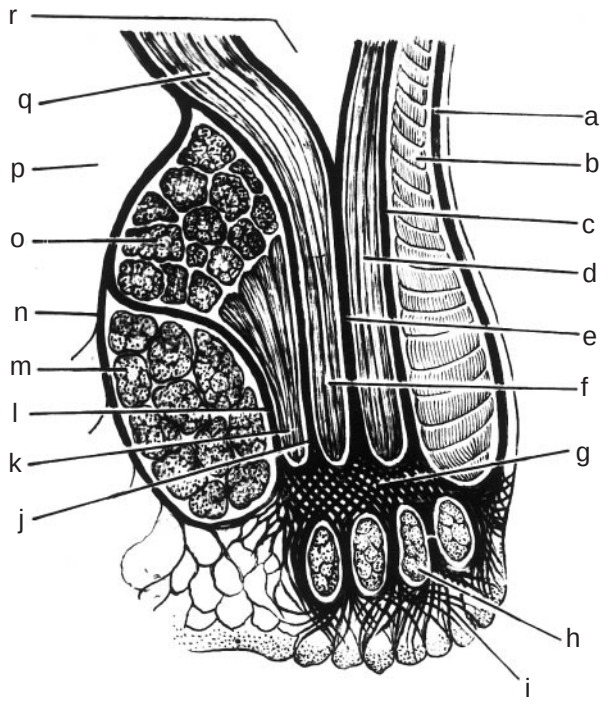


FIGURE 5 Diagram illustrating the rectal neck musculature and perirectal spaces. (a) Submucous space containing internal anal septum; (b) internal sphincter; (c, e, j, and l) four intersphincteric spaces; (d, f, and k) medial, intermediate, and lateral longitudinal muscles; (g) central space occupied by central tendon; (h, m, and o) base, intermediate, and top loop of external sphincter; (i) subcutaneous space containing corrugator ani cutis; (n) external fascial septum; (p) ischiorectal space; (q) levator plate; (r) pelvirectal space. Reprinted from Shafik (1987), with permission.

PERIANAL SPACES

Six perirectal spaces can be identified: subcutaneous, central, intersphincteric, pelvirectal, ischiorectal, and submucous (Figs. 2 and 5). The subcutaneous space was continuous with the ischiorectal space. The central space lies in the lower RN and is occupied by the central tendon. It is the main perirectal space; it communicates with all of the other spaces along the central tendon.

The central tendon gives rise to multiple fibrous septa. The medial tendon passes between the internal sphincter and the base loop to attach to the anal lining. The lateral septum passes between the intermediate and base loops into the ischiorectal fossa. The intermediate septa penetrate the base loop into the subcutaneous space. The central space thus communicates with all perianal spaces: the subcutaneous, submucous, ischiorectal, and intersphincteric, through which it communicates with the pelvirectal space.

There are four intersphincteric spaces that lie along the three layers of the longitudinal muscle. The most medial space communicates with the submucous space, whereas the lateral two spaces communicate with the ischiorectal space. The intermediate space communicates directly with the pelvirectal space. The intersphincteric spaces communicate inferiorly with the central space, through which they are connected to the subcutaneous space and perianal skin and to the ischiorectal space.

Fistula Classification

A new fistula classification was put forward based on pathoanatomical studies.

The route adopted by the pus to any of the six perianal spaces defines the type of fistula. According to the relation of the fistulous track to the EAS, two main types of fistulas could be recognized: intrasphincteric and extrasphincteric (Fig. 6).

Intrasphincteric Fistula

The track is medial to the EAS and the external opening is usually close to the anal orifice within the perianal skin corrugations. It starts as a central space infection that spreads either down to the subcutaneous space and perianal skin, forming a central fistula, or up into the intersphincteric spaces, forming an intersphincteric fistula (Fig. 6).

Extrasphincteric Fistula

The track is lateral to the EAS and the external opening usually overlies the base of the ischiorectal fossa away from the perianal corrugations. It arises as a central space infection that spreads laterally to the ischiorectal space (Fig. 6).

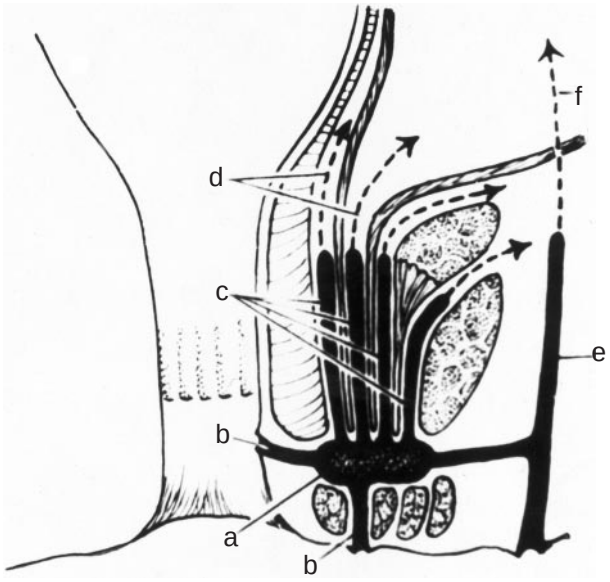


FIGURE 6 Fistula classification. (a) Central abscess; (b) central fistula, (c) low intersphincteric fistula; (d) high intersphincteric fistula; (e) low extrasphincteric fistula; (f) high extrasphincteric fistula. Reprinted from Shafik (1987), with permission.

LEVATOR HIATUS AND TUNNEL

The levator ani (LA) consists essentially of the pubococcygeus, the iliococcygeus being rudimentary in humans. The puborectalis is not a part of the LA; both differ in morphology, innervation, and function. The pubococcygeus is funnel-shaped with a transverse portion called the levator plate and a vertical portion called the suspensory sling (Figs. 2 and 5). The levator plate is an oval cone, which stretches across the pelvis; the levator hiatus occupies its anterior portion and the rectococcygeal raphe exists posteriorly (Fig. 7). Two patterns of the rectococcygeal raphe can be identified: single- and triple-decussation patterns. The latter seems to give firmness to the levator plate and might be a factor in resisting rectal prolapse. The hiatal ligament connects the medial border of the levator plate to the anorectal junction. The levator plate consists of two "crura," which bind the levator hiatus, and two lateral masses (Fig. 7). Three crural patterns have been identified: classic, crural overlap, and crural scissor. The lateral masses function as visceral support and the crura are the functional mobile parts of the LA.

The LA is the principal muscle of defecation. On contraction at defecation, it opens the RN for the

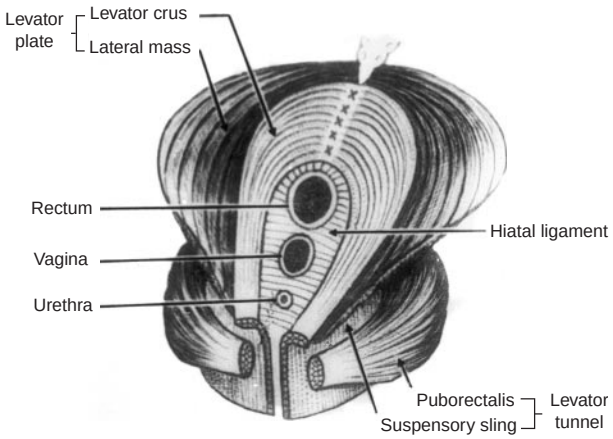


FIGURE 7 Diagram illustrating the levator plate and tunnel. Reprinted from Shafik (1981), with permission from Excerpta Medica.

stool to descend. Any interference with levator function results in disturbance of the act of defecation and leads to levator dysfunction syndrome, which presents as descending perineum, intussusception, solitary ulcer syndrome, and rectal prolapse. The different patterns of the rectococcygeal raphe and levator crura play an important role in RN support; their subluxation may eventually lead to rectal prolapse.

The "levator tunnel" is a muscular tube that surrounds the intrahiatal organs (RN, prostate in males or vagina and urethra in females) along their way down from the levator hiatus to the perineum (Fig. 7). The posterior tunnel wall (3–4 cm) is longer than the anterior tunnel wall (2.5–3 cm). The tunnel is double-sheathed with an inner coat of the suspensory sling and an outer coat of the puborectalis. Both coats are composed of striated muscle bundles. The inner coat is a tunnel dilator, which opens the RN at defecation, whereas the outer coat is a tunnel constrictor. The tunnel septum, a grayish white membrane, lines the inner aspect of the levator tunnel and separates it from the fascia propria of the intrahiatal organs. It separates the voluntary components from the involuntary components of the levator tunnel. It serves as an important landmark during mobilization of the intrahiatal organs from within the levator tunnel, e.g., in the operation of anorectal mobilization for rectal cancer.

Hiatal Ligament

The levator plate is connected to the intrahiatal organs by a fascial condensation called the hiatal ligament (Fig. 2). It arises from the inner edge of the levator plate and splits fanwise into multiple septa to insert into the upper RN, into the vesicle neck, and into the upper vaginal end. Anteriorly, the ligament fills the gap between the two levator crura at their origin, forming the puboprostatic or pubovesical ligament. The hiatal ligament plays a vital role in harmonizing the action between the levator plate and the intrahiatal organs during evacuation of their contents (defecation and urination). Hiatal ligament subluxation would interfere not only with the act of evacuation but would also lead to prolapse of the intrahiatal organs.

Puborectalis and the Double-Sphincter Control

The puborectalis (PR), as it proceeds backward from its origin in the symphysis pubis, gives off muscle bundles to each intrahiatal organ, forming "individual" voluntary sphincters for these organs (Fig. 8). It gives rise to the external urethral sphincter and deep EAS in both sexes, as well as to the vaginal sphincter in the female and the prostatic sphincter in the male. However, the PR and deep EAS were found fused together, the conjoint muscle being termed the top loop. Each intrahiatal organ is thus provided with a double voluntary sphincteric apparatus: (1) an individual organ sphincter, derived from the PR and specific for the organ, and (2) a "common" tunnel sphincter, the PR itself, which acts on the intrahiatal organs collectively. This separate sphincteric activity for the individual organs under the control of a common continent muscle secures not only an immune sphincteric function for the organ, but a harmonized action among the structures enclosed within the levator tunnel. Furthermore, the double sphincteric mechanism provided to each organ could be a guarantee of functional maintenance in case either of the two sphincters is damaged. Injury to either sphincter alone does not induce incontinence of the organ involved. Unless both the individual and the common sphincters are destroyed, continence can be maintained by either one alone.

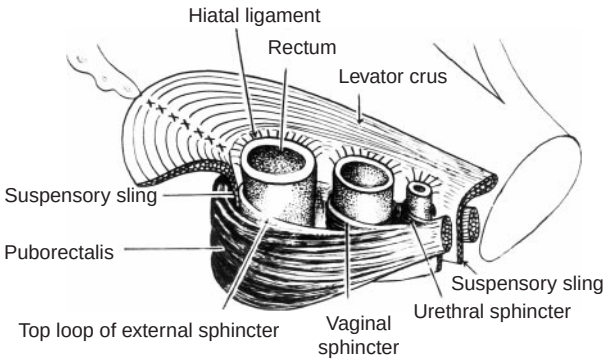


FIGURE 8 Diagram illustrating the "individual" sphincters arising from the puborectalis, which acts as a common sphincter for the intrahiatal structures. Reprinted from Shafik (1998), with permission of Springer-Verlag.

MECHANISM OF DEFECATION

Muscles of Defecation

The muscles that act on the RN are the external and internal anal sphincters, PR, LA, and longitudinal muscle. The external and internal sphincters as well as the PR are muscles of continence. Their role is to contract in order to interrupt or terminate the act of defecation. However, the principal muscles of defecation are the LA and the longitudinal muscles. They act jointly to open the RN at defecation. The two muscles are interrelated due to the fact that the suspensory sling, a part of the levator, constitutes the middle layer of the longitudinal muscle (Figs. 2 and 5).

Anatomical Mechanism of Defecation

With knowledge of the physioanatomical aspects of the pelvic floor muscles and assisted by manometric, electromyographic (EMG), and barium enema studies, the precise mechanism of defecation could be explored. As stools enter the rectum, reflex detrusor contraction and internal sphincter relaxation occur. The continuation of defecation depends on two factors: (1) EAS relaxation and (2) straining. If defecation is acceded to, the EAS is voluntarily relaxed. Straining is necessary to maintain defecation as it raises the intra-abdominal pressure, which serves a double purpose: it compresses the detrusor, which helps evacuation, and it stimulates levator contraction through the straining levator reflex. Although the intra-abdominal pressure compresses the detrusor, the RN is spared, owing to its protected location below the levator plate. When the levator plate contracts, it moves from the cone to the flat position and is elevated and laterally retracted (Fig. 9). This results in pulling on the hiatal ligament, which in turn pulls open the anorectal junction and partially opens the rectal angle. Simultaneously, the suspensory sling contracts and not only pulls up the base loop to unseal the anal orifice, but also partially opens the RN (Fig. 9).

The longitudinal muscle joins the detrusor in contraction, which results in shortening and opening of the RN as well as in complete straightening of the rectal angle. This brings the RN into alignment with the detrusor so that efficient fecal pumping occurs. The final result of the joint contraction of the detrusor, longitudinal muscle, and LA is the opening of the RN for the rectum to evacuate its contents.

Physiologic Mechanism of Defecation

The concerted functions of the anorectal musculature at defecation are initiated and harmonized by voluntary impulses and reflex actions. When the rectal detrusor is distended with fecal mass and the stretch receptors are stimulated, the recto-anal inhibitory reflex is initiated, whereby the rectal detrusor contracts and the internal sphincter relaxes. Detrusor contraction triggers two reflexes: the recto-puborectalis reflex and the recto-levator reflex. These two reflexes act simultaneously, yet have opposite functions; on detrusor contraction, the recto-levator reflex effects a reflex levator contraction, which opens the RN. At the same time, the reflex PR contraction, actuated by the rectopuborectalis reflex, functions to close or keep closed the RN as impulses reach the conscious level to evaluate the

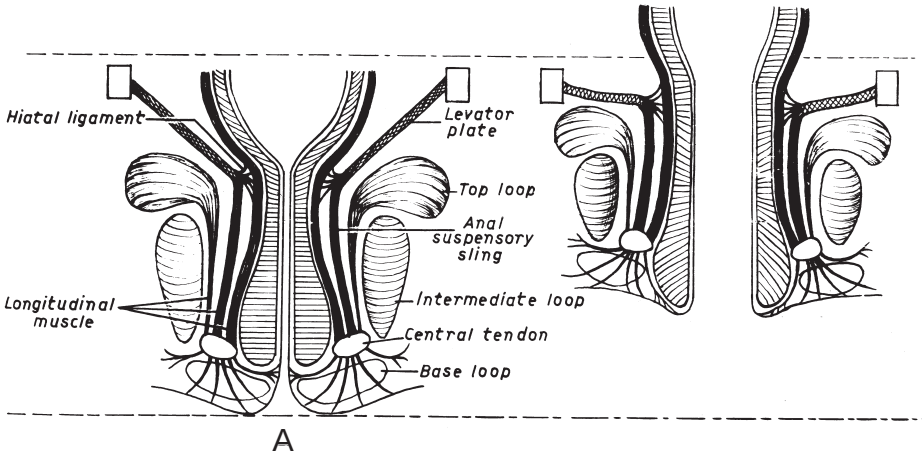


FIGURE 9 Mechanism of defecation. (A) At rest. (B) At defecation: flattening of levator cone and suspensory sling contraction result in opening of the anal canal. From Shafik (1983), with permission. Copyright Urban and Vogel. Reproduced with permission.

circumstances for defecation. If the circumstances are inopportune, the PR continues voluntary contraction.

Voluntary PR contraction evokes the voluntary inhibition reflex, which effects detrusor relaxation. Meanwhile, it aborts the recto-anal inhibitory reflex, which relaxes the internal sphincter. Hence, voluntary PR contraction, through the voluntary inhibition reflex, prevents internal sphincter relaxation, which results in reflex detrusor relaxation and waning of the urge to defecate. However, as soon as circumstances would allow defecation and the sensation of desire to defecate is perceived, the PR muscle relaxes voluntarily and the detrusor evacuates its contents. This demonstrates that the act of defecation is under voluntary control despite the presence of reflex actions sharing in the mechanism of defecation. Thus, although the recto-anal inhibitory and rectolevator reflexes function to open the RN, the recto-puborectalis reflex keeps the RN closed until the decision to defecate has been made.

Mild straining at the start of defecation is a normal physiological process and as such is part of the mechanism of defecation. By elevating the intra-abdominal pressure, it triggers the straining levator reflex, which effects levator contraction and the opening of the RN for the spontaneous evacuation of the stools.

See Also the Following Articles

Anal Sphincter • Constipation • Defecation • Fecal Incontinence • Fistula • Rectum, Anatomy • Sphincters

Further Reading

Shafik, A. (1981). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. XII. Anorectal mobilization: A new surgical access to rectal lesions. *Am. J. Surg.* **142**, 629–635.

Shafik, A. (1982). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. XIV. Anal canal: A fallacious embryological and anatomical entity. *Am. J. Proct. Gastroent. Colon Rect. Surg.* **33**, 14–23.

Shafik, A. (1983). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. XVIII. The levator dysfunction syndrome: A new syndrome with report of 7 cases. *Coloproctology*. **5**, 158–165.

Shafik, A. (1987). A concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. *Dis. Colon Rectum* **30**, 970–982.

Shafik, A. (1991). Constipation: Some provocative thoughts. *J. Clin. Gastroenterol.* **13**, 259–267.

Shafik, A. (1998). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation: Mass contraction of the pelvic floor muscles. *Int. Urogynecol. J.* **9**, 28–32.

Shafik, A., and Doss, S. (1999). Study of the surgical anatomy of the somatic terminal innervation to the anal and urethral sphincters: Role in anal and urethral surgery. *J. Urol.* **161**, 85–89.

Shafik, A. (2000). Injured external anal sphincter in erectile dysfunction. *Andrologia.* **33**, 35–41.

Shafik, A., and El-Sibai, O. (2000). The anocavernosal erectile dysfunction syndrome. II. Anal fissure and erectile dysfunction. *Int. J. Impot. Res.* **12**, 279–283.

Shafik, A., and El-Sibai, O. (2001). Rectal inhibition by inferior rectal nerve stimulation in dogs: Recognition of a new reflex—the "voluntary anorectal inhibition reflex." *Eur. J. Gastroenterol. Hepatol.* **13**, 413–418.

Anal Cancer

PASCAL GERVAZ
University Hospital Geneva, Switzerland

abdomino-perineal resection Surgical procedure involving removal of the rectum and anal sphincter muscles, resulting in the creation of a definitive colostomy. The dissection is achieved through a combined approach, with opening of the abdominal cavity and excision of the anus.

human papillomavirus Family of 60 subtypes of sexually transmitted viruses responsible for genital tract infections, such as condylomata (genital warts). Chronic infection with subtypes 16 and 18 has been identified as a strong risk factor for the development of cervix cancer and anal cancer.

Anal cancer is an uncommon condition that has served during the past two decades as a paradigm for the successful application of chemoradiation to solid tumors. This type of neoplasm also provides a good model to study the contribution of human papillomaviruses and immunodeficiency to the development of cancers.

ANATOMY AND HISTOLOGY

The various definitions of the anal area coexisting in the medical literature have been the source of considerable confusion. Anal cancer may arise from the anal canal or from the anal margin; 85% of the anal cancers occur in the anal canal and 15% occur in the anal margin. The anal canal is about 3.5 cm long and extends from the upper to the lower border of the anal sphincter. The anal margin corresponds to a 5-cm area of perianal skin, measured form the anal verge (Fig. 1). The anal verge is a visible landmark, corresponding to the external margin of the anus, which delineates the junction between the skin epithelium and the hairless and non-pigmented epithelium of the anal canal. Tumors arising within the anal canal are either squamous cell or cloacogenic carcinomas and are characterized by aggressive local growth, including extension to the sphincter muscles. Conversely, cancers originating from the perianal skin have a more favorable prognosis and tend to behave more like other skin cancers. Thus, the anal verge is an important anatomical landmark, separating two histologically distinct epithelial structures that give rise to two types of cancers with different natural histories, prognoses, and treatment.

EPIDEMIOLOGY AND ETIOLOGY

Of all digestive system cancers in the United States, anal cancer comprises 1.5%, with an estimated 3400 new cases and 500 deaths in 1999. The peak incidence is during the sixth decade, but the incidence of these tumors has markedly increased in younger males during the past three decades. In the San Francisco Bay area, the incidence of anal cancer in gay men was 10 times higher than expected in 1973–1979. In addition, the increased incidence of these neoplasms in renal transplant patients clearly suggests that immunosuppression may play a role in tumor development.

Population studies from California and Europe have documented well the relation between anal cancer and receptive anal intercourse in men. In addition, numerous reports have linked anal cancer and cervical cancer in women. Thus, in most cases, anal cancer is now considered a sexually transmitted disease, which is clinically related to the development of anal warts and infection with human papillomavirus (HPV). At the molecular level, the most well-characterized factor in the development of anal cancer is the integration of human papillomavirus DNA into anal canal cell chromosomes. It has been hypothesized that the

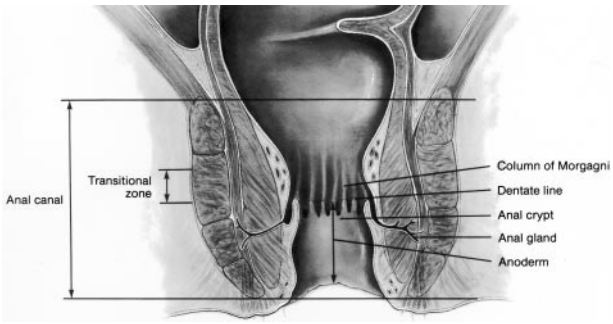


FIGURE 1 Anatomy of the anal canal.

 Copyright 2004, Elsevier (USA). All rights reserved.

increased risk for anal cancer among individuals with AIDS results from HPV infection, and not from immune deficiency. However, the fact that HIV infection seems to contribute to HPV persistence within the anal canal strongly suggests a dual viral etiology in the development of some of these neoplasms.

CLINICAL FEATURES

Tumors from the anal area typically present as a mass associated with bleeding and pain (Fig. 2). Even small lesions can produce significant local symptomatology when they protrude within the anal canal, a blessing in disguise if it facilitates early detection. Unfortunately, these symptoms are often erroneously attributed to "hemorrhoids," with subsequent delay in the diagnosis and treatment. It is important for the clinician to remember that hemorrhoids rarely cause pain (unless thrombosed), thus patients presenting with anal pain should be carefully evaluated, if possible under anesthesia, and biopsies should be obtained. Untreated anal cancer spreads by local extension to adjacent tissues and organs of the pelvic floor, including sphincter muscles, vagina, or prostate. When present, tenesmus, the painful urgency to defecate, suggests that the tumor has spread through the sphincter muscles. Anal margin and anal canal tumors have different patterns of lymphatic drainage. Cancers arising from the more proximal anal canal drain predominantly into the perirectal lymph nodes, whereas tumors within the distal canal and the anal margin drain exclusively into inguinal lymph nodes. Roughly 90% of anal tumors present with locoregionally confined disease, whereas 10% have distant metastases at the time of diagnosis.

Once the diagnosis of anal cancer has been proved by biopsy, it is important to assess the locoregional extension of the tumor and to further characterize its stage. Prognosis for anal cancer is related to tumor size, but it is unclear whether the independent variable is the actual tumor size or the depth of invasion. Neither the histology type nor the degree of differentiation has major prognostic significance, and therefore these criteria have not been included in the American Joint Committee on Cancer (AJCC)/International Union Against Cancer (UICC) staging system. Tumors larger than 5 cm in greatest dimension (T3) and lesions with metastases in regional lymph nodes (N1–3) have an increased risk for tumor recurrence after chemoradiation. Of note, the staging system for most solid tumors rests on the pathologic analysis of a complete surgical specimen. Because most cases of anal cancer are treated initially without surgery, a satisfactory staging method for these cancers remains to be developed.

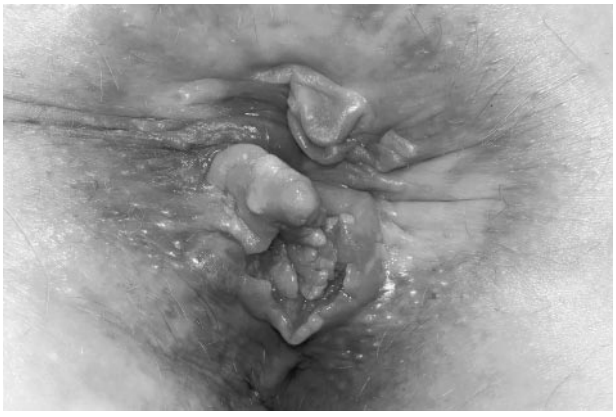


FIGURE 2 Anal cancer.

TREATMENT AND OUTCOME

Anal Margin Cancer

The rationale behind the therapeutic strategies for anal margin cancer derives from the proportionally increased risk for metastases with increasing tumor size. Wide local excision is usually recommended for early-stage lesions of less than 5 cm, and excellent results are reported. As for most skin cancers, the prognosis is good, with 5-year survival rates of 80 to 90%. Recurrences rarely occur and can be managed by reexcision in most cases. Larger (>5 cm) and node-positive lesions are best managed with combined radiation and chemotherapy (as for anal canal cancers). Abdominoperineal resection (APR) remains an option in patients with locally advanced tumors.

Anal Canal Cancer

Although small tumors (<2 cm) arising in the anal canal may be amenable to local excision, most patients present with lesions of greater dimensions. Thirty years ago, these neoplasms were treated routinely with an abdomino-perineal resection, resulting in a definitive colostomy. Thanks to the pioneering work of Nigro in the 1980s, it is now widely accepted that the majority of cases can be cured by a combination of chemotherapy and radiation therapy (CRT). The standard of care consists in the combination of radiation and chemotherapy; external-beam radiation (60 Gy over 5 weeks) with fluorouracil and mitomycin-C results in local eradication of tumors in 70 to 80% of patients. Complete response usually (but not always) results in cure, and

no additional treatment is required. Following chemoradiation, patients should be examined at 3-months intervals; local inspection, digital examination, anoscopy, and biopsy of any suspicious areas are recommended. It is important to differentiate residual disease (positive biopsies less than 6 months after completion of chemoradiation) from tumor recurrence (more than 6 months after cessation of treatment). In most cases, recurrence occurs within 2 years after chemoradiotherapy. The functional results are usually good, and most patients retain a functional anal sphincter, with no alteration in their lifestyle. However, it is known that radiation negatively affects the internal sphincter function, and in a small percentage of cases radiation-induced anorectal dysfunction may result in severe fecal incontinence.

For patients with locally recurrent anal canal cancer, an abdomino-perineal resection remains the treatment of choice. Relatively good results can be achieved with salvage APR in patients with recurrent anal canal cancer; 5-year survival rates up to 50 to 60% have been reported in patients who had evidence of residual disease after CRT. Research is ongoing to determine whether recurrent anal cancer is amenable to a cisplatin-based chemotherapy associated with additional radiotherapy.

SUMMARY

Anal cancer shares many features with cancer of the uterine cervix, including its association with human papillomavirus infection. As such, this type of tumor is now considered a sexually transmitted disease, with homosexual men being a high-risk population. Screening for anal cancer with anal Papanicolaou tests has proved clinically and cost-effective. Future research should investigate the molecular biology of these neoplasms and the genetic features associated with resistance to chemoradiation. In addition, the complex role of immunosuppression in facilitating persistence of HPV infection in the anal canal should be further investigated.

See Also the Following Articles

Anal Canal • Anal Sphincter • Cancer, Overview • Hemorrhoids

Further Reading

Corman, M. L. (1998). Malignant tumors of the anal canal. *In* "Colon & Rectal Surgery," 4th Ed., pp. 863–883. Lippincott-Raven Publ, Philadelphia, PA.

Frisch, M., Glimelius, B., van den Brule, A. J. C., Wohlfart, J., Meijer, C. J., *et al.* (1997). Sexually transmitted infection as a cause of anal cancer. *N. Engl. J. Med.* **337**, 1350–1358.

Fuchshuber, P. R., Rodriguez-Bigas, M., Weber, T., and Petrelli, N. J. (1997). Anal canal and perianal epidermoid cancers. *J. Am. Coll. Surg.* **185**, 494–505.

Goldie, S. J., Kuntz, K. M., Weinstein, M. C., Freedberg, K. A., Welton, M. L., and Palefsky, J. M. (1999). The clinical effectiveness and cost-effectiveness of screening for anal squamous intraepithelial lesions in homosexual and bisexual HIV-positive men. *JAMA* **281**, 1822–1829.

Myerson, R. J., Karnell, L. H., and Menck, H. R. (1977). The National Cancer Database report on carcinoma of the anus. *Cancer* **80**, 805–815.

Nigro, N. D., Vaitkevicius, V. K., Buroker, T., Bradley, G. T., and Considine, B. (1981). Combined therapy for cancer of the anal canal. *Dis. Colon Rectum* **24**, 73–75.

Ryan, D. P., Compton, C., and Mayer, R. J. (2000). Carcinoma of the anal canal. *N. Engl. J. Med.* **342**, 792–800.

Anal Sphincter

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

anal sphincters Rings of smooth muscle or skeletal muscle that surround the distal rectum and anal canal; contract to obstruct the passage of feces.
enteric nervous system Independent integrative nervous system in the gastrointestinal tract; sometimes referred to as the "brain in the gut."
external anal sphincter Rings of skeletal muscle that surround the distal rectum and anal canal; respond to volitional commands to maintain continence when feces have moved into the rectum.
internal anal sphincter Ring of smooth muscle that surrounds the distal rectum and anal canal; obstructs the passage of feces when contracted and permits passage when relaxed.
rectoanal reflex Relaxation of the internal anal sphincter in response to distension of the rectum.

Anal sphincters are rings of muscle surrounding the distal rectum and anal canal. There are two anal sphincters in this region, the internal anal sphincter and the external anal sphincter. Either sphincter, when contracted, assists in closing the anal canal to the passage of feces and flatus. The anal sphincters, in concert with the musculature of the pelvic floor, are responsible for the maintenance of fecal continence.

INTERNAL ANAL SPHINCTER

The internal anal sphincter is an extension of the circular muscle layer of the colon and rectum (Fig. 1A). The sphincter is a ring of smooth muscle that surrounds the terminal rectum and proximal anal canal. Contraction of the internal sphincter closes the anal canal and opposes the passage of feces. Like most smooth muscle sphincters in the alimentary canal, the internal anal sphincter contracts continuously and relaxes only in response to inhibitory neural input. It is the primary barrier to leakage of feces between acts of defecation.

Specialized physiology of the smooth muscle in the sphincter accounts for its ability to exist in a state of sustained contraction. The contractile apparatus of the muscles consists of the two proteins, actin and myosin. Contraction occurs during formation of cross-bridges between actin and myosin filaments. Contractile tension is maintained by a "catch" mechanism that latches the cross-bridges in place without expenditure of additional energy. Inhibitory input from the nervous system leads to uncoupling of the cross-bridges, relaxation of contractile tension, and opening of the sphincter. Excitatory input from the nervous system facilitates the latch mechanism and increases contractile strength and efficiency of the sphincter.

The internal anal sphincter is innervated by motor neurons in the enteric nervous system and postganglionic neurons of the sympathetic nervous system. Most of the enteric motor innervation is inhibitory. Firing of nerve impulses by the inhibitory motor neurons releases inhibitory neurotransmitters at their junctions with the smooth muscle of the sphincter. Two important inhibitory neurotransmitters are vasoactive intestinal polypeptide and nitric oxide. These neurotransmitters act to inhibit contraction of the smooth muscle and open the sphincter. Activation of the sympathetic innervation of the sphincter releases norepinephrine at the neuromuscular junctions. Norepinephrine acts at α-adrenergic receptors on the muscle to increase contractile force and "tighten" the sphincter.

EXTERNAL ANAL SPHINCTER

The external anal sphincter is composed of skeletal muscle that functions like other postural muscles in the body. It is a voluntary muscle that responds to volitional commands to contract and maintain continence when feces are propelled through a relaxed internal anal sphincter into the rectum. The external anal sphincter is composed of a deep external sphincter, a superficial external sphincter, and a subcutaneous external sphincter. These muscle groups form three loops surrounding the anus (Fig. 1C). The deep external sphincter attaches anteriorly to the pubis and forms the uppermost loop. The superficial external sphincter attaches posteriorly to the coccyx and forms the intermediate loop. The subcutaneous external anal sphincter attaches to the

 Copyright 2004, Elsevier (USA). All rights reserved.

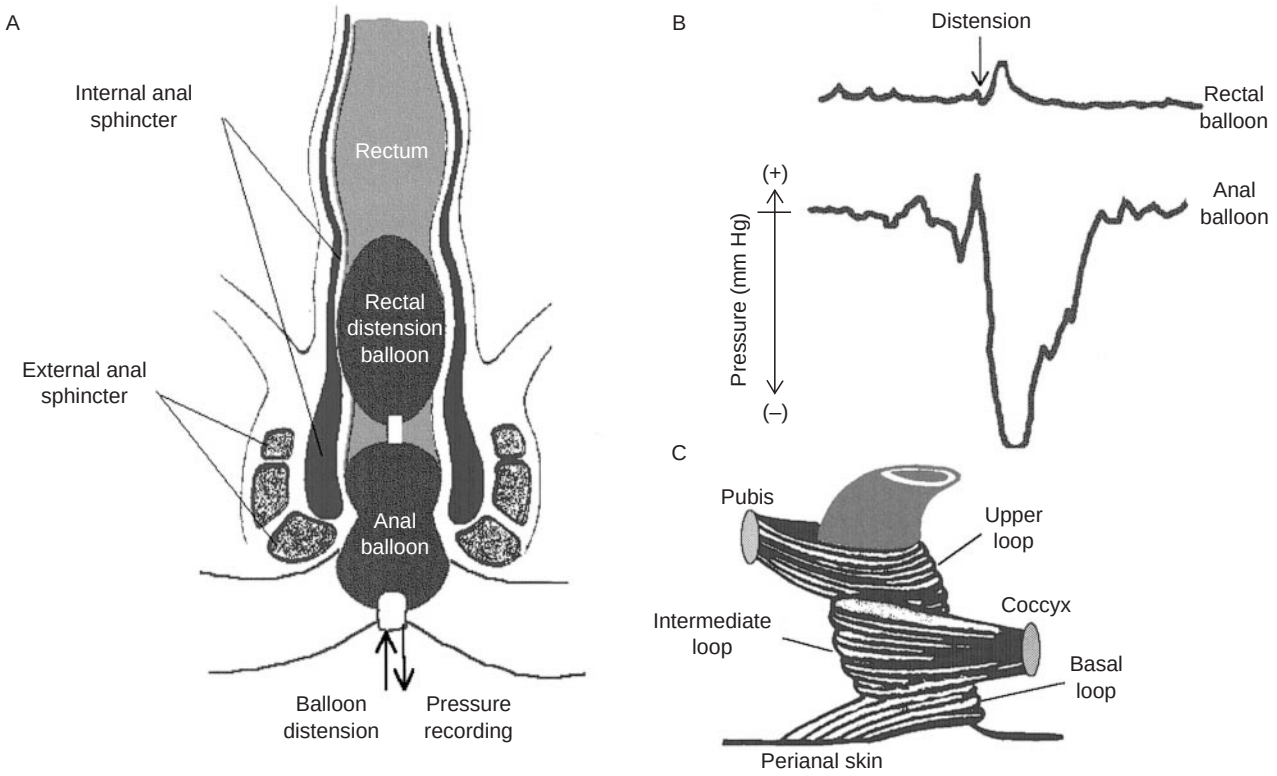


FIGURE 1 Anal sphincters are involved in the rectoanal reflex and are responsible for maintenance of fecal continence. (A) The internal and external anal sphincters surround the anal canal. Distension of a balloon in the rectum evokes reflex relaxation of the internal anal sphincter. (B) Recording of pressure in a balloon placed in the anal canal shows reflex relaxation of pressure in a balloon in the anal canal in response to distension of the rectum. (C) The external anal sphincter is a skeletal muscle consisting of three parts that loop around the anal canal. The upper loop inserts anteriorly on the pubis; the intermediate loop inserts on the coccyx, and the basal loop inserts in the perianal skin. Simultaneous contraction of the three loops closes the anal canal.

perianal skin and forms the lower loop. Simultaneous contraction of the three loops effectively occludes the anal canal.

RECTOANAL REFLEX

Distension of the rectum activates a rectoanal reflex (sometimes called rectosphincteric reflex) to relax the internal anal sphincter. Like other enteric reflexes, this involves a stretch receptor, interneurons in the enteric nervous system, and ultimately activation of firing of inhibitory motor neurons to the smooth muscle sphincter. The rectoanal reflex can be evoked experimentally by inflating a balloon in the rectum and recording changes in pressures in a balloon positioned in the anal canal (Fig. 1A). Distension of the first rectal balloon is followed by a reflex relaxation of pressure in the second balloon in the anal canal. This can be a diagnostic test of the functional integrity of the enteric nervous system in the anorectum. Due to loss of the enteric nervous system in the distal large intestine in Hirschsprung's disease, absence of the rectoanal reflex is a diagnostic hallmark for this disorder.

Reflex relaxation of the internal sphincter allows contact of the rectal contents with the sensory receptors in the lining of the anal canal. This is believed to be the mechanism by which an individual discriminates between gas, liquid, and solid in the anorectum. It also provides an early warning of the possibility of a breakdown of the mechanisms of continence. When this occurs, continence is protected by voluntary contraction of the external anal sphincter, which serves to close the anal canal. Whereas the rectoanal reflex is mediated by the integrative neural networks of the enteric nervous system, the neural circuitry for reflexes

of the external anal sphincter resides in the sacral portion of the spinal cord.

See Also the Following Articles

Achalasia • Anal Canal • Defecation • Enteric Nervous System

Further Reading

Diamant, N., Kamm, M., Wald, A., and Whitehead, W. (1999). American gastroenterological association medical position statement on anorectal testing techniques. *Gastroenterology* **116**, 732–760.

Shafik, A. (1975). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. The external anal sphincter. A triple-loop system. *Invest. Urol.* **12**, 412–449.

Anorexia Nervosa

JULIE E. B. NOLAN AND THOMAS D. GERACIOTI, JR.
University of Cincinnati and Veterans Affairs Medical Center, Ohio

amenorrhea Absence or abnormal stoppage of the menses.
arrhythmia An abnormal rhythm of the heartbeat.
body dysmorphic disorder A preoccupation with what is perceived to be a defect in one's own body or appearance. The defect is imagined entirely or, if an anomaly is present, the concern is excessive or extreme. The distress related to the misperceived defect is so great that significant social, occupational, or other impairment results.
cachexia Severe malnutrition.
compulsion A repetitive, seemingly involuntary, nonpreventable behavior.
genetic polymorphism A variant in the genetic material that codes for a specific biological protein (such as for a neurochemical receptor on the surface of neurons).
hypokalemia Abnormally low potassium concentrations in the blood.
hypophasphatemia Low blood phosphorus concentration.
leuokopenia A low white blood cell count.
nasogastric tube A long, thin tube that is inserted through the nose, down the throat, and into the stomach for the purpose of introducing liquid nourishment ("tube feeding").
obsession A repetitive, seemingly involuntary, nonpreventable thought.
ostopenia Reduced bone formation.
osteoporosis Diminished mineralization of bone.
postprandial After eating.
purge The elimination of undigested or partly digested food or feces by self-induced vomiting or with laxative or enema use.

As eating is necessary for survival, the neuronal mechanisms that regulate appetite and eating behavior are deeply embedded in the brain and are closely linked with neuronal systems that regulate other physiologic mechanisms of fundamental evolutionary significance, including those that underlie emotions, thirst, metabolism, body temperature, bonding, sexuality, reproduction, and sleep. It is not surprising, then, that emotional stresses and conflicts are often manifested in disorders of eating and that disorders of eating are commonly associated with emotional, metabolic, and behavioral symptoms. Although the ability to survive without food for short periods during times of hardship has survival value and, consequently, has been "hard-wired" into the human species over millennia of evolution, the failure to recognize nutritional needs, the relentless, obsessional pursuit of thinness, the sustained refusal to eat, and the failure to recognize self-starvation constitute a disorder of feeding behavior of pervasive life significance. Anorexia nervosa is such a disorder.

CORE SYMPTOMS OF ANOREXIA NERVOSA

Anorexia nervosa, typically a syndrome of adolescent females, is characterized by the intense fear of gaining weight, impairment in self-assessment of body appearance and shape, and the refusal (or inability) to eat

Copyright 2004, Elsevier (USA). All rights reserved.

of the external anal sphincter resides in the sacral portion of the spinal cord.

See Also the Following Articles

Achalasia • Anal Canal • Defecation • Enteric Nervous System

Further Reading

Diamant, N., Kamm, M., Wald, A., and Whitehead, W. (1999). American gastroenterological association medical position statement on anorectal testing techniques. *Gastroenterology* **116**, 732–760.

Shafik, A. (1975). A new concept of the anatomy of the anal sphincter mechanism and the physiology of defecation. The external anal sphincter. A triple-loop system. *Invest. Urol.* **12**, 412–449.

Anorexia Nervosa

JULIE E. B. NOLAN AND THOMAS D. GERACIOTI, JR.
University of Cincinnati and Veterans Affairs Medical Center, Ohio

amenorrhea Absence or abnormal stoppage of the menses.
arrhythmia An abnormal rhythm of the heartbeat.
body dysmorphic disorder A preoccupation with what is perceived to be a defect in one's own body or appearance. The defect is imagined entirely or, if an anomaly is present, the concern is excessive or extreme. The distress related to the misperceived defect is so great that significant social, occupational, or other impairment results.
cachexia Severe malnutrition.
compulsion A repetitive, seemingly involuntary, nonpreventable behavior.
genetic polymorphism A variant in the genetic material that codes for a specific biological protein (such as for a neurochemical receptor on the surface of neurons).
hypokalemia Abnormally low potassium concentrations in the blood.
hypophasphatemia Low blood phosphorus concentration.
leuokopenia A low white blood cell count.
nasogastric tube A long, thin tube that is inserted through the nose, down the throat, and into the stomach for the purpose of introducing liquid nourishment ("tube feeding").
obsession A repetitive, seemingly involuntary, nonpreventable thought.
ostopenia Reduced bone formation.
osteoporosis Diminished mineralization of bone.
postprandial After eating.
purge The elimination of undigested or partly digested food or feces by self-induced vomiting or with laxative or enema use.

As eating is necessary for survival, the neuronal mechanisms that regulate appetite and eating behavior are deeply embedded in the brain and are closely linked with neuronal systems that regulate other physiologic mechanisms of fundamental evolutionary significance, including those that underlie emotions, thirst, metabolism, body temperature, bonding, sexuality, reproduction, and sleep. It is not surprising, then, that emotional stresses and conflicts are often manifested in disorders of eating and that disorders of eating are commonly associated with emotional, metabolic, and behavioral symptoms. Although the ability to survive without food for short periods during times of hardship has survival value and, consequently, has been "hard-wired" into the human species over millennia of evolution, the failure to recognize nutritional needs, the relentless, obsessional pursuit of thinness, the sustained refusal to eat, and the failure to recognize self-starvation constitute a disorder of feeding behavior of pervasive life significance. Anorexia nervosa is such a disorder.

CORE SYMPTOMS OF ANOREXIA NERVOSA

Anorexia nervosa, typically a syndrome of adolescent females, is characterized by the intense fear of gaining weight, impairment in self-assessment of body appearance and shape, and the refusal (or inability) to eat

Copyright 2004, Elsevier (USA). All rights reserved.

enough to maintain even a minimally normal body weight for age and height. The formal diagnosis of anorexia nervosa requires that the body weight be below 85% of the average, expected weight for age, height, and bone structure as determined by standardized scales, such as pediatric growth charts or actuarial tables. As an alternative method to assessing body weight, body mass index (BMI) can be used (BMI = weight in kilograms divided by height in meters squared). A BMI below 17.5 is regarded as in the anorexic range. At an anorexic body mass, ovulation and menstruation (or pubertal development) cease and secondary sexual characteristics regress.

The individual with anorexia nervosa is excessively influenced by her body shape and weight and does not "see" her body accurately (for example, the anorexic adolescent girl who should optimally weigh 135 pounds but thinks that she is fat when she weighs only 85 pounds). In an emaciated, starving individual, the intense fear of becoming fat in combination with the belief that one's body is already fat (or perilously close to fatness) constitutes a delusional-like self-assessment. Some individuals with anorexia nervosa feel globally fat, whereas others are particularly concerned about parts of their bodies, such as the arms, abdomen, or thighs. Individuals with anorexia nervosa usually weigh themselves excessively, frequently view themselves in mirrors, and even measure body parts. Self-esteem becomes linked to self-perception of body weight wherein weight loss is perceived as an extraordinary accomplishment to be proud of and weight gain is perceived as a despicable failure of self-control.

TABLE I DSM-IV Criteria for Anorexia Nervosa

A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected).
B. Intense fear of gaining weight or becoming fat, even though underweight.
C. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight.
D. In postmenarcheal, amenorrhea, i.e., the absence of at least three consecutive menstrual cycles (a woman is considered to have amenorrhea if her periods occur only following hormone, e.g., estrogen, administration).

Restricting type

During the current episode of anorexia nervosa, the person has not regularly engaged in binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Binge-eating/purging type

During the current episode of anorexia nervosa, the person has regularly engaged in binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

ANOREXIA NERVOSA SUBTYPES

In 1980, a distinction was formally made between two subtypes of anorexia nervosa, the restricting type and the bulimic (binge-eating/purging) type (or so-called "bulimarexia") (Table I).

The restricting anorexic individual engages in a range of "restricting" behaviors including fasting, skipping meals, stringent dieting, food avoidance, and strict exclusion of certain foods from the diet, but does not purge. The restricting anorexic tends to be closed and guarded, with strict boundaries between herself and others, as if guarding against excessive permeability of the self to outside influence and, ultimately, against dissolution of the self. Superficially, however, an apparent smugness or arrogance is sometimes manifested as a strong feeling of accomplishment for being able to refrain from eating, unlike many other individuals.

The binge-eating/purging type is characterized by anorexia combined with regular episodes of binge eating and purging or, in some cases, purging without binge eating or after eating only small amounts of food. Additional methods of weight loss are sometimes employed by individuals with the binge-eating/purging type (bulimic anorexia or bulimarexia) of anorexia nervosa. Like persons with bulimia nervosa, individuals with the bulimic type of anorexia nervosa may self-induce vomiting, misuse laxatives, diuretics, and enemas, and exercise excessively.

As an overgeneralization, patients with the bulimic subtype of anorexia nervosa have a higher premorbid weight, a stronger family history of obesity, and higher impulsivity than restricting anorexic patients. Patients with the restricting subtype have greater levels of obsession and may be more isolated socially than those with the binge-eating/purging subtype.

VARIANTS OF ANOREXIA NERVOSA

In addition to the full syndrome of anorexia nervosa, which is in itself seen in forms ranging from the moderate to the severe, low-grade (subclinical) anorexia nervosa, mild anorexic-like syndromes, and other variants of anorexia nervosa also exist. The combination of weight preoccupation and chronic or excessive dieting

is in itself related to the syndrome of anorexia nervosa and, moreover, constitutes a major risk factor for the development of the eating disorder (especially in adolescents). However, severe dieting does not in itself equate with anorexia nervosa. If dieting is excessive, leading to a BMI 5–10% below ideal in the presence of a morbid fear of fatness and unrealistic overestimation of one's own body size, then low-grade or subclinical anorexia nervosa is present. Another example of low-grade or subclinical anorexia nervosa is illustrated by a thin young woman, weighing perhaps 90% of expected, who is intensely afraid of becoming fat, who is rigorously controlled in her emotions and behavior, and who is preoccupied with food preparation. Yet another example of a mild variant of anorexia nervosa, or of a mild anorexic-like syndrome, includes the athletic young woman who trains rigorously and compulsively, is amenorrheic, strictly restricts fat intake, is careful about what she eats, and is at the low margin (90%) of normal body weight. Similarly, slightly underweight persons with bulimia nervosa are not anorexic per se, but have a syndrome that is related to anorexia nervosa. Of course, bulimia nervosa shares in common with anorexia nervosa intense preoccupation with body shape and size and the fear of becoming fat. The bulimic type of anorexic patient—the bulimarexic—further shares compensatory behaviors, such as postprandial vomiting, and even binge eating itself, with the patient with bulimia nervosa. The distinction between bulimia nervosa and the binge-eating/purging type of anorexia nervosa revolves around the degree of weight loss and food restriction. Some individuals show alternating anorexia nervosa and bulimia nervosa.

Misperception or distorted self-evaluation of one's body, or one of its parts, is present not only in anorexia nervosa, but also in body dysmorphic disorder. The essential criterion for body dysmorphic disorder is preoccupation with an imagined or exaggerated defect in appearance.

Another clinical situation related to anorexia nervosa is the food-obsessed individual with obsessive–compulsive disorder, who may show obsessions and/or compulsions related to food, food intake, and food preparation. The essential diagnostic feature of obsessive–compulsive disorder is the presence of either recurrent obsessions or compulsions that are severe enough to cause significant suffering or functional impairment. However, if cachexia and distorted body image are not present, then neither is anorexia nervosa proper.

Individuals with a range of abnormal eating patterns that do not meet diagnostic criteria for bulimia nervosa, binge eating disorder, night eating syndrome (itself a provisional diagnosis), or anorexia nervosa are given the nonspecific diagnosis "Eating Disorder Not Otherwise Specified" according to the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association.

EPIDEMIOLOGY

Anorexia nervosa has an estimated prevalence of 0.2 to 1% in adolescent females; there is an even higher prevalence if milder, so-called subclinical forms of the illness are taken into account. Anorexia nervosa is an uncommon condition in males, accounting for only approximately 5% of patients with this illness. Most disordered eating emerges during adolescence. Weight and body shape concerns commonly develop in late childhood and in puberty. The typical age of onset of anorexia nervosa is between early adolescence and early adulthood (ages 12–24 years), although there are childhood onset cases and some case reports of anorexia nervosa in elderly persons.

Though very little is known about the true prevalence of eating disorders in non-Western and nonwhite populations, ethnic and cultural differences seem to play a role in their development. Anorexia nervosa appears to be rare in developing or non-Western cultures, more commonly emerges from a background of affluence, and is rarely encountered in persons from impoverished circumstances. In this regard, people who were raised in impoverished environments and who were chronically hungry as children are not likely to value fasting as adults. Young females who emigrate from a non-Western culture into a Western, industrialized environment increase their risk of developing an eating disorder. Anorexia nervosa is less prevalent among blacks than among whites and Hispanics; although substantial intraracial variability exists and this heterogeneity should not be minimized, in general thinness as an element of feminine beauty is more valued in white culture than in black culture.

RISK FACTORS, COMORBIDITIES, AND ASSOCIATED FEATURES

Anorexia nervosa, like bulimia nervosa, clusters in families; however, not enough data are available to apportion inheritable and environmental causality. An interaction between the environment and genes most likely confers vulnerability to develop the eating disorder. However, the limited data that are available indicate that monozygotic twins are at least two to four times more likely to show concordance for anorexia

nervosa than dizygotic twins. Elucidation of inheritable factors, such as genetic polymorphisms in neurochemical receptors, may eventually help explain why a small minority of girls, and an even smaller minority of men, deteriorate into anorexia following a period of successful dieting, whereas most individuals can diet without worry of becoming anorexic.

Strict dieting is a major risk factor for the development of anorexia nervosa, with severe dieting most likely to evolve into anorexia in vulnerable individuals. However, pinpointing dieting as a risk factor can be misleading, because, although dieting predates the onset of the illness, dieting is also one of the eating behaviors of the illness, once the illness becomes established. Furthermore, dieting is not the sufficient cause of an eating disorder. Nevertheless, the cultural impact of the high acceptability and popularity of dieting, combined with the social status associated with thinness in women, cannot be ignored. The internalization of the thin ideal for feminine beauty introduces a risk for anorexia in females, wherein the successful pursuit of thinness seems to (superficially) solve life's problems. Moreover, many anorexics do not reach a point where enough body satisfaction occurs to permit nutritious eating. Achieving an ideal body is nearly an impossibility given that many media portrayals of cover girls and professional models use special lighting, makeup, and airbrushes to remove imperfections, thus creating unrealistic body standards and increasing body dissatisfaction among many in the general population. Body dissatisfaction is itself another risk factor for the development of anorexia nervosa.

The presence of psychiatric morbidity is quite common among anorexia nervosa patients (especially among those with the bulimic type of anorexia) and, if preexisting, increases the risk for developing anorexia nervosa. Anorexia nervosa is often accompanied by depressive symptoms such as depressed mood, social withdrawal, irritability, insomnia, and obliteration of interest in sex. Apparent depression—with apathy, sadness, decreased energy, and anxiety—may indicate that the anorexic individual suffers from comorbid major depressive disorder or may be manifesting symptoms of semistarvation. Therefore, after partial or complete weight gain, the possibility of a comorbid full-blown depression should be reevaluated.

Comorbid features of anorexia nervosa commonly include obsessive–compulsive symptoms, which may or may not warrant an additional diagnosis of obsessive–compulsive disorder. At the very least, most individuals with anorexia nervosa are preoccupied with or obsessed with thoughts of food. According to DSM-IV criteria, the diagnosis of obsessive–compulsive disorder requires a pervasive pattern of preoccupation with orderliness, perfectionism, and mental and interpersonal control, at the expense of flexibility, openness, and efficiency. These obsessive–compulsive symptoms are frequently components of anorexic symptomatology and premorbid obsessive–compulsive perfectionism increases the chances of developing anorexia nervosa.

Exercise in individuals with anorexia nervosa is often excessive and compulsive, directed mainly at losing weight and molding body shape, rather than at maintaining fitness or at improving athletic performance in itself. It is remarkable to see emaciated girls with metabolic disturbances and other physiologic signs of starvation vigorously perform pushups and run in place in their hospital rooms.

Within the family, apparent or subtle focus on food, appearance, or weight, the reinforcement of superficially perfectionistic behavior, avoidance of intense feelings and conflict, chaos, and unpredictability, overinvolvement or enmeshment between family members, and lack of respect for boundaries and privacy (or, conversely, neglect) all incrementally conspire to create the context wherein anorexia nervosa can develop. Restricting food intake is one way that an anorexic can control her environment, construct boundaries, assert her individuality, gain a sense of self-satisfaction and esteem, and become a focus of attention. Importantly, anorexia prevents unwanted sexual maturity. Maternal body dissatisfaction increases the chances that her children will develop an eating disorder. Any type of abuse, whether sexual, physical, or emotional, increases the odds of developing an eating disorder.

COMPLICATIONS, PHYSICAL STIGMATA, AND MORBIDITY

Malnutrition is an essential component of anorexia nervosa and starvation-related adaptations or attempted adaptations, complications, and morbidity are routinely encountered. Mortality is high in the disorder, much higher than in uncomplicated bulimia nervosa, with starvation-related cardiac failure often the terminal event. The suicide rate, too, is high in anorexia nervosa. Available data suggest that gross mortality in anorexia nervosa, from all causes, is between 5 and 15%. The mortality rate for actively anorexic patients is roughly 1% per year—somewhat lower among those patients showing recognition of having an eating disorder and somewhat higher among those patients who deny weight abnormalities and who have required involuntary hospital commitment to prevent starvation.

The adverse effects of chronic cachexia are many and essentially involve cessation of growth and development in combination with an attempt to preserve function and life, in the absence of nutrients or fuel. More specifically, these adaptations to starvation involve slowing of metabolism and slowing of the heartbeat, the shutting down of systems related to sex and reproduction, and the activation of the neurohormonal stress response system (of which the hypothalamic–pituitary–adrenal axis is an important component) in a kind of physiologic fight-or-flight reaction. Thus, on some level, the anorexic's body recognizes and reacts to mortal threat even while she blithely denies being underweight. Shrinkage of the cerebrum (brain) occurs during severe anorexia and this contributes to impairments in thinking, reduced ability to achieve insight, and the worsening of any delusional beliefs about body shape, size, and weight. These aforementioned complications, or adaptations, are reversible upon, or within a few months of, weight renormalization.

Common complications of anorexia that are not necessarily reversible are osteopenia and osteoporosis, which can persist throughout life even after recovery from anorexia. In this regard, the risk of bone fractures is greatly increased in former anorexics, hence, the critical importance of vitamin D and calcium in the treatment of this condition.

The many anorexia-related cardiac abnormalities include arrythmias, bradycardia, with heart rates as low as 30 beats per minute not uncommon, decreased blood pressure, and dehydration. Low blood volume causes posture-related (orthostatic) hypotension (low blood pressure) that, in turn, often results in dizziness and even fainting.

Gastrointestinal adaptations or complications include delayed gastric emptying after eating, decreased intestinal movement, postprandial abdominal pain, and constipation. Acute gastric dilation (stomach enlargment), manifested in nausea, vomiting, and painful abdominal distension, can occur upon refeeding.

Other complications of anorexia nervosa include hair loss, brittle nails, reduced secretory activity of the sebaceous glands with dry yellowish skin, pruritis, and sometimes the growth of soft, downy hair on the torso, limbs, and face (so-called lanugo hair).

Anemia, leukopenia, and hypoglycemia are among the many laboratory abnormalities seen (Table II).

TABLE II Some Potential Abnormalities, Complications, or Adaptations in Anorexia Nervosa

Arrythmia (occasionally sudden death)
Bradycardia
Cardiac arrhythmias
Cardiac myopathies
Cold intolerance
Constipation/delayed gut transit time
Dehydration
Delayed gastric emptying
Dizziness or fainting spells
Dry skin
Esophageal reflux, spasms, tears
Foul breath odor (purging type)
Hyperamylasemia
Hypotension
Hypothermia
Increased postprandial satiety
Laboratory
Anemia
Leukopenia
Hypoglycemia
Hypothroxinemia
Hypophosphatemia
Hypokalemia
Hypocalcemia
Loss of hair
Lowered seizure threshold
Muscle weakness
Orthostatic hypotension with dizziness and fainting spells
Osteopenia and osteoporosis
Rectal bleeding
Reduced basal metabolic rate

Refeeding Syndrome

Ironically, it is when the anorexic patient finally begins to ingest adequate nourishment that major anorexia-related complications often ensue. These complications in aggregate are regarded as a "refeeding syndrome" and include hypophosphatemia (wherein phosphorus utilization by tissues increases dramatically and depletes circulating phosphate stores), edema, bloating, constipation, abdominal pain, and, rarely, heart failure due to the inability of the weakened cardiac muscle to cope with sudden increases in body fluid and metabolic demands.

PATHOPHYSIOLOGY AND PSYCHOPATHOLOGY

Anorexia in young women has been observed for centuries and in the 19th century was attributed to hysteria. Hysteria itself has long been associated with repressed sexuality or sexual conflict, albeit less so in modern times. Nevertheless, it is significant that anorexia nervosa typically has its onset during early adolescence, slows or halts development during a normally rapid growth phase, and serves to suppress the development of secondary sexual characteristics at a time when genital sexuality normally begins to emerge.

Patients with the restricting form of anorexia nervosa generally show a personality structure that is more emotionally, cognitively, and behaviorally controlled than do both bulimic patients and healthy persons. Patients with bulimic anorexia, however, are more likely to show impulsivity, depression, and substance abuse.

Reduced metabolic rate is achieved in part through reduced activity of the hypothalamic–pituitary–thyroid system. The hypothalamic–pituitary–gonadal endocrine axis is impaired, with major reductions in the circulating sex hormones testosterone and estradiol. Sexual development ceases in adolescents and sexuality is impaired in adults. Due to impairment of the brain–gonad system, amenorrhea and infertility are present in the overwhelming majority of anorexic females. The hyperactivation of the hypothalamic–pituitary–adrenocortical "stress axis" is manifested by increased urinary free-cortisol secretion and increased brain-derived corticotropin-releasing hormone in the cerebrospinal fluid. Derangements in appetitive sensations are also commonplace.

COURSE AND TREATMENT

Approximately one-third to one-half of individuals with anorexia nervosa recover more or less completely, whereas up to two-thirds of patients continue to have problems including morbid food and body weight preoccupation, amenorrhea, and chronic anorexia. Full recovery often requires years of treatment, but is highly variable and more likely in younger patients, those with the restricting type of anorexia, those with strong support systems (including understanding, supportive families), and those who receive adequate treatment early in the course of illness. The amount and duration of weight loss, the severity of the fear of fatness, and the degree of misperception the patient has of her own body are also major prognostic factors. Severe familial psychopathology and major psychiatric comorbidities are associated with more negative outcomes. Very-low-grade cases of anorexia nervosa, or subclinical anorexia nervosa, have been known to resolve with general support or care but without specialized treatment. The anorexic individual whose maximal weight loss leaves her still close to a minimally acceptable body weight and who acknowledges that she is not fat has, of course, a better prognosis than the chronically skeletal patient who insists, against all evidence to the contrary, that she is fat.

The treatment plan of anorexia nervosa is inextricably tied to the severity of the syndrome. Inpatient hospitalizations are used mainly in dangerous or life-threatening situations to stabilize the patient at an acceptable body weight and to devise the outpatient treatment plan that will ultimately be the vehicle for extended recovery. Severe anorexia nervosa—wherein body weight or body mass index has fallen below 75% of that expected, has fallen very rapidly, or is associated with serious complications or comorbidities—virtually demands inpatient treatment, with supervised, structured, or, if need be, forced refeeding, via a small nasogastric tube, to protect against death. Once an acceptable target weight is reached, most often regarded to be 90% of the expected body mass, and briefly maintained, treatment can be transferred to an outpatient basis. The return of menses (or the resumption of sexual development in the youngest patients) usually occurs if a body mass of at least 90% of expected is sustained.

"Moderate" anorexia nervosa, with body weight 15–25% below what would be expected for height and bone structure, also requires intensive treatment, usually in a structured, inpatient or partial hospitalization setting. Outpatient treatment can be used alone if weight loss can be reversed or stemmed before dangerous cachexia is reached. Mild anorexia nervosa, wherein weight loss is less than 15% of the expected body mass, is by definition subclinical, not because it is not of clinical significance but because it does not satisfy the weight loss criterion necessary to receive a formal diagnosis of anorexia nervosa. Outpatient treatment is usually the treatment of choice here, unless supervening factors, such as a highly unstable mood disorder or severe dehydration and electrolyte abnormalities, exist.

The inpatient treatment of anorexia nervosa is highly structured and best accomplished by an experienced team that ideally includes a psychiatrist, internist, dietician, psychologist and/or social worker, and nurses. Calculation of the precise nutritional requirements necessary to promote slow tissue regrowth and development is necessary to avoid overwhelming the depleted organism and precipitating a refeeding syndrome. Supervised, oral intake of food by the anorexic patient is greatly preferred, with liquid feeding via a thin nasogastric tube used if necessary (parenteral nutrition is practically never indicated)—although some very young patients appear relieved to turn over the responsibility of eating to hospital-provided nasogastric tube feedings. Weight gains of 1 to 3 pounds per week are generally targeted in inpatient refeeding programs, whereas somewhat more modest weight gains are generally accepted in outpatient settings. Calcium (and vitamin D) supplementation is administered to counter decreased bone density and its long-term negative consequences. Mineral and electrolyte levels, among other measures, are closely monitored and corrected

if abnormal. Weight gain often precipitates acute depressive and anxiety symptoms in anorexic patients.

In most cases, anorexic patients will consent to voluntary hospital admission and treatment, but compulsory admissions are sometimes necessary for sicker patients (including those with a history of childhood abuse, self-harm, psychosis, and treatment resistance). If the medical and nursing staff, parents, relatives, and judicial system take the patient's life-threatening illness seriously even when the patient does not seem to, this ultimately appears to have a therapeutic effect on the patient, who ultimately understands the admission as an act of compassion even if it is initially experienced as very distressing. Importantly, that an anorexic patient is admitted for refeeding against her will does not mandate that tube feeding be instituted.

Once weight stabilization occurs, intensive psychotherapy, often for a period of years, becomes a critical therapeutic undertaking. Psychoanalytic psychotherapy or cognitive behavioral therapy (CBT), or a combination of both, is an effective option for treating a patient with anorexia nervosa. CBT focuses on eliminating maladaptive thoughts about body shape and weight and seeks to replace food restriction with normal eating habits. Psychoanalytic psychotherapy—or exploratory, psychodynamic psychotherapy—is valuable because it provides a presuppositionless context for the deep exploration of unconscious needs, desires, and conflicts.

In conjunction with psychotherapy, antidepressant medications (specifically selective serotonin reuptake inhibitors) have been shown to be effective in treating comorbid symptoms, such as depressed mood, low energy, poor concentration, impaired cognition, and obsessive–compulsive symptoms, such as preoccupations with weight and shape. Anxiolytics, mood stabilizers, and other psychotropic drugs are used as needed on a case-by-case basis. Psychopharmacologic interventions are generally less helpful during the acute refeeding phase and more helpful during later stages of recovery and relapse prevention.

SUMMARY

Anorexia nervosa, one of the most common causes of nonaccidental death among young women, is a complex, serious, often long-term condition that may require a variety of treatment modalities (e.g., individual psychotherapy, hospitalization, outpatient treatment, nutritional rehabilitation, family psychotherapy, and pharmacotherapy) that may change throughout the course of illness and recovery. Early intervention is a key in preventing negative sequelae and chronic problems. Variable issues related to self-understanding and self-expression, conflicts regarding sexual maturity, a tendency toward isolation and guardedness, perfectionism, and the pressure of psychosocial and cultural influences that promote an overfocus and overvaluation of thinness, long-term psychotherapy is almost always indicated. Despite substantial progress in the diagnosis and treatment of anorexia nervosa, its cause remains to be determined.

See Also the Following Articles

Appetite • Bulimia Nervosa • Emesis • Nausea • Nutritional Assessment

Further Reading

Becker, A. E., Grinspoon, S. K., Klibanski, A., and Jerzog, D. B. (1999). Eating disorders. *N. Engl. J. Med.* **340**, 1092–1098.

Bruch, H. (1978). "The Golden Cage: The Enigma of Anorexia Nervosa." Harvard University Press, Cambridge MA.

Chial, H. J., McAlpine, D. E., and Camilleri, M. (2002). Anorexia nervosa: Manifestations and management for the gastroenterologist. *Am. J. Gastroenterol.* **97**(2), 255–268.

Heatherington, M. M. (2000). Eating disorders: Diagnosis, etiology, and prevention. *Nutrition* **16**(7/8), 547–551.

Leonard, D., and Mehler, P. S. (2001). Medical issues in the patient with anorexia nervosa. *Eating Behav.* **2**, 293–305.

Neumarker, K.-J. (1997). Mortality and sudden death in anorexia nervosa. *Int. J. Eating Dis.* **21**, 205–212.

Russel, G. F. M. (2001). Involuntary treatment in anorexia nervosa. *Psychiatric Clin. N. Am.* **24**, 337–349.

Stunkard, A. (1997). Eating disorders: The last 25 years. *Appetite.* **29**, 181–190.

Yager, J., *et al.* (2000). American Psychiatric Association Practice Guidelines: Practice Guidelines for the Treatment of Patients with Eating Disorders (Revision). *Am. J. Psychiatry* **157**(Suppl. 1), 1–39.

Yates, A. (1991). "Compulsive Exercise and the Eating Disorders." Bruner Mazel, New York.

Antacids

CHARLES T. RICHARDSON
Baylor University Medical Center, Dallas

antacid A medication that neutralizes hydrochloric acid in the lumen of the stomach.
pepsin A proteolytic enzyme produced by chief cells in the stomach.

Antacids reduce hydrogen ion concentration (gastric acidity) and increase intragastric pH. This mechanism of action is in contrast to medications that inhibit acid secretion such as histamine 2-receptor antagonists and proton pump inhibitors.

PHARMACOLOGY

Most antacids contain magnesium hydroxide and/or aluminum hydroxide. Some antacids contain calcium carbonate. Sodium bicarbonate in the form of baking soda is used also as an antacid. A few commercial antacids contain sodium bicarbonate. The duration of action of sodium bicarbonate is less than that of many antacids because sodium bicarbonate reacts rapidly with hydrochloric acid and the mixture empties quickly from the stomach.

The duration of action of an antacid is related primarily to the length of time a dose of medication remains in the stomach. In the fasting state, the duration of action is relatively short. Doubling the dose of antacid will lengthen only slightly the time during which gastric acidity is reduced. The duration of action can be prolonged by giving an antacid after food. For example, if a dose of antacid is taken at 1 and 3 h after a meal, the duration of effect is prolonged for up to 4 h.

Chemical equations illustrating the reaction between an aluminum hydroxide [$Al(OH)_3$]-containing antacid or a magnesium hydroxide [$Mg(OH)_2$]-containing antacid are shown as follows:

$$Al(OH)_3 + 3HCl \rightarrow AlCl_3 + 3H_2O$$
$$Mg(OH)_2 + 2HCl \rightarrow MgCl_2 + 2H_2O$$

CLINICAL EFFICACY

The clinical benefit of antacids in treating patients is likely due to their ability to reduce gastric acidity and, perhaps, peptic activity. Pepsin, a proteolytic enzyme produced by chief cells in the stomach, is not active at pH levels above 4.0. Thus, when an antacid reduces gastric acidity and increases intragastric pH, peptic activity also is reduced. This, along with the reduction in gastric acidity, may be beneficial in treating some so-called acid/peptic diseases. There is some evidence that aluminum-containing antacids may also absorb pepsin, which also may reduce the harmful effect of pepsin.

Antacids may be useful in the symptomatic therapy of gastroesophageal reflux disease, nonulcer dyspepsia, and gastric and duodenal ulcer disease. An antacid regimen, when compared with a placebo regimen, given during a 4-week period, has been shown to be effective in healing duodenal ulcers in a statistically significant number of patients. In practice, antacids have been replaced by histamine 2 (H2)-receptor antagonists and proton pump inhibitors in the healing of gastroesophageal reflux disease and gastric or duodenal ulcers. Products combining antacids with H2-receptor antagonists have been developed for over-the-counter use in dyspepsia and heartburn. Pure aluminum hydroxide-containing antacids may be useful in reducing phosphate absorption and serum phosphate levels in patients with renal disease.

SIDE EFFECTS

Diarrhea is the major side effect of antacids containing magnesium hydroxide. Magnesium hydroxide-containing antacids should not be given to patients with reduced renal function as these antacids may lead to magnesium toxicity. High magnesium levels also have been reported in rare patients with normal renal function. Antacids can interact with other medications and may alter the absorption or renal elimination of some medications. Sodium bicarbonate can cause metabolic alkalosis and contribute to renal stone formation.

See Also the Following Articles

Duodenal Ulcer • Functional (Non-Ulcer) Dyspepsia • Gastric Ulcer • Gastroesophageal Reflux Disease (GERD) •

 Copyright 2004, Elsevier (USA). All rights reserved.

H2-Receptor Antagonists • Over-the-Counter Drugs • Pepsin • Pharmacology, Overview • Proton Pump Inhibitors

Further Reading

Fordtran, J. S., Morawski, S. G., and Richardson, C. T. (1973). *In vivo* and *in vitro* evaluation of liquid antacids. *N. Engl. J. Med.* **288**, 923–928.

Goves, J., Oldring, J. K., Kerr, D., Dallara, R. G., Roffe, E. J., Powell, J. A., and Taylor, M. D. (1998). First line treatment with omeprozole provides an effective and superior alternative strategy in the management of dyspepsia compared to antacid/alginate liquid: A multicentre study in general practice. *Aliment. Pharmacol. Ther.* **12**(2), 147–157.

Peterson, W. L., and Richardson, C. T. (1983). Pharmacology and side effects of drugs used to treat peptic ulcer. *In* "Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (M. H. Sleisenger and J. S. Fordtran, eds.), 3rd Ed., pp. 708–725. W. B. Saunders, Philadelphia.

Antibiotic-Associated Diarrhea

STAVROS SOUGIOULTZIS AND CHARALABOS POTHOULAKIS
Beth Israel Deaconess Medical Center and Harvard University

antibiotic-associated diarrhea Any case of diarrhea associated with the use of antibiotics.

DNA fingerprinting A laboratory method elaborating a battery of molecular biology techniques in order to generate a pattern of DNA restriction fragments that is unique to an individual or a microbe. In the latter case, one of the aims is microbial source tracking. The method has been used to compare the identity of *Clostridium difficile* strains isolated from both the environment and infected patients.

pseudomembranes Adherent inflammatory exudates overlying sites of colonic mucosal injury; they are characteristic lesions of *Clostridium difficile* colitis overlying sites with dense polymorphonuclear infiltration of the colonic lamina propria. In cases where glandular disruption accompanies the polymorphonuclear cell infiltrate, the histologic picture may resemble a volcanic eruption (volcano lesion), characteristic of pseudomembranous colitis.

pseudomembranous colitis Acute colitis associated primarily with *Clostridium difficile* infection.

toxins A and B High-molecular-weight protein exotoxins released from toxigenic strains of *Clostridium difficile*. The genes for toxins A and B have been cloned and found to be separated by only 1.2 kb on the *C. difficile* chromosome. Both toxins have cytotoxic properties and cause cytoskeletal disorganization and cell rounding in human colonocytes. Toxin A, but not toxin B, is an enterotoxin in animal intestine and causes inflammatory diarrhea in intestinal loops of experimental animals.

Diarrhea associated with antibiotic intake is a common clinical entity with an estimated incidence of 5–25%, depending on the antibiotic used. It is generally accepted that antibiotics cause alterations in the intestinal microecology, which subsequently lead to the diarrheal syndrome. In approximately 20% of cases, it is well established that antibiotic-mediated alterations in intestinal microflora allow *Clostridium difficile*, a toxigenic bacterium, to colonize the intestine and develop a diarrheal infection. *C. difficile* is considered the most commonly identifiable cause of antibiotic-associated diarrhea and represents the offending factor in almost all cases of antibiotic-associated colitis. Thus, *C. difficile* has been studied extensively by both basic and clinical research groups. However, several other causes have been implicated in the pathogenesis of this clinical entity and these are also discussed in this article.

SYMPTOMATOLOGY AND CLINICAL EVALUATION

The clinical evaluation of a patient with suspected antibiotic-associated diarrhea (AAD) involves a detailed

 Copyright 2004, Elsevier (USA). All rights reserved.

Anti-Diarrheal Drugs

MATTHEW BANKS AND DAVID BURLEIGH

St. Bartholomew's and The Royal London School of Medicine and Dentistry, Queen Mary College, University of London

calmodulin An intracellular high-affinity calcium-binding polypeptide forming a calcium/calmodulin complex that regulates ion transport through the modification of cellular regulatory proteins.

enteric nervous system Largely autonomous nervous system, which regulates intestinal function. The cell bodies of enteric neurons lie in plexi within the intestine.

enterochromaffin cell Specialized intestinal epithelial cell that detects lumenal stimuli and activates signaling cascades through the release of mediators such as 5-hydroxytryptamine.

enterocytes Specialized epithelial ion-transporting cells, which collectively form a continuous monolayer lining the intestinal lumen. Tight junctions separate enterocytes, hence creating a selective barrier to electrolytes and nutrients.

secretagogue Substance that induces intestinal secretion.

To those who suffer from diarrhea, the descriptions "too fluid" and "too frequent" stool must seem self-evident. However, the precise definition of this condition requires that the daily stool output exceed 200 g for adults in the developed world. Anti-diarrheal drugs are intended to ameliorate this pathology.

INTRODUCTION

Diarrhea may be acute and a consequence of viral or bacterial infection or it may be chronic and a symptom of an underlying, ongoing disorder, such as malabsorption or inflammatory bowel disease. Diarrhea may also be present where there is a bowel tumor, diabetic autonomic neuropathy, hyperthyroidism, or a number of other conditions. It can also be drug-induced, such as in laxative abuse.

Such variable pathophysiologies will necessitate different therapeutic approaches, often directed at correcting the underlying primary cause rather than the resulting symptom of diarrhea. Pharmacotherapy for diarrhea per se is in the main restricted to irritable bowel syndrome, where the predominant symptom is diarrhea, and in the mild acute infective condition typified by traveler's diarrhea.

The mainstay of anti-diarrheal drug therapy, even after centuries of use, remains the opiate group of drugs. Surprisingly, despite the devastating effects of this condition in the underdeveloped world, there is still no effective, safe, anti-secretory drug for the treatment of acute secretory diarrhea.

PATHOPHYSIOLOGY

Normally, approximately 8–10 liters of fluid enters the lumen of the small intestine daily both from ingestion and from endogenous salivary, gastrointestinal, and pancreatic secretions. Net absorption occurs in the small intestine, driven by osmotic gradients that result from the transport of electrolytes (mainly Na^+ and Cl^-), sugars, and amino acids across the epithelial monolayer lining the length of the small intestinal lumen. Approximately 1 to 1.5 liters of fluid enters the colon and from there approximately 100 ml will be excreted in the feces. Both the small intestine and large intestine have spare absorptive capacity and can absorb a maximum of 16 and 5 liters, respectively. As stool consists mainly of water, especially in diarrhea, most cases of diarrhea can be ascribed to disorders of intestinal electrolyte and water transport. From a mechanistic viewpoint, diarrhea may have several causes.

Secretory Diarrhea

Secretory diarrhea results from excessive secretion or diminished absorption, or both, of electrolytes (mainly Na^+ and Cl^-) and water by epithelial cells of the intestinal mucosa. Maturing cells in the crypts of Lieberkuhn of the small intestine normally secrete electrolytes and water into the lumen. This process, stimulated by gastrointestinal hormones and neurotransmitters released after a meal, is essential for maintaining the liquidity of small intestinal contents, thereby allowing efficient digestion and absorption of nutrients. Mature epithelial cells on the intestinal villi are responsible for absorption of electrolytes, water, and nutrients. Secretory diarrhea can be produced by microbial

 Copyright 2004, Elsevier (USA). All rights reserved.

infection, gastrointestinal hormone-producing tumors, bile salt malabsorption, and inflammatory mediators, such as prostaglandins and leukotrienes.

Osmotic Diarrhea

Osmotic diarrhea occurs when an increased lumenal osmotic load results in retention of fluid in the intestinal lumen. Both electrolytes and nutrients, such as sugars and amino acids, can contribute to osmosis. If their absorption is reduced, then osmotic diarrhea can occur. Causes include celiac disease, laxative use, and lactase deficiency.

Increased Propulsive Motility

Increased propulsive motility may result in diarrhea as the extent of absorption can be diminished by a decreased transit time of intestinal contents. This may occur in hyperthyroidism and also in irritable bowel syndrome.

ORAL REHYDRATION THERAPY

In acute infectious diarrhea, rehydration is the first priority of treatment. This is particularly important in infants, the frail, and the elderly, where the risks of dehydration and electrolyte or pH imbalance are greatest. Promptly administered oral rehydration therapy (ORT) saves many lives and is the only therapy needed for the diarrhea of viral gastroenteritis in the young. Recourse to intravenous electrolyte and fluid replacement may be required in severe cases. ORT will not immediately reduce the volume of diarrhea, but absorption of the glucose electrolyte solution leads to the correction of fluid and electrolyte imbalances. ORT utilizes the ability of small intestinal epithelial cells to absorb sodium and glucose even in secretory states. This cotransport mechanism sets up an osmotic gradient across the intestinal mucosa, resulting in the transport of water from the intestinal lumen to the bloodstream. There are a number of approved formulations for ORT. Most have, as their basis, NaCl, KCl, sodium citrate, and glucose.

ANTI-MOTILITY, ANTI-SECRETORY AGENTS

Opioids and Enkephalinase Inhibitors

Opioids are compounds that are chemically related to opium, which is obtained from the juice of the opium poppy. Opiate drugs include those derived from opium, such as morphine and codeine, as well as synthetic compounds, such as loperamide and diphenoxylate. The realization that these chemicals were reacting with receptors in the body led to the discovery of endogenous opioid peptides, Met- and Leu-enkephalin, which are the naturally occurring ligands for opioid receptors. Opioid receptors are divided into three main groups, designated mu (μ), delta (δ) and kappa (κ). Morphine acts primarily on μ receptors, whereas the enkephalins act on μ and δ receptors. Historically, morphine was the first drug widely used for the treatment of diarrhea and today opiates still remain the most useful class of anti-diarrheal drug. Opiates slow intestinal transit, thereby allowing increased time for the absorption of lumenal contents. Their action is complex, involving increased tone (muscle contraction) and rhythmic contraction of intestinal smooth muscle but diminished propulsive activity. The pyloric, ileocecal, and internal anal sphincters are all contracted. The overall effect in healthy individuals is to produce an increase in "nonpropulsive" motility, resulting in constipation. The main opiates used for diarrhea are loperamide and diphenoxylate. Unlike morphine, very little of these drugs crosses the blood–brain barrier and they are given exclusively for their actions on the gastrointestinal tract. In addition to its anti-motility effects, loperamide acts as an anti-secretory agent through the inhibition of calmodulin (see below). Extensive studies on the action of morphine reveal that it can act centrally where stimulation of μ receptors in the central nervous system results in the slowing of transit throughout the intestine. Within the intestinal wall, opioid receptors have been found on enteric nerves, epithelial cells, and smooth muscle cells. Although most cases of diarrhea result from reduced absorption or increased secretion, or both, by the intestine, there is still much debate as to whether anti-secretory actions or slowing of intestinal transit is the primary mechanism underlying the anti-diarrheal actions of opiate drugs. Certainly there appear to be differences in the pharmacological profiles of morphine compared to the enkephalins and loperamide. The peptides act on δ receptors and may thus produce anti-secretory effects and a similar action would result from the inhibition of calmodulin by loperamide (see below).

Although slowing intestinal transit relieves the symptoms of diarrhea, there is a risk that this may prolong the contact time of infectious agents with the intestinal mucosa. Such concerns turned out to be largely unfounded, but stimulated a search for compounds with a more favorable therapeutic profile. Despite their anti-secretory effects, enkephalins were unsuitable due to a short duration of action. However, inhibitors of

enkephalinase, the enzyme that metabolizes enkephalins, are a novel class of drugs with marked anti-secretory actions. Racecadotril (acetorphan) is a selective inhibitor of enkephalinase and reduces stool volume and disease duration in infectious diarrhea without the side effect of constipation.

α-Adrenergic Receptor Agonists

Noradrenergic neurons with projections to the gastrointestinal tract influence motility, mucosal transport, and blood flow. Reduction of gastrointestinal motility occurs as a result of presynaptic inhibition of acetylcholine release from cholinergic neurons. Some sphincters are also contracted by a direct action on the sphincteric smooth muscle. Noradrenergic neurons also stimulate absorption and inhibit electrolyte and water secretion by an indirect inhibitory action on cell bodies of submucosal secretomotor neurons and also by a direct action on enterocytes. All these actions are mediated through α2-adrenergic receptor stimulation. Neural α2 receptors are located presynaptically on nerve terminals and inhibit the release of neurotransmitters. Not surprisingly, α2 agonists, such as clonidine, have proved effective anti-diarrheal agents in some circumstances, such as diabetic neuropathy where there was a loss of noradrenergic innervation. A disadvantage of clonidine is that it readily crosses the blood–brain barrier, causing hypotension.

Somatostatin Analogues

Immunohistochemical and radioimmunoassay techniques have demonstrated the presence of a number of peptides in enteric nerves where they presumably function as neurotransmitters. Many of these neurons either directly innervate the gastrointestinal mucosa or terminate around neural cell bodies in the submucosal plexus, thus indicating a role in the control of electrolyte and water transport. Somatostatin, a 14-amino-acid peptide, serves a number of physiological functions in the gastrointestinal tract, including inhibition of exocrine, gastric, and pancreatic secretions, inhibition of secretion of several gastrointestinal hormones, decrease of gastrointestinal motility, and inhibition of chloride secretion by epithelial cells. The usefulness of somatostatin as an anti-diarrheal agent is limited by a short half-life, necessitating administration by intravenous infusion. Newer agents such as octreotide, a synthetic analogue of somatostatin, have a longer duration of action and increased potency, providing greater therapeutic potential. Octreotide has been used successfully in cases of diarrhea due to peptide-secreting tumors such as 5-hyroxytryptamine (5-HT)-secreting carcinoid tumors or vasoactive intestinal peptide (VIP)-secreting VIPomas. Octreotide has also been used in dumping syndrome, where rapid emptying of gastric contents, usually seen postoperatively, results in excessive release of enteric peptides into the general circulation. The beneficial effects of octreotide probably result more from its actions on decreasing gut motility and release of gastrointestinal peptides than from a direct pro-absorptive or anti-secretory action.

Cholestyramine

After surgical resection of the terminal ileum, the normal enterohepatic circulation of bile salts is interrupted. Loss of bile salt absorption results in elevated lumenal concentrations in the colon, where they stimulate electrolyte and water secretion. Cholestyramine is an anion exchanger that is not absorbed and effectively binds and neutralizes bile salts.

Bismuth

Bismuth has anti-microbial properties, which probably account for the majority of its anti-diarrheal actions. Bismuth subsalicylate has been used successfully for the treatment of traveler's diarrhea.

5-Hydroxytryptamine Receptor Antagonists

Intestinal secretion induced by 5-HT is complex. It may be released from neurons and enterochromaffin cells and can act on enteric nerves, smooth muscle, and epithelial cells to produce its effects. 5-Hydroxytryptamine type 3 (5-HT_3) receptors are located on sensory neurons and 5-HT_3 antagonists have been used in the treatment of chemotherapy-induced vomiting as well as diarrhea associated with irritable bowel syndrome. 5-HT_3 antagonists do not prevent all the intestinal secretory effects of 5-HT. 5-HT is also implicated in the pathogenesis of cholera toxin-induced diarrhea. Cholera toxin releases 5-HT from enterochromaffin cells; 5-HT then activates the sensory neurons of a neurosecretory reflex.

Berberine

Berberine is a plant alkaloid that may have anti-microbial as well as anti-secretory and anti-motility effects. In China and India, it has been used as an anti-diarrheal agent for centuries. In humans, it has

been shown to be effective against enterotoxigenic *Escherichia coli*-induced diarrhea.

Calmodulin Inhibitors

Some secretagogues such as acetylcholine and 5-HT act by increasing intracellular calcium. Calcium combines with calmodulin, a calcium-binding protein, and the complex increases adenylate and guanylate cyclase activity, resulting in elevated concentrations of cyclic AMP and cyclic GMP. These nucleotides are fundamental intracellular messengers in electrolyte and water transport, causing enhanced chloride secretion and diminished sodium chloride absorption. A positive correlation between calmodulin binding and anti-diarrheal activity has been demonstrated for loperamide, chlorpromazine, and a number of other compounds. Zaldaride maleate, a selective and potent inhibitor of intestinal calmodulin, demonstrated a reduction of stool frequency and duration in traveler's diarrhea.

FUTURE DIRECTIONS

Potassium Channel Inhibition

The opening of chloride channels in the apical membrane plays a central role in the promotion of intestinal secretion and is the final common pathway of agents that cause acute secretory diarrhea (Fig. 1). Chloride secretion is an electrogenic process and the intestinal lumen becomes more negative, causing paracellular flow of sodium ions and a concomitant osmotic flow of water. It is possible to reduce chloride secretion by reducing the intracellular negative potential of the epithelial cell as this cellular hyperpolarization normally drives chloride ions through apical chloride channels into the intestinal lumen. Hyperpolarization is produced by the exit of potassium ions from the cell through basolaterally located (in the small intestine) potassium channels. It follows, therefore, that drugs blocking or inhibiting such channels have anti-secretory potential. In the human colon, patch-clamp single-channel recording and the use of agents that block potassium channels have demonstrated more than one type of potassium channel. More recent investigations have revealed that the established anti-secretory agents berberine, clonidine, and loperamide possess potassium channel-blocking effects. Finally, secretory responses of human ileal mucosa to VIP and *E. coli* heat-stable enterotoxin can be inhibited by potassium channel blockade. With the multiplicity of potassium channels in existence, it is possible that those on basolateral membranes of intestinal epithelial cells may have individual characteristics that can be exploited for the development of selective anti-diarrheal agents.

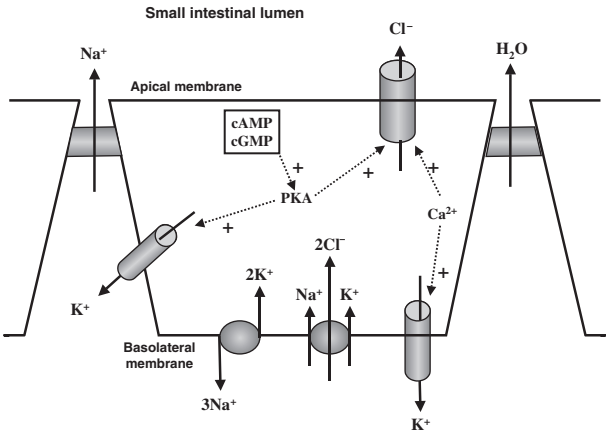


FIGURE 1 Chloride ion secretion and its regulation. The effects of cyclic AMP (cAMP) and cyclic GMP (cGMP) are mediated through the actions of protein kinase A (PKA). This enzyme phosphorylates intestinal membrane proteins of ion channels, leading to increased apical conductance of chloride ions and basolateral conductance of potassium ions. The basolateral exit of potassium ions hyperpolarizes the cell, providing an electrical driving force for the exit of chloride ions. Not shown are receptors for secretagogues located on both the apical and basolateral membranes. Cholera toxin and *Escherichia coli* heat-stable toxin act apically to elevate cAMP and cGMP, respectively. Acetylcholine and vasoactive intestinal polypeptide act basolaterally to elevate calcium and cAMP, respectively.

Vasoactive Intestinal Polypeptide Receptor Antagonists

The demonstration of VIP receptors on basolateral membranes of epithelial cells and the presence of VIP within enteric neurons innervating mucosal epithelial cells and blood vessels indicate a physiological role for this peptide in the control of mucosal ion transport. There are two main types of secretomotor neuron: cholinergic and noncholinergic. The noncholinergic neurons appear to mediate most of the local reflex responses and utilize VIP as their neurotransmitter.

VIP causes an increase in intestinal secretion in humans, both in normal volunteers and in those individuals with VIPomas. It is thought that the enteric nervous system is responsible for a significant proportion of the change in fluid transport induced by cholera toxin (Fig. 2) and antagonism of VIP receptors has been shown to reduce intestinal secretion produced by the toxin *in vivo*.

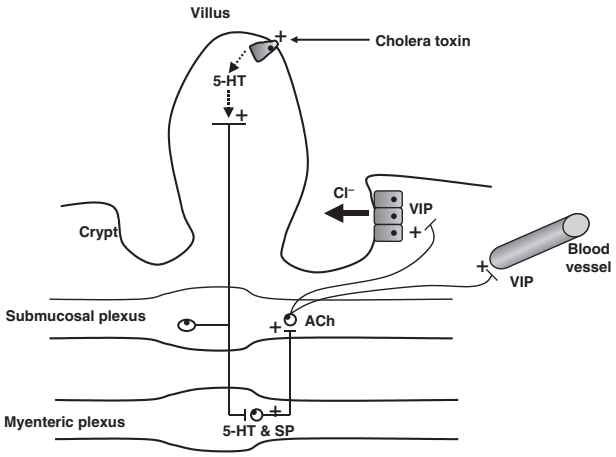


FIGURE 2 Putative neurosecretory reflex of cholera toxin. Cholera toxin stimulates 5-hydroxytryptamine (5-HT) release from enterochromaffin cells. 5-HT activates sensory neurons that synapse with cholinergic interneurons in the myenteric plexus. Secretomotor neurons that project to both epithelial cells and mucosal blood vessels are stimulated by acetylcholine (ACh) to release vasoactive intestinal polypeptide (VIP). VIP causes intestinal secretion and vasodilation of mucosal blood vessels.

See Also the Following Articles

Antibiotic-Associated Diarrhea • Diarrhea • Diarrhea, Infectious • Diarrhea, Pediatric • Over-the-Counter Drugs • Pharmacology, Overview • Somatostatin • Traveler's Diarrhea • Vasoactive Intestinal Peptide (VIP) • Vipoma

Further Reading

Banks, M. R., and Farthing, M. J. G. (2001). Current management of acute diarrhoea. *Prescriber* **19**, 83–86.

Burleigh, D. E. (2003). Involvement of inwardly rectifying K^+ channels in secretory responses of human ileal mucosa. *J. Pharmac. Pharmacol.* **55**, p. 527–531.

Chang, E. B., and Rao, M. C. (1994). Mechanisms of physiological and adaptive responses. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), pp. 2027–2069. Raven Press, New York.

Epple, H. J., Fromm, M., Riecken, E. O., and Schulzke, J. D. (2001). Antisecretory effect of loperamide in colon epithelial cells by inhibition of basolateral K^+ conductance. *Scand. J. Gastroenterol.* **36**, 731–737.

Jodal, M., Holmgren, S., Lundgren, O., and Sjoqvist, A. (1993). Involvement of the myenteric plexus in the cholera toxin-induced net fluid secretion in the rat small intestine. *Gastroenterology* **105**, 1286–1293.

Mourad, F. H., and Nassar, C. F. (2000). Effect of VIP antagonism on rat jejunal fluid and electrolyte secretion induced by cholera and *Escherichia* enterotoxins. *Gut* **47**, 382–386.

Sandle, G. I., McNicholas, C. M., and Lomax, R. B. (1994). Potassium channels in colonic crypts. *Lancet* **343**, 23–25.

Schiller, L. R. (1995). Review article: Anti-diarrhoeal pharmacology and therapeutics. *Aliment. Pharmacol.* **9**, 87–106.

H2-Receptor Antagonists • Over-the-Counter Drugs • Pepsin • Pharmacology, Overview • Proton Pump Inhibitors

Further Reading

Fordtran, J. S., Morawski, S. G., and Richardson, C. T. (1973). *In vivo* and *in vitro* evaluation of liquid antacids. *N. Engl. J. Med.* **288**, 923–928.

Goves, J., Oldring, J. K., Kerr, D., Dallara, R. G., Roffe, E. J., Powell, J. A., and Taylor, M. D. (1998). First line treatment with omeprozole provides an effective and superior alternative strategy in the management of dyspepsia compared to antacid/alginate liquid: A multicentre study in general practice. *Aliment. Pharmacol. Ther.* **12**(2), 147–157.

Peterson, W. L., and Richardson, C. T. (1983). Pharmacology and side effects of drugs used to treat peptic ulcer. *In* "Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (M. H. Sleisenger and J. S. Fordtran, eds.), 3rd Ed., pp. 708–725. W. B. Saunders, Philadelphia.

Antibiotic-Associated Diarrhea

STAVROS SOUGIOULTZIS AND CHARALABOS POTHOULAKIS
Beth Israel Deaconess Medical Center and Harvard University

antibiotic-associated diarrhea Any case of diarrhea associated with the use of antibiotics.

DNA fingerprinting A laboratory method elaborating a battery of molecular biology techniques in order to generate a pattern of DNA restriction fragments that is unique to an individual or a microbe. In the latter case, one of the aims is microbial source tracking. The method has been used to compare the identity of *Clostridium difficile* strains isolated from both the environment and infected patients.

pseudomembranes Adherent inflammatory exudates overlying sites of colonic mucosal injury; they are characteristic lesions of *Clostridium difficile* colitis overlying sites with dense polymorphonuclear infiltration of the colonic lamina propria. In cases where glandular disruption accompanies the polymorphonuclear cell infiltrate, the histologic picture may resemble a volcanic eruption (volcano lesion), characteristic of pseudomembranous colitis.

pseudomembranous colitis Acute colitis associated primarily with *Clostridium difficile* infection.

toxins A and B High-molecular-weight protein exotoxins released from toxigenic strains of *Clostridium difficile*. The genes for toxins A and B have been cloned and found to be separated by only 1.2 kb on the *C. difficile* chromosome. Both toxins have cytotoxic properties and cause cytoskeletal disorganization and cell rounding in human colonocytes. Toxin A, but not toxin B, is an enterotoxin in animal intestine and causes inflammatory diarrhea in intestinal loops of experimental animals.

Diarrhea associated with antibiotic intake is a common clinical entity with an estimated incidence of 5–25%, depending on the antibiotic used. It is generally accepted that antibiotics cause alterations in the intestinal microecology, which subsequently lead to the diarrheal syndrome. In approximately 20% of cases, it is well established that antibiotic-mediated alterations in intestinal microflora allow *Clostridium difficile*, a toxigenic bacterium, to colonize the intestine and develop a diarrheal infection. *C. difficile* is considered the most commonly identifiable cause of antibiotic-associated diarrhea and represents the offending factor in almost all cases of antibiotic-associated colitis. Thus, *C. difficile* has been studied extensively by both basic and clinical research groups. However, several other causes have been implicated in the pathogenesis of this clinical entity and these are also discussed in this article.

SYMPTOMATOLOGY AND CLINICAL EVALUATION

The clinical evaluation of a patient with suspected antibiotic-associated diarrhea (AAD) involves a detailed

 Copyright 2004, Elsevier (USA). All rights reserved.

history to identify other etiologies of diarrhea, as well as careful review of the concomitant medications and nutritional supplements. A long list of drugs such as antacids, anti-arrhythmics, anti-neoplastics, antihypertensives, cholinergics, laxatives, caffeine, and magnesium and potassium supplements can cause diarrhea that usually starts shortly after initiation of therapy, but can also occur following chronic drug treatment or after an increase in the dose of the drug. Antibiotics such as clindamycin, lincomycin, ampicillin, and cephalosporines are more likely to cause AAD, whereas aminoglycosides, erythromycin, and trimethoprim-sulfamethoxazole are considered safer. However, the clinician should be aware that virtually all antimicrobial agents, including antifungals and antivirals, can cause AAD, irrespective of the route of administration. In addition, several case reports implicate even metronidazole and vancomycin, the antibiotics most commonly used to treat *Clostridium difficile* infection, in the development of *C. difficile*-induced AAD.

AAD causes mild symptoms and it is clinically significant when there are more than three mushy or watery stools per day. However, when *C. difficile* is the underlying causative agent, the clinical presentation may vary from mild diarrhea to a severe bloody diarrheal syndrome accompanied by abdominal pain and systemic complications such as dehydration, fever, leukocytosis, and hypoalbuminemia.

Sigmoidoscopic examination is not always warranted in AAD, but, if performed, reveals either normal mucosa or mild edema or hyperemia. In cases of suspected *C. difficile* infection, endoscopy is not always necessary since the presence of *C. difficile* toxins in the stool establishes the diagnosis. However, in highly suspicious cases, and when toxin assays are negative, endoscopy is indicated. In these cases, colonoscopy should be performed instead of sigmoidoscopy, because *C. difficile*-induced mucosal lesions may occasionally spare the rectum and sigmoid colon.

Development of pseudomembranes, identified as raised yellow small plaques on the colorectal mucosa, is the characteristic endoscopic finding in *C. difficile* colitis and a sign of severe infection. In milder cases, nonspecific, diffuse, or patchy erythematous colitis without pseudomembranes may be observed.

Endoscopy should be either avoided in patients with suspected fulminant *C. difficile* colitis or cautiously performed with minimal air inflation to avoid bowel perforation. In such cases, clinical signs such as high fever, diffuse abdominal pain, involuntary guarding, or rebound tenderness together with radiologic findings should guide management decisions. Computed tomography findings in severe cases include thickening of the bowel wall in affected sites, signs of intestinal obstruction in cases of ileus, or intraperitoneal fluid when perforation has occurred. Dilated colon (>7 cm in its greatest diameter) in plain abdominal images can safely diagnose toxic megacolon.

PATHOPHYSIOLOGY OF AAD

Microbes Other Than *C. difficile* as Causative Agents of Antibiotic-Associated Diarrhea

Overall, it is generally accepted that 2–3% of AAD cases may be due to infections other than *C. difficile*. Several reports implicate the overgrowth of specific microorganisms following antibiotic therapy as the cause of AAD. *Staphylococcus aureus* was reported to be the cause of AAD in the 1950s and 1960s. However, it is now believed that the vast majority of these cases were actually caused by *C. difficile* infection, which had not been connected to the etiology of AAD at that time. *Clostridium perfringens* enterotoxin and high counts of enterotoxigenic strains of *C. perfringens* have been found in the stools of patients with diarrhea that developed after antibiotic treatment. It must be pointed out, however, that although *C. perfringens* spores can be found in the hospital environment, nosocomial *C. perfringens* infections can occur without prior use of antibiotics, thus calling into question the association of the microorganism with AAD.

Klebsiella oxytoca has been also implicated as the causative agent of acute segmental hemorrhagic penicillin-associated colitis. This is a rare complication of penicillin treatment, characterized by acute hemorrhagic diarrhea and painful abdominal cramps. The disease is thought to result from infection by cytotoxin-producing *K. oxytoca* strains.

A single study found that multidrug-resistant *Salmonella newport* was the cause of AAD in 12 patients who had consumed contaminated beef from animals fed subtherapeutic doses of antimicrobials. Although this is the only report, the findings are well documented and *Salmonella* may be considered a rare causative agent of AAD.

Overgrowth of *Candida* species (*Candida albicans* and *Candida tropicalis*) as a result of antibiotic treatment has also been associated with the development of AAD in pediatric and adult patients. Animal studies have shown that the yeast is able to depress lactase activity, potentially leading to lactose intolerance and malabsorption and, through release of endotoxin-like substances, to lumenal secretion of ions into the jejunum. Irrespective of the underlying mechanism(s),

Candida species should be considered a putative cause of AAD, especially in immunosuppressed patients.

Mechanisms Related to Direct or Indirect Effects of Antibiotics on Intestinal Functions

Antibiotics may exert various effects on the gastrointestinal tract, independent of their antimicrobial activity. Erythromycin, for instance, is a motilin-receptor agonist and can accelerate the rate of gastric emptying, which has been reported to cause functional diarrhea. Amoxicillin/clavulanate, another commonly prescribed antibiotic combination, mediates nonspecific motility disturbances in the gastrointestinal tract, leading to diarrhea. Moreover, neomycin, an aminoglycoside antibiotic frequently used as an adjunct in the treatment of hepatic encephalopathy, induces morphologic alterations of the intestinal mucosa, leading to malabsorption and diarrhea. In addition to these "direct" effects on intestinal motility or absorptive capacity, antibiotics cause alterations in the bacterial population of the colon, mainly by decreasing the number of anaerobes such as *Bacteroides* species. Anaerobic bacteria metabolize carbohydrates that escape absorption, reach the colon, and produce short-chain fatty acids such as acetate and butyrate. Short-chain fatty acids are subsequently absorbed by the colonic epithelium together with fluids and electrolytes. Antibiotics reduce the number of colonic anaerobes, thereby decreasing the fermentation of nonabsorbable carbohydrates, which eventually accumulate in the colonic lumen and may lead to osmotic diarrhea. Although this mechanism may explain some published observations, other studies found discrepancies between suppression of carbohydrate metabolism and manifestations of diarrhea. Therefore, it seems likely that other mechanisms, such as an adaptive increase in colonic transit time that has been described in cases of osmotic diarrhea, may operate to compensate for the increased osmotic load in the colonic lumen.

Another metabolic function of colonic anaerobes is the dehydroxylation of unabsorbed bile acids. It is well known that hydroxylated bile acids, especially the dihydroxy bile acids, chenodeoxycholic acid and deoxycholic acid, exert a cathartic effect on colonic mucosa. Thus, defective bile acid dehydroxylation, due to suppression of bacterial populations after antibiotic therapy, might represent a mechanism that contributes to AAD.

CLOSTRIDIUM DIFFICILE INFECTION

C. difficile is a gram-positive anaerobic bacterium and the most commonly identified causative agent of AAD. *C. difficile* infection represents one of the commonest nosocomial infections around the world and is responsible for approximately three million cases of diarrhea and colitis per year in the United States. The bacterium is primarily acquired in hospitals and chronic care facilities following broad-spectrum antibiotic therapy. It has been estimated that 20% of patients admitted to a general hospital either are colonized on admission or acquire the microorganism during their hospital stay, but only one-third of them develop AAD. Although it is not clear why only a fraction of the colonized patients develop AAD, risk factors associated with disease expression include advanced age, severity of the patient's clinical condition, and most likely defective immunity against *C. difficile* toxins.

Contamination of the hospital environment appears to be the main reservoir of *C. difficile* and is considered to be responsible for perpetuating human infections. This notion is reinforced by recent prospective epidemiological studies reporting the same *C. difficile* strains in both symptomatic patients and ward environments, after molecular characterization of the isolates by DNA fingerprinting. Moreover, molecular characterization revealed that some *C. difficile* strains have a greater propensity for nosocomial transmission (epidemic clones).

C. difficile exerts its effects in the human intestine by the release of two large toxins, toxins A and B. *C. difficile* toxins bind to and damage colonocytes by disrupting the cellular cytoskeleton and increase intestinal permeability by altering tight junctional integrity. The toxins also lead to a severe inflammatory reaction, which, in concert with colonocyte damage, results in micro-ulcerations and the volcano lesion seen in pseudomembranous colitis patients. These micro-ulcerations are usually covered by inflammatory pseudomembranes, containing cellular debris, leukocytes, fibrin, and mucin. A large number of animal studies also indicate that intestinal nerves, by release of neuropeptides, are important mediators of both diarrhea and inflammation in response to *C. difficile* toxins *in vivo*.

Several studies have shown that the majority of healthy adults and children have serum and mucosal antibodies directed against *C. difficile* toxins that may play a protective role in *C. difficile* disease expression. Animal studies showed that immunization against toxin A protects against *C. difficile* infection. It is therefore believed that humoral immunity is important in the development of *C. difficile* infection. A recent prospective study examined whether the presence of serum antitoxin antibodies is associated with the development of *C. difficile* infection in hospitalized patients. The results indicated that antibodies against *C. difficile* toxins or nontoxin antigens did not affect the development of colonization. In contrast, patients who were

colonized and had low serum immunoglobulin G (IgG) antibodies to toxin A were more likely to develop diarrhea. Another prospective study showed that low serum IgG antitoxin A antibodies after an initial episode of diarrhea predispose to recurrent *C. difficile* infection. These studies suggest that antibodies play a significant role in disease development and that vaccination or passive immunotherapy may be an effective strategy to control *C. difficile*-associated diarrhea and colitis.

DIAGNOSIS OF ANTIBIOTIC-ASSOCIATED DIARRHEA

The challenge for a clinician facing a patient with suspected ADD is to document whether this particular case is caused by *C. difficile* infection, a potentially serious but also manageable condition. Laboratory diagnosis of *C. difficile* is based on the demonstration of *C. difficile* toxins in stool. The commonest methods used for diagnosis are summarized in Table I. The most accurate diagnostic assay is the tissue culture assay for the detection of *C. difficile* cytotoxin (toxin B) in stool samples. In brief, phosphate-buffered saline stool suspensions are applied to monolayers of human fibroblasts after centrifugation and filtration. A positive test is indicated by cell rounding, which can be easily seen by light microscopy after 24 or 48 h. Results are reported as positive or negative since there is no clinical correlation between levels of *C. difficile* toxins and severity of disease. Moreover, a recent study suggests that the yield of toxin detection increases by 15% if *C. difficile* is isolated and cultured from stool samples before the cytotoxicity assay ("second-look" cytotoxicity assay). Although tissue culture assays are highly sensitive and specific tests, they are time-consuming and therefore rapid immunoassays have been developed with the goal of replacing them.

The latex agglutination test is not sufficiently sensitive or specific to justify routine use. On the other hand, enzyme-linked immunosorbent assays (ELISAs) that detect toxins A and B in stool samples represent quick and reliable methods. They have high specificity and although they are not as sensitive as the tissue culture assay, their sensitivity improves if more than one stool sample is tested. ELISA is now the method most frequently used by clinical laboratories for diagnosis of *C. difficile* infection, because it is quick, is easy to perform, is acceptably accurate, and does not require highly trained personnel or tissue culture facilities. Specific tests to target the other causes of AAD reported above are not readily available in clinical laboratories. However, in suspicious cases, a quantitative test for *Candida* in stool should be performed.

THERAPY OF ANTIBIOTIC-ASSOCIATED DIARRHEA

In cases of non-*C. difficile* AAD, discontinuation of the inciting antibiotics with or without administration of anti-peristaltic agents is usually sufficient. If continuation of antibiotic therapy is necessary, switching to antibiotics thought to have less risk of causing AAD, such as fluoroquinolones, macrolides, and tetracycline, together with the administration of a probiotic agent, such as *Saccharomyces boulardii* or *Lactobacillus* species, is a therapeutic option. Probiotics have been shown to reduce the risk of AAD, although their mechanism of action is not completely understood. If *Candida* is documented as the causative agent for the AAD, nystatin (250,000 or 1,000,000 U orally three to four times daily) should be given for 7 days.

In approximately 25% of patients with *C. difficile*-associated AAD, symptoms resolve after antibiotic discontinuation, corresponding to reconstitution of

TABLE I Diagnostic Tests for *Clostridium difficile*

Type of assay	Advantages	Disadvantages
Anaerobic culture	Identify carriers in *C. difficile* outbreaks	3–5 days; expensive; also detects nontoxigenic strains
Cytotoxicity assay for toxin B	Excellent sensitivity and specificity	1–3 days; expensive; requires tissue culture facilities
"Second-look" cytotoxicity assay	Excellent sensitivity and specificity	At least 3 days; expensive; requires tissue culture facilities
Enzyme immunoassays for toxins A and B	Very good sensitivity and specificity; takes 4 h; does not require special facilities; less expensive	Less sensitive than cytotoxicity assays

Note. Adapted from Pothoulakis C. *et al*. (1993). *Clostridium difficile* colitis and diarrhea. *J. Gastroenterol. Hepatol.* **8**, 311–312.

normal bowel flora. Therefore, specific therapy should be given when conservative therapy fails, when the administration of the offending antibiotic must continue, or when symptoms are severe.

No therapy is required for asymptomatic carriers. Per oral metronidazole (250–500 mg four times daily) and vancomycin (125 mg four times daily), for 10–14 days, are equally effective in *C. difficile* AAD, with response rates ranging between 95 and 100%. Metronidazole is usually the first-line treatment due to its lower cost. Moreover, metronidazole, in contrast to vancomycin, is effective when there is a need for intravenous administration, as in patients with ileus and *C. difficile* infection. The use of oral vancomycin recently has been restricted because of the growing rates of resistance among gut enterococci; metronidazole has side effects including nausea, abdominal pain, and an unpleasant metallic taste.

Despite the excellent initial response to specific treatment, 20–25% of patients will relapse, usually 3 to 21 days after discontinuation of metronidazole or vancomycin. This is a serious and difficult to manage complication of *C. difficile* AAD. Prolonged treatment with the same antibiotic is usually what is attempted initially. Unfortunately, a substantial fraction of patients relapse again. In cases of multiple relapses, several therapeutic schemes have been developed and exhibit varying degrees of success, underscoring the difficulty of treating this condition. Many experts suggest prolonged administration of vancomycin followed by gradual tapering; others advocate the coadministration of probiotics, such as *S. boulardii*, in an attempt to help in the reconstitution of normal colonic microflora. Intravenous immunoglobulin has also been given, based on putative defective antitoxin immunity in these subjects, with promising results. However, this approach is expensive and requires hospitalization. In severe cases that are resistant to aggressive medical treatment or when bowel perforation is suspected, subtotal or total colectomy with temporary ileostomy may be required.

As indicated above, therapy of AAD is simple and effective in the majority of patients. However, some cases may prove very difficult to treat and represent therapeutic dilemmas. It is anticipated that future research will provide more rational therapeutic alternatives, by means of preventive or therapeutic immunization against *C. difficile* infection.

See Also the Following Articles

Anti-Diarrheal Drugs • Bacterial Toxins • Colitis, Radiation, Chemical, and Drug-Induced • Diarrhea • Diarrhea, Infectious

Further Reading

Bartlett, J. G. (2002). Clinical Practice. Antibiotic-associated diarrhea. *N. Engl. J. Med.* **346**, 334–339.

Bouza, E., Pelaez, T., Alonso, R., Catalan, P., Munoz, P., and Creixems, M. R. (2001). "Second-look" cytotoxicity: An evaluation of culture plus cytotoxin assay of *Clostridium difficile* isolates in the laboratory diagnosis of CDAD. *J. Hosp. Infect.* **48**, 233–237.

Elmer, G. W., and McFarland, L. V. (2001). Biotherapeutic agents in the treatment of infectious diarrhea. *Gastroenterol. Clin. North Am.* **30**, 837–854.

Fawley, W. N., and Wilcox, M. H. (2001). Molecular epidemiology of endemic *Clostridium difficile* infection. *Epidemiol. Infect.* **126**, 343–350.

Gerding, D. N. (2000). Treatment of *Clostridium difficile*-associated diarrhea and colitis. *Curr. Top. Microbiol. Immunol.* **250**, 127–139.

Hogenauer, C., Hammer, H. F., Krejs, G. J., and Reisinger, E. C. (1998). Mechanisms and management of antibiotic-associated diarrhea. *Clin. Infect. Dis.* **27**, 702–710.

Kelly, C. P., and LaMont, J. T. (2002). Treatment of recurrent *Clostridium difficile* infection. *In* "UpToDate" (B. D. Rose, ed.). UpToDate, Wellesley, MA.

Kyne, L., Warny, M., Qamar, A., and Kelly, C. P. (2000). Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. *N. Engl. J. Med.* **342**, 390–397.

LaMont, J. T. (2002). Pathophysiology and epidemiology of *Clostridium difficile* infection. *In* "UpToDate" (B. D. Rose, ed.). UpToDate, Wellesley, MA.

Pothoulakis, C., and LaMont, J. T. (2001). Microbes and microbial toxins: paradigms for microbial–mucosal interactions. II. The integrated response of the intestine to *Clostridium difficile* toxins (2001). *Am. J. Physiol. (Gastrointest. Liver Physiol.)*, **280**, G178–G183.

Apoproteins

CHUN MIN LO*, JIAN-YING WANG†, AND PATRICK TSO*
*University of Cincinnati Medical Center, Ohio and †Veterans Administration Medical Center, Baltimore

Caco-2 cells Human colon adenocarcinoma cell line used to study intestinal epithelial cell functions.

chylomicrons Triglyceride-rich lipoprotein particles produced by the small intestine after a meal; larger in size than very-low-density lipoproteins, chylomicrons transport absorbed cholesterol and triglycerides to different parts of the body.

high-density lipoproteins Class of serum lipoproteins; like other lipoproteins, their core consists of neutral lipid surrounded by an envelope of polar lipid and specific proteins called apoproteins. High-density lipoproteins protect against atherosclerosis through a process known as reverse cholesterol transport, the pathway responsible for transporting excess cholesterol from the peripheral tissues back to the liver for excretion.

lecithin cholesterol acyltransferase Enzyme involved in the esterification of free cholesterol present in circulating plasma lipoproteins; a major determinant of the circulating level of high-density lipoproteins (for instance, overexpression of this enzyme in animals significantly increases the circulating plasma high-density lipoprotein levels).

peroxisome proliferator-activated receptors Regulators of differentiation and homeostasis; the γ form has been found to function as a regulator of adipogenesis and lipid homeostasis.

Pluronic L-81 Member of a family of nonionic detergent surfactants that contains block copolymers of propylene oxide and ethylene oxide. Duodenal infusion of Pluronic L-81 inhibits the formation of chylomicrons by the small intestine.

very-low-density lipoproteins Triglyceride-rich lipoproteins produced by the gut and liver. The gut produces apolipoprotein B48 and the liver produces apolipoprotein B100, both of which contain very-low-density lipoproteins.

The small intestine synthesizes several of the apoproteins associated with chylomicrons, very-low-density lipoproteins (VLDLs), and high-density lipoproteins (HDLs), which are all secreted by the small intestinal epithelial cells (enterocytes). Chylomicrons are secreted only during active lipid absorption, whereas VLDLs and HDLs are secreted all of the time. The major apoproteins synthesized by enterocytes are apolipoprotein (apo) AI, apo AIV, apo B48, and apo CIII. Regulation of the production of these apoproteins by the gut is specific to each apoprotein. For instance, it has been demonstrated that the synthesis and secretion of apo AIV are markedly stimulated during lipid absorption, whereas the synthesis of apo AI and apo B48 is stimulated only marginally. Less information is known about how apo CIII responds to lipid absorption.

APOLIPOPROTEIN AI

Apolipoprotein AI is abundantly expressed in the small intestine, and as much as 2% of total mRNA in intestinal enterocytes is ascribed to this protein. It has a molecular mass of about 28 kDa and plays an extremely important role in lipid metabolism because it is a major protein associated with circulating HDLs and is a cofactor for the circulating enzyme lecithin cholesterol acyltransferase. Both the liver and small intestine are major sources of circulating apo AI, and animal studies have indicated that as much as half of the circulating apo AI comes from the gut. Numerous animal and human studies have demonstrated that an inverse relationship exists between circulating HDL levels and the risk of coronary arterial disease, which is thought to result from the involvement of HDLs in "reverse cholesterol transport." High-density lipoproteins desorb free cholesterol from the membranes of cells and return it to the liver for storage or secretion.

Regulation of Intestinal Apo AI Synthesis and Secretion

The synthesis of apo AI by the small intestine has been studied quite thoroughly, because it is an integral component of HDL and because of the role it plays in reverse cholesterol transport. The absorption of fat seems to affect the synthesis and secretion of apo AI by the small intestine only marginally. For instance, it has been demonstrated in both rats and suckling pigs that active lipid absorption results in only a small increase (between 10 and 20%) in apo AI production by the small intestine, as measured by [^{3}H]leucine incorporation studies. In Caco-2 cells, it has been demonstrated that oxidized fatty acid, a ligand for

 Copyright 2004, Elsevier (USA). All rights reserved.

peroxisome proliferator-activated receptor-γ (PPAR-γ), stimulates both the mRNA level and the synthesis of apo AI. Incubation of porcine enterocytes with fatty acids stimulates the production of apo AI. Unsaturated fatty acids are more potent than saturated fatty acids in stimulating cellular apo AI production, and polyunsaturated fatty acids are more efficient in enhancing the cellular production of apo AI compared to monounsaturated fatty acids. Regulation of apo AI production following dietary consumption of different fatty acids by animals, however, is far more complex. For example, it has been shown that intestinal apo AI synthesis is twofold higher in preweaning piglets than in adult pigs, mainly as a result of the effect of saturated fats. Following weaning, however, polyunsaturated fats appear to play a more important role in maintaining apo AI production by the gut. In many of these studies, however, the fact that polyunsaturated fatty acids are taken up by enterocytes more efficiently than are saturated fatty acids may have been ignored. Therefore, distinguishing between the effects of polyunsaturated and saturated fatty acids and determining the differing amounts of fatty acids that are taken up by the cells following ingestion of either polyunsaturated or saturated fats may be difficult.

It has been reported that intestinal apo AI production is altered during different physiological and disease states. For instance, the production of apo AI appears to increase during aging; studies in rodents have demonstrated that apo AI synthesis by the small intestine is increased in older animals compared to younger animals. Increasing the circulating level of ethinylestradiol appears to down-regulate gene expression of apo AI in the gut whereas ovariectomy increases the apo AI mRNA level. As would be expected, supplementing ovariectomized animals with ethinylestradiol decreases the apo AI mRNA level in the intestine. Thyroidal status also impacts the synthesis of apo AI by the gut. Hyperthyroidism is associated with increased apo AI synthesis and secretion by the gut, but this increase is not associated with increased triglyceride secretion. How gastrointestinal (GI) hormones affect apo AI gene expression is an exciting area of investigation yet to be explored.

Clinical implications of intestinal apo AI production have been demonstrated in humans as well as in animals. Patients suffering from abetalipoproteinemia have reduced apo AI synthesis. It has been reported that apo AI production is stimulated by copper deficiency and that tumor necrosis factor α (TNFα) decreases apo AI production by the small intestine. Both mRNA levels and synthesis of apo AI are stimulated in nephritic animals. Further, the intestinal mucosa of patients with celiac disease show decreased apo AI synthesis, which is reflected by a marked reduction in circulating apo AI (∼40%).

APOLIPOPROTEIN AIV

Apolipoprotein AIV was discovered in the 1970s, but its physiological role was not firmly established until recently. Apo AIV is a protein secreted by the human small intestine; in rodents, apo AIV is secreted by both the small intestine and liver; although the small intestine is the organ most responsible for circulating levels of apo AIV. It has been recently demonstrated in rodents that the brain, specifically the arcuate nucleus, produces apo AIV as well. The production of apo AIV is regulated physiologically, i.e., it decreases with fasting and increases with feeding.

Regulation of Intestinal Apo AIV Synthesis and Secretion

Circulating apo AIV in rodents exhibits a circadian rhythm—the level increases at the beginning of the dark period and peaks midway through the dark cycle. Fasting significantly reduces the circulating level of apo AIV but it does not affect the pattern of the circulating apo AIV circadian rhythm. This circadian rhythm is abolished in animals when enterohepatic circulation is interrupted by bile diversion. Furthermore, it has been demonstrated that bile diversion significantly reduces apo AIV synthesis by the intestinal mucosa. Thus, intact enterohepatic circulation is necessary both for normal basal lymphatic output of apo AIV and for maintaining its circadian rhythm.

In vivo studies have demonstrated that the synthesis and secretion of apo AIV by the GI tract are regulated by fat absorption as well as by peptides secreted by the lower small intestine, e.g., peptide YY. It has been demonstrated by numerous groups that intestinal lipid absorption stimulates the synthesis and secretion of apo AIV in a dose-dependent manner. Apo AI and apo B48, however, are not stimulated in this way. Evidence to date suggests that increased synthesis of apo AIV following fat absorption results via a transcriptional mechanism, although the precise cellular mechanism responsible has not yet been determined. Increased apo AIV production probably occurs locally and does not involve the nervous system, because complete vagotomy fails to inhibit an increase. It has been demonstrated that intestinal lymphatic apo AIV transport increases in a gradient manner with increasing steady-state levels of intestinal triglyceride transport. It is tempting to think that measuring the circulating level of apo AIV may be an effective way of determining

fat ingestion. If this correlation can be demonstrated, measuring the circulating apo AIV level may be a convenient and effective way of following fat absorption by the gut. It may also be a convenient way of monitoring the intestinal absorptive function; formation and secretion of chylomicrons by the small intestine are among the most complex functions performed by the enterocytes, involving lipid synthesis, apoprotein synthesis, and carbohydrate synthesis (glycosylation of apoproteins).

Understanding the mechanisms responsible for stimulating intestinal apo AIV synthesis and secretion during fat absorption is crucial. Evidence gathered thus far suggests that the assembly and transport of chylomicrons are necessary for stimulating intestinal apo AIV synthesis. A unique inhibitor of chylomicron formation, Pluronic L-81 (L-81), a hydrophobic surfactant that specifically inhibits the formation of chylomicrons, has been used to explore the relationship between chylomicron formation and the stimulation of apo AIV synthesis and secretion by the gut. L-81 does not inhibit the digestion, uptake, or reesterification of absorbed lipid to form triglyceride inside the enterocytes. Because of the inhibition of chylomicron formation by enterocytes, the absorbed lipid accumulates in the intestinal mucosa. When L-81 is removed, lymphatic lipid transport as chylomicrons resumes because the accumulated mucosal lipid is cleared from the mucosa. These results demonstrate that L-81 inhibits lymphatic transport of chylomicrons and abolishes the increase in lymphatic secretion of apo AIV normally associated with active fat absorption. As expected, removing L-81 causes chylomicron formation as well as the stimulation of lymphatic apo AIV secretion, suggesting that the formation of chylomicrons stimulates apo AIV synthesis and secretion.

Further evidence of the dependency of lymphatic apo AIV output on chylomicron transport comes from animal studies examining intestinal synthesis and lymphatic secretion of apo AIV in response to intestinal infusion of fatty acids of different chain lengths (and therefore the route of transport from the intestine, i.e., lymph vs. blood). Infusion of long-chain fatty acids (oleic, C_{18}; arachidonic, C_{20}), which are transported via the lymph in chylomicrons, stimulates synthesis and output of apo AIV. However, medium- and short-chain fatty acids (caprylic, C_8; butyric, C_4), primarily transported as free fatty acids in the portal vein, did not elicit an apo AIV response. This finding in rodents differs from findings in neonatal swine, in which similar increases in jejunal apo AIV mRNA expression and synthesis in response to infusions of both medium- ($C_{8:0}$ and $C_{10:0}$) and long-chain triglyceride mixtures are observed. Additional studies are warranted to determine if the relationship between chylomicrons and apo AIV synthesis and secretion is common to all species and developmental stages. An extremely interesting but still unanswered question is how the formation of chylomicrons is signal transduced to stimulate intestinal apo AIV synthesis and whether this event occurs intracellularly.

It has been demonstrated that lipid infused into the ileum of rats stimulates both ileal and jejunal apo AIV synthesis. This finding is different from that of lipid infused into the duodenum, because only duodenal and jejunal synthesis of apo AIV is stimulated. Animals equipped with jejunal or ileal Thiry—Vella fistulas (segment of the intestine isolated luminally from the rest of the GI tract, but still connected to the body via the circulation and nervous system) increase proximal jejunal apo AIV synthesis when lipid is infused into the ileum. This stimulation of jejunal apo AIV synthesis by ileal lipid absorption is independent of the presence of lipid in the jejunum. These results strongly suggest that a signal released by the ileum in response to the presence of lipids stimulates apo AIV synthesis in the proximal gut. Recent evidence suggests that peptide YY (PYY) is potentially a primary factor contributing to the stimulation of jejunal apo AIV synthesis and secretion by the presence of lipids in the ileum. Continuous intravenous infusion of physiological doses of PYY in fasting animals elicits significant increases in both synthesis and lymphatic transport of apo AIV; this process is mediated through the vagus nerve, because vagotomy abolishes a jejunal increase in apo AIV synthesis. Thus, jejunal stimulation of apo AIV synthesis and secretion by fat absorption in the ileum is mediated by PYY, which, in turn, acts centrally to send a signal via the vagus nerve to the gut. This is the first demonstration of GI hormone involvement in controlling the expression and secretion of an intestinal apolipoprotein, thus bringing together two areas of research in GI physiology.

Physiological Functions of Apo AIV

In vitro studies have suggested the role of apo AIV in lipoprotein metabolism. Whether apo AIV plays the same role *in vivo* is uncertain, but the fact that apo AI can also perform such a role casts doubt on this physiological function of apo AIV. Here we discuss only *in vivo* functions of apo AIV that are not shared by apo AI. Apo AIV has been shown to protect apo E knockout mice from developing atherosclerosis despite their atherogenic plasma lipid profile (increased total plasma cholesterol with no significant change in HDL cholesterol).

Moreover, apo AIV deters the formation of diet-induced atherosclerotic lesions. These studies therefore suggest that apo AIV protects against atherosclerosis, possibly as a result of the ability of apo AIV to protect against lipid oxidation. Presently, the extent of this antioxidant effect is unclear.

Apo AIV also plays a unique role in the regulation of food intake. It is probably a satiety signal released by the GI tract following the ingestion of fat. Animals administered intravenous intestinal lymph from fasted rats did not respond normally to food intake. However, animals administered intravenous intestinal lymph from rats actively absorbing lipid demonstrated a marked suppression in food intake. These data implicate the presence of an active component in chylous lymph responsible for inhibiting food intake, and this active component is apo AIV. This is supported by a study demonstrating that the infusion of 200 μg of apo AIV (a physiological dose) inhibits food intake to the same degree as chylous lymph in 24-hour food-deprived rats. This unique function of apo AIV is not shared by apo AI. Thus, it has been proposed that apo AIV is a circulating signal released by the small intestine in response to fat feeding and likely mediates the anorectic effect associated with the ingestion of lipids. That the inhibiting effect of apo AIV on food intake is centrally mediated is based on a number of observations. First, administration of apo AIV in the brain results in an inhibitory effect on food intake. Second, removing apo AIV in the cerebrospinal fluid via apo AIV antibody results in feeding.

As well as inhibiting food intake, apo AIV has also been shown to inhibit gastric acid secretion and gastric motility. Administered doses of apo AIV, thought to reproduce the levels of apo AIV in cerebrospinal fluid after lipid feeding, markedly inhibit both gastric acid secretion and gastric motility. This suggests that apo AIV acts as an enterogastrone, a humoral mediator released by the intestine that mediates the humoral inhibition of gastric acid secretion and gastric motility following the ingestion of fat. Recent data from investigators interested in vagal neural activities have provided convincing evidence showing that apo AIV also inhibits intestinal motility. These findings have important physiological implications. The distal intestine is known to play an important role in the control of GI function. Nutrients, including lipid, delivered to the ileum result in inhibited gastric emptying, decreased intestinal motility and transit, and decreased pancreatic secretion. Ileal nutrients also inhibit food intake. The mechanism responsible for producing these effects has been collectively termed the "ileal brake" and is believed to be related to the release of one or more peptide hormones from the distal intestine. These effects have traditionally been considered operative only in the abnormal delivery of undigested nutrients to the distal gut, such as the malabsorptive state. Most GI physiologists as well as gastroenterologists interested in lipid absorption consider the upper small intestine the primary site for absorbing fat, with the lower small intestine absorbing fat only in the case of fat malabsorption by the upper small intestine. This notion has been recently challenged by investigators demonstrating that nutrients frequently reach the distal gut, even under normal conditions, due to rapid gastric emptying during the early phases of food ingestion. Consequently, it appears that a much greater length of intestine is involved in the absorption of lipid and in the control of gastric and upper gut functions, even under normal conditions, than has been previously recognized. Thus, the ileal brake may play an important role in the normal control of gut function and the control of lipid absorption. Once the ileal brake sets in, the upper small intestine unquestionably is the primary site for fat absorption.

As mentioned earlier, a potential peptide mediating the ileal brake phenomenon is peptide tyrosine-tyrosine (PYY). It is synthesized by the endocrine cells in the ileum and large intestine and is released in response to intestinal nutrients, especially long-chain fatty acids. PYY is also a potent stimulator of the synthesis and secretion of apo AIV by the jejunum. Because apo AIV exerts many of the actions associated with the ileal brake, and PYY stimulates the synthesis and secretion of apo AIV and PYY, apo AIV and the ileal brake phenomenon may possibly be related. Recently, PYY has been demonstrated to inhibit food intake, with this action centrally mediated. Interestingly, stimulation of apo AIV synthesis by PYY also involves the central nervous system, and this stimulation can be abolished by total vagotomy. The question of whether the action of PYY on food intake and upper GI function is mediated through apo AIV can be tested in apo AIV knockout mice.

APOLIPOPROTEIN B

Two different forms of apo B are formed in humans—enterocytes produce apo B48 and hepatocytes produce apo B100. Apo B48 is a 264-kDa protein that is colinear with the apo B100 amino terminus and is 48% of the apo B100 molecule, hence the name apo B48. Both apo B48 and apo B100 utilize the same apo B gene. In the production of apo B48, however, translation is stopped by substituting the CAA codon that encodes glutamine with the codon UAA, which specifies an in-frame stop codon. This interesting mechanism is referred to

as apo B mRNA editing and has since been found in other proteins as well. Apo B editing has a very important consequence. Apo B48 lacks the functional domain for the binding of low-density lipoproteins. The lack of LDL binding domain has led investigators to postulate that a different set of receptors, other than LDL receptors, is involved in the metabolism of chylomicrons. In contrast, apo B100 containing VLDLs produced by the liver, are catalyzed to form LDLs, which are subsequently taken up by the liver and other peripheral organs via the LDL receptors.

Synthesis of Apo B48 by the Small Intestine

Adult human small intestinal cells produce mostly apo B48, suggesting that the editing efficiency is very high in the adult intestine. Evidence to date indicates that neither apo B editing in the human small intestine nor apo B synthesis by the rat small intestine is regulated by fat absorption. Even chronic feeding of a high-fat diet fails to stimulate intestinal apo B synthesis in rodents. Not only are apo B editing and apo B synthesis unaffected by fat absorption, but the secretion of apo B into lymph as chylomicrons is also unaffected by the absorption of fat. This is very different from apo AIV, which is markedly stimulated by fat absorption. This is also very different from the synthesis of apo B100 in the liver, the secretion of which by the hepatocytes is markedly increased by the increase in VLDL secretion. Whether this is related to the fact that the liver makes apo B100 and the intestine makes apo B48 has not been studied.

Bile is required for the secretion of apo B48 by the small intestine, because the diversion of bile results in a marked reduction in the synthesis and secretion of apo B48 by the small intestine. The reintroduction of bile salts, fatty acids, or phospholipids into the intestinal lumen restores apo B biosynthesis by the enterocytes. Using current molecular and cellular biology techniques, the precise involvement of these various components of bile in regulating apo B synthesis will become clear. Perhaps the regulation of intestinal apo B synthesis by bile components is species specific, because bile diversion in the suckling pig has little or no effect on intestinal apo B synthesis.

Physiological Functions of Apo B

The clinical disorder abetalipoproteinemia, an autosomal recessive disorder, provided the first clue that apo B is important for the secretion of triglyceride-rich lipoproteins by the gut as well as by the liver. Patients suffering from abetalipoproteinemia experience fat malabsorption and the accumulation of large lipid droplets in both enterocytes and hepatocytes. Other symptoms associated with this disease are acanthocytosis and retinitis pigmentosa, mostly resulting from fat-soluble vitamin deficiency. However, some studies have questioned whether the mechanism for abetalipoproteinemia is caused by failure to produce apo B. For instance, immunohistochemistry has shown that enterocytes of abetalipoproteinemic patients contain apo B. Furthermore, it has been determined that abetalipoproteinemic patients have elevated levels of apo B mRNA. Thus, it is either the elevated levels of posttranslational apo B degradation or some other factor(s) critical to the association of apo B and absorbed lipids to form intestinal chylomicrons and liver VLDLs that are absent. The "other factor(s)" may be the microsomal triglyceride transfer protein (MTP), because convincing evidence points to either the absence or the mutation of MTPs in abetalipoproteinemic patients. Modern tools of molecular biology should provide a better understanding of the relationship between apo B, MTP, and chylomicron formation.

Because apo B is critical to chylomicron formation, it might be assumed that the availability of apo B is rate limiting for the formation of chylomicrons and, consequently, for the absorption of fat by the small intestine. However, this is not the case. A number of studies have found that apo B48 synthesis by the small intestine and its subsequent secretion into lymph are unaffected by fat absorption. This is different from the process in the liver, where increased VLDL secretion is associated with increased apo B100 secretion. Whether the relationship between lipid transport and apo B secretion is different in the small intestine than it is in the liver is because the gut makes apo B48 and the liver makes apo B100 is yet to be determined. The availability of the apoBEC-1 knockout mouse, whose small intestine produces only apo B100 due to its inability to perform apo B editing, will open this field of study.

APOLIPOPROTEIN CIII

Apolipoprotein CIII is another apoprotein synthesized by the small intestine. It is a much smaller protein compared to apo AI, apo AIV, and apo B48. The mature apo CIII is a 79-amino-acid protein with a molecular mass of 9 kDa. Apo CIII is secreted by the small intestine and the liver. It is probably secreted in association with chylomicrons, but it may also be secreted as a free protein by the enterocytes of the small intestine. In the circulation, it is mostly associated with VLDLs and HDLs.

Synthesis by Enterocytes

Relative to the other apolipoproteins, little is known about the regulation of apo CIII. Whether fasting and

lipid feeding affect the intestinal gene expression of apo CIII is unclear, although results of an older, isolated study show that fat absorption is inversely related to the subjects' fasting plasma HDL cholesterol and HDL apo CIII and directly related to plasma triglycerides. Vitamin A has been reported to regulate apo CIII positively in the small intestine. Apo CIII gene expression is down-regulated by insulin both in animals and in cultured cells.

Physiological Function of Apo CIII

The role of apo CIII in lipoprotein metabolism is best illustrated by the occurrence of hypertriglyceridemia, which is associated with increased expression of apo CIII, as in the case of transgenic animals. Two plausible explanations for the occurrence of hypertriglyceridemia might be that (1) apo CIII inhibits lipoprotein lipase activity or (2) apo CIII decreases tissue uptake of triglyceride-rich particles from the circulation as a result of increased apo CIII and decreased apo E in VLDL particles.

See Also the Following Articles

Barrier Function in Lipid Absorption • Cholesterol Absorption • Lipoproteins • Pancreatic Polypeptide Family • Protein Digestion and Absorption of Amino Acids and Peptides • Small Intestine, Absorption and Secretion

Further Reading

Davidson, N. O. (1994). Cellular and molecular mechanisms of small intestinal lipid transport. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1909–1934. Raven Press, New York.

Levy, E. (1996). The genetic basis of primary disorders of intestinal fat transport. *Clin. Invest. Med.* **19**, 317–324.

Plump, A. (1997). Atherosclerosis and the mouse: A decade of experience. *Ann. Med.* **29**, 193–198.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1867–1908. Raven Press, New York.

Tso, P., Liu, M., and Kalogeris, T. J. (2001). The role of apolipoprotein A-IV in the regulation of food intake. *Annu. Rev. Nutr.* **21**, 231–254.

Appendicitis

CHRISTINE HSU AND STEPHEN JOHN FERZOCO
Brigham and Women's Hospital, Boston

appendicalith A fecalith located in the appendix.
fecalith A small hard mass of feces.
incidental appendectomy Removal of the appendix when the laparotomy or laparoscopy is performed for another clinical reason.
McBurney's point The point in the right lower quadrant of maximal tenderness (overlying the appendix), described by McBurney as being between 1.5 and 2 inches. from the anterior iliac spine along the oblique line to the umbilicus.
negative appendectomy Removal of a grossly and histologically normal appendix when appendicitis is expected.
peritoneum The serous lining of the abdominal cavity; the parietal peritoneum lines the abdominal wall and the visceral peritoneum covers the organs.
vermiform Wormlike, a term often used to describe the appendix.

Appendicitis, inflammation of the appendix, is one of the most common surgical diseases affecting young people. Although it can affect infants and the elderly, it generally becomes manifest in young, otherwise healthy individuals. It can pose a diagnostic dilemma since many other abdominal processes can mimic the findings. With prompt diagnosis and treatment, the morbidity and mortality of appendicitis have been greatly reduced in the past century.

INTRODUCTION

Appendicitis is mostly a disease of the Western world, with lower dietary fiber presumably predisposing to appendiceal inflammation. Approximately 7% of the

Copyright 2004, Elsevier (USA). All rights reserved.

lipid feeding affect the intestinal gene expression of apo CIII is unclear, although results of an older, isolated study show that fat absorption is inversely related to the subjects' fasting plasma HDL cholesterol and HDL apo CIII and directly related to plasma triglycerides. Vitamin A has been reported to regulate apo CIII positively in the small intestine. Apo CIII gene expression is down-regulated by insulin both in animals and in cultured cells.

Physiological Function of Apo CIII

The role of apo CIII in lipoprotein metabolism is best illustrated by the occurrence of hypertriglyceridemia, which is associated with increased expression of apo CIII, as in the case of transgenic animals. Two plausible explanations for the occurrence of hypertriglyceridemia might be that (1) apo CIII inhibits lipoprotein lipase activity or (2) apo CIII decreases tissue uptake of triglyceride-rich particles from the circulation as a result of increased apo CIII and decreased apo E in VLDL particles.

See Also the Following Articles

Barrier Function in Lipid Absorption • Cholesterol Absorption • Lipoproteins • Pancreatic Polypeptide Family • Protein Digestion and Absorption of Amino Acids and Peptides • Small Intestine, Absorption and Secretion

Further Reading

Davidson, N. O. (1994). Cellular and molecular mechanisms of small intestinal lipid transport. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1909–1934. Raven Press, New York.

Levy, E. (1996). The genetic basis of primary disorders of intestinal fat transport. *Clin. Invest. Med.* **19**, 317–324.

Plump, A. (1997). Atherosclerosis and the mouse: A decade of experience. *Ann. Med.* **29**, 193–198.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1867–1908. Raven Press, New York.

Tso, P., Liu, M., and Kalogeris, T. J. (2001). The role of apolipoprotein A-IV in the regulation of food intake. *Annu. Rev. Nutr.* **21**, 231–254.

Appendicitis

CHRISTINE HSU AND STEPHEN JOHN FERZOCO
Brigham and Women's Hospital, Boston

appendicalith A fecalith located in the appendix.
fecalith A small hard mass of feces.
incidental appendectomy Removal of the appendix when the laparotomy or laparoscopy is performed for another clinical reason.
McBurney's point The point in the right lower quadrant of maximal tenderness (overlying the appendix), described by McBurney as being between 1.5 and 2 inches. from the anterior iliac spine along the oblique line to the umbilicus.
negative appendectomy Removal of a grossly and histologically normal appendix when appendicitis is expected.
peritoneum The serous lining of the abdominal cavity; the parietal peritoneum lines the abdominal wall and the visceral peritoneum covers the organs.
vermiform Wormlike, a term often used to describe the appendix.

Appendicitis, inflammation of the appendix, is one of the most common surgical diseases affecting young people. Although it can affect infants and the elderly, it generally becomes manifest in young, otherwise healthy individuals. It can pose a diagnostic dilemma since many other abdominal processes can mimic the findings. With prompt diagnosis and treatment, the morbidity and mortality of appendicitis have been greatly reduced in the past century.

INTRODUCTION

Appendicitis is mostly a disease of the Western world, with lower dietary fiber presumably predisposing to appendiceal inflammation. Approximately 7% of the

Copyright 2004, Elsevier (USA). All rights reserved.

population will have appendicitis in their lifetime, with the peak between the ages of 10 and 30. Males are affected more often than females (1.6 : 1) and perforation rates are higher in the elderly. The very young and the elderly also incur a higher rate of rarer complications such as fistula formation, intestinal obstruction, or systemic sepsis. Mortality of appendicitis is low (0.6% in nonperforated), but delay in appropriate treatment has a significant morbidity of 15–20%, especially considering the generally young population it most commonly affects.

HISTORY

In the early 1800s, right lower quadrant inflammation was thought to arise from near the cecum and the term "perityphlitis" was used to describe this entity. Amyand performed the first recorded appendectomy in 1736 when he removed the appendix upon exploration of a hernia sac. Melier described autopsy cases of inflammation of the appendix and outlined a pathophysiology similar to the present understanding. Then in 1886, Reginald Fitz first used the term "appendicitis" and recommended surgical treatment for the inflammatory disease. Several years later, Chester McBurney described the classic pain of appendicitis and the localization of pain to the point that bears his name, on an oblique line from the anterior superior iliac spine to the umbilicus. As the understanding of appendicitis as a surgical disease evolved and antibiotics became readily available by the mid-1900s, the mortality decreased to less than 2%. Early recognition and surgical therapy then greatly reduced the morbidity.

ANATOMY

The appendix is believed to be a vestigial organ, without a functional purpose in humans. However, because of its lymphoid aggregates, there are some who speculate that the appendix may be involved in immune surveillance. Embryologically, the cecal diverticulum (the origin of the cecum and appendix) presents in the sixth gestational week as a swelling on the antimesenteric border of the midgut loop. The appendix, which is initially a small diverticulum of the cecum, arises from the inferior tip of the cecum, becoming delineated during the fifth gestational month. The anatomic position of the appendix is variable depending on embryologic development, contributing to difficulty in diagnosis. After birth, the appendix rests on the medial side of the cecum because the wall of the cecum grows unequally. During elongation of the colon, the appendix may settle posterior to the cecum, anterior or posterior to the ileum, below the cecum, or over the brim of the pelvis. In over half of the population, the appendix is not fixed in one position but can move freely with postural and colonic changes.

The appendix opens to the cecum with a valve formed by a fold of mucosa. Average appendiceal length is 9 cm, and the width of the lumen is generally between 1 and 3 mm. The appendiceal artery, off the ileocolic artery, provides the blood supply to the appendix. The lymphatics drain to the ileocolic nodes along the ileocolic artery. Sympathetic fibers from the superior mesenteric plexus and parasympathetic fibers from the vagus provide innervation. Histologically, from outside to inside, the appendix consists of the serosa; longitudinal muscle fibers, which form taeniae continuous with the colonic taeniae coli; circular muscle; the submucosa, which contains lymphoid tissue; and the mucosal lumen. The taeniae coli of the ascending colon converge at the base of the appendix, enabling localization of the appendix during appendectomy by following the cecal taeniae.

PATHOPHYSIOLOGY

Most commonly, appendicitis results from obstruction of the appendiceal lumen. In older adults, fecaliths are a frequent cause of obstruction, whereas the most common cause in children and young adults is lymphoid hyperplasia. Fecaliths and appendicitis occur more frequently in populations that have a lower fiber and higher fat content in the diet, suggesting that diet plays a role in disease development. In less than 2% of cases, a foreign body, carcinoma, or carcinoid causes obstruction.

The appendix secretes an increased amount of mucus in response to the obstruction. Bacterial overgrowth also develops secondary to obstruction and subsequent stasis. The increased mucus and bacterial load lead to appendiceal dilation, eventually causing compromise of venous and lymphatic flow. The appendiceal artery can thrombose as the appendix enlarges. Ischemic injury can then lead to necrosis, and if the gangrene becomes full-thickness, perforation occurs. With rupture, fecal contents enter the abdominal cavity and produce peritoneal irritation. An abscess results when local inflammation produces adhesions and walls off the perforated appendix. If the rupture is not contained within an abscess, gross spillage and generalized peritonitis result.

Histopathologically, the gross specimen shows edema of serosal vessels early in appendicitis. The appendix then develops a dilated lumen, thickened wall, dusky serosa, and fibrinous serosal exudates. Late in the progression of the disease, there is mucosal necrosis, wall gangrene and softening, and purulent

serosal exudates. Microscopically, appendicitis demonstrates a neutrophilic infiltrate early, which progresses to mucosal necrosis and eventually to muscularis necrosis and microabscesses in the appendiceal wall. If allowed to progress untreated, the appendix ultimately perforates and leads to a walled-off abscess or generalized peritonitis.

DIAGNOSIS

The classic history of appendicitis is one of dull periumbilical pain followed by anorexia and then localized pain developing in the right lower quadrant (RLQ). Nausea and then vomiting generally precede localization of pain to the right lower quadrant, although nausea is present at some time during the clinical course in 90% of patients with appendicitis (Table I). The absence of anorexia makes the diagnosis of appendicitis questionable and the "hamburger sign" (asking the patient if he or she would want to eat a favorite food) is one way to assess this.

The usual order of events is as follows: epigastric or periumbilical pain, anorexia, nausea, vomiting, RLQ tenderness, fever (usually low grade), and leukocytosis. Since the innervation of the appendix migrates to the right lower quadrant from autonomic efferents associated with the spinal cord around T10 and the abdominal organs lack direct innervation by pain fibers, dull pain is initially felt around the umbilicus. As inflammation progresses, the parietal peritoneum in the right lower quadrant becomes more irritated, resulting in localized pain. The pain becomes constant as the lumen of the appendix develops increasing distension and ischemia results. Approximately 40% of patients will present with atypical pain. Perforation is rare if symptoms have been present for less than 24 h.

The differential diagnosis of abdominal pain that can mimic appendicitis is broad (Table II). Conversely, because of its potential for variable presentation, appendicitis should always be considered when evaluating any acute abdominal pain. Tenderness to percussion demonstrates peritoneal irritation, as does guarding and pain with motion. Occasionally, a tender mass,

TABLE I Common Signs and Symptoms of Appendicitis

Symptoms
Abdominal pain, periumbilical
RLQ pain
Anorexia
Nausea/vomiting
Pain migration
Signs
RLQ tenderness
Guarding
Rebound tenderness
Low-grade fever
Elevated white blood cell count
Rovsing's sign (see text for description)
Psoas sign
Obturator sign
Dunphy's sign

TABLE II Differential Diagnosis of Acute Appendicitis

Gastrointestinal
Cholecystitis
Inflamed or leaking duodenal ulcer
Crohn's disease
Cecal cancer
Inflamed Meckel's diverticulum
Intestinal obstruction
Diverticulitis with or without abscess
Gastroenteritis
Typhlitis
Omental torsion
Perforated viscus
Pancreatitis
Intussusception
Musculoskeletal
Psoas abscess
Rectus sheath hematoma
Urologic
Ureteral stone
Urinary tract infection
Pyelonephritis
Nephrolithiasis
Perinephric abscess
Hydronephrosis
Prostatitis
Gynecologic
Tubo-ovarian abscess
Ectopic pregnancy
Endometriosis
Salpingitis
Ruptured ovarian cyst
Ovarian torsion
Ruptured follicular cyst (mittelschmerz)
Pyosalpinx
Pelvic inflammatory disease
Systemic
Diabetic ketoacidosis
Porphyria
Sickle cell disease
Henoch-Schonlein purpura
Tropical areas
Amebic typhlitis
Malaria
Leaking liver abscess
Yersinia infection

which is either the distended, inflamed appendix or an appendiceal abscess, can be palpated in the right lower quadrant. Hyperesthesia of the skin overlying the right lower quadrant is an occasional finding in the distribution of the 10th through 12th dorsal spinal segments and 1st lumbar spinal segment.

Various physical signs are often present in appendicitis. The psoas sign elicits pain on passive extension of the right thigh with the patient lying on the left side since the inflamed appendix overlies the psoas muscle. Pain on passive internal rotation of the flexed thigh is called the obturator sign. Additional findings include Rovsing's sign, which is pain in the right lower quadrant on palpation of the left lower quadrant, and Dunphy's sign, which is pain on coughing. The presence of these signs helps to confirm the diagnosis, but they are by no means sensitive or specific. Tenderness on rectal exam may be helpful in making the diagnosis if there is an abscess or pelvic appendicitis. Bimanual/speculum exam should be performed in women to evaluate for gynecologic causes of pain such as pelvic inflammatory disease.

Laboratory tests often demonstrate an elevated white blood cell count and/or bandemia. The BUN to creatinine ratio may be elevated if the patient is dehydrated from decreased intake and vomiting. Urine analysis is useful in assessing for ureteral stones, which produce hematuria, or urinary tract infections (bacteruria, pyuria). Local inflammatory reaction to appendicitis, however, can also give white blood cells and red blood cells in the urine. A test for human chorionic gonadotropin in the urine should be checked in all women of childbearing age, both to exclude normal or ectopic pregnancy as a cause of pain and to tailor diagnostics and treatment if necessary.

Although the diagnosis of appendicitis is generally made on the basis of history and physical exam, radiologic studies are effective tools in equivocal diagnoses. Plain film rarely contributes to the diagnosis, although findings in appendicitis can include a visible fecalith, localized ileus, loss of the peritoneal fat stripe from the right lower quadrant inflammatory process, and right psoas muscle deformation from the patient splinting away from right lower quadrant pain. Ultrasound and computed tomography (CT) scan are the most effective methods of elucidating a diagnosis of appendicitis. On ultrasound, a positive study shows a distended appendix that measures greater than 6 mm in diameter, is noncompressible, and is tender with probe compression. The transverse images show a "target" or "bullet" sign, which is the side view of the distended, thickened appendix (Fig. 1). A normal appendix must be seen to rule out appendicitis. Depending on the operator, appendiceal ultrasound has a sensitivity of approximately 85% and a specificity of 90%.

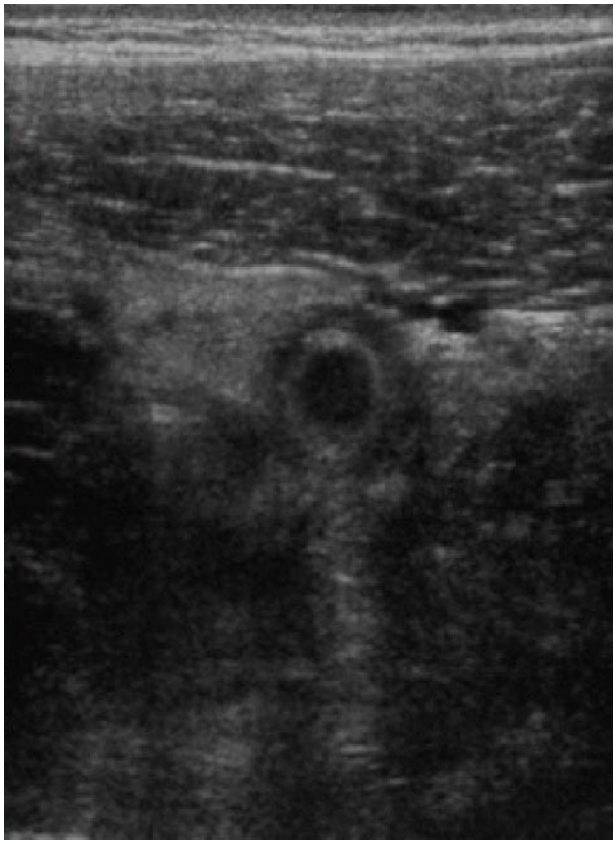


FIGURE 1 Ultrasound image of acute appendicitis. On transverse view, a "target" or "bullet" sign is seen underneath the abdominal wall. Figure courtesy of Stephen Ledbetter, M.D., Department of Radiology, Brigham and Women's Hospital, Boston, MA.

The use of CT images to diagnose appendicitis has greatly affected management of suspected appendicitis. Although the initial studies with CT-diagnosed appendicitis were performed with rectal contrast, Brigham and Women's Hospital favors a standard abdominal/pelvic CT scan with oral and intravenous contrast. If the radiologist is unable to make or exclude the diagnosis with the initial scan and there is a high index of suspicion, a rectal contrast CT scan can then be performed. An advantage of this technique is contrast-aided imaging of the entire abdomen for other possible diagnoses. Positive studies show a thick-walled, distended, non-contrast-filling appendix, periappendiceal inflammation, fluid collections or abscess, and often, an appendicalith (Fig. 2). If contrast fills the appendix and there is no evidence of surrounding inflammation, the diagnosis is considered to be excluded.

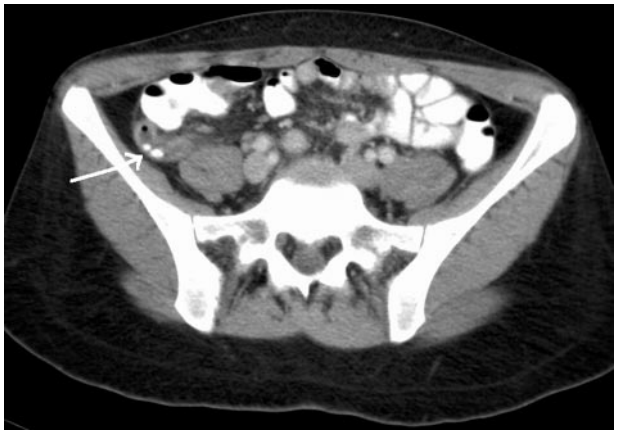


FIGURE 2 CT of appendicitis. A spiral CT with oral and intravenous contrast demonstrates two appendicaliths obstructing a fluid-filled, distended, thick-walled appendix (white arrow) with surrounding fat stranding consistent with local inflammation. Figure courtesy of Stephen Ledbetter, M.D., Department of Radiology, Brigham and Women's Hospital, Boston, MA.

Depending on the radiologist, appendiceal CT scans have a sensitivity of approximately 90–99% and a specificity of 97%. Multiple studies have demonstrated the cost effectiveness of CT when used in equivocal cases of appendicitis. In general, patients with a questionable history and physical who would have been observed closely in the past are now undergoing CT scans and either being discharged or taken to the operating room. As clinicians become more comfortable with using CT scans as a diagnostic tool, the rate of negative exploration is expected to decrease.

TREATMENT

Operative removal of the inflamed appendix is the accepted treatment for appendicitis. In nonperforated cases, broad-spectrum antibiotics are usually administered prior to the start of the operation. No further antibiotics are routinely necessary. However, in cases of perforation, the broad-spectrum antibiotics are generally continued to complete a course of 5 to 7 days.

The appendix can be removed through an open incision or using a laparoscope. Incisions for the open approach include a transverse right lower quadrant (Rockey–Davis), oblique RLQ (McArthur–McBurney) with muscle splitting in the direction of the fibers, or less commonly, a paramedial incision. If the diagnosis is in question, an abdominal exploration can be performed through a periumbilical midline incision.

The appendix is located by following the taeniae coli to the cecum and lifting the appendix out of the incision. After it is dissected off the mesoappendix, the specimen can be removed by stapling across its base or suture ligation. Some surgeons use either a Z-stitch or a purse-string stitch to invert the stump. There is no evidence, however, that this maneuver reduces the incidence of stump leak or fistulae. The wound is closed in nonperforated cases. In perforated appendicitis, a Penrose or other drain is often placed in the wound and/or the wound is left open with packing. Neither practice has been definitively shown to reduce the incidence of postoperative wound infection. The patient then completes a course of broad-spectrum antibiotics.

The laparoscopic approach involves longer operative time but is supposed to result in less pain and a quicker recovery time. However, a significant overall advantage of the laparoscopic procedure compared to the open procedure has not been shown in routine appendicitis. The main benefit of the laparoscopic approach is that it can be helpful initially as a diagnostic tool and then the surgeon can proceed to appendectomy in positive cases. The appendix is divided at its base using a linear stapler and there is no inversion of the stump.

Classically, a 15–20% rate of removal of normal appendix was acceptable given the broad differential diagnosis and high morbidity of missed appendicitis. With the use of ultrasound and CT, rates of negative appendectomy are expected to decrease. Despite this expectation with technological advances, recent studies have not yet demonstrated a reduction in the rate of perforation or negative exploration.

SPECIAL CASES

The existence of "chronic appendicitis" is controversial but generally involves chronic waxing-and-waning RLQ pain from intermittent obstruction and inflammation. Patients usually feel better after the appendix is removed and almost all surgical specimens demonstrate abnormal histology.

Rarely, patients may present with a RLQ mass that is an abscess or a phlegmon associated with a perforated appendix. Standard treatment is to percutaneously drain the abscess using image guidance and give intravenous antibiotics. An interval appendectomy, using either the open or the laparascopic procedure, can then be performed 6 to 8 weeks later.

Incidental appendectomy (removal of the appendix when an operation is being performed in the abdomen for another reason) is not warranted unless future diagnostic difficulties are anticipated. It can lead to increased infectious complications during certain procedures such as those involving vascular grafts.

Even in operations not associated with increased complications from incidental appendectomy, it is associated with increased operative time and cost with questionable benefit. Most patients who develop appendicitis are young, whereas most people who undergo other intra-abdominal operations are elderly and have a much smaller chance of ever needing an appendectomy.

Appendicitis in pregnancy poses a particularly difficult diagnostic problem because the enlarging uterus displaces the appendix. Appendicitis can occur during any trimester, although perforation is more common in the third trimester. The pregnant woman may complain of abdominal pain at any location to which the appendix is shifted. However, right lower quadrant pain is still the most common location of discomfort in pregnant women with appendicitis. Symptoms such as nausea, vomiting, and anorexia occur in both appendicitis and pregnancy and thus do not contribute to differentiating the diagnosis. Ultrasound is particularly helpful in making a diagnosis in pregnant women, both because it does not pose a radiation risk and because it enables visualization of pelvic pathology. Early appendectomy is preferable because ruptured appendicitis can lead to fetal death and maternal morbidity, whereas appendectomy prior to perforation and negative explorations generally pose less risk to the fetus.

See Also the Following Articles

Bacterial Overgrowth • Computed Tomography • Dietary Fiber • Emesis • Laparoscopy • Nausea • Pylephlebitis

Further Reading

Hale, D. A. (1997). Appendectomy: A contemporary appraisal. *Ann. Surg.* **225**, 252–261.

Hardin, D. M. (1999). Acute appendicitis: Review and update. *Am. Fam. Physician* **60**, 2027–2034.

Rao, P. M. (1998). Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *N. Engl. J. Med.* **338**, 141–146.

Sabiston, D. C. (1997). Appendicitis. *In* "Textbook of Surgery: The Biologic Basis of Modern Surgical Practice" (D. Sabiston and H. Lyerly, eds.), 15th Ed., pp. 964–970. W. B. Saunders, Philadelphia.

Schrock, T. R. (1998). Appendicitis. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman *et al.*, eds.), 6th Ed., pp. 1778–1787. W. B. Saunders, Philadelphia.

Silen, W. (1996). "Cope's Early Diagnosis of the Acute Abdomen," 19th Ed. Oxford University Press, New York.

Soybel, D. (2001). Appendix. *In* "Surgery: Basic Science and Clinical Evidence" (J. Norton *et al.*, eds.), pp. 647–665. Springer-Verlag, New York.

Appetite

MARGO A. DENKE
University of Texas Southwestern Medical Center, Dallas

anorexigenic Appetite suppressing.
appetite Instinctive desire to eat. Appetite promotes eating behaviors to sustain life.
leptin Adipocyte peptide hormone that serves to decrease appetite.
orexigenic Appetite stimulating.

Appetite dysregulation exacerbates disease, thus mechanisms that regulate appetite are a major research focus in the fields of obesity, cancer, eating disorders, and AIDS.

REGULATION BY THE HYPOTHALAMUS

Eating behaviors are chemically encoded in the hypothalamus. Orexigenic signals of neuropeptide Y, galanin, endogenous opioid peptides, melanin-concentrating hormone, glutamate, and γ-aminobutyric acid promote food consumption behavior. Anorexigenic signals, including the entire family of corticotropin-releasing hormone (CRH)-related peptides, neurotensin, glucagon-like peptide-1, melanocortin, and agoutiprotein, promote the cessation of food consumption. Each neuropeptide has its own specific cellular receptors, occurring in high concentration in the paraventricular nucleus of the hypothalamus but present in other areas of the brain. All appear interconnected with feedback loops whereby one signal peptide can alter the secretion of another signal peptide. No single peptide is the gatekeeper to turning on or off appetite; what is apparent is an entire network of signals, and their frequency and amplitude are responsible for triggering behaviors.

The network of appetite signals accounts for the behavioral observations that appetite and food consumption patterns are dynamic. Biological, environmental, and psychological events readily influence behavior. Habitual intake, memories of food-related activities, and the sheer anticipation of consumption have been shown to influence single meal consumption of specific foods. External clues, such as the appearance of food, aroma, anticipated palatability, and the number of food choices, have been shown to modify the perception of appetite as well as the behaviors of eating. Psychosomatic consequences of eating, such as reduction in anxiety, can exert additional influences on behavior. Appetite appears analogous to memory; although memory and appetite are chemically encoded, every individual has their own unique signal circuitry underlying their eating behaviors. Just as memories change over time, the circuitry for appetite can also be modified.

REGULATION BY FAT CELLS

A major breakthrough in the physiology of appetite regulation came with the discovery of leptin and resistin, two hormones synthesized by adipocytes.

Leptin secretion increases as adipocytes enlarge, and decreases during fasting. Identification of leptin receptors in the hypothalamus has provided an intriguing biochemical explanation for the ability of an animal to regulate body weight tightly within a fairly narrow set point range. The leptin signal may serve as an anorexin by its ability to alter secretion of orexins and anorexins. Obese persons have appropriately elevated leptin levels, but whether this is an epiphenomenon of obesity or a clue to its pathologic cause is uncertain.

Resistin secretion increases during feeding and during adipose tissue exposure to insulin. In contrast to leptin, a hypothalamic receptor for resistin has not yet been identified. Instead, resistin appears to induce adipocyte resistance to insulin. Resistin also inhibits adipocyte differentiation. Rosiglitazone, a drug classified as an "insulin sensitizer," reduces resistin levels, suggesting that resistin plays a key role in determining insulin resistance.

GENETIC DISORDERS OF APPETITE REGULATION

The importance of the orexigenic and anorexigenic signals and their receptors has been highlighted by the identification of rare families with specific genetic defects associated with childhood obesity. Mutations in leptin, the leptin receptor, prohormone convertase 1 (PC1), pro-opiomelanocortin (POMC), melanocortin 4 receptor (MC4-R), and peroxisome

 Copyright 2004, Elsevier (USA). All rights reserved.

proliferator-activated receptor (PPAR) γ2 genes have been described in children with severe obesity.

Prader–Willi syndrome is a rare disorder characterized by a preoccupation with food, lack of satiation, and incessant food-seeking behaviors due to loss of paternal gene expression from chromosome 15q11–q13. The dysregulation of appetite in Prader–Willi patients may be due to deletion of key genes that alter synthesis, release, metabolism, binding, intrinsic activity, or reuptake of appetite-regulating neurotransmitters.

See Also the Following Articles

Obesity, Treatment of • Prader–Willi Syndrome • Satiety

Further Reading

Chen, D., and Garg, A. (1999). Monogenic disorders of obesity and body fat distribution. *J. Lipid Res.* **40**(10), 1735–1746.

Dimitropoulos, A., Feurer, I. D., Roof, E., *et al.* (2000). Appetite behavior, compulsivity, and neurochemistry in Prader–Willi syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **6**, 125–130.

Kalra, S. P., Dube, M. G., Pu, S., *et al.* (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocr. Rev.* **20**, 68–100.

Rogers, P. J. (1999). Eating habits and appetite control: A psychological perspective. *Proc. Nutr. Soc.* **58**, 59–67.

Steppan, C. M., Bailey, S. T., Bhat, S., *et al.* (2001). The hormone resistin links obesity to diabetes. *Nature.* **409**, 307–312.

Arachidonic Acid

STEFANO FIORE
University of Illinois, Chicago

nonsteroidal anti-inflammatory drugs (NSAIDs) Chemicals endowed with pharmaceutical anti-inflammatory activities. Unlike glucocorticoids, NSAIDs do not impact endocrine and immunological functions and are therefore favored as therapeutic agents for long-term treatment of inflammatory conditions. Some of the most widely used NSAIDs include oxicam, salicylate (aspirin), acetic acid (indomethacin, diclofenac), fenamate, propionic acid (ibuprofen, naproxen), and pyrazole.

polyunsaturated fatty acids (PUFAs) A particular class of fatty acid sharing the general characteristics of a linear carbon backbone substituted with hydrogen atoms and bearing a carboxylic group at one end of the molecule (C1). The distinguishing features of specific PUFAs are the dietary source and the presence of double bonds within the molecule. Another characteristic that demonstrates a correlation with distinct biological actions is the distance of the last double bond from the last carbon atom of the molecule (C20 in the case of AA).

Arachidonic acid is the most extensively studied of the polyunsaturated fatty acids (PUFAs) present in eukaryotic cell membranes and it is typically found esterified to membrane phospholipids in the *sn*-2 position of the glycerol backbone. Like all eicosanoids (fatty acids composed of a 20-carbon-atom backbone, noted as C20) and other essential PUFAs, such as linoleic acid and linolenic acid, arachidonate can be subjected to a complex rearrangement via remodeling of its all-*cis* double-bond configuration and chemical modification by insertion of chemical groups.

INTRODUCTION

Various pathways using arachidonic acid (AA) as the initial substrate are composed of dioxygenases that carry out a complex reaction involving abstraction of selected hydrogens and insertion of molecular oxygen. Two major classes of enzymes, cyclooxygenases (COX) and lipoxygenases (LOX), are recognized for their prominent role in generating a number of important biological mediators. Among these, prostaglandins (PGs) and leukotrienes (LTs) are widely studied given their recognized role in human disease conditions as well as physiological and/or pathophysiological activities. Of

 Copyright 2004, Elsevier (USA). All rights reserved.

proliferator-activated receptor (PPAR) γ2 genes have been described in children with severe obesity.

Prader–Willi syndrome is a rare disorder characterized by a preoccupation with food, lack of satiation, and incessant food-seeking behaviors due to loss of paternal gene expression from chromosome 15q11–q13. The dysregulation of appetite in Prader–Willi patients may be due to deletion of key genes that alter synthesis, release, metabolism, binding, intrinsic activity, or reuptake of appetite-regulating neurotransmitters.

See Also the Following Articles

Obesity, Treatment of • Prader–Willi Syndrome • Satiety

Further Reading

Chen, D., and Garg, A. (1999). Monogenic disorders of obesity and body fat distribution. *J. Lipid Res.* **40**(10), 1735–1746.

Dimitropoulos, A., Feurer, I. D., Roof, E., *et al.* (2000). Appetite behavior, compulsivity, and neurochemistry in Prader–Willi syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **6**, 125–130.

Kalra, S. P., Dube, M. G., Pu, S., *et al.* (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocr. Rev.* **20**, 68–100.

Rogers, P. J. (1999). Eating habits and appetite control: A psychological perspective. *Proc. Nutr. Soc.* **58**, 59–67.

Steppan, C. M., Bailey, S. T., Bhat, S., *et al.* (2001). The hormone resistin links obesity to diabetes. *Nature.* **409**, 307–312.

Arachidonic Acid

STEFANO FIORE
University of Illinois, Chicago

nonsteroidal anti-inflammatory drugs (NSAIDs) Chemicals endowed with pharmaceutical anti-inflammatory activities. Unlike glucocorticoids, NSAIDs do not impact endocrine and immunological functions and are therefore favored as therapeutic agents for long-term treatment of inflammatory conditions. Some of the most widely used NSAIDs include oxicam, salicylate (aspirin), acetic acid (indomethacin, diclofenac), fenamate, propionic acid (ibuprofen, naproxen), and pyrazole.

polyunsaturated fatty acids (PUFAs) A particular class of fatty acid sharing the general characteristics of a linear carbon backbone substituted with hydrogen atoms and bearing a carboxylic group at one end of the molecule (C1). The distinguishing features of specific PUFAs are the dietary source and the presence of double bonds within the molecule. Another characteristic that demonstrates a correlation with distinct biological actions is the distance of the last double bond from the last carbon atom of the molecule (C20 in the case of AA).

Arachidonic acid is the most extensively studied of the polyunsaturated fatty acids (PUFAs) present in eukaryotic cell membranes and it is typically found esterified to membrane phospholipids in the *sn*-2 position of the glycerol backbone. Like all eicosanoids (fatty acids composed of a 20-carbon-atom backbone, noted as C20) and other essential PUFAs, such as linoleic acid and linolenic acid, arachidonate can be subjected to a complex rearrangement via remodeling of its all-*cis* double-bond configuration and chemical modification by insertion of chemical groups.

INTRODUCTION

Various pathways using arachidonic acid (AA) as the initial substrate are composed of dioxygenases that carry out a complex reaction involving abstraction of selected hydrogens and insertion of molecular oxygen. Two major classes of enzymes, cyclooxygenases (COX) and lipoxygenases (LOX), are recognized for their prominent role in generating a number of important biological mediators. Among these, prostaglandins (PGs) and leukotrienes (LTs) are widely studied given their recognized role in human disease conditions as well as physiological and/or pathophysiological activities. Of

 Copyright 2004, Elsevier (USA). All rights reserved.

these biological actions, one of the most significant is the major role played by eicosanoids in inflammation, where they contribute to all of the clinical symptoms associated with the inflammatory condition, namely, pain, redness, and swelling. The ever-growing number of molecules derived from AA includes other families such as lipoxins (LXs), hepoxilins, hepoxides, monohydroxyeicosatretraenoic acids (HETEs), dihydroxyeicosatretraenoic acids, and their hydroperoxy precursors. Whereas synthesis of most of these mediators involves the non-heme iron catalytic center typical of cyclo- and lipoxygenases, hepoxilins and hepoxides originate via heme proteins such as hematin and cytochrome P450.

ARACHIDONATE RELEASE

Although the number of compounds originating from AA exceeds the hundreds, an overarching consistency is found as to the carefully orchestrated synthesis of these mediators. A hallmark of these tightly regulated syntheses is the controlled release of esterified AA from the membrane phospholipid storage, a task accomplished via stimulus-induced activation of cytosolic and secreted forms of phospholipase A_2 ($cPLA_2$, $sPLA_2$) that cleave AA from the *sn*-2 position and lead to accumulation of the free acid form, the initial substrate for subsequent enzymatic catalysis. Upon availability of substrate, a second level of regulation involves the COX and LOX enzymes that require several cofactors and are subject to so-called "suicide inactivation," a self-inactivation process that effectively limits the amount of bioactive mediators that can be generated.

ARACHIDONATE BIOLOGY

AA per se is endowed with biological activities such as activation of leukocyte responses and is known to interact and modulate the function of several intracellular targets such as guanosine 5′-triphosphate-binding proteins. Structurally related molecules such as anandemide are known to play an important role in regulating specific central nervous system activities, and specific gastrointestinal (GI) activities for AA include the modulation of gallbladder contraction.

THE CYCLOOXYGENASE PATHWAYS: PROSTANOIDS

An impressive list of gastroenterological-specific actions can be assigned to the COX- and LOX-derived AA products. Prostaglandins (PGs), for example, in addition to acting as cellular and physiological mediators (i.e., their receptor-mediated regulation of cyclic AMP levels and their role in parturition), are known to carry out important physiological tasks in mucosal secretion and motility as well as pathophysiological activities contributing to mucosal inflammation, injury, and tumor progression. PG-mediated gastric protective activities include mucus production, production of surface-active phospholipids, and bicarbonate secretion as well as regulation of mucosal blood flow and cell proliferation. The physiological role of PGs in mucosal protection is exemplified by the well-known gastrointestinal adverse events (dyspepsia, bleeding, and ulceration) caused by inhibitors of PG synthesis. The largest and most widely used class of drugs inhibiting PG synthesis targets the COX enzymes and is better known as the family of nonsteroidal anti-inflammatory drugs (NSAIDs), which includes agents such as aspirin and ibuprofen. Only in the past 10 years has the discovery of two COX isoforms, the constitutive form, or type 1 (COX-1), and the inducible form, or type 2 (COX-2), allowed the issue of gastric toxicity of NSAIDs to be addressed. In an ongoing effort to elucidate the specific role of these two enzymes, it is now generally assumed that COX-1 is mainly involved in the physiological aspects of prostanoid biology, whereas the inducible COX-2 is associated with expression of inflammatory events and the pathophysiology of several disease conditions. In addition to the GI toxicity of NSAIDs, which has at least been partially addressed by the development of new drugs such as rofecoxib and colecoxib that take advantage of the slightly larger substrate pocket of COX-2 to selectively inhibit this enzyme sparing the COX-1 isoform, recent epidemiological studies suggest that selective NSAIDs have chemopreventive properties toward solid tumors including colon adenocarcinoma. At present, however, it is highly debated whether this NSAID property is linked to suppressed synthesis of PGs [and most notably the major product prostaglandin E2 (PGE2)] or involves an as yet to be identified mechanism of action. Conversely, PGE2 is known to contribute to accelerated tumor progression, with marked pro-angiogenic properties, and animal models show a positive correlation between overexpression of COX-2 and progression from early stages of adenoma and polyp formation to adenocarcinoma. These active areas of investigation will likely provide critical new knowledge about the extensive networks of interactions mediated by PGs in the GI tract including inflammatory bowel disease (IBD), for which, in contrast to the *ex juvantibus* criteria by which the beneficial use of anti-inflammatory drugs such as sulfosalazine suggests that eicosanoids are relevant to IBD pathogenesis, it has not yet been

demonstrated that excess synthesis of lipid mediators plays a major role in establishing or maintaining the disease condition. In addition to PGE2, the prostaglandin family of mediators in the GI tract includes PGI2, $PGF2_{\alpha}$, and thromboxanes. To each of these compounds have been attributed specific actions within various segments of the GI system. The response to each group of PGs is determined not only by their structural specificity but also by the specific form of the receptors that are expressed. Four major subtypes of prostaglandin receptors, EP_1-EP_4, have been identified to date, in addition to a thromboxane receptor and a PGI2 receptor. The tissue distribution and the intrinsic properties of these G-protein-coupled receptors are under active investigation and selective agonists as well as antagonists are in early or late stages of *in vivo* animal studies, clinical trials, or clinical use (i.e., misoprostol, iloprost).

THE LIPOXYGENASE PATHWAYS: LEUKOTRIENES AND LIPOXINS

Although the experience in GI clinical practice highlights the role of COX-derived prostanoids, other families of eicosanoids, such as the LOX-derived LTs, play a significant role in modulating GI functions. Elevated LTB_4 levels, for example, are found in GI tissues affected by inflammatory conditions and peptido-leukotrienes LTC_4 and D_4 are known to be among the most potent stimuli causing intestinal smooth muscle contraction. Based on this evidence, inhibitors of the 5-lipoxigenase enzyme, a common biosynthetic element in LTB_4, $-C_4$, $-D_4$, and $-E_4$ synthesis, have been used in an IBD clinical trial leading to improvement of symptoms and sigmoidoscopic appearance. In manner analogous to PGs, many of the LT-induced responses are mediated by specific G-protein-coupled receptors. Two forms have been identified thus far for both peptido-leukotrienes (cys-LT1 and cys-LT2) and LTB_4 (BLTR1 and BLTR2) and they are characterized by specific tissue distributions. For example, cys-LT2 seems to be the main receptor involved in peptido-leukotriene-induced contraction of ileum smooth muscle cells and BLTR1 and BLTR2 are present on leukocytes and lymphocytes, respectively. BLTR2 is of particular interest because of its expression on T cells and its role in mediating the immune-regulatory effects of HETEs and LTB_4. The activation of these signaling pathways and the resulting interactions between immune competent cells and inflammatory infiltrates and the gastrointestinal tissues are recognized as an increasingly important contribution to the pathophysiology of GI disorders and may play an important role in IBD.

The interactions between leukocytes and lymphocytes and the intestinal mucosa cell components also play a major role in eicosanoid production. In fact, it is mostly via transcellular routes of biosynthesis that specific classes of AA-derived mediators are formed. Among these, the LXs, a class of compound generally endowed with anti-inflammatory/inhibitory effects originated by sequential recruitment of multiple LOX enzymes (three major LOX enzymes are expressed in mammalian cells: 5-, 12-, and 15-LOX), have been the subject of specific studies examining their gastrointestinal effects. Inhibition of neutrophil transmigration across intestinal epithelial monolayers, inhibition of pathogen-induced release of cytokines from the intestinal epithelium, and reduced severity and mortality in an animal model of dextran sodium sulfate-induced colitis are some of the GI-specific effects observed to date. As with the previously described PGs and LTs, LX specificity is also based on the interaction with specific G-protein-coupled receptors (ALXR, LXA_4R).

See Also the Following Articles

Non-Steroidal Anti-Inflammatory Drugs (NSAIDS) • NSAID-Induced Injury

Further Reading

Eberhart, C. E., and DuBois, R. N. (1995). Eicosanoids and the gastrointestinal tract. *Gastroenterology* **109**, 285–301.

Goetzl, E. J., An, S., and Smith, W. L. (1995). Specificity of expression and effects of eicosanoid mediators in normal physiology and human diseases. *FASEB J.* **11**, 1051–1058.

Krause, W., and DuBois, R. N. (2000). Eicosanoids and the large intestine. *Prostaglandins Other Lipid Mediat.* **61**, 145–161.

Jones, M. K., *et al.* (1999). Inhibition of angiogenesis by nonsteroidal anti-inflammatory drugs: Insight into mechanisms and implications for cancer growth and ulcer healing. *Nat. Med.* **5**, 1318–1323.

Serhan, C. N., Haeggstrom, J. Z., and Leslie, C. C. (1996). Lipid mediator networks in cell signaling: Update and impact of cytokines. *FASEB J.* **10**, 1147–1158.

Vane, J. R., and Botting, R. M. (1998). Anti-inflammatory drugs and their mechanism of action. *Inflamm. Res.* **47**, S78–S87.

Ascites

SHABANA F. PASHA AND PATRICK S. KAMATH
Mayo Clinic and Foundation

azotemia Elevation of blood urea nitrogen due to impaired renal function.
Budd–Chiari syndrome Obstruction of the hepatic venous outflow at the level of the large hepatic veins or the suprahepatic or intrahepatic segment of the inferior vena cava.
Child–Pugh classification Grouping based on determination of hepatic functional reserve; a useful predictor of surgical morbidity and mortality.
cirrhosis Advanced liver disease characterized by distorted architecture secondary to hepatic fibrosis and regenerative nodules.
constrictive pericarditis Fibrous scarring and noncompliance of the pericardium that results from causes of chronic pericarditis, including tuberculosis, malignancy, or radiation.
hepatic hydrothorax Pleural effusion (more than 500 ml) in patients with cirrhosis, in the absence of cardiopulmonary or subdiaphragmatic pathology.
hepatojugular reflux Sustained increase in jugular venous pressure elicited by compression of the abdomen in patients with right heart failure.
hepatorenal syndrome Impaired renal function in the presence of advanced liver disease.
Meig's syndrome Triad of benign ovarian fibroma with ascites and right-sided pleural effusion.
myxedema Thyroid deficiency in adults associated with skin and soft tissue edema.
pseudomyxoma peritonii Metastatic peritoneal tumor that results in gelatinous implants on the peritoneum.
pulsus paradoxicus Exaggerated decrease (greater than 20 mmHg) in inspiratory systolic blood pressure.
renin–angiotensin–aldosterone system Vasoactive system that causes renal vasoconstriction and retention of sodium and water.
spontaneous bacterial peritonitis Primary infection of ascitic fluid in patients with advanced liver disease.
venoocclusive disease Hepatic venous outflow obstruction that occurs in patients undergoing bone marrow transplantation, radiation therapy, liver transplantation, or ingestion of alkaloid toxins; the result of occlusion of hepatic sinusoids and small venules.

Ascites is defined as the excessive accumulation of fluid in the peritoneal cavity. Cirrhosis is the most common cause of ascites, followed by malignancy and cardiac failure. Patients with advanced liver disease develop infections of the ascitic fluid, a condition known as spontaneous bacterial peritonitis. Moreover, these patients can develop hepatorenal syndrome, a functional renal failure. Finally, patients may have hepatic hydrothorax, which involves symptomatic pleural effusions.

MALIGNANT ETIOLOGY

Cirrhosis is the cause of ascites in up to 80% of cases; malignancy and cardiac failure are the causes in 10% and 5%, respectively. Other causes account for fewer than 5% of cases of ascites. About 5% of patients have ascites due to more than one cause. Fifty percent of cirrhotic patients eventually develop ascites; up to 10% of these have ascites refractory to treatment. Ovarian cancer is the most common cause of malignant ascites and accounts for almost 50% of cases of the disease. Occult malignancies account for 20% of cases of malignant ascites and the remaining 30% of cases result from pancreatic cancer, gastric cancer, colon cancer, lung cancer, breast cancer, or lymphoma. The causes of ascites are summarized in Table I.

PATHOGENESIS

Ascites in cirrhotic patients results from a combination of portal hypertension and renal retention of sodium. As a result of factors such as nitric oxide (NO), which are present in excess in cirrhosis, there is splanchnic and peripheral vasodilatation. This results in a decrease in the effective arterial blood volume (EABV). In an attempt to correct the EABV, there is stimulation of the renin–angiotensin–aldosterone system (RAAS), the sympathetic nervous system (SNS), and vasopressin. These vasoactive systems work in concert to cause renal retention of sodium and water, as well as renal vasoconstriction. The result is an increase in plasma volume. The excessive fluid retained is compartmentalized to the peritoneal cavity as a result of portal hypertension, which is an increase in pressure within the hepatic

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Etiology of Ascites

Hepatic	
	Cirrhosis
	Budd–Chiari syndrome
	Liver metastases
	Alcoholic hepatitis
	Venoocclusive disease
	Portal vein thrombosis
Cardiac	
	Congestive heart failure
	Constrictive pericarditis
	Right atrial myxoma
Malignant	
	Peritoneal carcinomatosis
	Pseudomyxoma peritonii
Infectious	
	Tuberculous peritonitis
	HIV infection
Renal	
	Nephrotic syndrome
	Continuous ambulatory peritoneal dialysis
Other	
	Pancreatitis
	Myxedema
	Meig's syndrome
	Lymphatic obstruction/disruption
	Collagen vascular diseases
	Protein-losing enteropathy

sinusoids. A hepatic sinusoid pressure greater than 12 mmHg is usually required for ascites to develop.

Malignant ascites results from exudation of fluid from peritoneal carcinomatosis and occlusion of diaphragmatic lymphatics, with impairment of peritoneal fluid absorption. Tumor infiltration in the liver, leading to hepatic venous obstruction, is a less common cause of malignant ascites. Ascites in patients with hepatocellular carcinoma may be secondary to portal hypertension or to portal vein thrombosis due to tumor.

Tuberculous peritonitis leads to exudation of proteinaceous fluid into the peritoneal cavity, and resultant ascites. Pancreatic ascites can be seen in both acute and chronic pancreatitis and results from disruption of the pancreatic duct and leakage of pancreatic secretions into the peritoneum. Chylous ascites occurs from a lymphatic disruption as a result of trauma or malignant obstruction of the lymphatic ducts.

PHYSICAL EXAMINATION

Normally there is less than 75–100 ml of fluid in the peritoneal cavity. Ascites can be detected by eliciting shifting dullness when peritoneal fluid collection exceeds 500 ml. Fluid wave is positive in the presence of tense ascites; patients with tense ascites may also have concomitant lower extremity edema. Ultrasonography, which is frequently used to confirm the diagnosis, can detect as little as 100 ml of peritoneal fluid.

Peripheral stigmata of chronic liver disease, including spider angiomata, palmar erythema, caput medusae, gynecomastia, and testicular atrophy, may be seen when ascites results from cirrhosis and portal hypertension. Some patients may have generalized anasarca. The cirrhotic liver in patients with an advanced stage of disease is usually shrunken and may not be palpable. The spleen may be palpable following a therapeutic paracentesis.

Patients with congestive heart failure and ascites have elevated jugular venous pressure, peripheral edema, and presence of S3 or S4 (low-pitched sounds detected using the stethoscope) on cardiac examination. Constrictive pericarditis is characterized by presence of pulsus paradoxicus, rapid X and Y descents of the jugular venous pulse, pericardial knock, and ascites out of proportion to peripheral edema. The presence of hepatojugular reflux confirms a cardiac cause for the ascites.

DIFFERENTIAL DIAGNOSIS

Other etiologies of a distended abdomen, including pregnancy, ovarian mass, gaseous distension from bowel obstruction, and obesity, must be excluded in a patient suspected to have ascites. On percussion of the abdomen, ascites presents with flank dullness, whereas an ovarian mass typically presents with central dullness and tympanitic flanks.

DIAGNOSIS

Abdominal paracentesis is the most rapid and cost-effective method of diagnosing the etiology of ascites. Paracentesis should be performed in all patients with new-onset ascites and at the time of every hospital admission in all patients with ascites. A low threshold must be maintained for repeating the paracentesis, because infection may present with only minimal symptoms. The only absolute contraindication to paracentesis is an uncooperative patient. Complications occur in fewer than 1% of cases and include abdominal wall hematomas. Serious complications such as hemoperitoneum and bowel perforation occur in less than 1 in 1000 paracenteses. Ascitic fluid analysis should include total protein and albumin concentration, total and differential cell count, and, in selected patients, gram stain, bacterial culture, and cytology. The serum-ascitic fluid albumin gradient, which is the difference between the serum albumin and ascitic fluid albumin, has a

TABLE II Serum and Ascitic Fluid Albumin Gradient

Gradient	Cause of ascites	
	Ascitic fluid total protein <2.5 g/dl	Ascitic fluid total protein >2.5 g/dl
>1.1	Cirrhosis Fulminant hepatic failure	Congestive heart failure Constrictive pericarditis Budd–Chiari syndrome Venoocclusive disease
<1.1	Nephrotic syndrome Myxedema	Peritoneal carcinomatosis Tuberculous peritonitis Pancreatic ascites Chylous ascites

sensitivity and specificity of greater than 95% in differentiating ascites secondary to portal hypertension from other causes (Table II).

Other tests include ascitic fluid amylase, triglyceride, glucose, and lactate dehydrogenase (LDH) concentration. The nucleated cell count is less than 500 cells/mm^3, with fewer than 250 neutrophils/mm^3 in uninfected ascites secondary to portal hypertension. Ascitic fluid cytology has a low sensitivity in the diagnosis of malignancy but is highly specific. Ascitic fluid culture has a sensitivity of only 50% in the diagnosis of tuberculous peritonitis. Chylous ascites is diagnosed when ascitic fluid triglyceride concentrations are higher than simultaneously drawn serum triglyceride concentrations. Grossly hemorrhagic ascites may occur in 22% of cases of malignant ascites and in 50% of cases of ascites secondary to metastatic hepatocellular carcinoma. Hemorrhagic ascites due to malignancy is differentiated from hemorrhage caused by a needle trauma during paracentesis by the absence of clotting of the sample in malignant ascites.

MANAGEMENT

The goal of the management of ascites is to maintain a negative sodium balance. This is achieved with dietary sodium restriction and the use of diuretics. Response to dietary sodium restriction alone, manifesting as weight loss of between 250 and 500 g/day and a decrease in ascites, occurs in 10–20% of patients. Most patients with ascites can be managed with dietary sodium restriction (90 mEq/day) along with use of a diuretic. Spironolactone is the diuretic of choice administered as a single dose. The initial dose is 50 mg/day and can be increased to a maximum dose of 400 mg/day as tolerated. Weight loss should not exceed 0.5–0.75 kg/day in patients without pedal edema, because this can lead to intravascular volume depletion and azotemia. Furosemide can be used in combination with spironolactone, but is associated with a greater likelihood of azotemia. The metabolic alkalosis and hypokalemia induced by excessive diuretic use may be associated with precipitation or worsening of hepatic encephalopathy.

REFRACTORY ASCITES

Most patients with ascites that is difficult to manage are noncompliant with their sodium restriction. A 24-hour urine sodium in excess of the restriction, in the presence of increasing ascites, indicates noncompliance. Refractory ascites includes diuretic-resistant ascites and diuretic-intractable ascites. Diuretic-resistant ascites is defined as ascites that cannot be managed with dietary sodium restriction and intensive diuretic use of spironolactone (400 mg/day) and furosemide (160 mg/day) for at least 1 week, or ascites that reaccumulates early despite this treatment regime. Diuretic-intractable ascites is defined as ascites that cannot be treated or that recurs as a result of inability to use an adequate diuretic regime due to diuretic-induced complications.

The mortality rates for patients with refractory ascites exceed 50% in 1 year and 80% in 2 years. Most patients with refractory ascites belong to Child class C. Medications that inhibit prostaglandin synthesis, such as nonsteroidal antiinflammatory drugs (NSAIDs), worsen renal function in patients with ascites and should not be used.

HEPATORENAL SYNDROME

Hepatorenal syndrome is the presence of impaired renal function as demonstrated by a glomerular filtration rate below 40 ml/minute or a serum creatinine >1.5 mg/dl in the presence of advanced liver disease and portal hypertension. Shock, bacterial sepsis, nephrotoxic agents, fluid loss, or excessive diuretic use should be excluded before the diagnosis is made. Patients with hepatorenal syndrome have proteinuria <500 mg/day with normal renal structure on light microscopy. Hepatorenal syndrome is of two types. In type 1 hepatorenal syndrome, there is a rapid progression in renal dysfunction, with the creatinine reaching a level >2.5 mg/dl or the creatinine clearance decreasing to <20 ml/minute in a period of less than 2 weeks. The progression in renal dysfunction is more gradual in type 2 hepatorenal syndrome. Type 1 hepatorenal syndrome is probably a more advanced stage of type 2 hepatorenal syndrome. Type 1 hepatorenal syndrome is seen predominantly in patients with alcoholic hepatitis and in patients with spontaneous bacterial peritonitis or other infections.

On the other hand, type 2 hepatorenal syndrome presents predominantly with refractory ascites.

The initial step in the approach to a patient suspected to have hepatorenal syndrome is to demonstrate a reduction in glomerular filtration rate or worsening in renal function by measuring serum creatinine or creatinine clearance. The next step is to rule out other causes of renal dysfunction. The fractional excretion of sodium helps in these cases, with a fractional excretion of sodium of <1.0 being seen in prerenal azotemia, hepatorenal syndrome, and glomerulonephritis. In glomerulonephritis, the urine sediment is grossly abnormal, whereas in prerenal azotemia, renal function improves with volume expansion. It must be emphasized that in the presence of aggressive use of diuretics, the fractional excretion of sodium can be >1.0, even when the underlying diagnosis is hepatorenal syndrome.

MANAGEMENT OF REFRACTORY ASCITES AND HEPATORENAL SYNDROME

Therapeutic Paracentesis

Refractory ascites can be managed with repeated large-volume paracentesis. However, removal of excessive peritoneal fluid can result in decreased effective intravascular volume, with decreased pulmonary capillary wedge pressure and atrial natriuretic peptide levels, and increased renin–angiotensin–aldosterone activity, the so-called postparacentesis circulatory dysfunction. Therefore, 6–8 grams of albumin should be infused intravenously for every liter of ascitic fluid that is removed to counteract this circulatory dysfunction. Patients undergoing large-volume paracenteses have been shown to have a higher response rate, shorter hospital stay, and fewer complications, with similar survival as compared to patients treated with diuretics.

Peritoneovenous Shunt

The LeVeen shunt is a subcutaneous peritoneovenous shunt placed between the superior vena cava and peritoneum via the internal jugular vein. It has no benefit in the reduction of mortality or complication rate as compared to patients undergoing repeated paracenteses for the treatment of refractory ascites. Complications include catheter infection, thrombosis, occlusion, and low-grade disseminated intravascular coagulation. Peritoneovenous shunts are currently seldom used because the alternative of transjugular intrahepatic portosystemic shunts is available for patients who fail paracentesis therapy.

Transjugular Intrahepatic Portosystemic Shunt

The transjugular intrahepatic portosystemic shunt (TIPS), placed by interventional radiologists, has largely replaced surgical shunts in the management of refractory ascites. The TIPS is effective in reducing activity of the renin–angiotensin–aldosterone system and hence leads to diuresis and natriuresis. Relatively normal renal function is required for a TIPS procedure to be effective in reducing ascites. The TIPS has been shown to be more effective than paracentesis in the treatment of ascites, but without a significant survival benefit and at increased costs. Small studies suggest a potential benefit of using the TIPS for type 1 hepatorenal syndrome, but the results may be influenced by patient selection. Important predictors of survival after a TIPS procedure include etiology of liver disease, prothrombin time, serum bilirubin, and serum creatinine. Complications include a high incidence of encephalopathy and shunt stenosis. The recommended indication for the TIPS in the management of refractory ascites is for patients who have an unsatisfactory response to paracentesis.

Surgical Portosystemic Shunts

Creating a side-to-side anastomosis between the portal vein and inferior vena cava leads to a reduction in portal pressure, natriuresis, diuresis, and relief of ascites. However, the high rates of mortality and complications from the procedure have led to the abandonment of the procedure for refractory ascites.

Orthotopic Liver Transplantation

Liver transplant is the only therapy for refractory ascites and hepatorenal syndrome associated with an improvement in long-term survival.

Summary

In summary, treatment of type 1 hepatorenal syndrome involves treatment of the underlying sepsis. A few studies have shown the benefit of using volume expansion with albumin in association with splanchnic vasoconstrictor agents such as terlipressin or midodrine, octreatide, or norepinephrine. The treatment of type 2 hepatorenal syndrome, which in effect is treatment of refractory ascites, is with large-volume paracentesis and, if this fails, transjugular intrahepatic portosystemic shunts. The role of a TIPS in the treatment of type 1 hepatorenal syndrome is not established. Liver transplantation is the only treatment option associated with long-term improvement in survival in patients with hepatorenal syndrome.

SPONTANEOUS BACTERIAL PERITONITIS

Ascitic fluid paracentesis should be carried out in all patients with new-onset ascites, at the time of hospital admission, and in patients with fever, abdominal pain, worsening encephalopathy, or renal insufficiency. An ascitic fluid absolute neutrophil count of $\geq 250/mm^3$ (neutrocytic ascites) is required for the diagnosis of spontaneous bacterial peritonitis (SBP). The ascitic fluid protein concentration is usually less than 1 g/dl. Most common organisms seen in SBP include *Escherichia coli, Streptococcus pneumoniae,* and *Klebsiella pneumoniae*. Diagnostic yield of ascitic culture is increased if ascitic fluid is inoculated into blood culture bottles immediately on paracentesis. Secondary bacterial peritonitis is polymicrobial, as opposed to monomicrobial SBP. SBP is characterized by a very low bacterial count and thus gram stain carries a sensitivity of 10% or lower in the diagnosis of spontaneous bacterial peritonitis. Secondary bacterial peritonitis should be suspected if ascitic fluid protein is more than 1 g/dl, glucose is less than 50 mg/dl, and ascitic fluid LDH is greater than serum LDH. The antibiotic of choice for the treatment of SBP is a third-generation cephalosporin, usually cefotaxime. Response to treatment is determined by demonstrating a reduction in the ascitic fluid neutrophil count of greater than 25%. Patients who have experienced even one episode of SBP should be treated indefinitely with quinolones such as norfloxacin to prevent subsequent episodes. The only other indication for antibiotic prophylaxis is in the patient with cirrhosis who has acute gastrointestinal bleeding, for which norfloxacin is given for up to 7 days. SBP is associated with a 1-year survival of between 40 and 70%, and consequently all patients with SBP should be evaluated for liver transplantation.

HEPATIC HYDROTHORAX

Hepatic hydrothorax is defined as pleural effusions >500 ml in patients with cirrhosis of the liver in the absence of cardiopulmonary or subdiaphragmatic pathology. The pathophysiology of hepatic hydrothorax is similar to the pathophysiology of ascites. In fact, it is ascitic fluid that moves from the peritoneal cavity into the pleural space through diaphragmatic defects. Because intrathoracic pressure is negative, this favors movement of fluid into the thoracic space. Thus, patients may have hepatic hydrothorax even in the absence of ascites. The right pleural space is more commonly involved. Similar to patients with ascites, patients with hepatic hydrothorax can have spontaneous bacterial infection of the fluid that results in spontaneous bacterial empyema.

Patients with hepatic hydrothorax have advanced liver disease and are usually candidates for liver transplantation. Therapy in such patients is directed at relieving symptoms and preventing pulmonary complications until such time that a liver transplant can be carried out. The initial management is with sodium restriction and diuretics, similar to the management of patients with ascites. Therapeutic thoracocentesis is carried out to relieve symptoms of dyspnea. Pleurodesis and chest tube placement should be avoided at all costs. If sodium restriction and diuretics fail, then transjugular intrahepatic portosystemic shunts can be placed. These shunts are usually effective in preventing the accumulation of fluid in the pleural space.

MANAGEMENT OF MALIGNANT ASCITES

Diuretics are not effective in the treatment of malignant ascites. Therapeutic paracenteses may be employed to alleviate pressure symptoms related to large-volume ascites. Peritoneovenous shunts have occasionally been placed in patients with malignant ascites and may have a role in treating patients with a life expectancy greater than a few months. There is no evidence to support the theory of widespread metastases due to dissemination of malignant cells via the shunt, leading to decreased survival rates.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Budd–Chiari Syndrome • Cirrhosis • Fulminant Hepatic Failure • Hepatorenal Syndrome • Portal Hypertension and Esophageal Varices • Portal Vein Thrombosis

Further Reading

Arroyo, V., Gines, P., *et al.* (1996). Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. *Hepatology* **23**, 164–176.

Gines, P., Uriz, J., Calahorra, G., *et al.* (2002). Transjugular intrahepatic portosystemic shunting versus paracentesis plus albumin for refractory ascites in cirrhosis. *Gastroenterology* **123**, 1839–1847.

Lazaridis, K. N., Frank, J. W., Krowka, M. J., and Kamath, P. S. (1999). Hepatic hydrothorax: Pathogenesis, diagnosis, and management. *Am. J. Med.* **107**, 267–272.

Malinchoc, M., Kamath, P. S., Gordon, F. D., Peine, C. J., Rank, J., and ter Borg, P. C. (2000). A model to predict survival in patients undergoing transjugular intrahepatic porto-systemic shunts. *Hepatology* **31**, 864–871.

Rimola, A., Garcia-Tsao, G., and Navasa, M. (2000). Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: A consensus document. International Ascites Club. *J. Hepatol.* **32**, 142–153.

Runyon, B. A. (1994). Care of the patient with ascites. *N. Engl. J. Med.*, **330**, 337–342.

Runyon, B. A. (1998). Management of adult patients with ascites caused by cirrhosis. *Hepatology* **27**, 264–272.

Atrophic Gastritis

ERNST J. KUIPERS

Erasmus MC University Medical Center, Rotterdam, The Netherlands

atrophic gastritis Loss of the glandular structures and a collapse of the reticulin skeleton of the gastric mucosa, with thinning of the glandular layer of the mucosa and replacement of glands by fibrosis and sometimes by intestinal-type cells.

chronic active gastritis Condition of chronic inflammation of the gastric mucosa associated with neutrophilic infiltration.

Helicobacter pylori Gram-negative rod-shaped bacterium that is able colonize the human gastric mucosa and induce chronic active gastritis.

intestinal metaplasia Condition in which the intestinal cells in the gastric mucosa become metaplastic.

The anatomy and function of the gastric mucosa remain unchanged throughout life, but can be disturbed by the occurrence of chronic active gastritis. Although this can be due to a variety of conditions, the most common cause is colonization with *Helicobacter pylori*. In a considerable proportion of affected subjects, chronic active gastritis leads to a loss of the glandular structures and a collapse of the reticulin skeleton of the mucosa, a condition called atrophic gastritis. As a result, the glandular layer of the mucosa becomes thinner and glands are replaced by fibrosis and sometimes by intestinal-type cells. This intestinal metaplasia may resemble either small bowel mucosa with sialomucin-containing goblet cells or colonic mucosa with sulfomucin-containing columnar cells, or a mixture of both.

INTRODUCTION

The exact mechanisms by which inflammation leads to atrophic gastritis and intestinal metaplasia are still unknown, but toxic bacterial and inflammatory products, dysregulated cell turnover, and autoimmunity may all play a role. In those with *Helicobacter pylori* gastritis, *H. pylori* eradication may halt the process, or even lead to a partial regression of atrophic gastritis. The risk for atrophic gastritis in subjects with chronic active gastritis particularly depends on the severity and distribution of inflammation.

SYMPTOMS

Chronic active gastritis and atrophic gastritis are mostly symptomless conditions, although some *H. pylori*-positive subjects may suffer from dyspepsia and may possibly benefit from *H. pylori* eradication. Subjects with severe atrophic gastritis of the corpus mucosa may ultimately develop megaloblastic anemia or myelopathy; these signs of vitamin B_{12} deficiency are due to impaired vitamin uptake as a result of insufficient intrinsic factor secretion and reduced peptic digestion of food–B_{12} complexes.

DIAGNOSIS

Atrophic gastritis is usually diagnosed by means of microscopic evaluation of gastric biopsy specimens obtained during endoscopy. Endoscopy without biopsy sampling is insufficient to diagnose or rule out atrophic gastritis. The demonstration of reduced basal and stimulated gastric acid output may support a diagnosis of atrophic gastritis, but this test is not routinely used. Serum pepsinogen and gastrin levels can also help to

 Copyright 2004, Elsevier (USA). All rights reserved.

Rimola, A., Garcia-Tsao, G., and Navasa, M. (2000). Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: A consensus document. International Ascites Club. *J. Hepatol.* **32**, 142–153.

Runyon, B. A. (1994). Care of the patient with ascites. *N. Engl. J. Med.*, **330**, 337–342.

Runyon, B. A. (1998). Management of adult patients with ascites caused by cirrhosis. *Hepatology* **27**, 264–272.

Atrophic Gastritis

ERNST J. KUIPERS

Erasmus MC University Medical Center, Rotterdam, The Netherlands

atrophic gastritis Loss of the glandular structures and a collapse of the reticulin skeleton of the gastric mucosa, with thinning of the glandular layer of the mucosa and replacement of glands by fibrosis and sometimes by intestinal-type cells.

chronic active gastritis Condition of chronic inflammation of the gastric mucosa associated with neutrophilic infiltration.

Helicobacter pylori Gram-negative rod-shaped bacterium that is able colonize the human gastric mucosa and induce chronic active gastritis.

intestinal metaplasia Condition in which the intestinal cells in the gastric mucosa become metaplastic.

The anatomy and function of the gastric mucosa remain unchanged throughout life, but can be disturbed by the occurrence of chronic active gastritis. Although this can be due to a variety of conditions, the most common cause is colonization with *Helicobacter pylori*. In a considerable proportion of affected subjects, chronic active gastritis leads to a loss of the glandular structures and a collapse of the reticulin skeleton of the mucosa, a condition called atrophic gastritis. As a result, the glandular layer of the mucosa becomes thinner and glands are replaced by fibrosis and sometimes by intestinal-type cells. This intestinal metaplasia may resemble either small bowel mucosa with sialomucin-containing goblet cells or colonic mucosa with sulfomucin-containing columnar cells, or a mixture of both.

INTRODUCTION

The exact mechanisms by which inflammation leads to atrophic gastritis and intestinal metaplasia are still unknown, but toxic bacterial and inflammatory products, dysregulated cell turnover, and autoimmunity may all play a role. In those with *Helicobacter pylori* gastritis, *H. pylori* eradication may halt the process, or even lead to a partial regression of atrophic gastritis. The risk for atrophic gastritis in subjects with chronic active gastritis particularly depends on the severity and distribution of inflammation.

SYMPTOMS

Chronic active gastritis and atrophic gastritis are mostly symptomless conditions, although some *H. pylori*-positive subjects may suffer from dyspepsia and may possibly benefit from *H. pylori* eradication. Subjects with severe atrophic gastritis of the corpus mucosa may ultimately develop megaloblastic anemia or myelopathy; these signs of vitamin B_{12} deficiency are due to impaired vitamin uptake as a result of insufficient intrinsic factor secretion and reduced peptic digestion of food–B_{12} complexes.

DIAGNOSIS

Atrophic gastritis is usually diagnosed by means of microscopic evaluation of gastric biopsy specimens obtained during endoscopy. Endoscopy without biopsy sampling is insufficient to diagnose or rule out atrophic gastritis. The demonstration of reduced basal and stimulated gastric acid output may support a diagnosis of atrophic gastritis, but this test is not routinely used. Serum pepsinogen and gastrin levels can also help to

 Copyright 2004, Elsevier (USA). All rights reserved.

diagnose atrophic gastritis, in particular for the purpose of population screening. A diagnosis of chronic active gastritis is then usually based on a positive serology for *H. pylori*, and a diagnosis of atrophic gastritis is based on a combination of increased serum gastrin levels and decreased serum pepsinogen levels, in particular a decrease of the serum pepsinogen I/pepsinogen II ratio.

ASSOCIATION WITH GASTRIC CANCER

The major clinical importance of atrophic gastritis is the increased risk for the intestinal type of gastric cancer. This risk may be elevated up to 90-fold in subjects with severe atrophic gastritis throughout the entire stomach. The annual incidence of gastric cancer among patients with atrophic gastritis is estimated to be between 0.3 and 1.0%.

Different factors may play a role in the increased risk for gastric cancer. Among these are the persistently increased cell turnover in atrophic mucosa, the increased mutagenesis caused by higher levels of nitrite and decreased levels of ascorbic acid in the gastric juice, and the lack of acid secretion, leading to decreased clearance of cells from newly arising carcinoma *in situ* microlesions.

See Also the Following Articles

Gastritis • *Helicobacter pylori*

Further Reading

Dixon, M. F., Genta, R. M., Yardley, J. H., and Correa, P. (1996). Classification and grading of gastritis. The updated Sydney system. *Am. J. Surg. Pathol.* **20**, 1161–1181.

Kuipers, E. J. (1998). Review article: Relationship between *Helicobacter pylori*, atrophic gastritis and gastric cancer. *Aliment. Pharmacol. Ther.* **12**(Suppl. 1), 25–36.

Kuipers, E. J., Uyterlinde, A. M., Peña, A. S., Roosendaal, R., Pals, G, Nelis, G. F., Festen, H. P. M., and Meuwissen, S. G. M. (1995). Long-term sequelae of *Helicobacter pylori* gastritis. *Lancet* **345**, 1525–1528.

Offerhaus, G. J. A., Price, A. B., Haot, J., ten Kate, F. W. J., Sipponen, P., Fiocca, R., Stolte, M., and Dixon, M. F. (1999). Observer agreement on the grading of gastric atrophy. *Histopathology* **34**, 320–325.

Sung, J. J., Lin, S. R., Ching, J. Y., Zhou, L. Y., To, K. F., Wang, R. T., Leung, W. K., Ng, E. K., Lau, J. Y., Lee, Y. T., Yeung, C. K., Chao, W., and Chung, S. C. (2000). Atrophy and intestinal metaplasia one year after cure of *H. pylori* infection: A prospective, randomized study. *Gastroenterology* **119**, 7–14.

Autoimmune Liver Disease

ALBERT J. CZAJA
Mayo Clinic and Foundation

autoimmune hepatitis Self-perpetuated liver inflammation of unknown cause associated with autoantibodies, increased serum gamma globulin level, interface hepatitis on histological examination, and responsiveness to corticosteroid therapy.

bridging necrosis Histological pattern of severe liver inflammation in which the inflammatory infiltrate and evidence of hepatocyte damage extend from portal tract to portal tract or portal tract to central vein.

interface hepatitis The *sine qua non* for the histological diagnosis of autoimmune hepatitis involves disruption of the limiting plate of the portal tract with extension of a mononuclear, frequently plasmacytic, infiltrate into acinar tissue in association with evidence of hepatocyte damage.

multiacinar necrosis Histological pattern of severe liver inflammation in which the inflammatory infiltrate and evidence of hepatocyte damage extend across and collapse lobules of liver tissue.

primary biliary cirrhosis Autoimmune liver disease characterized by antimitochondrial antibodies, cholestatic laboratory indices, and histological features of bile duct injury, including destructive or granulomatous cholangitis ("florid duct lesions").

primary sclerosing cholangitis Autoimmune liver disease, frequently associated with inflammatory bowel disease, that is characterized by cholangiographic changes of bile duct narrowing, cholestatic laboratory indices, and histological features of bile duct injury or biliary obstruction.

relapse Recrudescence of clinical, biochemical, and histological activity after discontinuation of corticosteroid therapy following induction of remission.

remission Disappearance of symptoms, improvement in serum aspartate aminotransferase activity to normal or less than twice normal, normal serum bilirubin and gamma globulin levels, and histological resolution to normal or near normal.

treatment failure Worsening of symptoms, laboratory indices of liver inflammation, and/or histological features despite compliance with therapy.

variant Lack of classical features of a single disease.

Autoimmune hepatitis is a self-perpetuated inflammation of the liver of unknown cause; it is characterized by interface hepatitis on histological examination, hypergammaglobulinemia, autoantibodies, and responsiveness to corticosteroid therapy. There are no pathognomonic features, and the diagnosis requires exclusion of virus-related, drug-induced, and hereditary conditions that may resemble it.

PREVALENCE

In the United States, 100,000–200,000 persons have autoimmune hepatitis and the condition accounts for 5.9% of the liver transplantations. The incidence of autoimmune hepatitis among Caucasian Northern Europeans is 1.9 per 100,000 population and its point prevalence is 16.9 per 100,000 population. Women have a threefold greater incidence of the disease compared to men, and the disease occurs in all age groups, in diverse ethnic populations, and in far-flung geographic regions.

DIAGNOSIS

The diagnostic criteria for autoimmune hepatitis have been codified by an international panel. An acute, occasionally fulminant, presentation has been recognized, and the requirement for 6 months of disease activity to establish chronicity has been waived. Cholestatic biochemical changes and histological features of bile duct destruction dissuade the diagnosis, whereas viral infection or drug-induced injury precludes it.

Clinical Criteria

The definite diagnosis requires a normal α1-antitrypsin phenotype; normal serum ceruloplasmin, iron, and ferritin levels; absence of markers of active hepatitis A, B, and C virus infection; daily alcohol intake of less than 25 g; no recent exposure to hepatotoxic drugs; predominant serum aminotransferase abnormality; serum globulin, gamma globulin, or immunoglobulin G levels of at least 1.5 times the upper limit of normal; presence of antinuclear antibodies (ANAs), smooth muscle antibodies (SMAs), or antibodies to liver/kidney microsome type 1 (anti-LKM1) in titers $\geq$1 : 80; absence of antimitochondrial antibodies (AMAs); at

 Copyright 2004, Elsevier (USA). All rights reserved.

least moderate to severe interface hepatitis on histological examination; and no histological evidence of biliary lesions, granulomas, or other prominent changes suggestive of another disease.

The probable diagnosis is justified by partial α1-antitrypsin deficiency; nonspecific abnormalities in serum ceruloplasmin, iron, and ferritin levels; alcohol intake of up to 50 g/day; low-level hypergammaglobulinemia; and low titers ($<1:80$) of ANA, SMA, or anti-LKM1, or the presence of other autoantibodies, including antibodies to soluble liver antigen/liver pancreas (anti-SLA/LP), asialoglycoprotein receptor (anti-ASGPR), actin (anti-actin), neutrophil cytoplasm (perinuclear anti-neutrophil cytoplasm antibodies, pANCAs), or liver cytosol type 1 (anti-LC1). Moderate to severe interface hepatitis must be present in the liver tissue, and prominent biliary lesions, granulomas, or features suggesting another diagnosis must be absent.

Scoring Criteria

A scoring system has been promulgated by the International Autoimmune Hepatitis Group to quantify the strength of the diagnosis, prevent isolated inconsistent findings from discounting the disease, and ensure the homogeneity of patient populations in clinical reports and treatment trials (Table I). The score based on pretreatment features can be upgraded or downgraded by the response to treatment. The sensitivity of the scoring system for definite autoimmune hepatitis ranges from 97 to 100%, and its specificity for excluding the disease in patients with chronic hepatitis C ranges from 66 to 92%. Its major weakness has been in discounting cholestatic syndromes with autoimmune features. In most instances, the scoring system is unnecessary for diagnosis, and its major value may be in the objective assessment of variant or atypical syndromes.

TYPES

Three types of autoimmune hepatitis have been proposed based on their autoantibody profiles. This subclassification has not been established because the various types do not have different etiologies, outcomes, or treatment requirements.

Type 1 Autoimmune Hepatitis

Type 1 autoimmune hepatitis is the most common form of autoimmune hepatitis worldwide, affecting at least 80% of all patients. ANAs and/or SMAs are its serological hallmarks. Concurrent immune diseases, especially autoimmune thyroiditis, are present in 38% of patients, and cirrhosis is established at presentation in 25% of patients. The target autoantigen is unknown.

Type 2 Autoimmune Hepatitis

Type 2 autoimmune hepatitis is characterized by anti-LKM1. It affects mainly children from ages 2 to 14 years, and it can occur in the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

TABLE I Scoring Criteria for Definite and Probable Diagnoses of Autoimmune Hepatitis[a,b]

Feature	Score	Feature	Score
Female	+2	Hepatotoxic drug exposure	−4
AP:AST ratio		No drug or toxin exposure	+1
Low ratio (<1.5)	+2	Average alcohol <25 g/day	+2
High ratio AP (>3.0)	−2	Average alcohol >60 g/day	−2
Globulin, gamma globulin, or immunoglobulin G level		Interface hepatitis	+3
		Lymphoplasmacytic infiltrate	+1
High level (>2 ULN)	+3	Rosette formation	+1
Moderate level (1.5−2 ULN)	+2	No typical histologic features	−5
Mild level (1−1.5 ULN)	+1	Biliary changes	−3
Autoantibodies (conventional)		Other features (fat, granulomas)	−3
Titer > 1:80	+3	Concurrent immune disease	+2
Titer 1:80	+2	Other autoantibodies	+2
Titer 1:40	+1	HLA DR3 or DR4	+1
Antimitochondrial antibodies	−4	Response to corticosteroids	+2
Active hepatitis A, B, or C	−3	Relapse after drug withdrawal	+3
Absent hepatitis markers	+3		

[a] Abbreviations: AP, alkaline phosphatase; AST, aspartate aminotransferase; ULN, upper limit of normal; HLA, human leukocyte antigen.
[b] Diagnostic scores pretreatment, definite, >15; probable, 10−15; diagnostic scores posttreatment, definite, >17; probable, 12−17.

(APECED). The disease is most common in Germany and France, where 20% of patients are adults. The target autoantigen is the cytochrome monooxygenase, CYP 2D6. Concurrent immune diseases include insulin-dependent diabetes mellitus, autoimmune thyroiditis, and vitiligo.

Type 3 Autoimmune Hepatitis

Type 3 autoimmune hepatitis is characterized by the presence of anti-SLA/LP. The target autoantigen may be a transfer ribonucleoprotein complex involved in the incorporation of selenocysteine into polypeptide chains. The clinical features and outcome are indistinguishable from type 1 autoimmune hepatitis.

PATHOGENESIS

Loss of Self-Tolerance

The negative selection of autoreactive immunocytes within the thymus can be defective during ontogeny. Polymorphisms in autoimmune regulator genes can enhance this defect. Molecular mimicry between foreign and self-antigens can result in cross-reactive immunological responses. Autoimmunity induced against one self-antigen can spread via intermolecular epitope mimicry to other homologous self-antigens and to anatomically distant tissue sites. The degree of antigenic homologies, the antigenic dose, and the genetic predisposition of the host determine the risk of disease.

Genetic Predisposition

The risk factors for type 1 autoimmune hepatitis in Caucasian North Americans and Northern Europeans are human leukocyte antigens (HLAs) HLA DR3 and HLA DR4. HLA DR3 is associated with early age-onset disease and a higher frequency of treatment failure and requirement for liver transplantation, compared to HLA DR4. In contrast, HLA DR4 is associated with late age onset, responsiveness to corticosteroid therapy, and frequent concurrent immune diseases.

The principal susceptibility allele of type 1 autoimmune hepatitis is *DRB1*0301*, and the secondary and independent susceptibility allele is *DRB1*0401*. Each allele encodes an identical six-amino-acid sequence (LLEQKR) between positions 67 and 72 on the DRβ polypeptide chain of the class II molecule of the major histocompatibility complex (MHC). A lysine residue (K) at position DRβ71 is the critical contact point between the antigenic peptide, class II MHC molecule, and T cell antigen receptor of the CD4 T helper cell. It may be the principal single determinant of susceptibility.

Other HLA DR4 alleles affect susceptibility in different ethnic groups, but they each encode an amino acid of charge and structure similar to that of lysine at position DRβ71. The principal susceptibility allele in Mestizo Mexicans is *DRB1*0404*, and it is *DRB1*0405* in Japanese and Argentine adult patients. The principal susceptibility allele among Argentine children and Brazilian patients is *DRB1*1301*, which may predispose to protracted infection with hepatitis A virus and prolonged exposure to hepatic self-antigens.

Polymorphisms of disease-nonspecific autoimmune promoters may act in synergy with the principal risk factors to affect disease susceptibility and outcome. Among Caucasian North Americans and Northern Europeans, polymorphisms of the tumor necrosis factor-α gene (*TNFA*2*), at position −308, and the cytotoxic T lymphocyte antigen-4 gene (*CTLA-4*) are associated with the risk of disease.

Cytodestructive Mechanisms

The principal effectors of liver cell destruction are antigen-specific clones of liver-infiltrating CD8 cytotoxic lymphocytes, the differentiation of which is modulated by type 1 cytokines, including interleukin-2, interferon γ, and tumor necrosis factor α. Antibody-dependent cell-mediated cytotoxicity is also involved under the mediation of type 2 cytokines, especially interleukin-10, and antigen−antibody complexes on the hepatocyte surface can attract natural killer cells that destroy the cell.

TREATMENT

Indications

Symptoms of fatigue, myalgia, and/or jaundice; serum (asparate aminotransferase, AST) levels of at least 10-fold normal or at least 5-fold normal in conjunction with serum gamma globulin levels of at least twice normal; and the presence of moderate to severe interface hepatitis, bridging necrosis, or multilobular necrosis on histological assessment are absolute indications for treatment. Less severe indices of disease activity are relative indications for treatment, and inactive disease does not require therapy.

Schedules

Regimens based on prednisone alone or a lower dose of prednisone in combination with azathioprine are equally effective and superior to no therapy or nonsteroidal schedules in managing all forms of autoimmune hepatitis (Table II).

TABLE II Treatment Regimens in Autoimmune Hepatitis

	Starting treatment dose (mg/day)			Withdrawing treatment dose (mg/day)		
Week	Prednisone only	Prednisone	Azathioprine	Prednisone only	Prednisone	Azathioprine
1	60	30	50	15	7.5	50
2	40	20	50	10	7.5	50
3	30	15	50	5	5	50
4	30	15	50	5	5	25
5	20	10	50	2.5	2.5	25
6		Fixed daily doses thereafter		2.5	2.5	25
7		Maintenance until end point		None	None	None

End Points

Treatment should be continued until disappearance of symptoms; normal or near normal serum AST, bilirubin, and gamma globulin levels; and inactive or minimally active histological findings. The average duration of treatment until remission in severe disease is 22 months. Therapy should then be withdrawn in a tapered fashion over a 6-week period while serum AST, bilirubin, and gamma globulin levels are monitored for relapse (Table II). Deterioration despite compliance with therapy (treatment failure) and drug intolerance are other end points of standard therapy.

TREATMENT FAILURE

Worsening symptoms, an increase in the serum AST and/or bilirubin levels by 67% of pretreatment values, and/or histological progression to bridging necrosis or multiacinar collapse indicate treatment failure. The management strategy must be modified by increasing the doses of medication or proceeding to liver transplantation.

Drugs

The dose of prednisone is increased to 60 mg daily if it is the sole drug or to 30 mg daily if it is combined with azathioprine. The dose of azathioprine is increased to 150 mg daily. The prednisone dose is then reduced by 10 mg and the azathioprine dose is reduced by 50 mg after each month of improvement until conventional maintenance levels are reached.

Liver Transplantation

Manifestations of decompensation despite therapy compel liver transplantation. Patient and graft survival rates after liver transplantation for autoimmune hepatitis range from 83 to 92%, and the actuarial 10-year survival is 75%. Autoimmune hepatitis recurs in 17% of patients, especially in those who are receiving inadequate immunosuppression, but it is usually managed satisfactorily by adjustments in the immunosuppressive regimen.

RELAPSE

Recrudescence of symptoms, increase in the serum AST level to more than threefold normal, and/or histological features of interface hepatitis after corticosteroid withdrawal connote relapse of the disease and the need for retreatment. Reapplication of the original treatment schedule reliably induces remission, but another relapse is common after drug withdrawal. Long-term maintenance schedules based on low-dose prednisone alone (less than 10 mg daily) or azathioprine alone (2 mg per kg daily) are effective indefinite management strategies after multiple relapses following conventional treatment and withdrawal.

PROMISING TREATMENTS

Drugs

Anecdotal successes with cyclosporine (5–6 mg per kg daily), tacrolimus (3 mg twice daily), mycophenolate mofetil (1 g twice daily), and 6-mercaptopurine (1.5 mg per kg daily) have been reported in the treatment of small numbers of patients recalcitrant to conventional corticosteroid therapies. Rigorous clinical trials are necessary to establish fully their benefit–risk ratios and their superiority to prednisone and azathioprine.

Site-Specific Interventions

Treatments based on targeted interruptions of the critical pathogenic pathways are the hopes for the future. These would include (1) peptides that could compete with self-antigens for the antigen-binding

groove of class II MHC molecules, (2) soluble cytotoxic lymphocyte antigen-4, which could dampen immunocyte activation, (3) T cell vaccines that could protect against clonal expansion of antigen-sensitized, liver-infiltrating CD8 cytotoxic T cells, (4) oral tolerance programs that could induce CD4 T helper cell anergy or apoptosis, (5) cytokine antibodies or recombinant products that could modulate the cytokine profile, and (6) gene therapies that could deliver regenerative growth factors and/or counterbalance polymorphic genes promoting immune reactivity.

VARIANT FORMS

Variant forms are conditions in which the classical features of autoimmune hepatitis are intermixed with features of primary biliary cirrhosis, primary sclerosing cholangitis, or chronic viral hepatitis. Eighteen percent of patients with autoimmune liver disease can be reclassified as having variant syndromes. Treatment is empiric and based on the predominant manifestations of the disease. Corticosteroids are used in the treatment of patients with serum alkaline phosphatase levels of less than twofold normal. Ursodeoxycholic acid alone or in combination with corticosteroids is appropriate for patients with pruritus and/or higher serum alkaline phosphatase levels. Patients with chronic viral hepatitis and autoimmune features must be distinguished from patients with autoimmune hepatitis and coincidental viral infection, because antiviral or immunosuppressive drugs are treatment options.

See Also the Following Articles

Alpha-1-Antitrypsin (α1AT) Deficiency • Cholangitis, Sclerosing • Cirrhosis • Liver Transplantation

Further Reading

Al-Khalidi, J. A., and Czaja, A. J. (2001). Current concepts in the diagnosis, pathogenesis, and treatment of autoimmune hepatitis. *Mayo Clin. Proc.* **76**, 1237–1252.

Ben-Ari, Z., and Czaja, A. J. (2001). Autoimmune hepatitis and its variant syndromes. *Gut* **49**, 589–594.

Czaja, A. J. (2001). Understanding the pathogenesis of autoimmune hepatitis. *Am. J. Gastroenterol.* **96**, 1224–1231.

Czaja, A. J. (1999). Drug therapy in the management of type 1 autoimmune hepatitis. *Drugs* **57**, 49–68.

Czaja, A. J., and Carpenter, H. A. (2001). Autoimmune hepatitis with incidental histologic features of bile duct injury. *Hepatology* **34**, 659–665.

Czaja, A. J., and Carpenter, H. A. (2002). Autoimmune hepatitis. *In* "Pathology of the Liver" (R. N. M. MacSween, P. P. Anthony, P. J. Scheuer, A. D. Burt, B. Portmann, and K. G. Ishak, eds.), 4th Ed., pp. 415–433. Churchill Livingstone, Edinburgh.

Czaja, A. J., Cookson, S., Constantini, P. K., Clare, M., Underhill, J. A., and Donaldson, P. T. (1999). Cytokine polymorphisms associated with clinical features and treatment outcome in type 1 autoimmune hepatitis. *Gastroenterology* **117**, 645–652.

Czaja, A. J., and Donaldson, P. T. (2000). Genetic susceptibilities for immune expression and liver cell injury in autoimmune hepatitis. *Immunol. Rev.* **174**, 250–259.

Czaja, A. J., and Homburger, H. A. (2001). Autoantibodies in liver disease. *Gastroenterology* **120**, 239–249.

Czaja, A. J., Strettell, M. D. J., Thomson, L. J., Santrach, P. J., Moore, S. B., Donaldson, P. T., and Williams, R. (1997). Associations between alleles of the major histocompatibility complex and type 1 autoimmune hepatitis. *Hepatology* **25**, 317–323.

Autonomic Innervation

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

ganglia Clusters of nerve cell bodies outside the central nervous system.

postganglionic neuron A ganglionic neuron that receives synaptic input from the central nervous system and sends its axon to innervate peripheral organs.

preganglionic neuron A neuron that projects from the brain or spinal cord to form synapses with neurons in ganglia.

Innervation of the digestive tract by the autonomic nervous system (ANS) automatically controls contractile behavior of the musculature, secretion and absorption across the mucosal lining, and blood flow inside the walls of the esophagus, stomach, intestines, and gallbladder. Autonomic control is ongoing and normally occurs below the level of conscious perception. Depending on the kind of neurotransmitter released, ANS motor neurons may evoke muscle contraction or actively inhibit contraction. Secretion of water, electrolytes, and mucus into the lumen and absorption from the lumen are determined by motor neurons of the ANS innervation. The amount of blood flow and the distribution of flow between the muscle layers and mucosa are also controlled by ANS nervous activity.

AUTONOMIC INTEGRATIVE CENTERS

Autonomic integrative centers are positioned both in the central nervous system (CNS) and in peripheral locations. The autonomic nervous system (ANS) control centers are hierarchically structured with five basic levels of integrative organization that can be identified (Fig. 1). Level 5 is the enteric nervous system (ENS), which behaves like a local "minibrain" within the gut itself. The next higher level of integrative organization is in the prevertebral ganglia of the sympathetic division of the ANS. Levels 1, 2, and 3 are within the CNS. Sympathetic and parasympathetic divisions of the ANS transmit signals to the digestive tract. The signals originate in central parasympathetic and sympathetic centers represented by Levels 2 and 3 in the medulla oblongata (Fig. 1). The nerves that carry the outflow from these medullary centers are the final common pathways out of the CNS en route to the gut. Level 1 includes higher brain centers that provide input for integrative functions at Levels 2 and 3. The peripheral ANS innervation, which connects the CNS to the gut, is subdivided into three parts consisting of the sympathetic, parasympathetic, and enteric divisions (Fig. 2).

The pathways formed by the sympathetic and parasympathetic divisions represent the extrinsic component of ANS innervation. Neurons of the enteric division form the local intramural control networks and are the intrinsic components of innervation. The parasympathetic and sympathetic divisions are identified by the location of the last neuron in the pathway to the gut and by the point of outflow from the CNS.

PERIPHERAL AUTONOMIC CONNECTIONS

The cell bodies of the last neurons in autonomic pathways to the gut are in structures called ganglia. A

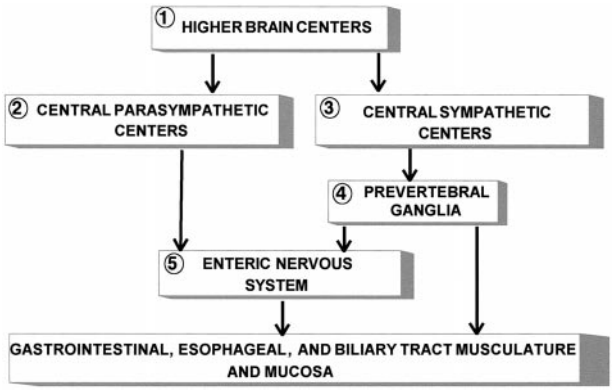


FIGURE 1 Autonomic neural control of the digestive tract is hierarchic with five basic integrative centers positioned as successively higher levels in the nervous system. Level 1 includes higher brain centers that provide input for integrative functions at Levels 2 and 3. Sympathetic and parasympathetic signals originate at Levels 2 and 3 in the brain's medulla oblongata and represent the final common pathways for outflow of information from the central nervous system to the gut. The fourth level of organization is in prevertebral sympathetic ganglia located in the abdomen. Level 5 is the enteric nervous system, which behaves like a local "minibrain."

 Copyright 2004, Elsevier (USA). All rights reserved.

ganglion is defined as a cluster of neuronal cell bodies located outside the ENS. Cell bodies of the next to last neurons in the autonomic pathways are in the CNS.

Sympathetic Division

Figure 3 illustrates the location of the ganglia of the sympathetic nervous system in relation to the spinal cord and ENS. The neurons in the spinal cord are called preganglionic neurons. Preganglionic sympathetic neurons have their cell bodies in the intermediolateral horn of the spinal cord between the first thoracic and the third lumbar spinal segments. The locations and areas of emergence of sympathetic preganglionic neurons from the spinal cord are the basis for sometimes referring to the sympathetic nervous system as the ANS thoraco-lumbar division. Preganglionic sympathetic neurons project their axons out of the spinal cord to form synapses with neurons in ganglia that are located in the abdomen. The ganglia in the abdomen are called prevertebral sympathetic ganglia. There are three prevertebral ganglia. One is the celiac ganglion, the second is the superior mesenteric, and the third is the inferior mesenteric ganglion. Neurons in the prevertebral ganglia are postganglionic neurons. These postganglionic sympathetic neurons form synapses with neurons in the ENS. The primary neurotransmitter released at synapses between pre- and postganglionic neurons is acetylcholine. The main neurotransmitter released at synapses with enteric neurons is norepinephrine.

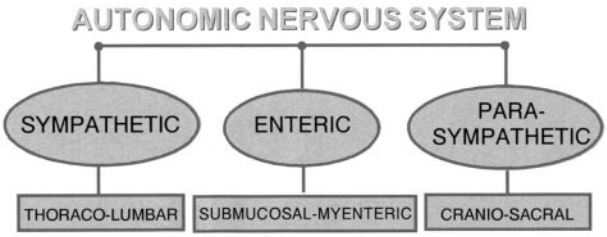


FIGURE 2 Three divisions of the autonomic nervous system innervate the digestive tract. The digestive tract is innervated by the autonomic nervous system and by sensory nerves that project from the gut to the brainstem and spinal cord. The cell bodies of autonomic neurons are found in the brainstem, in the sacral region of the spinal cord, in ganglia within the walls of the digestive tract, and outside the digestive tract in the abdomen. The sympathetic, parasympathetic, and enteric divisions make up the autonomic innervation. Sympathetic and parasympathetic pathways transmit signals from the brain and spinal cord to the gut. This is called the extrinsic component of innervation. Neurons of the enteric division form the local intramural control networks making up the intrinsic component of innervation. The parasympathetic and sympathetic divisions are identified by the positions of the ganglia containing the cell bodies of the second-order postganglionic neurons and by the point of outflow from the brain or spinal cord. Postganglionic neurons of the sympathetic division are located in prevertebral sympathetic ganglia located in the abdomen and the outflow is from the spinal cord between the first thoracic and third lumbar segments. The sympathetic division can therefore be referred to as the thoraco-lumbar division of the autonomic nervous system. Second-order parasympathetic neurons are part of the neural networks that make up the enteric division and the outflow is from the medulla oblongata of the brain and the sacral segments of the spinal cord. Another term for the parasympathetic division is therefore the cranio-sacral division of the autonomic nervous system. The neurons of the enteric division are synaptically interconnected into local neural networks that behave like a "brain-in-the-gut." The cell bodies of the neurons of the enteric division are positioned in the submucosal and myenteric plexuses within the walls of the gut.

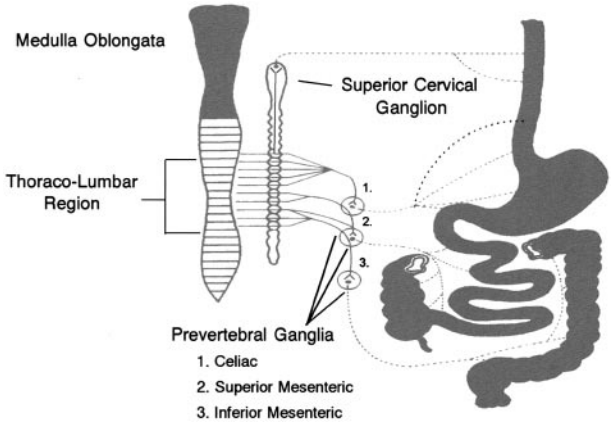


FIGURE 3 Sympathetic neural pathways to the digestive tract consist of preganglionic neurons that have their cell bodies in the spinal cord and postganglionic neurons with their cell bodies located in prevertebral sympathetic ganglia in the abdomen. There are three prevertebral ganglia known as the celiac, superior mesenteric, and inferior mesenteric. There are two synapses in the sympathetic pathways. One is a synapse in the prevertebral ganglia between pre- and postganglionic neurons that uses acetylcholine and nicotinic receptors in neurotransmission. The second is a synapse between postganglionic sympathetic neurons from the sympathetic ganglia and neurons of the enteric nervous system that uses norepinephrine as a neurotransmitter.

Parasympathetic Division

Cell bodies of neurons of the parasympathetic division of the ANS are located either in the brainstem (medulla oblongata) or in the sacral region of the spinal cord. The parasympathetic division is sometimes referred to as the ANS cranio-sacral division, based on the anatomical distribution of the parasympathetic neuronal cell bodies. These neurons project their axons to form synapses with neurons in the ganglia of the ENS (Fig. 4). The terms "preganglionic" and "postganglionic"

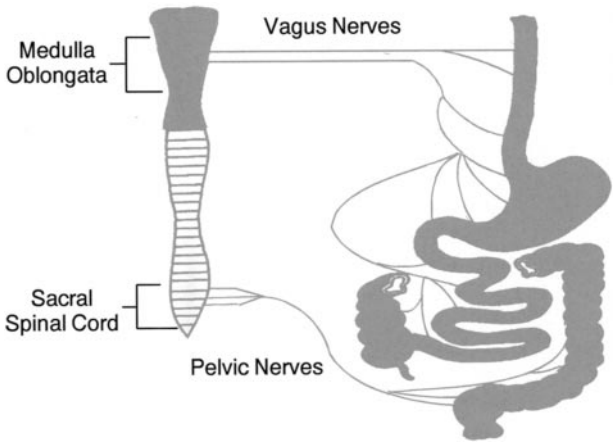


FIGURE 4 Neurons of the autonomic parasympathetic division project from the medulla oblongata and sacral regions of the spinal cord. The parasympathetic division of the autonomic innervation is subdivided anatomically into cranial and sacral divisions due to the neuroanatomic organization in which neurons that project to the digestive tract are located both in the brainstem and in the sacral region of the spinal cord. Neuronal cell bodies of the cranial division reside in the medulla oblongata and project in the vagus nerves. Cell bodies of the sacral division are located in the sacral regions of the spinal cord and project in the pelvic nerves to the large intestine. Efferent vagal fibers form synapses with neurons in the enteric nervous system of the esophagus, stomach, small intestine, colon, gallbladder, and pancreas. Efferent fibers in the pelvic nerves form synapses with neurons in the enteric nervous system.

are no longer used in discussion of the parasympathetic innervation of the gastrointestinal tract. These terms imply that ganglia of the enteric nervous system are like parasympathetic ganglia in other organs and function as simple relay-distribution centers for information from the central nervous system. Ample evidence suggests that the neurophysiology of enteric ganglia is more complex than that of other parasympathetic ganglia that function as simple relay-distribution stations.

ENTERIC DIVISION

The vagus nerves are a major parasympathetic transmission pathway for control signals from the brain to the digestive tract as far down as the proximal to mid large intestine and the sacral nerves represent the major pathway to the distal large intestine. Both vagal and sacral parasympathetic nerves are a source of synaptic input to neurons in the ENS. Earlier concepts portrayed the ganglia of the ENS as a simple relay-distribution function for transmission from the CNS to the gut musculature and secretory glands. This concept has been modernized and changed.

New awareness of the independent integrative properties of the ENS has led to revision of earlier concepts of mechanisms of vagal and sacral nerve influence. Earlier concepts of parasympathetic innervation presumed that ganglia of the digestive tract were the same as parasympathetic ganglia in other visceral systems where the ganglia generally have a relay-distribution function. The previous concepts assumed that parasympathetic innervation of the gut was similar. Preganglionic parasympathetic fibers were believed to form synapses directly with ganglion cells that innervated the muscles and thereby evoke muscle contractions. This concept, illustrated in Fig. 5, is inconsistent with later evidence and has been abandoned.

The earlier concept placed the "computer" entirely within the brain. Current concepts place integrative neural networks in the ENS in close proximity to the motor and secretory systems that require control. Numbers of neurons equal to those of the spinal cord are present in the ENS. The large number of neurons required for program control of the digestive processes would greatly expand the volume of the CNS if situated there. Rather than having the neural control circuits packed exclusively within the CNS and transmitting control information over long transmission lines to the gut, vertebrate animals have most of the neural circuits for automatic feedback control (i.e., setpoint control) located in close apposition to the musculature, secretory glands, and blood vessels.

Figure 5 illustrates the current concept of CNS involvement in gut function. Local integrative circuits of the ENS are organized for program operations independent of input from the CNS. Subsets of neural circuits are preprogrammed for control of distinct patterns of behavior in each effector system (i.e., musculature, glands, and blood vascular system) and for the coordination of activity of multiple systems. Enteric motor neurons are the final common transmission pathways for the variety of different programs and reflex circuits required for ordered gut function.

Rather than controlling a multitude of motor neurons individually, messages transmitted by parasympathetic efferent fibers are command signals for the activation of expanded blocks of integrated circuits positioned in the gut wall. This explains the strong influence of a small number of parasympathetic efferent fibers (only 10% of fibers in the vagus nerves go to the gut and other viscera; 90% of the fibers are sensory afferents) on the musculature and glands over extended

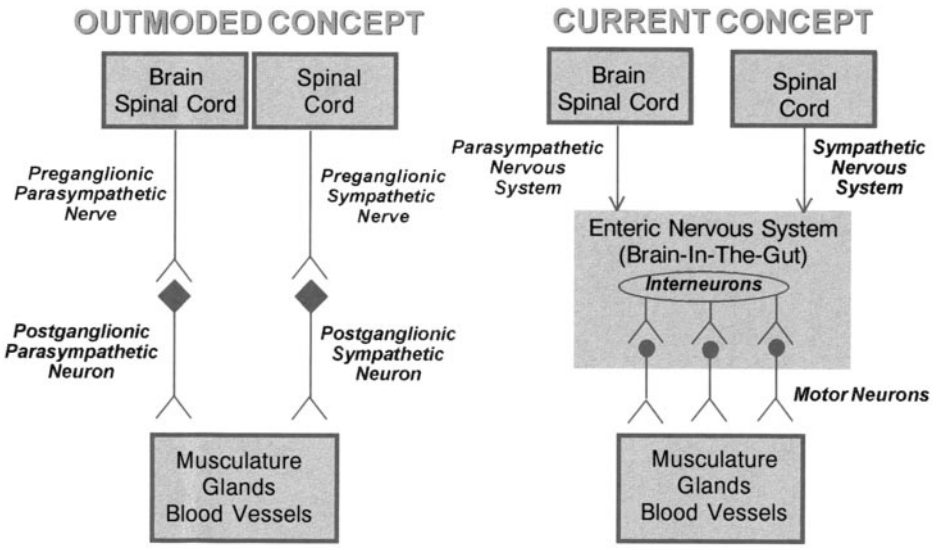


FIGURE 5 Concepts of how the autonomic nervous system controls the behavior of the digestive tract have been changed by improved knowledge of the neurophysiology of the enteric nervous system. The current concept (right) recognizes that the enteric nervous system, like the brain and spinal cord, consists of networks of interneurons and motor neurons with directional flow of neural information from interneurons to motor neurons to the effectors (i.e., the musculature, glands, and blood vessels). Commands to the enteric nervous system from the central nervous system are transmitted by parasympathetic and sympathetic neural pathways. The outmoded concept (left) assumed that all of the integrative circuitry for the minute-to-minute control of the digestive tract resided in the central nervous system and that parasympathetic and sympathetic postganglionic neurons were the motor neurons to the effector systems of the digestive tract.

regions of the stomach or intestine. In this respect, the ENS is analogous to a microcomputer with its own independent software, whereas the brain is like a larger mainframe with extended memory and processing circuits that receive information from and issue commands to the enteric computer.

SENSORY INNERVATION

The ANS is an automatic control system that maintains the many parameters of digestive function at constant levels, sometimes referred to as setpoint determinations. Examples are the determination of constant acidity (i.e., pH setpoint) in the small intestine, constant osmolarity in the small intestine, and a set amount of contractile tension in a given muscle. In order to control a parameter so that its value remains at or close to the setpoint, the integrative neural networks of the ANS must be able to make continuous calculations of the deviation of the actual value of the parameter from the setpoint. Calculation of the deviation from setpoint requires a continuous flow of information on the actual value of the parameter as various perturbations cause the parameter to increase or decrease. Information on parameters such as acidity, osmolarity, and muscle tension is generated by the sensory innervation of the gut. The sensory information is transmitted to the central nervous system by sensory afferent nerves that project to the spinal cord and the brainstem.

Sensory projections to regions of the spinal cord above the sacral region are called splanchnic afferents; the projections to the sacral cord are known as sacral afferents. Splanchnic afferent fibers from the intestines and sympathetic efferent fibers to the intestine occur in the same "mixed" nerves. The mixed nerves accompany the blood vessels in the mesentery and carry two-way traffic between the spinal cord and the gut. A common error of referring to the sensory nerves as "sympathetic afferents" has emerged from the anatomical presence of splanchnic afferents and sympathetic efferents in the same mixed nerves. The sympathetic division of the autonomic nervous system does not include sensory nerves and to apply the term sympathetic afferent is incorrect.

See Also the Following Articles

Brain–Gut Axis • Enteric Nervous System • Parasympathetic Innervation • Sensory Innervation • Sympathetic Innervation • Vagus Nerve

Further Reading

Szurszewski, J., and Miller, S. (1994). Physiology of prevertebral ganglia. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 795–877. Raven Press, New York.

Travagli, R., and Rogers, R. (2001). Receptors and transmission in the brain–gut axis: Potential for novel therapies. V. Fast and slow extrinsic modulation of dorsal vagal complex circuits. *Am. J. Physiol.* **281**, G595–G601.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Bacterial Overgrowth

PHILLIP P. TOSKES
University of Florida College of Medicine

Crohn's disease Chronic disorder characterized by patchy transmural inflammation; may affect any portion of the gastrointestinal tract, but most commonly involves the ileum and colon.
malabsorption Failure to digest and absorb dietary nutrients.
steatorrhea Increased fat in stool due to malabsorption.

The development of malabsorption in a patient with overgrowth of bacteria within the small intestine is known as bacterial overgrowth, a condition that develops when the mixture of bacterial flora of the proximal small intestine becomes more like that of the healthy colon. When colonic-type flora inhabit the proximal small intestine, the bacteria compete for the nutrients ingested by the human host. What ensues is a complex array of clinical problems resulting from intraluminal bacterial catabolism of nutrients, often leading to toxic metabolites and direct injury to the small intestinal enterocyte.

NORMAL ENTERIC FLORA

The proximal small intestine is normally inhabited by a few bacteria, usually lactobacilli, enterococci, gram-positive aerobes, or facultative anaerobes, present in concentrations of up to 10^3 viable organisms per milliliter of jejunal secretions. Qualitative and quantitative changes appear at the ileum and become quite striking in the colon. In the colon, the bacterial population increases up to 1 million times and reaches 10^9–10^{12} bacteria per gram of colonic content. In the colon, in contrast to the proximal small intestine, the anaerobic bacteria outnumber the aerobic bacteria by as much as 10,000 to 1. In bacterial overgrowth, *Bacteroides*, anaerobic lactobacilli, and *Clostridium* are prevalent. Enterobacteria, including coliforms, are abundant in the overgrowth flora.

CLINICAL CONDITIONS ASSOCIATED WITH BACTERIAL OVERGROWTH

Table I lists the recognized clinical conditions associated with bacterial overgrowth. In the past, bacterial overgrowth was usually associated with structural abnormalities of the gastrointestinal tract, such as Billroth II anastomosis, Crohn's disease, stagnant loops of intestine resulting from fistulae or surgical enterostomies, and multiple duodenal and/or jejunal diverticula. It became appreciated that obstruction of the small intestine caused by Crohn's disease, adhesions, radiation damage, lymphoma, or tuberculosis may lead to small bowel bacterial overgrowth. In those conditions, although a primary disease may lead to malabsorption, superimposed overgrowth may be the most treatable form of malabsorption. Currently, motility disturbances in the gastrointestinal tract are the most important settings for bacterial overgrowth to occur. This is particularly important if this dysmotility syndrome is associated with hypo- or achlorhydria. In this category, diseases such as scleroderma, intestinal pseudo-obstruction, and diabetic autonomic neuropathy are found.

Certain individuals have been demonstrated to have clinically important malabsorption associated with bacterial overgrowth, but no radiographic abnormalities. Some of these individuals have an absent or disordered migrating motor complex. Their intestinal "housekeeper" is not operating appropriately. Elderly patients may develop malabsorption secondary to bacterial overgrowth and many believe that bacterial overgrowth is the most frequent cause of clinically important malabsorption in the elderly. The elderly are at risk for bacterial overgrowth because of their often-associated motility disturbances, either from the aging gut or from previous gastrointestinal surgery, and the decreased acid secretion that many elderly patients manifest. The importance of having normal motility and appropriate acid secretion is underscored by the observation that patients with scleroderma, who are manifesting esophageal reflux and doing well on histamine-2 (H2) receptor antagonists, have developed severe malabsorption secondary to bacterial overgrowth when the H2 receptor antagonist was replaced with a proton pump inhibitor. Patients with chronic pancreatitis may have malabsorption secondary to bacterial overgrowth because they manifest hypomotility as a result of pain, the use of narcotics, and previous surgery. Management

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Clinical Conditions Associated with Bacterial Overgrowth

Site	Associated clinical condition
Gastric proliferation	Hypochlorhydria or achlorhydria, especially when combined with motor or anatomical disturbances
	Sustained hypochlorhydria induced by proton pump inhibitor
Small intestinal stagnation	
Anatomical	Afferent loop of Billroth II partial gastrectomy
	Duodenal–jejunal diverticulosis
	Surgical blind loop (end-to-side anastomosis)
	Surgical recirculating loop (end-to side anastomosis)
	Ileal anal pouch
	Obstruction (stricture, adhesion, inflammation, neoplasm)
Motor	Scleroderma
	Idiopathic intestinal pseudo-obstruction
	Absent or disordered migrating motor complex
	Diabetic autonomic neuropathy
Abnormal communication between proximal and distal gastrointestinal tract	Gastrocolic or jejunocolic fistula
	Resection of diseased ileocaecal valve
Miscellaneous	Chronic pancreatitis
	Immunodeficiency syndromes
	Cirrhosis

Modified with permission from Toskes, P. and Kumar, K. (1998). Enteric bacterial flora and bacterial overgrowth syndrome. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed., pp. 1523–1535. W. B. Saunders, Philadelphia, PA.

of such patients may be quite difficult unless pancreatic enzymes are given along with antibiotic therapy to treat the maldigestion of the chronic pancreatitis and the associated overgrowth.

Further observations have also confirmed that an appreciable number of patients with irritable bowel syndrome may develop bacterial overgrowth. Just how frequently this occurs is still being defined and is rather controversial. Several other clinical entities listed in Table I are associated with bacterial overgrowth. Their pathogenesis in respect to the overgrowth is ill understood. These entities include end-stage renal disease, cirrhosis, myotonic muscular dystrophy, fibromyalgia, chronic fatigue syndrome, and various immunodeficiency syndromes such as chronic lymphocytic leukemia, immunoglobulin deficiencies, and selected T cell deficiency.

CLINICAL FEATURES OF BACTERIAL OVERGROWTH

The clinical features noted in patients with bacterial overgrowth are listed in Table II. it is now rather common that the presenting symptoms of a patient with bacterial overgrowth are often very nonspecific. Diagnosis should not be delayed until cobalamin malabsorption or steatorrhea is present. Again, it is important to underscore the observation that bacterial overgrowth may be superimposed on a number of common clinical conditions that may be primary causes of malabsorption, but the overgrowth that is present may be the most easily treatable part of the patient's malabsorption. The clinical conditions that comprise dysmotility syndromes usually do not present with malabsorption due to the dysmotility per se but will have malabsorption, not infrequently, due to superimposed bacterial overgrowth. These would include patients with gastroparesis, irritable bowel syndrome, diabetes, and scleroderma. Weight loss associated with clinically apparent steatorrhea has been observed in about one-third of patients with bacterial overgrowth severe enough to cause cobalamin deficiency. Osteomalacia, vitamin K deficiency, night blindness, hypocalcemic tetany, and vitamin E deficiency may ensue.

TABLE II Clinical Features of Bacterial Overgrowth

Bloating
Abdominal distension
Abdominal pain
Diarrhea
Steatorrhea
Decreased urinary xylose excretion
Hypoalbuminemia
Cobalamin (vitamin B_{12}) deficiency

PATHOGENESIS OF METABOLIC ABNORMALITIES ASSOCIATED WITH BACTERIAL OVERGROWTH

The malabsorption that is observed in patients with bacterial overgrowth results from an abnormal intraluminal catabolism of substrates by the bacterial flora and direct injury to the small intestinal enterocyte induced by the overgrowth flora. A patchy, intestinal mucosal lesion has been demonstrated in both experimental animals and human subjects with bacterial overgrowth; at times, this overgrowth can obliterate the intestinal villi and lead to a flat biopsy. Steatorrhea in this condition may be due to bacterial alteration of bile salts, which leads to a decrease in micelle formation. The accumulation of toxic concentrations of free bile acids may also contribute to the steatorrhea by inducing a patchy intestinal mucosal lesion. The anemia of bacterial overgrowth is primarily a cobalamin deficiency. The anemia is megaloblastic and serum cobalamin levels are low. Patients may develop neurological abnormalities, both central and peripheral, from the cobalamin deficiency. The anemia can be corrected by the administration of cobalamin. The cobalamin deficiency is primarily due to gram-negative anaerobes. Iron deficiency may also occur in bacterial overgrowth due to blood loss resulting from the patchy ulcerated areas. Thus, in some patients, there may be two populations of red cells; those that are macrocytic and those that are microcytic. Folate deficiency, however, is not a common occurrence in bacterial overgrowth. Indeed, patients with bacterial overgrowth may manifest high levels of folate because the overgrowth flora can synthesize folate. Low levels of albumin in the serum occur and occasionally this is severe enough to lead to edema. The reasons for the hypoalbuminemia are multifactorial but include decreased uptake of amino acids by the damaged small intestinal enterocyte, intraluminal breakdown of protein and protein precursors by the bacteria of the overgrowth flora, and a protein-losing enteropathy. It has been appreciated for a long time that as many as 60% of patients with bacterial overgrowth may have a decreased urinary xylose excretion. The primary reason for the decreased urinary xylose excretion is catabolism of the xylose by the intraluminal bacterial flora. Diarrhea in the overgrowth state can come from the production of organic acids caused by the overgrowth flora, leading to an increased osmolarity of the small intestine and decreased intraluminal pH. Bacterial metabolites such as free bile acids, hydroxy fatty acids, and organic acids can stimulate water secretion and electrolytes into the lumen.

DIAGNOSIS OF BACTERIAL OVERGROWTH

In any patient who presents with unexplained diarrhea, steatorrhea, macrocytic anemia, or weight loss, bacterial overgrowth should be suspected. This suspicion should be very great if the patient is elderly or has had a previous abdominal surgery. At most medical centers, the most common causes of clinically important malabsorption are bacterial overgrowth or chronic pancreatitis. Figure 1 speaks to that issue. It presents an algorithm for the evaluation of patients with malabsorption, including those with bacterial overgrowth. The algorithm emphasizes the use of noninvasive and inexpensive tests. An attempt should be made to document the presence of steatorrhea. If the patient has clinically significant bacterial overgrowth, cobalamin absorption is usually impaired, even though the patient may not have yet developed cobalamin deficiency. Intrinsic factor administration will not improve the cobalamin malabsorption in patients with bacterial overgrowth. The urinary excretion of xylose is often decreased, in contrast to patients with pancreatic steatorrhea, for whom the urinary excretion of xylose is usually normal. It is remarkable that the serum folate level may be increased in some but not all patients with bacterial overgrowth.

The definitive diagnosis of bacterial overgrowth requires a properly collected and appropriately cultured aspirate from the proximal small intestine. This specimen should be collected under anaerobic conditions, serially diluted, and cultured on selected media. In patients with bacterial overgrowth, the total concentration of bacteria usually exceeds 10^5 organisms/ml of the jejunal secretions. However, it is important to stress that patients are now presenting with symptoms of bloating, abdominal distension, and pain with bacterial levels as low as 10^4/ml of jejunal secretions. Usually, these patients do not have cobalamin malabsorption or steatorrhea, compared to patients with higher amounts of organisms. Qualitatively, the organisms are colonic-like flora, i.e., *Bacteroides*, anaerobic lactobacilli, coliforms, and enterococci. Intestinal cultures, properly done, are time consuming, uncomfortable for the patient, and expensive. Consequently, they are not usually done in clinical practice. Therefore, a variety of surrogate tests for detecting bacterial overgrowth have been devised based on the various metabolic actions of the bacteria within the overgrowth flora. Table III lists various tests done worldwide and reports on their ease of performance, sensitivity, specificity, and safety. A number of these tests cannot distinguish malabsorption from other causes from that caused by bacterial overgrowth. A 1-g ^{14}C-labeled xylose breath test has been found to be

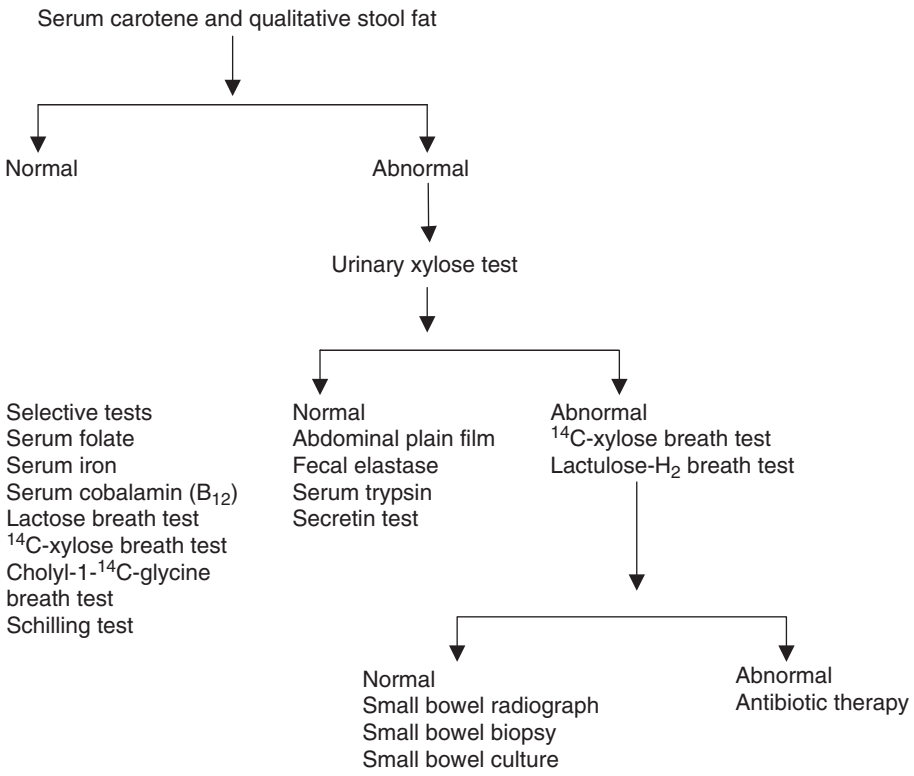


FIGURE 1 Algorithm for evaluation of malabsorption. From Toskes, P.P. (2002), with permission from Lippincott Williams & Wilkins.

a sensitive and specific test for detecting the presence of bacterial overgrowth. When this test is evaluated in comparison to a properly performed intestinal culture, the reliability is around 90%. It is not recommended that the ^{14}C-labeled xylose be used as a substrate in the diagnosis of bacterial overgrowth in children. Therefore, nonradioactive substrates such as ^{13}C-labeled sorbitol have been developed and appear to be excellent tests for detecting bacterial overgrowth while affording no radiation risk to the human host. These ^{13}C-labeled substrates have not yet been approved for clinical use. Analysis of breath hydrogen following the administration of various carbohydrate substrates has been employed as a test for overgrowth, but this test suffers greatly from a lack of sensitivity and specificity. It is noteworthy that an elevated fasting level of hydrogen is present in up to 30% of people with bacterial overgrowth, and if found, can be used to make the diagnosis.

MANAGEMENT OF BACTERIAL OVERGROWTH

The aim of therapy is to correct, when appropriate, the cause of the stasis, but surgery is often impractical for patients with scleroderma, multiple diverticula, diabetes, intestinal pseudo-obstruction, etc. Thus, antimicrobial therapy is the cornerstone of treatments. Remarkable improvements can be achieved in most patients such that they have a cessation of diarrhea and steatorrhea, and gain weight, leading to a better quality of life. Sensitivities of a culture obtained from the proximal intestine for selection of the proper therapy are not employed because of the many different bacterial species present, which often with very different antimicrobial sensitivities. Thus, it is important to select an antimicrobial agent that will be effective against the two main groups of bacteria responsible for the abnormalities observed in bacterial overgrowth. Antibiotics that have been shown to be effective, either by controlled trials or extensive clinical practice, against both the aerobic and anaerobic enteric bacteria are shown in Table IV. Antibiotics that are known to have poor activity against anaerobes should not be utilized in treating bacterial overgrowth. Such antibiotics include penicillin, ampicillin, the oral aminoglycosides, kanamycin, and neomycin. In most patients, a single 10-day course of therapy will markedly improve symptoms and the patient may remain symptom free for a considerable time. In others, the symptoms recur quickly after treatment and acceptable results can be obtained only

TABLE III Diagnostic Tests for Bacterial Overgrowth

Tests	Ease of performance	Sensitivity	Specificity	Safety
Culture	Poor	Excellent	Excellent	Good
Urinary indican	Good	Poor	Poor	Excellent
Jejunal fatty acids	Poor	Fair	Excellent	Good
Jejunal bile acids	Poor	Fair	Excellent	Good
Fasting breath H_2	Excellent	Poor	Excellent	Excellent
^{14}C-Labeled bile acid breath test	Excellent	Fair	Fair	Good
^{14}C-Labeled xylose breath test	Excellent	Excellent	Excellent	Good
Lactulose-H_2 breath test	Excellent	Fair	Fair	Excellent
Glucose-H_2 breath test	Excellent	Good	Fair	Excellent

Modified with permission from Toskes, P. P., and Kumar, K. (1998). Enteric bacterial flora and bacterial overgrowth syndrome. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed., pp. 1523–1535. W. B. Saunders, Philadelphia, PA.

with cyclic therapy, such as 10 days out of each month. In still others, continuous therapy may be needed for up to 30 days. If the antimicrobial agent is effective, there will be a resolution or marked diminution of symptoms within 7 days. Diarrhea and steatorrhea will decrease and cobalamin malabsorption will be corrected.

Prolonged antibiotic therapy poses potential clinical problems, including diarrhea, enterocolitis, patient intolerance, and bacterial resistance. A prokinetic agent that could help clear the intestine of the overgrowth flora would be most advantageous. Salutary results have been obtained with both cisapride and octreotide (50 μg, nightly) in a small number of studies. Hydroxytryptamine (HT3 and HT4) antagonists are available to use in patients with dysmotility syndromes. The value of these agents in patients with overgrowth remains to be defined, but their use is provocative. Other recommendations for managing the patient with bacterial overgrowth include supplementation with cobalamin, calcium, fat-soluble vitamins, and iron. A lactose-free diet is suggested because the patients often develop lactase deficiency secondary to mucosal injury, and this may linger on even though the patient's antimicrobial therapy has decreased most of their symptoms. The use of medium-chain triglycerides instead of long-chain triglycerides is recommended because the medium-chain triglycerides, in contrast to the long-chain triglycerides, are not rendered ineffective by the bile salt-deconjugating activity of the overgrowth flora. The use of probiotics for the treatment of bacterial overgrowth has been disappointing. Controlled trials comparing antibiotic therapy to treatment with *Saccharomyces boulardii* have demonstrated the effectiveness of antibiotics but the ineffectiveness of the administered probiotic supplements.

The clinician should have a very low threshold for suspecting bacterial overgrowth as a cause of malabsorption, because this condition can be easily diagnosed and well treated. An appropriate attempt should be made to document whether bacterial overgrowth is present and then to administer proper therapy.

TABLE IV Therapy for Bacterial Overgrowth

Antimicrobial agent	Dose/day (10-day course) [a]
Tetracycline	250 mg qid
Doxycycline	100 mg bid
Minocycline	100 mg bid
Amoxicillin–clavulanic acid	875 mg bid
Cephalexin+metronidazole	250 mg qid and 250 mg tid
Trimethoprim–sulfamethoxazole	One double-strength tab bid
Ciprofloxacin	500 mg bid
Norfloxacin	400 mg bid
Chloramphenicol	250 mg qid

[a] Abbreviations: qid, quarter in die (four times/day); bid, bis in die (twice/day); tid, ter in die (three times/day).

See Also the Following Articles

Breath Tests • Carbohydrate and Lactose Malabsorption • Cobalamin Deficiency • Crohn's Disease • Gastric Motility • Intestinal Pseudoobstruction • Malabsorption • Microflora, Overview • Migrating Motor Complex

Further Reading

Bouhnik, Y., Alain, S., Attar, A., *et al.* (1999). Bacterial populations contaminate the upper gut in patients with small intestinal bacterial overgrowth syndrome. *Am. J. Gastroenterol.* **94**, 1327–1331.

Gregg, C. R., and Toskes, P. P. (2002). Enteric bacterial flora and small bowel bacterial overgrowth syndrome. *In* "Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. S. Sleisenger, eds.), 7th Ed., Vol. 2, pp. 1783–1793. W. B. Saunders, Philadelphia, PA.

Meyers, J. S., Ehrenpreis, E. D., and Craig, R. M. (2001). Small intestinal bacterial overgrowth syndrome. *Curr. Treat. Opt. Gastroenterol.* **4**, 7–14.

Toskes, P. P. (2002). Small intestine bacterial overgrowth, including blind loop syndrome. *In* "Infections of the Gastrointestinal Tract" (M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant, eds.), 2nd Ed., pp. 291–300. Lippincott Williams & Wilkins, Philadelphia, PA.

Toskes, P. P. (2003). Small bowel bacterial overgrowth. *In* "Oxford Textbook of Medicine" (D. A. Warrell, T. M. Cox, J. D. Firth, and E. J. Benz, eds.), 4th Ed., Vol. 2, pp. 580–584.

Toskes, P. P., and Kumar, K. (1998). Enteric bacterial flora and bacterial overgrowth syndrome. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed., pp. 1523–1535. W. B. Saunders, Philadelphia, PA.

Bacterial Toxins

MING L. CHEN AND CHARALABOS POTHOULAKIS
Beth Israel Deaconess Medical Center and Harvard Medical School

cytotoxins Virulence factors secreted or released by pathogenic bacteria, causing cytoskeletal destruction of target cells.

enterochromaffin cells Largest endocrine cell types in the gastrointestinal tract; dispersed among mucosal epithelial cells. The cytoplasm of enterochromaffin cells is filled with a large number of secretory granules, characteristic of endocrine cells.

enterotoxins Virulence factors secreted or released by pathogenic bacteria, causing fluid accumulation in closed intestinal or colonic loops of experimental animals.

G proteins Heterotrimeric GTP-binding proteins composed of α, β, and γ subunits; usually coupled with cell surface receptors containing seven membrane-spanning domains. Agonist binding to these receptors stimulates the exchange of GDP for GTP on the G_α subunit, resulting in the dissociation of G_α from β and γ subunits. These subunits separately activate their downstream effectors, including adenylyl cyclases, phospholipases, and ion channels.

lipid rafts Plasma membrane microdomains enriched with cholesterol and glycosphingolipids; serve as entry routes for bacterial toxins in target cells.

tight junctions Cell–cell contact sites localized in the uppermost region of polarized intestinal epithelium; regulate paracellular flux of ion and solutes and consist of fibril proteins, claudins, and occludin, and peripheral proteins such as zonula occludens.

Bacterial toxins derived from enteric pathogens participate in the pathophysiology of intestinal infections; enterotoxins trigger intestinal fluid secretion and inflammation, cytotoxins cause cytopathogenic effects, and further damage involves stimulating host immune responses. Toxins are enzymes that can biochemically modify specific molecules on the plasma membrane or cytosol of target cells. They are secreted or released by pathogenic bacteria and often utilize existing host cellular pathways to cross the plasma membrane barrier. Enterotoxins traffic to subcellular organelles of intestinal epithelial cells, where they can be proteolytically processed, leading to the release of enzymatically active toxin fragments. These toxin fragments are able to alter specific host cell signaling cascades, leading to cell injury and inflammation. Some bacterial toxins share structural similarities in their enzymatically active toxin domains, which are different from the receptor-binding domains that are responsible for toxin internalization and trafficking to the target site. Various molecular mechanisms involving the receptors, enzymatic activity, and cell targets play a role in the pathophysiology of the most common bacterial toxins that affect the gastrointestinal tract (Table I).

VIBRIO CHOLERAE TOXINS

Vibrio cholerae toxin is the major virulence factor responsible for cholera epidemics, representing a major public health burden in developing countries. Patients with cholera experience voluminous loss of salt and

Copyright 2004, Elsevier (USA). All rights reserved.

Toskes, P. P. (2002). Small intestine bacterial overgrowth, including blind loop syndrome. *In* "Infections of the Gastrointestinal Tract" (M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant, eds.), 2nd Ed., pp. 291–300. Lippincott Williams & Wilkins, Philadelphia, PA.

Toskes, P. P. (2003). Small bowel bacterial overgrowth. *In* "Oxford Textbook of Medicine" (D. A. Warrell, T. M. Cox, J. D. Firth, and E. J. Benz, eds.), 4th Ed., Vol. 2, pp. 580–584.

Toskes, P. P., and Kumar, K. (1998). Enteric bacterial flora and bacterial overgrowth syndrome. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed., pp. 1523–1535. W. B. Saunders, Philadelphia, PA.

Bacterial Toxins

MING L. CHEN AND CHARALABOS POTHOULAKIS
Beth Israel Deaconess Medical Center and Harvard Medical School

cytotoxins Virulence factors secreted or released by pathogenic bacteria, causing cytoskeletal destruction of target cells.

enterochromaffin cells Largest endocrine cell types in the gastrointestinal tract; dispersed among mucosal epithelial cells. The cytoplasm of enterochromaffin cells is filled with a large number of secretory granules, characteristic of endocrine cells.

enterotoxins Virulence factors secreted or released by pathogenic bacteria, causing fluid accumulation in closed intestinal or colonic loops of experimental animals.

G proteins Heterotrimeric GTP-binding proteins composed of α, β, and γ subunits; usually coupled with cell surface receptors containing seven membrane-spanning domains. Agonist binding to these receptors stimulates the exchange of GDP for GTP on the G_α subunit, resulting in the dissociation of G_α from β and γ subunits. These subunits separately activate their downstream effectors, including adenylyl cyclases, phospholipases, and ion channels.

lipid rafts Plasma membrane microdomains enriched with cholesterol and glycosphingolipids; serve as entry routes for bacterial toxins in target cells.

tight junctions Cell–cell contact sites localized in the uppermost region of polarized intestinal epithelium; regulate paracellular flux of ion and solutes and consist of fibril proteins, claudins, and occludin, and peripheral proteins such as zonula occludens.

Bacterial toxins derived from enteric pathogens participate in the pathophysiology of intestinal infections; enterotoxins trigger intestinal fluid secretion and inflammation, cytotoxins cause cytopathogenic effects, and further damage involves stimulating host immune responses. Toxins are enzymes that can biochemically modify specific molecules on the plasma membrane or cytosol of target cells. They are secreted or released by pathogenic bacteria and often utilize existing host cellular pathways to cross the plasma membrane barrier. Enterotoxins traffic to subcellular organelles of intestinal epithelial cells, where they can be proteolytically processed, leading to the release of enzymatically active toxin fragments. These toxin fragments are able to alter specific host cell signaling cascades, leading to cell injury and inflammation. Some bacterial toxins share structural similarities in their enzymatically active toxin domains, which are different from the receptor-binding domains that are responsible for toxin internalization and trafficking to the target site. Various molecular mechanisms involving the receptors, enzymatic activity, and cell targets play a role in the pathophysiology of the most common bacterial toxins that affect the gastrointestinal tract (Table I).

VIBRIO CHOLERAE TOXINS

Vibrio cholerae toxin is the major virulence factor responsible for cholera epidemics, representing a major public health burden in developing countries. Patients with cholera experience voluminous loss of salt and

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Receptor, Enzymatic Activity, and Cellular Target of the Most Common Bacterial Toxins Affecting the Intestine

Enterotoxin	Receptor	Enzymatic activity	Cellular target
Vibrio cholerae toxin	GM1 ganglioside	ADP-ribosyltransferase	Heterotrimeric G protein ($G_{\alpha s}$)
Zonula occludens toxin	Unknown	Unknown	Phospholipase C/PKC-α
Heat-labile enterotoxins	GD1b ganglioside	ADP-ribosyltransferase	Heterotrimeric G protein ($G_{\alpha s}$)
Heat-stable enterotoxins	Unknown	Unknown	Receptor-coupled guanylyl cyclase
Shiga and shiga-like toxins	Gb_3 ceramide	*N*-Glycosidase	28S RNA
Clostridia toxins A and B	Unknown	Glucosyltransferase	Rho GTPases
Bacteroides fragilis enterotoxin	Unknown	Zinc metalloprotease	E-Cadherin

fluid that may lead to metabolic acidosis and death. Cholera toxin (CT) is encoded in the genome of a filamentous phage (CTXϕ) that is integrated into one of the two bacterial chromosomes. CT, which belongs to the AB_5 type family of toxins, consists of one A subunit, containing A1 and A2 peptide chains linked by a disulfide bond, and five identical B subunits. The AB_5 family of toxins also includes the *Escherichia coli* heat-labile enterotoxin and shiga and pertussis toxins. The binding domain, or the B subunit, of CT forms a pentamer and binds stoichiometrically to cell surface glycolipid receptor GM1 ganglioside. The internalization of CT into target cells is mediated via a lipid raft-dependent mechanism. CT, which is transcytosed in membrane vesicles pinched off form the plasma membrane, passes through the Golgi cisternae and enters the endoplasmic reticulum, where the holotoxin is unfolded. A luminal chaperone protein, protein disulfide isomerase, aids in the unfolding of CT and in the subsequent proteolytic release of its A1 chain. The A1 chain exits the endoplasmic reticulum through a protein-conducting channel (Sec61p) and travels back to the plasma membrane. This enzymatically active toxin peptide modifies (by adenosine diphosphate ribosylation) the heterotrimeric G protein, $G_{\alpha s}$, resulting in activation of adenylyl cyclase and increased cyclic adenosine monophosphate (cAMP) levels. This leads to the opening of chloride channels in intestinal epithelial cells and loss of salt and water, a major characteristic of cholera diarrhea.

Physiology studies in whole animals demonstrate that in addition to its cAMP stimulatory mechanism, CT-mediated intestinal fluid secretion *in vivo* may involve extensive interactions between enterochromaffin cells, intestinal nerves, and epithelial cells. Evidence indicates that CT binds to enterochromaffin and intestinal epithelial cells in the intestinal mucosa and stimulates release of serotonin and prostaglandins, respectively. These molecules activate neurons in the intestinal submucosa region, leading to release of substances that interact with crypt cells, i.e., acetylcholine, substance P, and vasoactive intestinal peptide (VIP), resulting in secretion of chloride and water. Thus, pharmacologic blockade of CT-associated neuroendocrine pathways may have a place in the treatment of cholera diarrhea.

In addition to the stimulation of secretory pathways, CT also induces the release from epithelial cells of several antiinflammatory cytokines, such as IL-1 antagonist, IL-6, and IL-10, and inhibits antigen presentation by macrophages. This is believed to increase the virulence potential of CT by compromising the host's immune defense. The binding subunit of CT also modulates host immune responses and includes the depletion of CD8+ T cells, alterations of CD4+ T-cell differentiation, activation of B cells, and promotion of antigen processing and presentation by macrophages. The potent capacity of the CT B subunit to modulate immune responses suggests its potential application in the treatment of inflammatory autoimmune diseases and/or as an adjuvant for mucosal or systemic delivery of specific antigens.

Zonula occludens toxin (Zot) is an enterotoxin produced by *V. cholerae*. Zot, like CT, is encoded in the genome of filamentous phage CTXϕ integrated into one of the two chromosomes of *V. cholerae*. Newly synthesized Zot (45 kDa) is transported across the inner membrane of the bacterium and undergoes proteolytic cleavage to an N-terminal polypeptide (33 KDa), and to a small C-terminal domain (12 KDa). The N terminus may aid in the assembly of the CTXϕ phage whereas the C terminus induces paracellular permeability changes in intestinal epithelial cells. Zot binding to cell surface receptors decreases along the axis of the intestinal tract and correlates with its effect on intestinal permeability, which is more pronounced in the jejunum and ileum and absent in colon. After binding, Zot is internalized and activates the signaling molecule phospholipase C. This leads to production of inositol 1,4,5-trisphosphates and diacylglycerol and activation of protein kinase C-α (PKC-α), leading

to actin polymerization and opening of epithelial tight junctions. Thus, both Zot and CT may be responsible for the diarrheal effects of *V. cholerae*.

Escherichia coli TOXINS

Heat-labile enterotoxin (LT) has been isolated from enterotoxigenic *E. coli* (ETEC), representing the most frequent cause of traveler's diarrhea worldwide. LT is highly homologous to CT, and the tertiary structures of these two toxins are almost superimposed. The mode of LT action is also similar to that of CT in that the A chain of LT possesses adenosine diphosphate (ADP)-ribosyltransferase activity against the $G_{\alpha s}$ substrate. The loss of GTPase activity results in activation of adenylyl cyclase followed by massive loss of fluid and solutes. The LT-mediated fluid secretion involves secretomotor neurons, but, in contrast to CT-mediated fluid secretion, serotonin and substance P do not appear to be involved in LT-induced intestinal fluid secretion. Moreover, intestinal diarrhea induced by LT is less severe than CT-induced diarrhea, possibly because of different intestinal cell receptors for these two toxins. Several LT variants (LTI, LTIIa, and LTIIb) have been isolated from different ETEC serogroups. LTIIa binds to ganglioside GD1b with high affinity, but it also binds to the gangliosides GD1a, GT1b, GQ1b, and GM1 with lower affinity. Binding of LTIIa to the CT receptor GM1 fails to produce CT-like effects in cultured colonocytes. In contrast, LTIIa binds to lipid raft-associated GD1b and mediates trafficking of the toxin into host cells, eliciting a chloride secretory response.

Heat-stable enterotoxins (STs) represent a family of cysteine-rich peptides that bind to specific intestinal receptors and cause diarrhea. STa and STb are two subfamilies of STs that display structural and immunological differences. STa consists of toxins secreted by enterotoxigenic *E. coli, V. cholerae* non-O1 strains, *Yersinia enterocolitica*, and *Citrobacter freundii*. STa from ETEC is synthesized as a 72-amino-acid precursor that is proteolytically cleaved to a mature C-terminal 23-amino-acid peptide. The conserved "toxic domain" of mature STa contains six cysteine residues, forming intramolecular disulfide bonds that are essential for the tertiary structure and the biological activity of toxin. STa binds to receptors belonging to the family of natriuretic peptide receptors characterized by an extracellular ligand binding site and a cytoplasmic guanylyl cyclase domain. Activation of guanylyl cyclase C increases intracellular cyclic GMP and cyclic GMP-dependent protein kinase activity, leading to the phosphorylation and opening of the cystic fibrosis transmembrane conductance regulator (CFTR), a chloride channel. This results in the net efflux of ions and water into the intestinal lumen.

ETEC STb contains 48 amino acids and resides in the C-terminal end of a precursor protein. The four conserved cysteine residues in STb form two disulfide bonds that are essential for toxicity. STb does not alter cAMP or cGMP levels in the intestinal mucosa. However, injection of STb into rabbit intestinal loops increases secretion of serotonin and prostaglandin E2, known intestinal secretagogues that may mediate STb-induced diarrhea.

Shigella dysenteriae TOXINS

Shiga and shiga-like toxins are produced by *Shigella dysenteriae* (type 1) and some strains of enterohemorrhagic *E. coli* (EHEC) O157:H7. Humans are often infected by these pathogens after ingestion of contaminated food and develop symptoms ranging from mild gastroenteritis to hemorrhagic colitis and hemolytic uremic syndrome. The amino acid sequence of shiga toxin is almost identical to that of shiga-like toxin-1, but shares only 56% sequence homology with shiga-like toxin-2. These toxins are members of the AB_5 family that include one enzymatic (A) and five binding (B) peptide chains. The B subunit binds with high binding affinity to its glycolipid receptor globotriasoyl ceramide (Gb_3), which is highly expressed in the gastrointestinal tract, kidney, and brain, reflecting the ability of the toxin to affect these organs. The holotoxin is internalized from the apical side of host cells via endocytosis, and travels in a retrograde fashion to the endoplasmic reticulum. The A polypeptide is then released into the cytosol whereas the B polypeptide may be delivered to the basolateral membrane and released into the subepithelial layer. The A chain is proteolytically cleaved to the A1 and A2 peptides. The cytosolic target of the A1 chain is the 28S RNA of the 60S ribosomal unit. The specific cleavage of an N-glycosidic bond in the 28S RNA by the A1 chain blocks the binding of aminoacyl-tRNA to its acceptor site, resulting in inhibition of cellular protein synthesis and cell death.

Shigella dysenteriae invades the intestinal mucosal epithelium and induces severe inflammation in the ileum and colon. Shiga toxin may penetrate the intestinal epithelial barrier either by transcytosis or via a paracellular route that can be compromised by epithelial cell damage during inflammation. On entering the intestinal microcirculation, shiga toxin affects vascular endothelial cells and causes vascular thrombosis. Subsequently, the toxin travels to the kidney via the systemic circulation and causes severe histopathological

kidney changes known to be associated with hemolytic uremic syndrome.

Clostridium difficile TOXINS

Clostridium difficile mediates diarrhea and pseudomembranous colitis in some patients receiving antibiotic therapy. *Clostridium difficile* is the most frequently diagnosed cause of infectious diarrhea in hospitalized patients and is associated with significant morbidity and mortality. This bacterium releases two exotoxins, A and B, which belong to the large clostridial toxin family that includes *Clostridium sordellii* and *Clostridium novii* toxins. Clostridial toxins have large molecular masses (>250 kDa), demonstrate extensive amino acid similarities, and are characterized by absence of subunits and similar cytopathogenic (cell rounding) effects on cultured cells. *Clostridium difficile* toxins A and B share 49% amino acid sequence homology and overall structural similarity. The cytotoxic effects of toxins A and B involve modification of the Ras superfamily small GTP-binding proteins that regulate cellular actin. The N-terminal domains of toxins A and B possess glucosyltransferase activity that catalyzes the transfer of a glucose moiety from UDP-glucose to the Rho family of the small GTP-binding proteins RhoA, Rac, and Cdc42, the major regulators of cellular actin dynamics. Glucosylated RhoA exhibits reduced intrinsic GTPase activity and loses its ability to couple with its downstream effectors, leading to actin depolymerization and cell death. The C-terminal receptor-binding domain contains several repetitive oligopeptide subunits that, in animal intestine, are responsible for toxin A binding to enterocyte receptors containing terminal α-galactose epitopes. The human intestinal receptor for these toxins, however, has yet to be identified.

Both toxins A and B can damage colonic epithelial cells and increase paracellular flux in human colonic explants. However, only toxin A has enterotoxic activities in animal intestine. Toxin A-induced paracellular permeability changes are mediated by a protein kinase C α/β-dependent pathway and involve translocation of the proteins ZO-1, occludin, and claudins from the tight junction to the cytoplasmic compartment. The *in vivo* pathophysiology of toxin A-induced inflammation involves activation of immune cells and neurons and secretion of proinflammatory cytokines from colonic epithelial and lamina propria cells. Pharmacologic blockade of receptors for the neuropeptides substance P, calcitonin gene-related peptide, and neurotensin inhibits toxin A-induced intestinal inflammation and fluid secretion, suggesting participation of intestinal neurons in toxin A-mediated host inflammatory responses. Toxin A triggers interleukin (IL-8) secretion via stimulation of mitogen-activated protein kinase cascades and activation of the nuclear transcription factor NF-κB in monocytes and colonic epithelial cells. Proinflammatory cytokines, released in response to *C. difficile* toxins from intestinal mucosal cells, may be responsible for activation of intestinal nerves observed in animal models of toxin A-induced enterocolitis.

Antibodies against *C. difficile* toxins A and B are present in the majority of healthy adults and children, and immunization against toxin A protects animals from *C. difficile* infection. Recent evidence indicates that the development of symptomatic *C. difficile* infection is well correlated with the immune response to *C. difficile* toxins and that a defective antibody response to toxin A is linked to recurrent *C. difficile* infection. Thus, vaccination or passive immunotherapy for *C. difficile*-associated diarrhea and colitis may represent exciting possibilities for prophylaxis and treatment of this disease.

Bacteroides fragilis ENTEROTOXIN

Bacteroides fragilis enterotoxin (BFT) has been associated with strains of *B. fragilis* isolated from animals and children with diarrheal disease. BFT, also known as fragilysin, is synthesized by enterotoxic strains of *B. fragilis* (ETBF) as a precursor protein (397 amino acids) and is proteolytically cleaved into a mature toxin (amino acids 212–397). Three distinct BFT-encoding genes (*bft-1, -2*, and *-3*) have been cloned from different ETBF strains with 92–96% amino acid homology. Sequence comparison reveals that BFT contains a zinc-binding metalloprotease motif. The C-terminal 20 amino acids of the *bft-2* gene product exhibits an amphipathic structure and may be responsible for membrane insertion of BFT-2 into host cells. The mature toxin cleaves the extracellular domain of the transmembrane adherence junction protein E-cadherin. This is followed by degradation of the remaining cytoplasmic domain of E-cadherin by cellular proteases. The degradation of E-cadherin apparently triggers reorganization of cell actin filaments, leading to increased cell volume and diminished tight junction barrier functions.

Purified BFT stimulates fluid secretion in ligated ileal and colonic loops in experimental animals. Analysis of the fluid content of intestinal loops reveals increased levels of sodium, chloride, albumin, and protein. In addition, mildly hemorrhagic fluid and patchy mucosal wall hemorrhage are observed, suggesting that BFT evokes an intestinal inflammatory response. BFT stimulates release of the proinflammatory cytokine IL-8, increases paracellular permeability, and causes

alterations of tight junctional proteins in cultured intestinal epithelial cells. These data indicate that BFT is the virulent factor responsible for ETBF-induced intestinal diarrhea and inflammation.

See Also the Following Articles

Cholera • Diarrhea • Diarrhea, Infectious • Food Poisoning • Shigella • Traveler's Diarrhea

Further Reading

Aktories, K. (1997). Bacterial toxins that target Rho proteins. *J. Clin. Invest.* **99**, 827–829.

Fasano, A. (2000). Regulation of intercellular tight junctions by zonula occludens toxin and its eukaryotic analogue zonulin. *Ann. N. Y. Acad. Sci.* **915**, 214–222.

Heyderman, R. S., Soriani, M., and Hirst, T. R. (2001). Is immune cell activation the missing link in the pathogenesis of post-diarrhoeal HUS? *Trends Microbiol.* **9**, 262–266.

Lencer, W. I. (2001). Microbes and microbial toxins: Paradigms for microbial–mucosal interactions. V. Cholera: Invasion of the intestinal epithelial barrier by a stably folded protein toxin. *Am. J. Physiol.* **280**, G781–G786.

Lungdren, O. (2002). Enteric nerves and diarrhoea. *Pharmacol. Toxicol.* **90**, 109–120.

Nair, G. B., and Takeda, Y. (1998). The heat-stable enterotoxins. *Microb. Pathog.* **24**, 123–131.

O'Loughlin, E. V., and Robins-Browne, R. M. (2001). Effect of Shiga toxin and Shiga like toxins on eukaryotic cells. *Microbes Infect.* **3**, 493–507.

Pothoulakis, C., and LaMont, J. T. (2001). Microbes and microbial toxins: Paradigms for microbial–mucosal interactions. II. The intergrated response of the intestine to *Clostridium difficile* toxins. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G178–G183.

Rappuoli, R., Pizza, M., Douce, G., and Dougan, G. (1999). Structure and mucosal adjuvanticity of cholera and *Escherichia coli* heat-labile enterotoxins. *Immunol. Today* **20**, 493–500.

Scott, R. O., Thelin, W. R., and Milgram, S. L. (2002). A novel PDZ protein regulates the activity of guanylyl cyclase C, the heat-stable enterotoxin receptor. *J. Biol. Chem.* **277**, 22934–22941.

Sears, C. L. (2001). The toxins of Bacteroides fragilis. *Toxicon* **39**, 1737–1746.

Uzzau, S., and Fasano, A. (2000). Cross talk between enteric pathogens and the intestine. *Cell. Microbiol.* **2**, 83–89.

Williams, N. A., Hirst, T. R., and Nashar, T. O. (1999). Immune modulation by the cholera-like enterotoxins: From adjuvant to therapeutic. *Immunol. Today* **20**, 95–100.

Wimer-Mackin, S., Holmes, R. K., Wolfe, A. A., Lencer, W. I., and Jobling, M. G. (2001). Characterization of receptor-mediated signal transduction by *Escherichia coli* type IIa heat-labile enterotoxin in the polarized human intestinal cell line T84. *Infect. Immun.* **69**, 7205–7212.

Wu, S., Lim, K.-C., Huang, J., Saidi, R. F., and Sears, C. L. (1998). Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14979–14984.

Barium Radiography

SUSAN TEEGER
Weill Medical College of Cornell University

aspiration Inhalation of ingested substances.
barium enema Radiographic examination of the colon and rectum.
barium sulfate White powdery substance used to facilitate radiographic visualization of internal structures of the body.
colonoscopy Examination of the colon using a colonoscope to visualize the lumen of the colon directly.
contrast media Substances used to facilitate radiographic visualization of internal structures of the body.
deglutition Swallowing.
diverticula Outpouchings of mucous membrane through the muscular wall of a tubular organ.
dysphagia Difficulty swallowing.
endoscopy Visual inspection of any cavity of the body by means of an endoscope, which is an instrument to visualize the interior of a hollow organ.
en face Face-on.
enteroclysis Type of radiographic examination of the small bowel.
fluoroscopy Examination by means of a fluoroscope, which is a device used for examining deep portions of the body.
gastric Pertaining to the stomach.
gastric fundus Upper portion of the stomach.
gastrointestinal tract Portion of the body comprising the pharynx, esophagus, stomach, small bowel, colon, and rectum.
gastrointestinal tract contrast examination Radiographic examination of the gastrointestinal tract utilizing a radiopaque substance to facilitate visualization of portions of the gastrointestinal tract.
intubation Obtaining access to the gastrointestinal tract with a tube.
lumen Cavity within a tubular organ, such as the gastrointestinal tract.
manometry Examination measuring pressure in the gastrointestinal tract.
mastication Chewing.
mucosal Pertaining to the mucous membrane.
pharyngoesophagram Radiographic examination of the pharynx and esophagus.
pharyngogram Radiographic examination of the pharynx.
pharynx Uppermost portion of the gastrointestinal tract between the mouth and the esophagus.
proctography Radiographic evaluation of the rectum and anus.
radiographic examination Using X rays to evaluate portions of the body.
small bowel series Radiographic examination of the small bowel.
upper gastrointestinal tract examination Assessment of the esophagus, stomach, and duodenum.
viscus Tubular portion of the gastrointestinal tract.

Barium radiographic examinations are often crucial in the investigation of patients with suspected gastrointestinal disease. The examination involves a certain element of routine, although every examination can and should be specifically tailored to answer the clinical question. Barium radiography should be tailored to the patient's ability to undergo the examination. It is therefore of extreme importance that the radiologist be given all of the appropriate clinical information to ensure that the examination yields useful diagnostic information.

INTRODUCTION

Gastrointestinal tract contrast examinations can be performed using what is commonly known as single- or double-contrast technique, or in some circumstances, a combination of both techniques (biphasic technique). The single-contrast technique is performed by filling the lumen with a relatively low-density barium preparation and utilizing compression for detection of contour irregularities or abnormalities or filling defects in the barium pool. Because of the density of the barium-filled organ, the compression helps to displace barium into ulcer craters and to better allow detection of polypoid and ulcerated lesions *en face*. Unfortunately, several technical problems occur with this technique. Superficial mucosal abnormalities cannot be easily recognized, the rectum and gastric fundus become very dense and are relatively inaccessible to compression, thereby precluding demonstration of *en face* lesions, and compression may be impossible because of obesity or because of recent abdominal surgery or trauma. Double-contrast examinations are performed with a smaller amount of high-density barium to coat the mucosal surface and

 Copyright 2004, Elsevier (USA). All rights reserved.

with swallowed effervescent carbon dioxide granules to distend the viscus with gas. This allows visualization of lesions *en face* and in profile. Compression is routinely performed and is especially helpful for anterior wall lesion demonstration and for lesions characterized by areas of narrowing. Double-contrast techniques allow detection of contour abnormalities as well as fine mucosal abnormalities and irregularities and are also often helpful in detecting lesions that cannot be detected by endoscopy, are missed or misinterpreted at endoscopy, and are often helpful in evaluating patients who refuse or are not candidates for endoscopy. Double-contrast examinations are, however, not adequate where there is poor coating, where there is insufficient distension (which may hide lesions), or where there is overdistension (which can obscure lesions). In some cases, in addition to the double-contrast examination, it may be necessary to perform a single-contrast examination afterward to gain additional information.

PHARYNX

Swallowing and feeding problems are common in all age groups, but particularly in older patients and are becoming an increasingly frequent problem in medical practice.

Applications

A radiographic pharyngeal examination is most frequently performed on patients with feeding difficulty or difficulty swallowing, and is also helpful in evaluating patients with respiratory problems, for whom laryngeal penetration and aspiration may result in aspiration pneumonia, chronic bronchitis, chronic coughing, or choking. Radiographic pharyngeal examination may also be helpful in detection of soft palate insufficiency, which may result in nasal regurgitation or a nasal quality to the voice. In addition, radiologic evaluation of the pharynx may be helpful in assessing pharyngeal morphology and function in patients with certain neurologic diseases and degenerative neurologic or muscular abnormalities, cerebrovascular accidents, head injury, pharyngeal neoplasia, and complications of head and neck surgery or radiation. Radiographic evaluation of swallowing is often needed to assist the swallowing therapist in the choice of an appropriate nutritional option in those patients suspected of aspiration with different nutritional consistencies. In addition, the swallowing therapist can visualize the effect of various therapeutic maneuvers that can be performed to help diminish laryngeal penetration and aspiration. Pharyngoesophagrams are useful in evaluating patients for suspected foreign body, fistula, or abscess. A prominent cricopharyngeus muscle, which may cause patients to complain of dysphagia, may be detected during a pharyngogram. Pharyngography may be helpful in diagnosing lateral pharyngeal pouches and diverticula, which may result in spillage of contents into the hypopharynx with resultant aspiration into the larynx and tracheobronchial tree, and may be complicated by ulceration or neoplasia. Zenker's diverticula are usually detectable with radiographic examinations of the pharynx (Fig. 1). Barium studies are of limited value in patients with viral, bacterial, or fungal infection of the pharynx because these patients usually have normal pharyngograms or lymphoid hyperplasia of the palatine tonsil or base of the tongue. Occasionally, *Candida* or herpes pharyngitis may be detected, especially in patients with AIDS. Barium studies may also be helpful in patients with chronic sore throat, in which underlying gastroesophageal reflux or reflux esophagitis may be the contributing factor. Double-contrast pharyngography plays a role in the detection and workup of pharyngeal neoplasms. The size, level, and extent of the lesion can

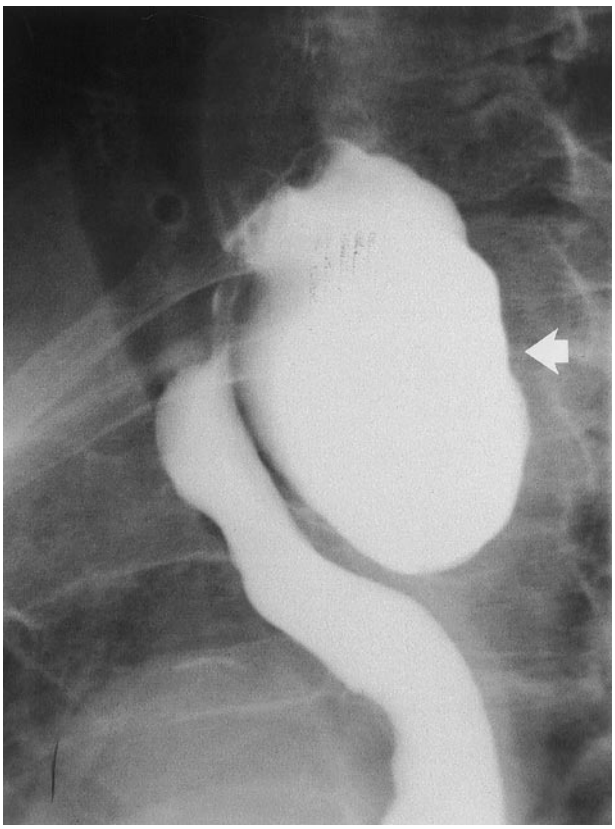


FIGURE 1 Fluoroscopic spot image from a barium esophagram showing a large Zenker's diverticulum (arrow) projecting from the posterior aspect of the proximal esophagus.

be determined and regions of the pharynx that are difficult to visualize or easily missed at endoscopy (such as submucosal masses) can be visualized.

Contraindications

Contrast examination of the pharynx may be dangerous in a patient with acute airway obstruction. The most notable clinical example for which barium studies are contraindicated is suspected epiglottitis. Patients with suspected airway obstruction should have plain films of the neck as the initial radiographic examination.

Technique

Radiographic examination of the pharynx includes a cine or video pharyngoesophagram to evaluate motility, and a series of spot films to evaluate structural abnormalities. In a patient with suspected foreign body, fistula, or abscess, frontal and lateral plain films of the neck should be obtained in addition.

The videofluoroscopic swallowing examination is normally limited to the oral cavity, pharynx, and cervical esophagus and is designed to study the anatomy and physiology of the oral preparatory, oral, pharyngeal, and cervical esophageal stages of deglutition. Small amounts of contrast material are used to minimize the risk while evaluating the physiology of the oral cavity and pharynx. Four consistencies of material are used to investigate patient complaints of variable swallowing disability: thin and thick liquid barium, barium paste mixed with applesauce, and material requiring mastication (crackers mixed with barium paste). Swallowing therapists who may be involved in the care of the patient often attend the radiographic examination and assist in deciding in the types and amounts of testing materials to be used. The cine or video recording of each swallow allows a frame-by-frame analysis of the various parameters of swallowing. Movement of the tongue, soft palate, and epiglottis, as well as laryngeal closure and cricopharyngeal opening, symmetry of tongue motion, pharyngeal peristalsis, and epiglottic tilt, can be assessed using a combination of lateral and frontal projections.

Double-contrast views of the pharynx are obtained in frontal and lateral projections to demonstrate the contours of the valleculae and piriform sinuses, lateral walls of the tonsillar fossae and hypopharynx, superior and inferior borders of the base of the tongue, the soft palate, the posterior pharyngeal wall, the anterior hypopharyngeal wall, the epiglottis, and the cricopharyngeus.

UPPER GASTROINTESTINAL TRACT

Currently, most patients presenting with epigastric pain or other gastrointestinal symptoms are initially treated empirically with a trial of medication without undergoing any diagnostic investigations. However, compliant patients who fail to respond to this therapy will usually undergo either endoscopy or an upper gastrointestinal radiographic examination as the first diagnostic test, with proponents existing for both endoscopy and upper gastrointestinal radiographic examinations. Those that advocate endoscopy feel that the additional expense and definite risk of the procedure are warranted because of its perceived greater accuracy and possibility for definitive treatment. With the use of the biphasic technique (combination of single- and double-contrast techniques), the ability to diagnose accurately disorders of the esophagus, stomach, and duodenum has dramatically improved. A meticulously performed biphasic examination, although not as sensitive for detection of mild inflammatory conditions, remains an excellent cheaper, less risky, and relatively accurate alternative. In addition, although the upper endoscopy can only directly evaluate the mucosa, the radiographic upper gastrointestinal examination is capable of evaluating the mucosa as well as the wall of the upper gastrointestinal tract and is helpful in evaluating motility disorders.

Applications

Esophagus

The esophagus should be evaluated with a dedicated examination directed to the esophagus, if patient symptoms suggest localized esophageal disease. Alternatively, if the symptoms cannot be specifically localized to the esophagus, the esophagus can be evaluated as part of a dedicated upper gastrointestinal examination.

In patients presenting with reflux symptoms, the barium examination may be helpful for documentation of a hiatal hernia (Fig. 2), for demonstration of suspected reflux, and for detection of possible reflux sequelae, such as reflux esophagitis and ulceration, strictures, Barrett's esophagus, and esophageal adenocarcinoma (Fig. 3).

Patients presenting with dysphagia often benefit from a barium examination, whereby abnormalities of both structure and function responsible for the dysphagia may be detected. Structural lesions include benign lesions such as rings, webs, and strictures; benign neoplasms; and malignancies such as adenocarcinoma, lymphoma, or Kaposi's sarcoma. In addition, esophageal motility disorders, both primary and secondary, can cause dysphagia. Primary motility disorders such as

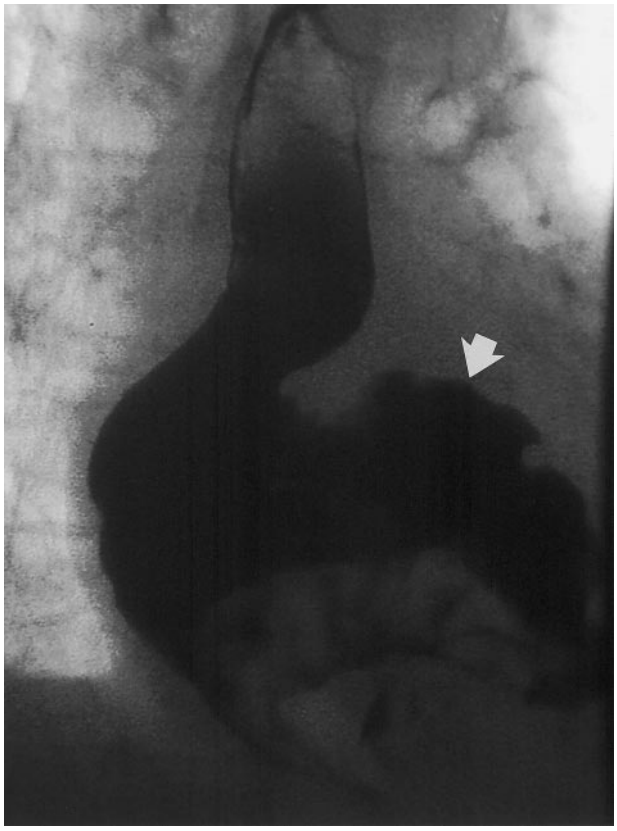


FIGURE 2 A large hiatal hernia (arrow) containing most of the gastric fundus in the chest.

achalasia (Fig. 4), diffuse esophageal spasm (Fig. 5), nonspecific esophageal motility disorder, and presbyesophagus can all be evaluated with barium examinations of the esophagus. Nutcracker esophagus, and hypertensive lower esophageal sphincter, which can cause dysphagia, are best diagnosed with manometry. Secondary motility disorders such as those occurring with collagen vascular disease, especially scleroderma, mixed connective tissue disease, dermatomyositis, and polymyositis, may be detected with barium radiography. Other secondary motility disorders that can be evaluated with barium studies include motility abnormality secondary to reflux esophagitis, caustic esophagitis, and radiation therapy. In addition, other secondary causes of motility abnormality and subsequent dysphagia include infectious disorders such as Chagas' disease, metabolic and endocrine disorders such as diabetes mellitus, and alcohol abuse, and many neuromuscular disorders, including cerebrovascular accidents, multiple sclerosis, amyotrophic lateral sclerosis, myasthenia gravis, and postpoliomyelitis. All these entities may be evaluated with barium studies of the esophagus.

Infectious esophagitis has become a relatively recent problem, particularly in patients who are immunocompromised and present with dysphagia. When a patient is suspected of having infectious esophagitis, double-contrast examination of the esophagus may confirm the diagnosis and often helps to diagnose the offending organism (Fig. 6).

Barium radiography is of use in evaluating patients with dysphagia caused by any of the other forms of clinically suspected esophagitides, including drug-induced and radiation-induced esophagitis, caustic esophagitis, Crohn's disease, epidermolysis bulosa, pemphigoid, nasogastric intubation esophagitis, eosinophilic esophagitis, alcohol-induced esophagitis, graft-versus-host disease, Behcet's disease, and esophageal intramural pseudodiverticulosis.

Various other abnormalities of the esophagus can be detected using barium radiographic examinations. These include Mallory–Weiss tears, varices, esophageal hematomas, esophageal perforations (Fig. 7), foreign body impactions, fistulas, diverticula, congenital esophageal stenosis, and extrinsic impressions on the esophagus. Water-soluble contrast media, rather than barium, are used as the initial materials to diagnose esophageal

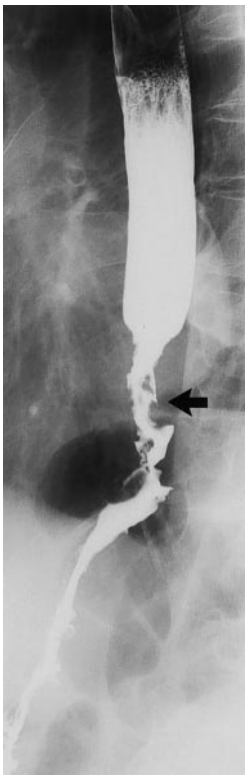


FIGURE 3 Spot image from a double-contrast esophagram revealing a markedly irregular area of narrowing in the distal esophagus (arrow), compatible with esophageal carcinoma.

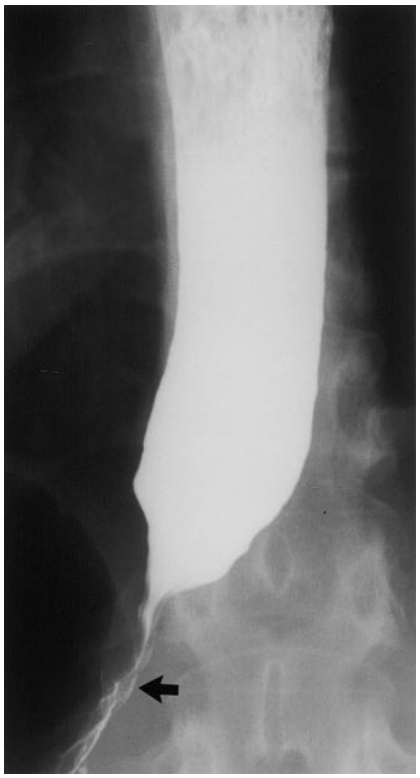


FIGURE 4 Fluoroscopic image from an esophagram showing marked narrowing in the distal esophagus with the "bird's beak appearance" (arrow), with dilatation of the esophagus proximally, all features characteristic of achalasia.

perforations. Barium has been shown to excite an inflammatory reaction in the mediastinum with subsequent fibrosis, whereas water-soluble media have not. Water-soluble media are also rapidly absorbed from the mediastinum so that followup studies are not compromised by residual contrast in the mediastinum. Because water-soluble contrast media are less dense than barium, some perforations may be missed solely with the water-soluble contrast materials. Therefore, if the initial study performed with water-soluble contrast media does not show a leak, the examination should be repeated with barium to detect subtle leaks. In these cases, the harmful effects of barium in the mediastinum are less serious than that of missing a potentially life-threatening condition.

The integrity of esophageal anastomoses in the postoperative patient can also be assessed with radiography by utilizing the same technique as that for detection of esophageal perforations.

Stomach and Duodenum

Radiographic examinations of the stomach and duodenum are helpful in the evaluation of patients who present with dyspepsia, bloating, belching, early satiety, weight loss, and signs and symptoms suggestive of upper gastrointestinal bleeding such as hematemesis, melena, and guaiac-positive stool. The double-contrast examination will often lead to a diagnosis of various gastric and duodenal processes that may explain the above symptoms. These processes include peptic disease, other inflammatory and infectious conditions such as erosive gastritis of any cause, antral gastritis, *Helicobacter pylori* gastritis, hypertrophic gastritis, Menetrier's disease, atrophic gastritis, eosinophilic gastritis, emphysematous gastritis, caustic ingestion, radiation-induced abnormalities, gastric ulcers of any cause (Fig. 8), duodenitis, granulomatous conditions such as Crohn's disease, sarcoidosis, tuberculosis, syphilis, fungal diseases, and other infectious diseases such as cytomegalovirus, cryptosporidiosis, toxoplasmosis, and strongyloidiasis. In addition, benign and malignant neoplasms of the stomach and duodenum can be detected with barium radiography. Other miscellaneous abnormalities, including varices, diverticula, diaphragms and webs, hypertrophic pyloric stenosis, gastric outlet obstruction, duodenal obstruction, gastric

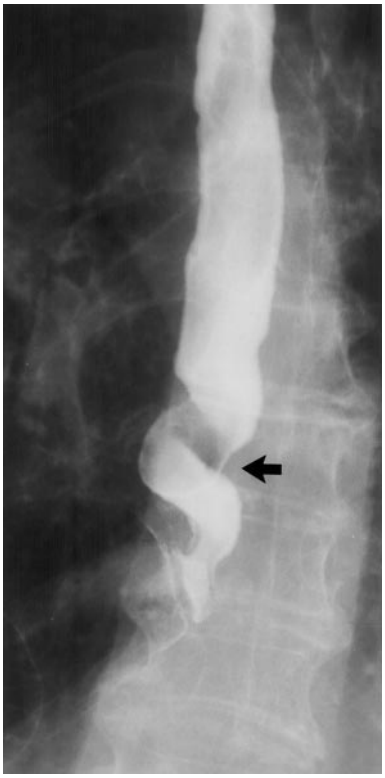


FIGURE 5 Spot film from an esophagram showing a "corkscrew" appearance (arrow) in a patient with diffuse esophageal spasm.

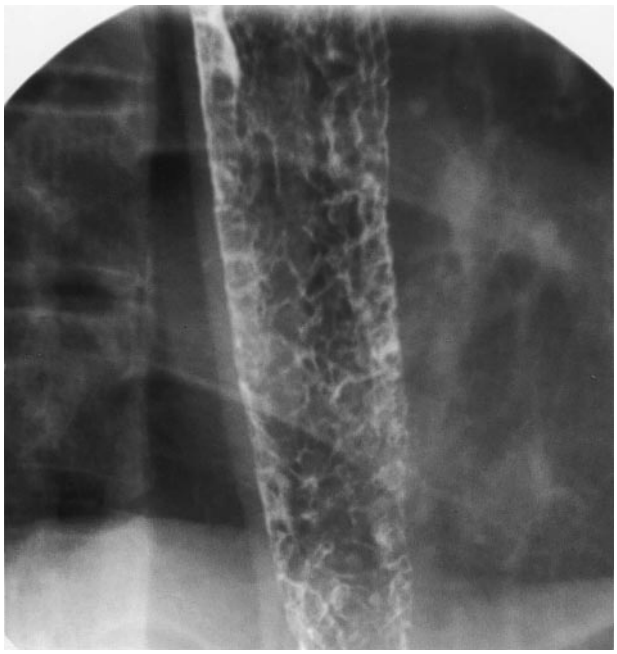


FIGURE 6 Spot fluoroscopic image from a double-contrast esophagram showing a "shaggy" esophagus, typical of candidal esophagitis in an AIDS patient.

dilatation without outlet obstruction, extrinsic processes such as pancreatic disease, renal and liver masses, gallbladder disease and colonic masses (all of which may cause mass effect on the stomach and duodenum), bezoars, foreign bodies, hematomas, gastric volvulus (Fig. 9), gastroduodenal and duodenojejunal intussusceptions, fistulas, gastric and duodenal perforations, and amyloidosis, can all be detectable with the upper gastrointestinal examination. Patients who have undergone surgery to the stomach and duodenum may be evaluated for postoperative problems using various upper gastrointestinal radiographic techniques. In these patients, it is essential that the operative procedure performed be described to the radiologist before the procedure is performed in order that the postoperative anatomy, the efficiency of the procedure, and possible postoperative complications can be accurately determined if necessary.

If a diagnosis can be confidently made from the upper gastrointestinal examination, there is no need for confirmation with endoscopy. If, however, the upper gastrointestinal examination reveals findings that cannot be confidently diagnosed definitively or if the findings are equivocal or suspicious for malignancy, upper endoscopy with biopsy, if indicated, should be performed. If the upper gastrointestinal examination is unrevealing, the decision to perform upper endoscopy should be dictated by the clinical scenario.

Contraindications

As mentioned previously, barium is contraindicated with suspected esophageal perforation, but is also contraindicated with suspected gastric or duodenal perforation. Water-soluble contrast media should be used instead. Double-contrast examinations may be difficult or impossible to perform on elderly or debilitated patients who are not able to perform the various turning maneuvers required for this technique. In those patients, single-contrast techniques can be performed, but the limitations of this technique have to be kept in mind. Patients who are unable to follow commands enabling them to ingest the contrast materials because of diminished consciousness or severe dementia are not candidates for barium radiography, unless the examination can be performed with a single-contrast technique through an indwelling nasogastric tube.

Technique

The routine upper gastrointestinal examination should ideally be performed as a biphasic study, utilizing double-contrast techniques followed by

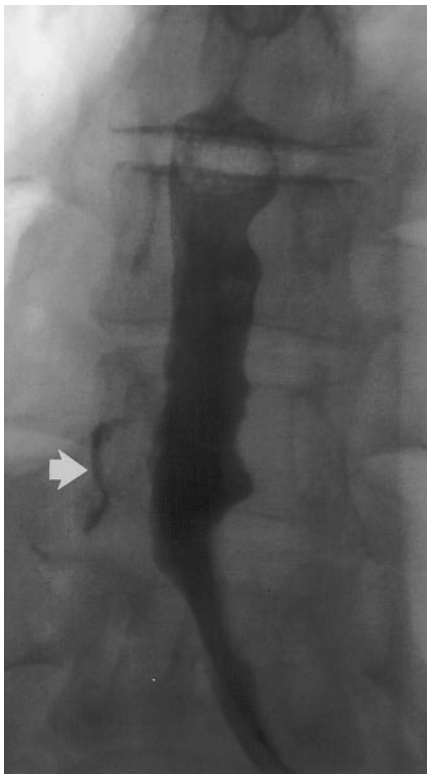


FIGURE 7 An image from an esophagram performed with water-soluble contrast showing extravasation of contrast to the right of the esophagus related to an esophageal perforation (arrow).

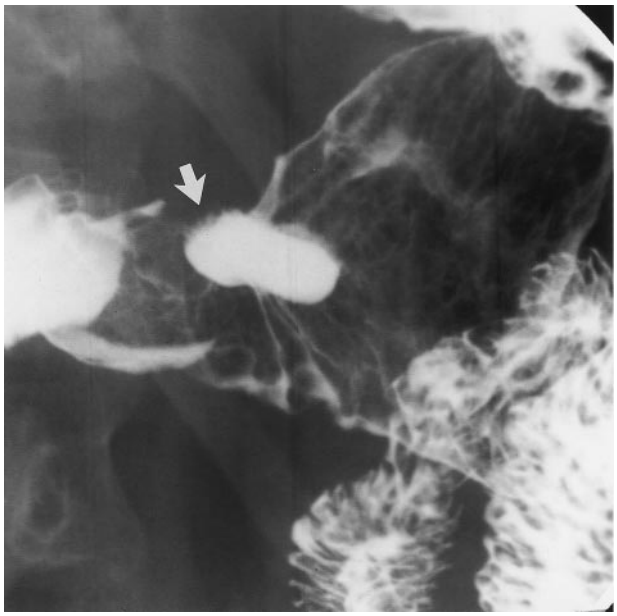


FIGURE 8 Double-contrast image of the stomach showing a large, contrast-filled ulcer in the gastric antrum (arrow).

single-contrast techniques. The routine double-contrast portion of the examination is designed to coat the mucosal surface with a thin layer of high-density barium, while the lumen is distended with gas obtained from effervescent granules. Double-contrast views of the esophagus, stomach, and duodenum are obtained in different projections utilizing various turning maneuvers to achieve adequate coating and distension. This is followed by single-contrast views of the esophagus, stomach, and duodenum, using low-density barium liquid and varying degrees of compression.

SMALL BOWEL

Barium radiography remains relatively unchallenged by endoscopy in evaluation of a large portion of the small bowel. Although small bowel enteroscopy is becoming more prevalent, for evaluating the proximal small bowel, this procedure can only detect mucosal disease, and cannot evaluate the distal small bowel, although a skilled colonoscopist can frequently evaluate the terminal ileum.

The entire small bowel is best evaluated radiographically with either a dedicated fluoroscopic small bowel series and/or enteroclysis. Barium small bowel examinations should be ordered only in the presence of clinical symptoms and signs specifically pointing to the possibility of small bowel disease. Vague, nonspecific symptoms have a very low yield of positive findings on small bowel examinations, and ordering of these examinations in this instance exposes patients to unnecessary radiation and anxiety.

Applications

In most radiologic practices in the United States, the majority of small bowel problems are evaluated with the small bowel series. The major advantage of the fluoroscopic small bowel series is that it does not require intubation and is therefore better tolerated. In addition, expertise in performing enteroclysis (small bowel enema) is not widespread. The enteroclysis provides a more detailed double-contrast examination of the small intestine. By better distending the lumen, the shape and thickness of the small bowel folds are better demonstrated, regions of partial obstruction are more readily delineated, and mucosal nodularity is more easily identified.

At our institution and at many others, the enteroclysis is not the initial examination in all patients but is reserved for the following indications when there remains a high index of suspicion for small bowel disease despite an unremarkable or inconclusive fluoroscopic

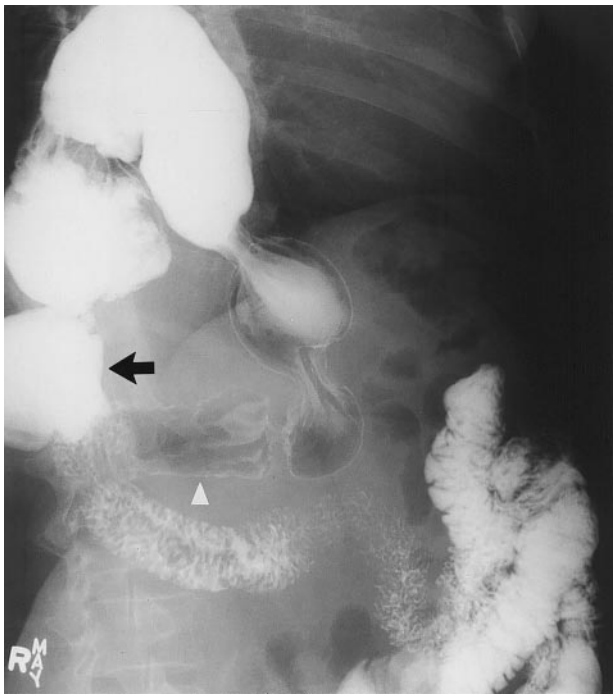


FIGURE 9 Spot film of the stomach showing an organo-axial volvulus, with gastric fundus (arrow) lying to the right of the gastric antrum (arrowhead).

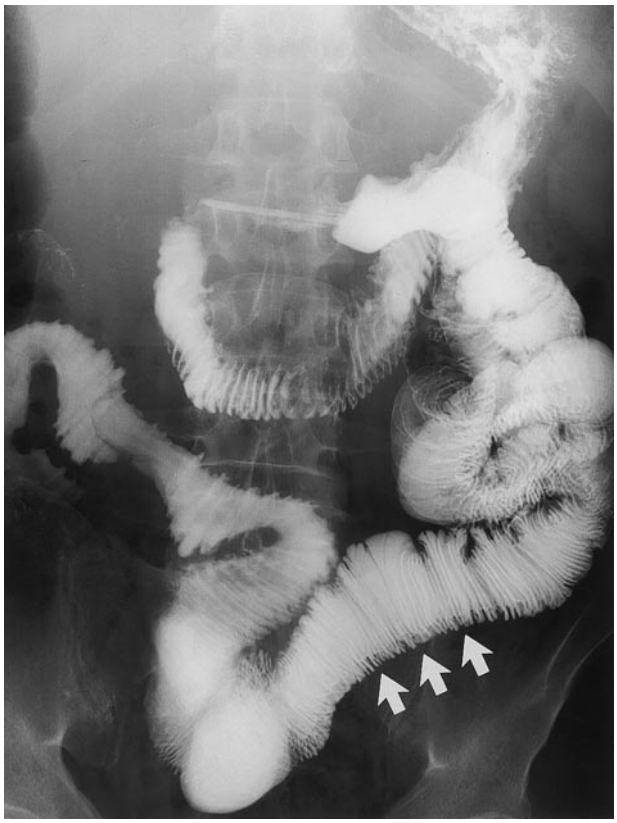


FIGURE 10 Overhead film from a small bowel series in a patient with scleroderma showing a dilated small bowel, with tightly packed folds (arrows).

small bowel series:

1. Evaluation of patients with suspected or known malabsorption states, such as celiac disease, tropical sprue, scleroderma (Fig. 10), lymphangiectasia, Whipple disease, bacterial overgrowth syndromes, short bowel syndrome, adult cystic fibrosis, abetalipoproteinemia, systemic mastocystosis, and Waldenström's macroglobulinemia.
2. Partial, low-grade, or intermittent small bowel obstruction.
3. Preoperative evaluation of Crohn's disease to determine extent of disease, and search for proximal segments of disease (Fig. 11).
4. Gastrointestinal blood loss of obscure origin, when upper gastrointestinal and colonic radiologic examinations, small bowel series, and/or endoscopic studies are negative.
5. Evaluation of patients with radiation-induced damage to the small bowel.
6. Detection of small bowel tumors, such as carcinoids, primary and secondary malignancies, polyps, and lymphoma.

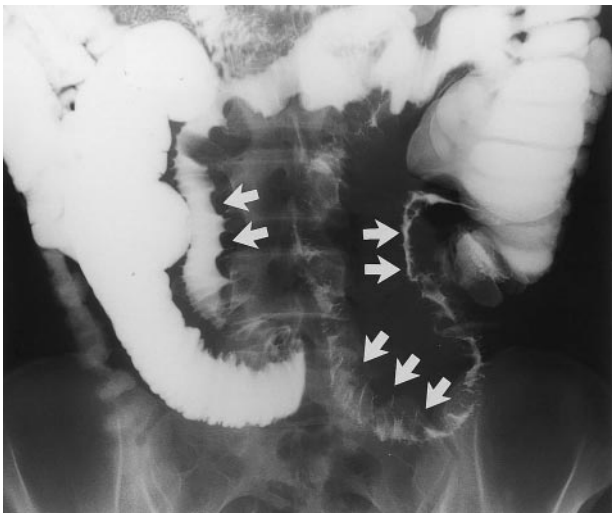


FIGURE 11 Overhead film from a small bowel series showing nodular, thickened, spiculated narrowed small bowel loops in a patient with extensive Crohn's disease (arrows).

7. Detection of Meckel diverticulum.
8. For patients, presently asymptomatic, in whom physicians suspect clinical symptoms were caused by partial small bowel obstruction(s) secondary to adhesions (Fig. 12).

Barium examinations of the small bowel may also be helpful in evaluation of various inflammatory conditions of the small bowel, such as parasitic infestations, bacterial and viral infections, and inflammatory

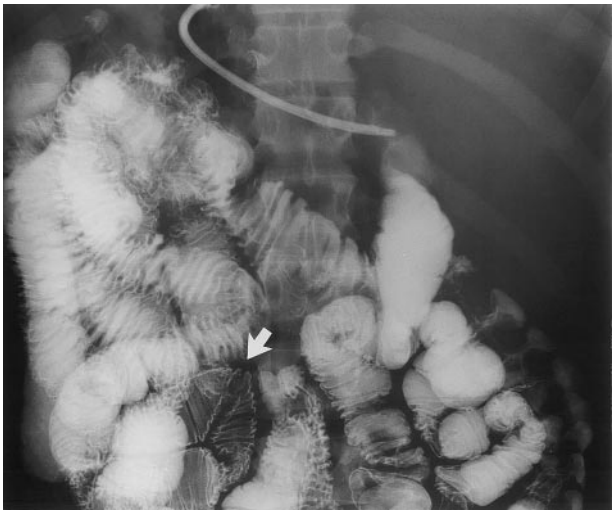


FIGURE 12 Overhead view from an enteroclysis showing an acutely angulated, narrowed small bowel loop, related to an adhesion (arrow).

conditions occurring in association with immunodeficiency. Vascular disorders of the small bowel, such as mesenteric ischemia and infarction, intramural hemorrhage, vasculitis and vascular malformations, may also be evaluated utilizing barium small bowel examinations. Post-operative complications, malrotations, hernias, gut duplications, intestinal edema, foreign bodies, enteroliths, bezoars, posttransplant lymphoproliferative disorder, and graft-versus-host disease are other miscellaneous small bowel processes that can be evaluated with barium radiography.

Contraindications

Suspected intestinal perforation is the only absolute contraindication to performance of a barium small bowel examination. Barium in the peritoneal cavity in these patients leads to a fulminant chemical peritonitis, which should be avoided whenever possible. Water-soluble contrast materials such as gastrografin should not be used in small bowel evaluation because they become diluted by the luminal fluid, thereby precluding adequate visualization. They may be satisfactorily used only in evaluating for suspected perforation of the very most proximal small bowel. In addition, they should not be used in patients with obstruction, because they are hyperosmolar and may draw fluid into the bowel lumen, thereby aggravating what may be an already altered fluid and electrolyte balance.

Technique

Dedicated Fluoroscopic Small Bowel Series

The patient ingests approximately 3 cups (500–600 ml) of 40–50% (weight/volume) barium sulfate solution, followed by abdominal radiographs obtained at 15-minute intervals for the first half-hour, and then every 30 minutes until the right colon is opacified. Each abdominal film is reviewed, followed by intermittent fluoroscopy with compression until all small bowel loops have been demonstrated and evaluated. The intermittent fluoroscopy with compression is an essential portion of the examination because lesions can be missed if only the overhead views are evaluated, due to the tendency of loops to overlap and obscure lesions. Periodic spot films are obtained to document any suspicious areas, and once the right colon is opacified, spot films of the terminal ileum are obtained.

Enteroclysis (Small Bowel Enema)

Small bowel transit is accelerated by administration of 20 mg of oral metoclopramide. Gastric intubation with a balloon catheter using either the oral or the transnasal route is then performed, followed by fluoroscopically directed manipulation of the catheter into either the distal duodenum or the proximal jejunum, just beyond the ligament of Treitz. The balloon is then inflated under fluoroscopy to prevent reflux of contrast into the stomach. Approximately 200–250 ml of a 75–80% (weight/volume) barium suspension is injected into the small bowel followed by approximately 1000–2000 ml of a 0.5% methylcellulose solution to obtain a double-contrast effect. The radiologist is present during the entire procedure, and the entire small bowel is examined with intermittent fluoroscopy utilizing graded abdominal compression. Periodic spot films of the small bowel are obtained to document the fluoroscopic findings. The examination is completed with an overhead radiograph of the abdomen to include the entire small bowel.

COLON AND RECTUM

Barium examination of the colon and rectum is predominantly utilized to detect mucosal abnormalities. Its greatest competitor is colonoscopy, whereby the mucosa can be directly visualized and diagnostic biopsies obtained. A carefully performed barium examination of the colon is, however, an excellent, less invasive, safer, and cheaper alternative for detection of most colonic and rectal lesions. Colonoscopy remains more accurate than barium radiography in detection of subtle mucosal abnormalities of the colon, such as polypoid lesions less than 1 cm in diameter, and the early preulcerative changes of inflammatory bowel disease. Because barium radiographic examination can also be used for the detection of mural colonic and rectal lesions, it has a competitive advantage over colonoscopy, because colonoscopy has not been proved to be an adequate test for mural or extrinsic lesion detection. Because colonoscopy and barium radiography both have limitations, the double-contrast barium enema and colonoscopy should be viewed as complementary techniques for the diagnosis of various colonic processes. Computer tomography (CT) and in some cases magnetic resonance imaging (MRI) and endoscopic ultrasound are the cross-sectional imaging modalities that directly compete with barium radiographic examination in the detection and evaluation of mural and rectal colonic entities. Although CT, MRI, and endoscopic ultrasound are superior to barium radiographic examination in the detection of extrinsic processes affecting the colon and rectum, these lesions can sometimes be detected and evaluated with barium examinations. Barium

radiography is often used to complete the colonic examination in those patients in whom the colonoscopist is unsuccessful in passing the endoscope to the cecum.

Adequate preparation is essential in the satisfactory performance of a colonic and rectal barium examination, because residual feces will obscure polypoid lesions and the early manifestations of inflammatory pathology. Although a variety of preparatory regimens exist, the most effective protocol consists of a clear liquid diet for 24 hours prior to the procedure, laxatives on the afternoon and evening before the test, and a suppository the morning of the examination. Cleansing enemas preferably should not be used, because they result in significant residual intraluminal fluid, which may degrade the quality of mucosal coating. In our institution, patients in renal failure are prepared with Golytely (polyethylene glycol solution), because magnesium citrate, which is the normally utilized laxative, is contraindicated in the presence of renal compromise. The one, but unavoidable, drawback of utilizing oral lavage regimens such as Golytely in renal failure patients is that the quality of mucosal coating may be degraded because of the presence of residual intraluminal fluid. In patients with inflammatory bowel disease, milder laxatives should be used, or the preparation may consist solely of clear liquids by mouth. Preparation may not be necessary in patients with acute colonic obstruction, in those for whom the study is simply being performed to evaluate the anatomy, or when there is a clinical suspicion of a perforation.

Applications

The most common indications for barium radiography include detection of colorectal polyps (Fig. 13) and carcinoma (Fig. 14), assessment of inflammatory bowel disease (Fig. 15), diagnosis and assessment of diverticular disease (Fig. 16), and evaluation of colonic extrinsic mass lesions. Barium radiography can also be used to evaluate other inflammatory conditions of the colon, including infectious processes such as bacterial, viral, parasitic, and fungal diseases and noninfectious colitides, when colonoscopy is not feasible. Other tumors of the colon, including lymphoma, hemangioma, lymphangioma, Kaposi's sarcoma, carcinoid tumors, lipomas, stromal tumors, squamous cell and cloacogenic carcinoma, and metastases, can be detected with double-contrast barium enemas. Other miscellaneous disorders of the colon, including colonic ischemia, radiation colitis, rectal hemorrhoids, rectal varices, and cathartic colon; functional disorders of the colon, including irritable bowel syndrome, chronic constipation, collagen vascular disorders, and Ogilvie's syndrome;

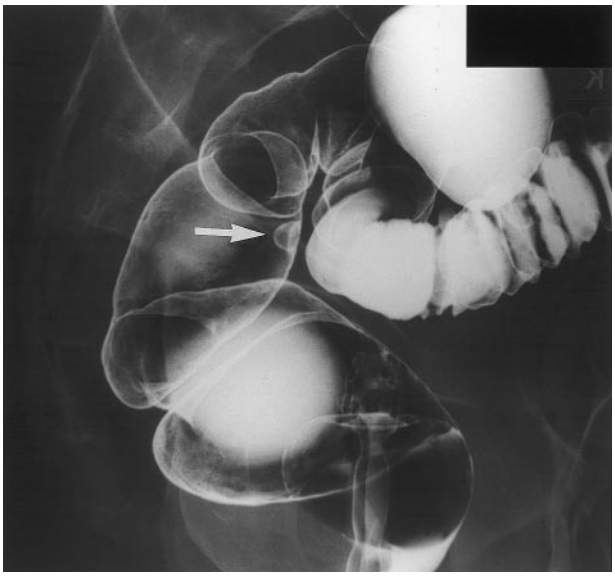


FIGURE 13 Double-contrast image of the rectum showing a polyp along the anterior wall (arrow).

and extracolonic diseases affecting the colon, such as endometriosis, benign gynecologic tumors, gynecologic malignancies, and pelvic inflammatory disease, can also be evaluated with barium enema examinations. Postoperative colonic abnormalities can be assessed and postoperative anatomic evaluations can be performed with barium radiography. Finally, evaluation of defecatory

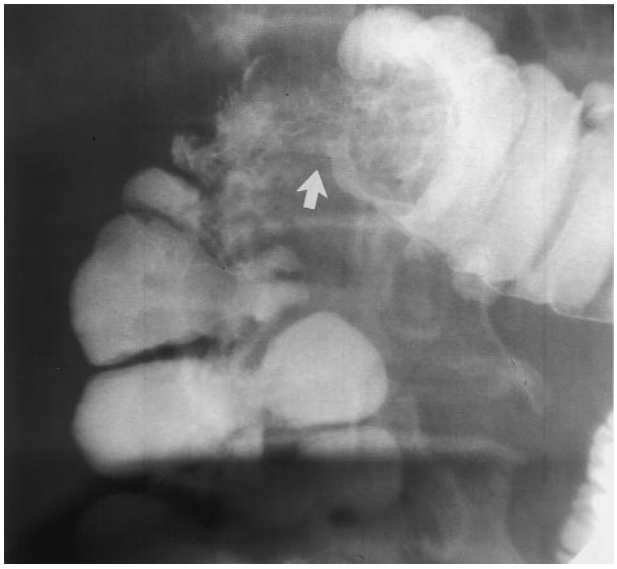


FIGURE 14 Image from a barium enema showing an "apple core" lesion in the hepatic flexure of the colon caused by a colon carcinoma (arrow).

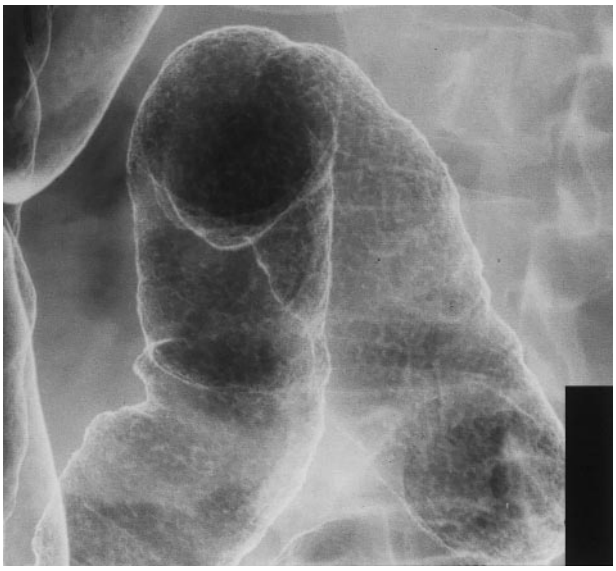


FIGURE 15 Double-contrast image from a barium enema revealing a finely granular mucosal pattern in a patient with ulcerative colitis.

disorders is performed with barium evacuation proctography as part of a multimodality evaluation.

Contraindications

Barium colonic examinations should not be performed in patients with toxic megacolon, ischemic colitis, suspected diverticulitis with signs of an acute abdomen, active colitis or with other diseases in which the bowel wall is potentially friable and could perforate during the procedure. In addition, barium examinations are contraindicated in patients with suspected colonic perforation. In these patients, the examination should be performed with water-soluble contrast materials, which do not incite a peritoneal reaction if extravasated and can be reabsorbed from the peritoneal cavity. Barium examinations should never be performed immediately postbiopsy with large biopsy forceps because this predisposes to perforation. In these cases, the barium examination should be postponed for at least 5 days. Barium enema should not be performed as the first test in patients with active bleeding to determine the site of bleeding. This is best shown with a nuclear medicine bleeding scan and/or angiography. The barium enema, if still necessary, should be postponed until adequate preparation can be performed.

Relative contraindications to barium radiography include inability of the patient to roll and turn during the examination, thereby not allowing adequate filling and coating of the colon, and colonic obstruction. Patients with colonic obstruction are better evaluated with water-soluble contrast media, which does not become inspissated proximal to an area of obstruction in the colon. Incomplete bowel preparation and residual oral contrast in the abdomen are other relative contraindications.

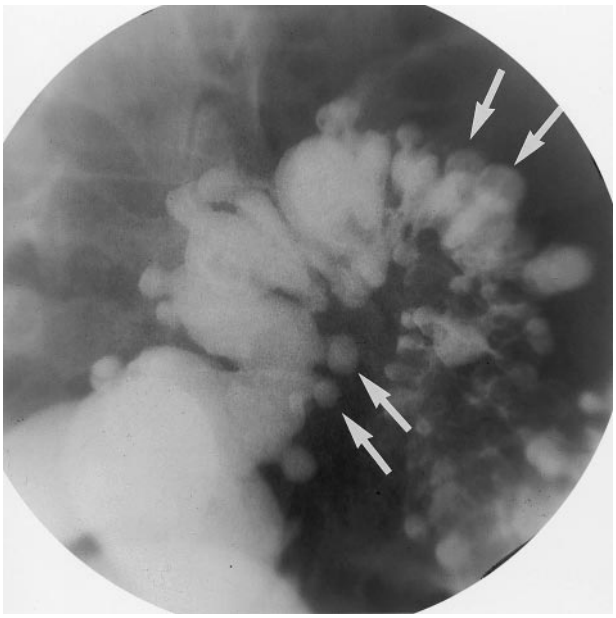


FIGURE 16 Spot radiograph shows extensive sigmoid and descending colonic diverticulosis (arrows).

Technique

As with most portions of the gastrointestinal tract, radiographic examination of the colon can be performed with a single- or double-contrast technique. In single contrast studies, the entire colon is filled with a relatively low-density barium suspension, with the filling monitored by fluoroscopic observation and the use of extensive palpation and compression of the colon during filling. Fluoroscopic spot films and overhead radiographs are performed with both techniques. In double-contrast examinations, the colonic mucosa is coated with a smaller volume of high-density barium, and luminal distension is obtained using air insufflation to generate the double-contrast effect.

The double-contrast technique is superior to the single-contrast method for detection of fine mucosal abnormalities such as small polypoid lesions and early inflammatory pathology. In addition, it is preferable for evaluation of the rectum, and of the splenic and hepatic flexures, which are relatively inaccessible to palpation and compression. The single-contrast examination is faster and less expensive and is preferred for

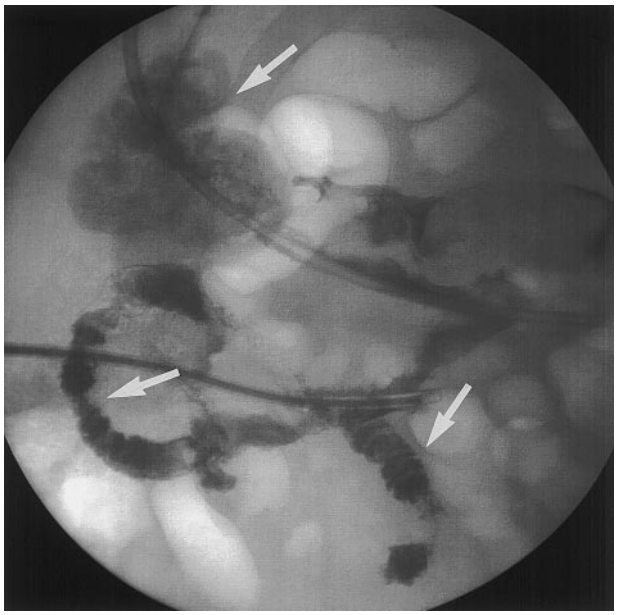


FIGURE 17 Spot image from a patient with two enterocutaneous fistulas shows communication between both catheterized fistulas and the small bowel and colon (arrows).

mechanical problems such as obstruction, fistulas, Hirschprung's disease, and diagnosis and reduction of intussusception, and in cases of suspected diverticulitis when there are no signs suggestive of an acute abdomen, which would preclude performance of a barium examination. Patients who are too old and debilitated are usually evaluated with the single-contrast technique, because they are unable to tolerate performance of the double-contrast technique.

Patients with a colostomy can be evaluated, if clinically indicated, through the colostomy, using a single- or double-contrast technique. When perforation of the large bowel is suspected, the examination should be performed with water-soluble contrast materials to avoid spillage of barium into the peritoneal cavity.

FISTULOGRAPHY AND SINUS TRACT EVALUATION

The gastrointestinal radiologist frequently can contribute significantly to the management of a patient with documentation of the presence of, and demonstration of the extent of, an enterocutaneous fistula (Fig. 17). In addition, a cutaneous sinus tract connected to an abscess cavity can be demonstrated. The sites of communication of a fistula or sinus tract may be demonstrated by catheterizing a cutaneous fistula or sinus tract and injecting the fistula or sinus tract with water-soluble contrast.

See Also the Following Articles

Colonoscopy • Computed Tomography (CT) • Endoscopic Ultrasonography • Magnetic Resonance Imaging (MRI) • Upper Gastrointestinal Endoscopy

Further Reading

Glick, S. N. (2000). Other inflammatory conditions of the colon. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 993–1008. W. B. Saunders, Philadelphia.

Gore, R. M., Laufer, I., and Berlin, J. W. (2000). Ulcerative and granulomatous colitis: Idiopathic inflammatory bowel disease. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 945–992. W. B. Saunders, Philadelphia.

Herlinger, H. (2000). Barium examinations of the small bowel. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 704–725. W. B. Saunders, Philadelphia.

Herlinger, H. (2000). Malabsorption. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 759–784. W. B. Saunders, Philadelphia.

Herlinger, H., and Ekberg, O. T. (2000). Other inflammatory conditions of the small bowel. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 746–758. W. B. Saunders, Philadelphia.

Herlinger, H., Jones, B., and Jacobs, J. (2000). Miscellaneous abnormalities of the small bowel. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 865–883. W. B. Saunders, Philadelphia.

Laufer, I. (1985). Barium contrast examinations. *In* "Gastroenterology" (J. E. Berk, ed.), 4th Ed., Vol. 1, pp. 456–469. W. B. Saunders, Philadelphia.

Laufer, I. (2000). Barium studies of the colon. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 892–904. W. B. Saunders, Philadelphia.

Levine, S. (2000). Other esophagitides. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 364–386. W. B. Saunders, Philadelphia.

Levine, M. S., Megibow, A. J., Kochman, M. L., Eisenberg, R. L., Smith, C. H., and Gore, R. M. (2000). Section on stomach and duodenum. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 514–702. W. B. Saunders, Philadelphia.

Levine, M. S., Rubesin, S. E., Herlinger, H., and Laufer, I. (1999). Plain and contrast radiology. *In* "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, L. Laine, C. Owang, and D. W. Powell, eds.), 3rd Ed., Vol. 2, pp. 2938–2960. Lippincott Williams & Wilkins, Philadelphia.

Nolan, D. J., and Herlinger, H. (2000). Vascular disorders of the small bowel. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 838–854. W. B. Saunders, Philadelphia.

Ott, D. J. (2000). Motility disorders of the esophagus. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 316–328. W. B. Saunders, Philadelphia.

Ott, D. J., Peele, V. N., Chen, Y. M., and Gelfand, D. W. (1990). Oropharyngeal function study: Radiologic means of evaluating swallowing difficulty. *South. Med. J.* **83**, 191–193.

Ott, D. J., and Pikna, L. A. (1993). Clinical and videofluoroscopic evaluation of swallowing disorders. *Am. J. Roentgenol.* **161**, 507–513.

Rubesin, S. (2000). Pharynx: Normal anatomy and examination techniques. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 190–211. W. B. Saunders, Philadelphia.

Thoeni, R. F., and Laufer, I. (2000). Polyps and carcinoma of the colon. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and, M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 1009–1048. W. B. Saunders, Philadelphia.

Zalev, A. H. (1992). Current concepts in radiology of the small bowel: Enteroclysis. *Appl. Radiol.* (Sept.) 71–78.

Barostat

FERNANDO AZPIROZ

University Hospital Vall d'Hebron and Autonomous University of Barcelona, Spain

gut manometry Evaluation of gut motility by measuring intraluminal pressures.

gut tone Tonic, i.e., sustained, muscular contraction of the gut wall.

tensostat Computerized pump that applies fixed tension levels to the wall of hollow viscera.

The barostat is an air pump that maintains a constant, predetermined pressure level within a flaccid bag introduced into a hollow muscular organ. Hence, the barostat is a pressure clamp that can be used either to measure muscular activity as volume variation at a low and constant pressure level or to distend the organ by increasing the intraluminal pressure. Distension in turn can be applied to measure the local distensibility of the organ (compliance) or to stimulate afferent pathways and measure either reflex responses (reflexes) or conscious perception (sensitivity).

INTRODUCTION

Some hollow muscular organs, such as the stomach, may exert phasic (i.e., brief) or tonic (i.e., sustained) muscular contractions. There is no strict definition to distinguish between phasic and tonic activity, but by convention, contractions lasting more than 1 min have been considered tonic. The tonic muscular activity of hollow organs is also called tone. In the gastrointestinal tract, phasic contractions produce intraluminal pressure changes that can be readily measured by manometry, a technique that has been available since the 1970s. However, tonic contractions outside sphincteric regions generally do not produce recordable pressure changes, probably because the changes are too small and are obliterated by respiratory movements and other artifacts. The barostat was developed in the early 1980s initially for the purpose of measuring gastric tone, but was later applied to measure tonic contraction of other hollow muscular organs.

The barostat is a feedback system that consists of a pressure sensor linked by either an electronic or a computerized relay to an air pump. When the organ relaxes, the barostat injects air to prevent a pressure drop, and when the organ contracts, air is aspirated to prevent a pressure rise.

TECHNICAL SPECIFICATIONS

The bag of the barostat should be flaccid and oversized, so that it does not interfere with the measurements. For reservoir organs, such as the stomach and the rectum, spheroidal bags can be used, and for tubular segments, such as the small intestine, cylindrical bags that are of a fixed length and radially oversized, compartmentalize the segment under study. The bag should be connected by a double-lumen tube to the pressure sensor and the air pump of the barostat. Otherwise, a single lumen tube damps the responsiveness of the system. In each barostat

Copyright 2004, Elsevier (USA). All rights reserved.

Ott, D. J., Peele, V. N., Chen, Y. M., and Gelfand, D. W. (1990). Oropharyngeal function study: Radiologic means of evaluating swallowing difficulty. *South. Med. J.* **83**, 191–193.

Ott, D. J., and Pikna, L. A. (1993). Clinical and videofluoroscopic evaluation of swallowing disorders. *Am. J. Roentgenol.* **161**, 507–513.

Rubesin, S. (2000). Pharynx: Normal anatomy and examination techniques. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 190–211. W. B. Saunders, Philadelphia.

Thoeni, R. F., and Laufer, I. (2000). Polyps and carcinoma of the colon. *In* "Textbook of Gastrointestinal Radiology" (R. M. Gore and, M. S. Levine, eds.), 2nd Ed., Vol. 1, pp. 1009–1048. W. B. Saunders, Philadelphia.

Zalev, A. H. (1992). Current concepts in radiology of the small bowel: Enteroclysis. *Appl. Radiol.* (Sept.) 71–78.

Barostat

FERNANDO AZPIROZ

University Hospital Vall d'Hebron and Autonomous University of Barcelona, Spain

gut manometry Evaluation of gut motility by measuring intraluminal pressures.

gut tone Tonic, i.e., sustained, muscular contraction of the gut wall.

tensostat Computerized pump that applies fixed tension levels to the wall of hollow viscera.

The barostat is an air pump that maintains a constant, predetermined pressure level within a flaccid bag introduced into a hollow muscular organ. Hence, the barostat is a pressure clamp that can be used either to measure muscular activity as volume variation at a low and constant pressure level or to distend the organ by increasing the intraluminal pressure. Distension in turn can be applied to measure the local distensibility of the organ (compliance) or to stimulate afferent pathways and measure either reflex responses (reflexes) or conscious perception (sensitivity).

INTRODUCTION

Some hollow muscular organs, such as the stomach, may exert phasic (i.e., brief) or tonic (i.e., sustained) muscular contractions. There is no strict definition to distinguish between phasic and tonic activity, but by convention, contractions lasting more than 1 min have been considered tonic. The tonic muscular activity of hollow organs is also called tone. In the gastrointestinal tract, phasic contractions produce intraluminal pressure changes that can be readily measured by manometry, a technique that has been available since the 1970s. However, tonic contractions outside sphincteric regions generally do not produce recordable pressure changes, probably because the changes are too small and are obliterated by respiratory movements and other artifacts. The barostat was developed in the early 1980s initially for the purpose of measuring gastric tone, but was later applied to measure tonic contraction of other hollow muscular organs.

The barostat is a feedback system that consists of a pressure sensor linked by either an electronic or a computerized relay to an air pump. When the organ relaxes, the barostat injects air to prevent a pressure drop, and when the organ contracts, air is aspirated to prevent a pressure rise.

TECHNICAL SPECIFICATIONS

The bag of the barostat should be flaccid and oversized, so that it does not interfere with the measurements. For reservoir organs, such as the stomach and the rectum, spheroidal bags can be used, and for tubular segments, such as the small intestine, cylindrical bags that are of a fixed length and radially oversized, compartmentalize the segment under study. The bag should be connected by a double-lumen tube to the pressure sensor and the air pump of the barostat. Otherwise, a single lumen tube damps the responsiveness of the system. In each barostat

Copyright 2004, Elsevier (USA). All rights reserved.

the speed of air flow and the pressure window for reaction should be determined (approximately 40 ml/s flow rate and 0.5 mm Hg window are appropriate). The compressibility of air and the deformation of the pump with pressure increments introduce an error in intraluminal volume measurements. This compression factor, which is usually linear, should be determined and carefully corrected for.

UTILITIES

Constant Pressure Studies

At a low constant pressure, the barostat measures tonic contraction as volume changes without disturbing the physiological activity of the gut or inducing perception. Measurements of tone may be applied to study (1) physiological events, such as fasting and postprandial changes, (2) reflex responses of the organ induced by distant stimuli, and (3) changes produced by pharmacological manipulations.

Intraluminal Pressure Manipulations

The barostat can be used to manipulate intraluminal pressure and produce distension. The local response to distension, measured as intraluminal volumes at different pressure levels, i.e., compliance, reflects both the capacity (size) of the organ and the distensibility of its walls; the distensibility, in turn, depends on the level of the muscular contraction in the walls. Distension stimulates receptors in the gut wall linked to reflex and sensory pathways that may induce various reflexes and conscious perception, respectively. These responses are stimulus-related, and since wall receptors operate as tension receptors, rather than as pressure or volume receptors, a modification of the barostat concept, the computerized tensostat, probably provides a better standardization of the stimuli under different experimental conditions. The tensostat is a computerized pump that maintains a fixed tension level within the gut wall; based on intraluminal pressure and volume, the system calculates wall tension by applying La Place's law and drives the pump to maintain the desired tension level.

APPLICATIONS

The barostat has been extensively used in both animal and human research. By contrast, its clinical applications are still limited, with very precise indications and restricted use in experienced laboratories. The barostat may provide useful information particularly on gastric and rectal function, specifically to evaluate compliance, pharmacological responses, and sensitivity, whereas the study of reflexes may be still more experimental. The barostat is particularly useful in the evaluation of some patients with severe postprandial symptoms. During ingestion, the stomach relaxes to accommodate the meal and subsequently the stomach progressively recontracts to produce gastric emptying. The barostat may elucidate whether symptoms after a meal are produced by poor accommodation due to defective gastric relaxation or by delayed emptying due to impaired contraction. Clinical symptoms for these two conditions may be the same, whereas therapeutic approaches used to correct defective gastric relaxation differ from the treatments used for delayed gastric emptying.

See Also the Following Articles

Gastric Motility • Manometry • Postprandial Motility

Further Reading

Azpiroz, F. (2002). Gastrointestinal perception: Pathophysiological implications. *Neurogastroenterol. Motil.* **14**, 1–11.

Azpiroz, F., and Malagelada, J.-R. (1985). Physiologic variations in canine gastric tone measured by an electronic barostat. *Am. J. Physiol.* **248**, G229–G237.

Azpiroz, F., and Salvioli, B. (2002). Barostat measurements. *In* "Schuster Atlas of Gastrointestinal Motility in Health and Disease" (M. M. Shuster, M. D. Crowel, and K. L. Koch, eds.), 2nd Ed., pp. 151–170. B. C. Decker, Hamilton, Canada.

Whitehead, W. E., Delvaux, M., and The Working Team (1999). Standardization of barostat procedures for testing smooth muscle tone and sensory thresholds in the gastrointestinal tract. *Digest. Dis. Sci.* **42**, 223–241.

Barrett's Esophagus

MICHAEL B. WALLACE AND DONALD O. CASTELL
Medical University of South Carolina, Charleston

gastroesophageal reflux disease Acid-related disorder.
goblet cells Mucus-secreting cells in the gastrointestinal tract.
Helicobacter pylori Bacterium that colonizes the mucosal lining of the stomach.

Barrett's esophagus is defined as the replacement of squamous tissue with columnar mucosa in the form of specialized intestinal metaplasia within the tubular esophagus. The critical ingredient in this process is the presence of goblet cells, the *sine qua non* of intestinal metaplasia, which can be confirmed only by endoscopic biopsy from the esophagus.

CLINICAL RELEVANCE

Barrett's esophagus (BE), as a complication of chronic gastroesophageal reflux disease (GERD), is the single best predictor for the development of adenocarcinoma of the esophagus. Esophageal adenocarcinoma is rising in incidence more rapidly than any other carcinoma in humans, although the cause of this rise is unknown.

EPIDEMIOLOGY

Rising Incidence

Despite the rising rate of adenocarcinoma of the esophagus, the absolute risk is low. There are approximately 12,000 carcinomas of the esophagus diagnosed annually in the United States, about half of which are adenocarcinomas. For a patient with Barrett's esophagus, the risk of developing carcinoma is estimated to be between 0.3 and 0.5% per year.

Risk Factors

GERD, Caucasian or Hispanic race, and male gender are the three strongest risk factors for developing BE. Well-controlled clinical studies have demonstrated that patients with long-standing, frequent, and severe GERD are at markedly increased risk of adenocarcinoma of the esophagus, presumably with BE as an intermediate step. The reason for the predominance among Caucasian or Hispanic males is unclear, although obesity and the declining prevalence of *Helicobacter* infection may play a role.

Role of *Helicobacter pylori*

There is increasing evidence that *H. pylori* may have a protective role against BE and carcinoma of the esophagus. The incidence of adenocarcinoma of the esophagus shows an inverse relationship with the decreasing prevalence of *H. pylori*. Chronic *H. pylori* infection leads to gastric atrophy and loss of acid production (achlorhydria) with decreased acid reflux. Patients with BE, especially those with dysplasia or carcinoma, are less likely to harbor *H. pylori* than are those without BE. The management of *H. pylori* is evolving, and it is premature at this time to recommend either eradication of *H. pylori* or leaving it *in situ* based on its potential role in esophageal adenocacinoma. There is little controversy, however, in recommending eradication of *H. pylori* in patients with peptic ulcer disease.

PATHOPHYSIOLOGY

Although BE has been clearly linked to GERD, little is known about the molecular and cellular mechanisms of the development of BE. Cross-sectional population studies suggest that BE develops early in life, with half of the maximum prevalence reached by 40 years of age and relatively little increase in prevalence after age 60. Once established, the length and presence of BE change little. The cell of origin for BE is unknown, with candidates including pluripotent stem cells in the esophagus, subsquamous glandular cells, or heterotopic gastric mucosa. Markers of inflammation, such as up-regulation of cylcooxygenase-2 (COX-2), are present in BE, offering hope that COX-2 inhibitors may blunt the development or progression of BE. It remains unclear if reflux of acid, reflux of bile, or both are required for the development of BE, although the experimental production of this lesion in a canine

 Copyright 2004, Elsevier (USA). All rights reserved.

model provides strong evidence for the critical role of acid.

DIAGNOSIS

Detection of Goblet Cells

The diagnosis of BE requires endoscopic biopsy of the esophagus and histological examination for the presence of goblet cells, the *sine qua non* of intestinal metaplasia. This seemingly simple criterion is nonetheless widely confused. This is due in large part to the changing definition of BE over time. Early literature defined BE simply as columnar epithelium, usually occupying at least 2–3 cm of esophagus. The modern definition is a functional one, based on the fact that only patients with intestinal metaplasia are at elevated risk of adenocarcinoma, regardless of the length of BE. The most difficult area of definition lies in distinguishing very short irregularities in the normal squamocolumar junction of the esophagus from true BE. Because intestinal metaplasia is present in the gastric cardia in up to 20% of asymptomatic individuals, it is critical to take biopsies from the tubular esophagus, which requires careful attention to anatomic landmarks.

TREATMENT

The treatment of BE should be distinguished from the treatment of GERD. Almost all BE patients have GERD, and these symptoms are usually controlled with acid suppression, surgical fundoplication, or lifestyle modifications. In contrast, no therapy has definitively been shown to alter the likelihood of BE progressing to adenocarcinoma.

Acid Suppression

Acid suppression has been a mainstay for treating patients with BE for two reasons. Acid suppression controls symptoms of GERD and reduces the degree of inflammation in the esophageal epithelium. Although this reduced inflammation has not been shown to reverse BE or change its natural history, it makes the histological interpretation of dysplasia considerably easier because inflammatory changes can be misinterpreted for low-grade dysplasia. Recent studies suggest that long-term, high-dose acid suppression with a proton pump inhibitor (titrated by pH monitoring) may cause a small reduction in the area of columnar epithelium, although there may be hidden columnar epithelium underneath new squamous cell islands. To date, it is unknown whether the overall risk of carcinoma is changed by this therapy.

Antireflux Surgery

Surgery offers the theoretical advantage of sustained reduction of both acid and bile reflux. However, surgical studies have not shown any significant effect on either the length of Barrett's or, more importantly, the likelihood of developing adenocarcinoma. A randomized controlled trial has compared surgery to medical therapy (histamine-2 receptor antagonists) with long-term outcomes. This Veteran's Administration (VA) cooperative study showed no significant difference in the development of Barrett's or adenocarcinoma, or change in the length of Barrett's. In fact, long-term mortality was lower in the medical therapy group, largely due to a higher cardiac death rate in the surgical group for unknown reasons. Long-term followup of BE patients after surgery suggests no overall effect on the likelihood of developing adenocarcinoma or dysplasia, likely due to ongoing low rates of acid and bile reflux.

Ablation of Barrett's Epithelium

Several investigators have evaluated methods to ablate the metaplastic epithelium of patients with BE either with or without dysplasia. Almost all methods employed (heat coagulation, laser therapy, cyrotherapy, and photodynamic therapy, all in conjunction with high-dose acid suppression) result in substantial replacement of the BE with squamous epithelium. Unfortunately, 20–30% of patients develop squamous epithelium overlying intestinal metaplasia, presenting a difficult clinical dilemma for surveillance and an unclear effect on long-term risk of carcinoma. At this point, attempts to ablate BE should be considered experimental, especially for nondysplastic BE.

ENDOSCOPIC SURVEILLANCE

Endoscopic surveillance of BE has been the cornerstone of management. Multiple uncontrolled observations have demonstrated that under surveillance, patients are more likely to be diagnosed with early-stage tumors or dysplasia, and have better survival compared to patients not under surveillance. No prospective controlled trial data are available to prove the effectiveness of BE surveillance. Decision analysis models have suggested that surveillance is cost effective compared to other commonly accepted medical practices, such as mammography and colon cancer screening. The most cost-effective interval for surveillance is every 5

years, although organizational recommendations have been more conservative, recommending surveillance every 2–3 years for patients without dysplasia, and every 6 months for the first year, then yearly for those with low-grade dysplasia. Surgical esophagectomy remains the standard of care for treatment of high-grade dysplasia, although photodynamic therapy is emerging as a possible alternative. In addition, several recent studies suggest that patients with focal high-grade dysplasia and those with normal DNA ploidy (as opposed to aneuploid) have a low risk of progression and can be followed by close surveillance (every 3–6 months). Newer techniques for optical detection of dysplasia, such as light-scattering spectroscopy and fluorescence spectroscopy, are promising tools in development, but remain experimental at this time.

CONCLUSIONS

Barrett's esophagus is the single most significant risk factor for adenocarcinoma of the esophagus. Patients with long-standing symptoms of GERD should undergo screening to detect Barrett's, and should be considered for regular surveillance to detect the early development of dysplasia or cancer. Areas of active investigation are identification of the cellular and molecular mechanisms for development of BE, better markers of risk among patients with BE, and nonsurgical therapies to eradicate BE once it has developed.

See Also the Following Articles

Esophageal Cancer • Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD • Gastroesophageal Reflux Disease (GERD) • H2-Receptor Antagonists • *Helicobacter pylori* • Proton Pump Inhibitors

Further Reading

Csendes, A., Braghetto, I., Burdiles, P., *et al.* (1998). Long-term results of classic antireflux surgery in 152 patients with Barrett's esophagus: Clinical, radiologic, endoscopic, manometric, and acid reflux test analysis before and late after operation. *Surgery* **123**, 645–657.

Lundell, L., Miettinen, P., Myrvold, H. E., *et al.* (2001). Continued (5-year) followup of a randomized clinical study comparing antireflux surgery and omeprazole in gastroesophageal reflux disease. *J. Am. Coll. Surg.* **192**, 172–179{discussion}, 179–181.

Schnell, T. G., Sontag, S. J., Chejfec, G., *et al.* (2001). Long-term nonsurgical management of Barrett's esophagus with high-grade dysplasia. *Gastroenterology* **120**, 1607–1619.

Spechler, S. J., Lee, E., Ahnen, D., *et al.* (2001). Long-term outcome of medical and surgical therapies for gastroesophageal reflux disease: Follow-up of a randomized controlled trial. *JAMA* **285**, 2331–2338.

Barrier Function in Lipid Absorption

A. B. R. THOMSON*, M. T. CLANDININ*, L. DROZDOWSKI*, T. WOUSTRA*, AND G. WILD†
*University of Alberta, Edmonton and †McGill University, Montreal

absorption The movement of nutrients and fluid from the intestinal lumen into the lymphatics or bloodstream. This comprises two steps, uptake from the intestinal lumen and transfer from the cytosol of the enterocyte into the lymphatics or portal circulation.

adaptation The process by which the morphology and/or absorptive function of the intestine changes in response to stimuli such as a change in the composition of the diet, the development of diabetes, or a loss of a portion of the intestine.

basolateral membrane The portion of the cell lining of the enterocyte that communicates with the submucosal tissue.

brush border membrane The membrane of the intestinal absorptive cell, the enterocyte, which faces the intestinal lumen.

enterocyte The major cell lining the intestinal villus that is responsible for the absorption of nutrients.

intravillous space The space between the long villi. Nutrients must diffuse into this space in order to be taken across the brush border membrane in portions of the villus away from the villus tip and toward the crypt.

lipid-binding proteins Proteins in the brush border membrane and the enterocyte cytosol that bind lipids. Their physiological role is unknown; they may influence or modify intestinal absorption, bind lipids as they move from the brush border membrane to microsomes, or possibly play a role in the handling of adaptation.

unstirred water layer Consists of a series of lamellae of water, progressively less and less stirred as one moves from the bulk phase in the intestinal lumen to the brush border membrane. The unstirred water layer reduces the concentration radiant between the bulk phase and the brush border membrane and thereby slows the uptake of nutrients.

Lipids represent a major source of calories and are thereby nutritionally important. Some lipids are essential as they are not synthesized in the body and play an important role in a variety of functions, such as integrity of membranes. Most lipids are emulsified in the stomach, digested in the intestinal lumen as the result of enzymes secreted from the pancreas in response to the presence of food in the intestinal lumen, and then solubilized by bile acids secreted by the liver. The long-chain fatty acids and cholesterol are solubilized in bile salt micelles, diffuse across the intestinal unstirred water layer, and then passively permeate the lipophilic brush border membrane (BBM). There are lipid-binding proteins in the BBM as well as in the enterocyte cytosol and these may contribute to the uptake of lipids. Once the lipids are in the enterocyte, they pass to the endoplasmic reticulum, where they are further metabolized before being excreted across the basolateral membrane into intestinal lymphatics or into the portal vein. Although the process of intestinal absorption of lipids has many steps, it is complex and integrated and it is a highly efficient process in which normally more than 95% of lipids in the diet are absorbed.

INTRODUCTION

Lipid absorption involves uptake across the brush border membrane (BBM) into the enterocyte, where the lipids must be transferred intracellularly to sites of metabolism or exit the cell across the basolateral membrane (BLM) into the lymph or portal blood. Fat digestion and absorption constitute a complex process involving insoluble substrates, neutral and amphipathic lipids, and lipases acting in the stomach and small intestine. Intestinal lipid absorption involves several coordinated steps, including digestion and solubilization of the lipid, diffusion across the unstirred water layer (UWL), mediated and nonmediated transport across the BBM, diffusion across the cytosol, intracellular metabolism, binding to lipoproteins, and exit across the basolateral membrane into the lymph or portal blood.

The uptake of lipids into the enterocyte occurs predominantly at the upper third of the villus. Nutrients must cross two barriers in series in order to be taken up by the enterocyte: the intestinal UWL and the BBM. The rate of nutrient uptake is determined by the dimensions and properties of these two barriers, as well as by the activity of protein-mediated components of transport.

 Copyright 2004, Elsevier (USA). All rights reserved.

LIPID UPTAKE ACROSS THE INTESTINAL BBM

Lipid uptake across the BBM is the rate-limiting step for uptake of short- and medium-chain fatty acids, whereas passive uptake of long-chain fatty acids is rate-limited by passage through the UWL. Three models of passive uptake of long-chain fatty acids (LCFA) and cholesterol have been outlined. The first model proposes that the entire mixed micelle is absorbed by the BBM. However, no experimental evidence supports this model. In the second model, the micelle collides with the BBM, allowing for the uptake of lipids to occur. Evidence for this model is suggested by the linear relationship between lipid uptake and bile acid concentration. In contrast, the third model proposes that the lipids dissociate from the micelle into the aqueous compartment of the UWL before being taken up by the BBM. Support for this model is suggested by the finding that fatty acid uptake decreases with an increase in the number of bile acid micelles, under conditions in which fatty acid concentration is kept constant. The dissociation of lipids from bile acid micelles is under the influence of the acidic microclimate adjacent to the BBM. Under these acidic conditions, the critical micellar concentration increases, the fatty acids become protonated, and this protonation increases their rate of permeation across the BBM. Increase in the fluidity of the BBM also increases the rate of permeation of lipids. Other factors influencing the rate of lipid uptake may be the lumenal lipid composition, lipid-binding proteins, and membrane potential. For example, polyunsaturated fatty acids and phophatidylcholine may inhibit cholesterol absorption, possibly by shifting the partition coefficient of cholesterol away from the cell membrane back to the micelle, thereby preventing the uptake of lipid. There are numerous lipid-binding proteins in the BBM and cytosol of the enterocyte (Table I) and these likely play a role in lipid absorption.

Short-chain fatty acids (SCFA) are produced by bacterial degradation of complex carbohydrates and proteins entering the colon. SCFA are rapidly absorbed. SCFA stimulate electroneutral sodium absorption via activation of the BBM Na^+/H^+ exchanger (NHE), which results from the SCFA changing the intracellular pH gradients. Medium-chain fatty acids are neutral lipids containing fatty acid molecules with chain lengths ranging from 6 to 12 carbon atoms. These fatty acids are more rapidly and completely hydrolyzed than are long-chain triglycerides (LCT), even in the absence of pancreatic lipase. Medium-chain fatty acids are absorbed at the same rate as SCFA in the human rectum. The absorption of medium-chain triglycerides (MCT) is less dependent on the action of bile salts, because MCT are more water soluble than LCT. Although MCT are better absorbed than LCT in the presence of pancreatic insufficiency, pancreatic extracts may increase their absorption. Thus, if patients with pancreatic insufficiency are given pancreatic supplements, there may be little additional benefit to lipid absorption when giving MCT compared with using LCT. The availability of orally administered hydrophilic drugs to the systemic circulation is generally limited by the barrier properties of the intestinal membrane. The sodium salts of medium-chain fatty acids (MCFA) enhance the absorption of hydrophilic drugs across the intestinal mucosa and structurally similar MCFA display differences in their mechanism of action.

The uptake of cholesterol into the intestine occurs by a process of passive permeation across the BBM, after dissociation of cholesterol from the micelle directly into the membrane or into the acidic aqueous phase external to the BBM. Proteins facilitate cholesterol and phosphatidylcholine absorption. Cholesterol esters may also be subject to protein-mediated uptake. Cholesterol absorption is a multistep process that includes hydrolysis of cholesterol esters in the gut lumen, formation of mixed micelles, transport of cholesterol into enterocytes, its reesterification, and then assembly and secretion of chylomicrons as well as nascent high-density lipoproteins. There may be lipid-binding proteins that facilitate cholesterol absorption.

Bile acids are water-soluble end products of cholesterol metabolism that participate in fat digestion in the gastrointestinal tract. Bile acids are synthesized from cholesterol in the liver and are secreted with bile into the small intestine. Lumenal bile acids form micellar structures that solubilize lipids and facilitate their absorption across the brush border membrane. In the terminal ileum, the lumenal bile acids are actively reabsorbed by the enterocytes and are returned to the liver via the portal circulation. This process is known as the enterohepatic circulation. Lipid absorption requires an optimal concentration of bile acids in the intestinal lumen and this concentration is maintained by the balance between ileal excretion and ileal active reabsorption of bile acids. The enterohepatic cycling of bile salts is a major factor in the maintenance of the bile salt pool. Bile salts are absorbed passively in the jejunum and are actively absorbed in the ileum. Bile acids are taken up by passive or protein-mediated processes. The BBM Na^+/bile acid cotransporter has been cloned from rat, hamster, and human ileum and is subject to both translational and posttranslational regulation. A 14 to 15 kDa cytosolic binding protein, the intestinal bile acid-binding protein, belongs to a family of hydrophobic

TABLE I Intestinal Lipid-Binding Proteins

Protein	Molecular weight (kDa)	Localization	Substrate
Brush border membrane proteins			
Caveolin	22	Small intestine	Cholesterol, LCFA
SR-BI	27	Liver, peripheral tissue	High-density lipoproteins, phospholipids, triacylglycerol, cholesterol, cholesterol esters
FABPpm	40	Adipose tissue, heart, liver, intestine	Cholesterol, LCFA
FAT/CD36	88	Adipose tissue, heart, skeletal, muscle, spleen, intestine	LCFA, triglycerides, cholesterol
FATP4	63	Small intestine	LCFA (oleate)
Cholesterol transport protein	145	Small intestine	Cholesterol
Intracellular proteins			
FABPc	14–15	Adipose tissue, muscle, heart, brain, kidney	LCFA
L-FABPc	14–15	Liver, small intestine	LCFA, heme, bile acids, acyl CoA
I-FABPc	14–15	Small intestine	LCFA
ILBP	14–15	Ileum (predominant in distal ileum)	Bile acids
Prechylomicron transport protein		Small intestine	Triacylglycerol
MTP	58	Small intestine	Triacylglycerol cholesterol ester

ligand-binding proteins, the fatty acid-binding proteins. The binding of bile acids to ileal lipid-binding protein (ILBP) increases the affinity of ILBP for bile acids. This may be a substrate-load modification of transport activity and a positive-feedback regulation for active uptake of bile acids in the ileum.

Low-molecular-weight cytosolic bile acid-binding proteins are also involved in the intracellular transport of bile acids and may be subject to transcriptional regulation. Bile acids may provide positive feedback regulation for active bile acid uptake by binding the ileal ILBP.

The Unstirred Water Layer

The passive and carrier-mediated transport processes are described kinetically on the basis of the passive permeability coefficient (P), the incremental Gibbs free energy values ($\delta\Delta F_{w\rightarrow 1}$) associated with the addition of specific substituent groups to a probe molecule, the Michaelis affinity constant (K_m), and the maximal transport rate (V_{max}). These measurements provide critical information about such important characteristics of the BBM as its relative hydrophobicity and permeability and about the number and characteristics of specific transporters in the BBM. Although it has been assumed that uptake is determined by the properties of the BBM, it is critical to correct uptake for potentially serious qualitative and quantitative errors arising from failure to account for the effect of the UWL resistance. Failure to correct for the effect of the UWL may lead to underestimation of passive permeability coefficients, incremental free-energy values and reflection coefficients for carrier-mediated processes, overestimation of Michaelis constants for carrier-mediated processes, or erroneous identification of "transition" temperatures in the membrane. Failure to correct for the UWL may lead to overestimation of the value of V_{max} if the contribution of concurrent passive uptake is not taken into account or if recruitment of the carrier along the intervillus space is not taken into consideration. These potential errors may invalidate the use of Michaelis-Menten kinetics for the analysis of carrier-mediated transport processes.

Adjacent to all biological membranes there is a layer of relatively unstirred water through which solute moves by diffusion. In the intestine, this UWL consists of a series of water lamellae that extends out from the BBM, each layer of which is progressively more stirred until it blends in with the bulk water phase in the intestinal contents. The UWL is an operational term, because the boundary between the bulk water phase and the UWL is not well defined. Therefore, the dimensions

of UWL thickness and surface area give rise to a value of "effective resistance."

The UWL is formed by "hydrated mucus and a series of water lamellae extending outward from the BBM, each progressively more stirred, until the layers blend imperceptibly with the bulk phase." It is unclear whether the UWL is the physical structure of water, mucus, glycocalyx, villi, and microvilli; laminar flow of the lumenal perfusate; or movement of the intervillous space. The thickness of the intestinal UWL has been estimated to be 30 to 1000 μm, depending on the methods used to make the measurement, the use of *in vitro* or *in vivo* tissue preparations to obtain varying degrees of stirring or mixing of the lumenal fluid, the shape of the villi, and the extent of villous motility observed in the different species studied. The thickness of the UWL may be 500 Å in humans, although much lower values have been reported.

The effective resistance of this UWL must be taken into consideration when describing nutrient absorption. The rate of solute diffusion across the UWL is determined by the thickness and surface area of the UWL, the aqueous diffusion constant of the solute, and the concentration of the solute. Lipids diffuse across the intestinal UWL before contacting the BBM. The effect of the resistance of the UWL is to reduce the concentration of the lipid presented to the BBM. Therefore, it is important to correct for the effect of the UWL on absorption, to assess the true permeability coefficient of the BBM. Failure to correct for the effective resistance of the UWL will result in underestimation of the true permeability properties of the BBM. Bile acid micelles greatly enhance the uptake of fatty acids and cholesterol from the small intestine, by helping to overcome the resistance of the UWL and by increasing monomer concentration at the aqueous–membrane interface.

A variety of approaches have been developed to estimate UWL resistance. These include techniques that are based on the following: (1) the morphometric analysis of small bowel dimensions; (2) the rates of carbon monoxide diffusion out of the intestinal lumen; (3) the effects of changes in viscosity of the perfusate on solute absorption; (4) the osmotic gradient technique to determine time required for development of diffusion potentials; (5) the kinetics of entry of macromolecules into the intervillus spaces; and (6) the comparison of the K_m values for solute transport or peptide digestion with the values observed in the intact intestine.

The osmotic transient technique has been used *in vitro* and it assumes the presence of a planar surface. The osmotic transient technique has also been used in intestinal perfusion studies in animals and in humans. UWL thickness estimates of 300 to 800 μm have been obtained from the determination of the half-time of changes in the potential difference in a perfused gut segment when one solution is rapidly substituted for another with a different osmolarity. This method probably does not yield accurate estimates of UWL thickness because the osmolarity of the lumenal bulk solution is achieved slowly. Furthermore, fluid perfused through the rat intestine moves by laminar flow, rather than segregating into a UWL and well-mixed bulk lumenal contents.

Another method used to assess the effective resistance of the UWL is to determine the rate of intestinal uptake of a homologous series of probes, such as saturated fatty acids. The fatty acid partition coefficient increases with its chain length by a factor corresponding to a decrease in the incremental change in free energy moving from an aqueous phase to a lipid phase. Values for the apparent passive permeability coefficient (P) increase with the chain length of the fatty acids, until the value of P becomes proportional to their free diffusion coefficients. This point indicates that uptake was limited by the rate of diffusion across the UWL up to the BBM. From rates of uptake of such diffusion-limited probes, the UWL resistance in rat jejunum *in vivo* may be calculated and closely approximates the value of the UWL resistance calculated from the half-time of change in diffusion potential. The diffusion barrier resistance falls by approximately 45% over a range of perfusion rates observed *in vivo* (1.5–15 ml/min).

In vivo, the magnitude of the UWL thickness varies depending on the method used to access this dimension: in rat jejunum the UWL is 700 to 800 μm thick in a 30 cm segment perfused in the conventional fashion on the abdominal cavity, falling to 200 to 400 μm when the segment is placed in the abdominal wall and falling to 32 to 68 μm with shaking of the intact rat. Values for the effective thickness of the UWL in rat perfusion experiments have been obtained by various methods: 410 to 430 μm by measuring the development of osmotically induced potential difference, 460 to 486 μm by the segmented flow technique, and 212 to 708 μm by changing the flow rate.

Another method used to obtain measures of UWL thickness *in vivo* involves the measurement of the rate of hydrolysis of probes *in vitro* and *in vivo*; from the difference in these rates, the UWL effect is calculated. Such estimates have yielded UWL thickness values of 48 μm in the human jejunum. This method assumes that the K_m of the disaccharides, measured *in vitro*, accurately reflects the K_m of these enzymes in the intact intestine. This is a reasonable assumption, given that the kinetics of these enzymes are not altered by dissociation from the BBM. This method also assumes that all

disaccharide hydrolysis occurs in the BBM rather than in the lumenal contents, again a reasonable assumption.

Measurements of UWL thickness made in the jejunum of conscious dogs by assessing the absorption rate of two rapidly absorbed probes, glucose and [^{14}C] warfarin, gave values of only approximately 35 and 50 μm for perfusion rates of 26 and 5 ml/min, respectively. Measurements of the maximal UWL thickness for the human jejunum calculated from previous studies of glucose absorption yielded a mean value of only 40 μm. These measurements are less than one-tenth of previously reported values obtained using the osmotic transient technique.

The effective resistance of the UWL is determined not just by its thickness, but also by its surface area and by the diffusion coefficient of the probe under consideration. The diffusion coefficient may vary with viscosity of the bulk phase or of the mucus in the UWL. Earlier physiological studies showed that the uptake of most nutrients occurs from the upper portion of the villus. The "recruitment" of additional transporters or membrane surface area available for uptake will depend on the rate of uptake of solute from the enterocytes at the top of the villus, their access to the intervillus space, and the movement of the villi. Thus, the surface area of the BBM used for nutrient uptake is much less than the total membrane surface area.

The viscosity of fluid influences diffusion resistance. Dietary fiber or its components, such as guar gum, may be used to increase the viscosity of the lumenal contents. Such changes in viscosity may reduce absorption by affecting UWL resistance. Fiber may also increase the pressure in the intestinal lumen, distend the bowel, and thereby increase the uptake of some substances such as antipyrine (a weak base, almost completely undissociated). *In vivo*, other aspects of fiber may play a role in nutrient absorption such as a slowing effect on gastric emptying or a prolonged mouth-to-cecum transit time.

There is evidence to suggest that the dimensions of the UWL may be specific to individual experimental conditions and cannot be predicted. Thus, to correct for effective resistance of the UWL and thereby to obtain valid estimates of the kinetic properties of the tissue in question, this resistance factor must be measured. Only in this way will it be possible to establish whether an adaptation in transport that occurs as a result of a dietary, pharmacological, or other experimental manipulation designed to mimic a disease state influences nutrient uptake as a result of alterations in the value of P, K_m, or V_{max}. Once the kinetic mechanism of altered transport is known, then mechanisms responsible for this adaptation can be determined. Thus, the value of the UWL resistance must be measured for each change in experimental design.

Intervillus Space and Villous Motility

Increasing the rate of stirring of the bulk phase enhances the *in vitro* uptake of nutrients. Faster rates of perfusion through the intestine enhance *in vivo* uptake of nutrients by a process that involves a reduction in the thickness of the UWL and an enhancement of mucosal surface area available for uptake. Distension of rat ileum leads to a reduction in villous height and a marked increase in width of the intervillous space (IVS) in both the transverse and the longitudinal dimensions. The net effect of this distension, however, is that there is no absolute change in total mucosal surface area.

Because the proportion of the IVS that is involved in solute uptake will depend on the extent to which cells at the villous tip exceed their capacity to transport the solute, the V_{max} or P may vary depending on the extent of the IVS used for uptake. With increasing recruitment of carriers along the IVS or recruitment of increasing amounts of cell membrane for passive uptake, the estimated value of these kinetic constants may vary.

Because the IVS is only approximately 50 μm wide in dogs and 15 μm in rats, 7 to 25 μm is the maximum UWL that could separate the solute from the absorptive epithelium in this space. The remaining UWL must be confined to the villous tips. Using measurements of maltose hydrolysis in the rat jejunum, hydrolysis was accurately predicted by a model in which the unstirred fluid extended from 20 μm over the villous tips throughout the IVS. In this model, the depth of diffusion into the IVS is inversely proportional to the efficiency of epithelial handling of the solute. As a result, both the aqueous barrier and the functional surface area are variables rather than constants.

If the thickness of the UWL *in vivo* is low, then what will be its impact on nutrient uptake? Levitt and colleagues have estimated for a rapidly absorbed nutrient such as glucose that the ratio of the cross-sectional surface area of the IVS to the lumenal mucosa surface is 1 : 3, that 50% of the glucose will be absorbed within 9 μm of the villous tips, and that this preepithelial diffusion barrier may remain the rate-limiting step for rapidly permeating probes such as glucose. A UWL of 35 μm would still produce approximately 75% of the total resistance to glucose absorption, because the total resistance (UWL plus epithelial cell) to transport of low concentrations of glucose is equivalent to a UWL of 48 μm. As the infusate concentration increases, the carriers at the villous tip become saturated and the probe must diffuse down the IVS.

When uptake occurs largely from cells along the upper portion of the villus, the thickness of the UWL can be approximated by simple laminar flow analysis and for modeling purposes the intestine can then be assumed to be a smooth cylinder. The IVS may become an important site of uptake when flow rate is high or when there is intestinal distension. However, the increase in the absorption rate after distension is smaller than the enlargement of the inner cylindrical surface area, presumably because of a change in supravillous diffusion resistance.

In canine jejunum *in vivo*, substances absorbed into the villus tips must penetrate an unstirred layer of 500 to 1000 μm, and for those substances absorbed into the lateral surfaces of the villi, an additional barrier of as much as 800 μm exists. The fluid in the IVS is poorly stirred and presumably movement of the villi has little impact on mixing of the IVS fluid. Westergaard and Dietschy, in 1974, and Winne, in 1978, also concluded that villus movement produced little stirring of the UWL. When water absorption occurs then penetration of the UWL by the process of solvent drag may enhance absorption into the tip region of the villus. However, because water absorption occurs largely in the uppermost portion of the villi, diffusion is likely the only process for permeation across the BBM of enterocytes further down the villus.

Intestinal villi exhibit spontaneous movement in the living animal. The villous movements are thought to be due to contraction of smooth muscle fibers from the muscularis mucosae that run longitudinally within the villous core. There is a piston-like retraction and extension of the villi, pendular side-to-side movement, and tonic contraction of several or all villi. Local, neural, and hormonal factors modulate villous motility.

A videomicroscopic method has been used to analyze quantitatively villous motility in the dog intestine. The villous retractions are most frequent and of longest duration in the duodenum, followed by the jejunum and the ileum. It is predicted that villi are in the retracted state for approximately 30, 13, and 6% of the time in these three sites, respectively. The frequency of pendular movements is greatest in the jejunum, followed by the ileum and the duodenum. Lumenal pH or the presence of glucose has no effect on villous motility, whereas amino acids and free fatty acids increase villous contraction frequency by 30–50 and 90%, respectively. It is unclear whether these changes in villous contraction are achieved by feeding alone or by specific nutrients. Increasing villous motility by fluid expansion in the intact canine intestine actually decreases absorption of water and lauric acid. Thus, it is unclear whether the villous motility perturbs the UWL sufficiently to actually modify nutrient absorption.

According to the principles of laminar flow, the flow rate of fluid is most rapid and most completely stirred in the center of the intestinal lumen, with progressively slower flow rates and incomplete stirring in fluid near the BBM. Laminar flow occurs in the gut. The two-dimensional laminar flow model is valid for determining kinetic parameters of carrier-mediated transport *in situ* and for predicting absorption rate *in situ* from uptake rate *in vitro*. Thus, in the constantly perfused intestine, fluid moves with laminar flow. This is different from the conventional unstirred water layer model, which proposes a thickness of totally unstirred water adjacent to the lumen with well-mixed contents in the center of the lumen.

The flow rate of fluid through the proximal small intestine varies widely from an average of 2.5 ml/min in fasting subjects to as high as 20 ml/min after meals. Increasing the flow rate increases the absorption of tritiated water and D-glucose and opens the IVS more widely. Increasing the jejunal flow rate from 5 to 20 ml/min decreases the permeability ratio of xylose/urea by approximately 30% and decreases the average calculated pore radius of the diffusion pathway. Presumably there is increased exposure of the lumenal fluid to the less permeable cells in the IVS.

Most studies of perfused bowel *in vivo* have been performed in anesthetized laparotomized rats, where there is little motility and near-perfect laminar flow, just as if the perfused fluid mass was moving through a pipe. In conscious rats, maximal preepithelial resistance is equivalent to an UWL thickness of only approximately 100 μm, with anesthesia doubling this resistance and anesthesia and laparotomy increasing resistance to approximately 600 μm. Clearly, corrections in UWL resistance become even more important under these experimental conditions.

Acid Microclimate

A layer juxtaposed to the mucosal surface where the proton concentration is higher than in the bulk phase of the intestinal lumen is designated the "acid microclimate." The existence of the acid microclimate was postulated on the basis of differences in the steady-state distribution of weak acids and bases from the values predicted by the pH-partition hypothesis. In a theoretical study in 1977, Winne concluded that the absorption of weak electrolytes could be modified by the existence of unstirred layers. The presence of the acid microclimate was demonstrated indirectly on the basis of acidification of the incubation medium by isolated

segments of rat jejunum, as well as by direct measurements using surface microelectrodes or tip microelectrodes on tissue *in vitro* or *in vivo*. Point-by-point determinations using 50 μm tip diameter antimony microelectrodes show that in rat jejunum *in vitro*, the highest proton concentration (24−224 nmol/liter = pH 6.67−6.65) was found 10 to 100 μm below the tip of the villus. No gradient is seen along the villi in the ileum. The thickness of the acid microclimate in rat jejunum *in vitro*, 700 μm, is similar to the thickness of the UWL. Such an acid microclimate has been described in human intestinal biopsy material and may change in disease states such as celiac disease.

Thus, the pH of the UWL is below 6, which is lower than in the bulk phase of the intestinal lumen, and is lowest in the mucus layer immediately adjacent to the BBM. The low pH of this acid microclimate is maintained by activity of the Na^+/H^+ exchanger in the BBM. The NHEs in the intestinal BBM are responsible for acidifying the UWL adjacent to the BBM. This facilitates partitioning of fatty acids out of the bile salt micelles, their protonation, and hence their greater permeation across the BBM. NHE appears to play a more important role when there is a H^+/Na^+ gradient across the BBM, whereas the fatty acid-binding protein in the BBM, FABPm, is important when there is less of such a gradient. Inhibition of NHE or FABPm results in an approximately 30 to 40% decline in the uptake of fatty acids. Mucus glycoproteins may contribute to this pH gradient by inhibiting the diffusion of protons into the bulk phase. The mucus may act as an ampholyte, restricting H^+ movement in its matrix. The acid microclimate is also maintained by physical properties of the mucus that retard diffusion of H^+ from the BBM to the bulk phase in the intestinal lumen. The acid microclimate may alter the proportion of ionized to non-ionized solutes, which may thereby influence the ability of the solutes to be taken up by the BBM. This is particularly important for fatty acids ($pK_a \sim 4.2$), since the majority would exist in the non-ionized form in the acidic microclimate of the UWL and thereby allow these protonated fatty acids to permeate the BBM more readily.

Brush Border Membrane

The second barrier to lipid uptake is the BBM of the enterocyte. The BBM of the enterocyte is polarized in composition and in function. Tight junctions divide the plasma membrane into two domains: the BBM and the BLM. The surface area of the BBM increases the surface area of the villus up to 40-fold and is subject to direct or indirect regulation by food or food substances. The BBM serves as a permeability barrier that separates the enterocyte from the intestinal lumen. Nutrient transport is determined by endocytosis, carrier proteins, and permeability properties of the membrane. Absorbed nutrients exit the enterocyte across the BLM into the portal blood and into the lymph. The BLM is also the entry site of nutrients, hormones, and ions from the blood. Lipids are presumably incorporated in the external lipid monolayer of the BBM; they subsequently diffuse through the BBM and are released from the inner lipid monolayer of the BBM into the cytosol of the enterocyte. The permeability properties of the BBM are subject to adaptation in health and disease.

Alterations in BBM lipid composition may modify physical properties of the membrane, resulting in alterations in activity of membrane-bound proteins. These changes in BBM composition may produce alterations in cell function, including nutrient transport. For example, alterations in BBM lipid composition with aging have been reported for rat intestine and rabbit intestine. The ratio of total phospholipid to cholesterol in the BBM increases with aging, primarily due to an increase in the phospholipid content of the membrane.

Changes in the fatty acid and cholesterol content of the diet result in alterations in BBM lipid composition and nutrient transport. The degree of unsaturation of dietary fatty acids influences lipid composition and function of membranes isolated from many tissues including intestine. In addition, dietary glycosphingolipid such as ganglioside may alter membrane fluidity and permeability. Nutrients also play a significant role in the regulation of gene expression. In a membrane permeability study using enzymatic and fluorescence spectrometry techniques, incorporation of exogenous gangliosides into the phospholipid vesicle was shown to increase the passive membrane permeability.

Fluidity is a property of membranes that describes the freedom of lipid molecules to move in the membrane. Alterations in fluidity are associated with changes in membrane lipid composition and function. Passive and carrier-mediated nutrient transport processes depend on the nature of the fluidity of the BBM and its lipid composition. The lipid content of intestinal BBM changes with fasting. There are decreased ratios of cholesterol/phospholipid, sphingomyelin/phosphatidylcholine, and protein/lipid, decreased oleic and linoleic acids, increased brush border membrane total phospholipid, an increased double-bond index, and an increased percentage of stearic and arachidonic acids. The low fluidity of the BBM reflects a low phospholipid to cholesterol ratio and may be functionally important for efficient nutrient transport. Enterocyte

BBM fluidity is greater in the proximal small intestine than in the distal small intestine, decreases as cells migrate from the crypt to the villus, and is associated with changes in lipid composition. Passive lipid permeability and carrier-mediated glucose uptake are influenced by changes in BBM fluidity.

Once lipids have diffused across the UWL, their uptake is mediated by passive diffusion, although lipid-binding proteins may contribute to uptake.

Lipid-Binding Proteins

There are numerous lipid-binding proteins in the BBM and cytosol of the enterocyte (Table I). The presence of BBM lipid transport proteins raises the possibility that once lipids have partitioned out of the bile acid micelle, their uptake may occur by this carrier-mediated transport as well as by passive diffusion. Lipid-binding proteins in the BBM may serve to transport lipids into the enterocyte, whereas lipid-binding proteins in the cytosol may remove lipids from the BBM and/or bind them in the cytosol, thereby changing the lipid concentration gradient across the BBM and enhancing further uptake by passive permeation.

Caveolin-1 is a 22 kDa integral membrane protein located in detergent-resistant microdomains of the BBM. The caveolins may act as a plasma membrane storage of cholesterol and may play a role in the sterol-sensing component of the BBM. Caveolin-1 also exhibits a binding affinity for long-chain fatty acids. The role of caveolin-1 in the intestine has not been established, but may be involved in the intracellular targeting of lipids.

The scavenger receptor of class B type I (SR-BI) is a 57 kDa integral membrane protein and is located in the BBM. SR-BI may act as a docking receptor for donor particles, such as bile acid micelles, followed by the transfer of lipids to the BBM. SR-BI may also play a role in cholesterol absorption, as well as lipoprotein transport. SR-BI is present on both the apical and basolateral surfaces of the jejunum villus, with little SR-BI being detectable on either apical or basolateral membranes in the ileum.

A 43 kDa protein, known as the plasma membrane fatty acid-binding protein (FABPpm), was identified in the BBM and BLM of intestinal cells. This protein binds LCFA, monoglycerides, and cholesterol. Incubation of rabbit jejunal BBM vesicles with anti-FABPpm antibody results in a reduction in oleic acid uptake. Polyclonal antibody to FABPm inhibits intestinal oleate uptake in a dose-dependent, noncompetitive fashion, with a reduction in the uptake of fatty acids, cholesterol, monoacylgycerol, and lyso-PC but not in the uptake of glucose, alanine, or bile acids.

Fatty acid translocase (FAT) is an 88 kDa transmembrane glycoprotein located in the BBM of enterocytes. Rat FAT is 85% homologous to the human scavenger receptor CD36, which is found in platelets, lactating mammary epithelium, monocytes, and adipocytes. FAT RNA is more abundant in the jejunum than in the ileum and is present in the upper two-thirds of the intestinal villi; the FAT protein is limited to the BBM. Dietary fat rich in polyunsaturated fatty acids up-regulates the expression of intestinal FAT mRNA. FAT is thought to account for approximately 35% of free cholesterol uptake by the BBM.

The fatty acid transport protein (FATP) is a 63 kDa membrane protein. There is a large family of FATPs, but only FATP4 is present in appreciable levels in the intestine. FATP4 mRNA is expressed in the enterocytes of the jejunum and ileum and at lower levels in the duodenum. Expression of FATP4 mRNA is absent from the crypts of the small intestine and from the colon. Fatty acids containing 10–26 carbon atoms are thought to be substrates for FATP4.

The passive absorption of sterols occurs as a result of collision between mixed bile salt micelles and the BBM, but the BBM vesicle transport is reduced following membrane digestion with proteases, suggesting the existence of a transport protein. This "cholesterol transport protein" is an integral membrane protein, with at least one hydrophobic domain. The multidrug resistance protein (MDR; MDR1 in humans) may be involved in the uptake of cholesterol into intestinal epithelial cells. A sterol glycoside derivative specifically binds to the BBM of enterocytes and blocks the absorption of cholesterol. A 145 kDa integral membrane protein in the BBM of rabbit enterocytes has also been identified and may contribute to intestinal cholesterol absorption.

In the cytosol of the enterocyte, fatty acids are bound and transported to their respective sites of metabolism by several cytosolic lipid-binding proteins including two fatty acid-binding proteins (FABPc), the intestinal type (I-FABPc), which is present exclusively in the intestine, and the liver type (L-FABPc), which is present in both liver and intestine. Rat intestinal I-FABPc and L-FABPc genes and human I-FABP have been cloned. I-FABP is localized in mature villus tip cells and is not usually present in the crypt cells, whereas L-FABP is confined to the crypt–villous junction and is not present in villus tip cells. L-FABPc appears in the crypt cells only under the conditions of fasting, likely in order to obtain fatty acids from the blood. I-FABPc binds palmitic, oleic, and arachidonic acid, whereas

L-FABPc binds both saturated and unsaturated fatty acids, monoacylglycerols, lysophospholipids, and bile salts, but not cholesterol. L-FABPc is a 14.1 kDa protein located in the duodenum and jejunum, with maximal expression in the proximal jejunum. I-FABPc is a 15.1 kDa protein that is expressed throughout the small intestine, with maximal expression in the distal ileum. The mRNAs for I-FABPc and L-FABPc are expressed throughout the small intestine and along the length of the villi. I-FABPc is thought to be important for binding lumenally derived fatty acids absorbed across the BBM, whereas L-FABPc binds fatty acids from the bloodstream. I-FABPc is pH-insensitive, which suggests that it binds protonated fatty acids, whereas L-FABPc binds only the unprotonated fatty acids. L-FABPc also binds growth factors, prostaglandins, and leukotrienes, which suggests that L-FABPc plays a role in enterocyte growth and differentiation. These differences in binding specificities and properties may aid in the targeting of lipids to their sites of metabolism.

Peroxisome-proliferator activator receptors (PPAR) may play an obligatory role in up-regulating the expression of L-FABP and I-FABP genes. The livers of mice fed bezafibrate, a PPAR hypolipidemic drug, showed a fourfold increase in L-FABPc protein and mRNA. Bezafibrate increases the mRNA expression of FAT, suggesting a complementary role of FAT and FABP proteins in lipid absorption.

The microsomal triglyceride transport protein (MTP) is an endoplasmic reticulum (ER)-localized cofactor required for the assembly of apolipoprotein B (apoB). MTP is a heterodimer consisting of a 58 kDa subunit of protein disulfide isomerase (a multifunctional ER protein) and a unique 97 kDa subunit. MTP is essential for the transfer of triglyceride and cholesterol esters into the hydrophobic core of apoB. Mansbach and co-workers have suggested that the rate-limiting step in lipid absorption is the trafficking of triacylglycerol from the ER to the Golgi. Specifically, the rate-limiting step may be formation of a prechylomicron vesicle that transports the developing chylomicron from the ER to the Golgi. This rate-limiting step in the complex process from fatty acid and monoacylglycerol entry to triacylglycerol export may involve a protein for the transport of triacylglycerol from the endoplasmic reticulum to the Golgi complex. This transport particle has been isolated and characterized; electron microscopy shows a 200 nm vesicle containing immunoidentifiable apoB48 and apoA-IV, but very little apoA-I. The surface of the chylomicron contains both exchangeable apoA-I, apoA-IV, apoC, and apoE and nonexchangeable apoB48.

See Also the Following Articles

Apoproteins • Bile Composition • Cholesterol Absorption • Dietary Fiber • Fat Digestion and Absorption • Hyperlipidemia • Lipoproteins • Pancreatic Triglyceride Lipase (PTL)

Further Reading

Clandinin, M. T., Cheema, S., Field, C. J., Garg, M. L., Venkatraman, J., and Clandinin, T. R. (1991). Dietary fat exogenous determination of membrane structure and cell function. *FASEB J.* **5**, 2761.

Clarke, S. D., and Jump, D. B. (1994). Dietary polyunsaturated fatty acid regulation of gene transcription. *Annu. Rev. Nutr.* **14**, 83–98.

Diamond, J. M., and Wright, E. M. (1969). Biological membranes: The physical basis of iron and nonelectrolyte selectivity. *Annu. Rev. Physiol.* **31**, 581–646.

Jump, D. B., and Clarke, S. D. (1999). Regulation of gene expression by dietary fat. *Annu. Rev. Nutr.* **19**, 63–90.

Keelan, M., Wierzbicki, A. A., Clandinin, M. T., Walker, K., Rajotte, R. V., and Thomson, A. B. R. (1990). Alterations in dietary fatty acid composition alter rat brush border membrane phospholipid fatty acid composition. *Diabetes Res.* **14**, 165–170.

Levitt, M. D., Furne, J. K., Strocchi, A., Anderson, B. W., and Levitt, D. G. (1990). Physiological measurements of luminal stirring in the dog and human small bowel. *J. Clin. Invest.* **86**, 1540–1547.

Meddings, J., and Thiessen, S. (1989). Development of rat jejunum: Lipid permeability, physical properties, and chemical composition. *Am. J. Physiol.* **256**, G931–G940.

Poirer, H., Degrace, P., and Niot, I. (1996). Localization and regulation of the putative membrane fatty acid transporter (FAT) in small intestine compared with fatty acid binding protein (FABP). *Eur. J. Biochem.* **238**, 368–373.

Schlegel, A., and Lisanti, M. P. (2001). The caveolin triad: Caveolae biogenesis, cholesterol trafficking, and signal transduction. *Cytokine Growth Factors Rev* **12**, 41–51.

Shiau, Y.-F. (1990). Mechanisms of intestinal fatty acid uptake in the rat: The role of an acidic microclimate. *J. Physiol.* **421**, 463–474.

Talmud, P. J., and Waterworth, D. M. (2000). *In vivo* and *in vitro* nutrient gene interactions. *Curr. Opin. Lipidol.* **11**, 31–36.

Thomson, A. B. R., and Wild, G. (1997). Adaptation of intestinal nutrient transport in health and disease. Part I. *Digest. Dis. Sci.* **42**, 453–469.

Thomson, A. B. R., Keelan, M., Clandinin, M. T., and Walker, K. (1986). Dietary fat selectively alters transport properties of rat jejunum. *J. Clin. Invest.* **77**, 279–288.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, D. H. Alpers, J. Christensen, E. D. Jacobson, and J. H. Walsh, eds.), pp. 1867–1907. Raven Press, New York.

Westergaard, H., and Dietschy, J. M. (1974). Delineation of the dimensions and permeability characteristics of the two major diffusion barriers to passive mucosal uptake in the rabbit intestine. *J. Clin. Invest.* **54**, 718–732.

Basic Electrical Rhythm

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

interstitial cells of Cajal Specialized nonneural cells that function as pacemakers for the gastric and intestinal musculature.
myogenic Originating in the musculature of an organ.

Electrodes attached to the serosal surface of the stomach or small and large intestines of mammals detect electrical activity that is continuous and rhythmic. Because of its cyclic periodicity, following its discovery, this electrical activity was called "basic electrical rhythm." The more modern term for the same kind of activity is "electrical slow wave." The time period from the onset of one wave to the onset of the next is designated as one cycle of the rhythm. Quantitative descriptions of the rhythm frequencies in the specialized regions of the gastrointestinal tract are usually expressed as cycles per minute.

INTRODUCTION

The basic electrical rhythm (BER) in the stomach or small and large intestines of mammals is a property of the musculature. The rhythm continues after blockade of all of the innervation of the gut muscles and continues in isolated preparations of stomach or intestine placed in warmed physiologic solutions outside of the body. This is the basis for describing the slow waves as myogenic. Depolarization of the muscle membrane potential accounts for the rising phase of the slow wave and repolarization accounts for the falling phase.

Specialized pacemaker cells determine the frequency at which the slow waves occur. The specialized pacemaker cells are the interstitial cells of Cajal; named after the famous Spanish neuroanatomist, Ramón y Cajal. Interstitial cells of Cajal are electrically coupled to the bulk of the smooth muscle in the stomach and intestine. Electrical slow waves generated by the pacemaker cells spread passively into the bulk musculature. The slow waves do not trigger contractions of the muscle. Invasion of the bulk musculature by the slow waves depolarizes the membrane potential of the muscle cells to the threshold for discharge of action potentials, and the action potentials trigger contraction. Contractions and electrical slow waves occur at the same frequency when action potentials are triggered by each and every slow wave. No contractions occur when the slow waves fail to trigger action potentials in the muscle. The nervous system of the gut determines when the ongoing slow waves trigger action potentials and their accompanying contractions.

BASIC ELECTRICAL RHYTHM OF THE STOMACH

The waveforms (Fig. 1) of the BER, as recorded with intracellular microelectrodes in the stomach and small and large intestines, are similar in shape. Frequencies of occurrence of the BER cycles in the stomach are slower than in the small intestine. The waveform of the gastric BER consists of a sharply rising depolarizing phase, a plateau phase, and a repolarization phase (Fig. 1D). Action potentials may appear on the plateau phase in the distal regions of the stomach (Fig. 1E). A dominant pacemaker located in the midregion of the stomach initiates each slow wave. Once started at the pacemaker site, the slow waves travel rapidly around the gastric circumference and trigger the characteristic ringlike contractions of the distal stomach. The slow waves and associated ringlike contractions then travel more slowly toward the gastroduodenal junction.

Electrical connections between smooth muscle cells account for the propagation of the slow waves from the pacemaker site to the junction with the small intestine. The pacemaker region generates slow waves and contractions at a frequency of 3 cycles/min in humans, 5 cycles/min in dogs, 3–4 cycles/min in cats, 6–7 cycles/min in guinea pigs, and 4–5 cycles/min in rats. The duration of each gastric electrical slow wave in humans is about 10 sec.

BASIC ELECTRICAL RHYTHM OF THE SMALL INTESTINE

Electrical slow waves in the intestine, like the BER in the stomach, occur spontaneously and are always present. Action potentials are triggered at the crests

 Copyright 2004, Elsevier (USA). All rights reserved.

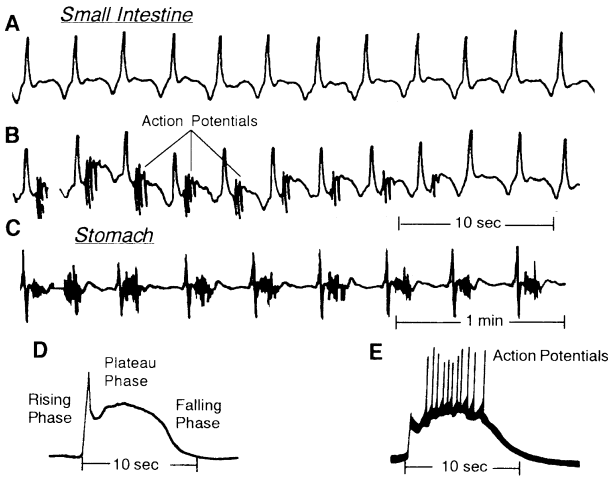


FIGURE 1 Electrical slow waves constitute the basic electrical rhythm in the small intestine and stomach. In the small intestine, the electrical slow waves that constitute the rhythm did not trigger action potentials (A) and exhibit action potentials at the crests of the waves (B). (C) Basic electrical rhythm in distal region of the stomach. (D) Waveform of the slow wave in the midregion of the stomach exhibits a rising phase (depolarization), plateau phase, and falling phase (repolarization). (E) Waveform of the slow wave in a more distal region of the stomach. Action potentials appear in association with the plateau phase. An electrode was attached to the serosal surface of the intestine or stomach for recording of the basic electrical rhythm in A–C. An intracellular microelectrode in a single gastric muscle cell recorded the waveforms in D and E.

of the depolarization phase of the slow wave and are followed by the associated muscle contraction (Fig. 1A). Consequently, the highest frequency of contractions is the same as the frequency of the slow waves. The small intestinal BER frequency in humans is highest in the duodenum, at about 12 cycles/min and progressively decreases more distally, with the ileum having the lowest frequency. Every slow wave may not trigger action potentials and the frequency of contractions in an intestinal segment may not be the same as the BER frequency. When not all slow waves are generating contractions, the intervals between contractions are multiples of the shortest slow-wave interval.

Intestinal slow waves occur synchronously around the circumference of the intestine. They behave as if they either travel very rapidly or are triggered simultaneously at all points around the intestinal segment. At the same time, they appear to travel at much slower velocities in the longitudinal direction along the intestine. The direction of travel is in the caudal direction in the small intestine. Velocity of travel in the duodenum is 5–15 cm/sec, depending on the animal species. Velocity of propagation decreases from the proximal to distal small intestine. Contraction waves tend to travel in the caudal direction with the same velocity as the slow waves because the slow waves trigger the action potentials that trigger contractions.

BASIC ELECTRICAL RHYTHM OF THE LARGE INTESTINE

Specific properties of slow waves in the large intestine are different from the properties of gastric and small intestinal slow waves. The frequency gradient for the large intestinal slow waves is reversed relative to the small intestine. The lowest frequency of slow waves in the large intestine occurs in the proximal colon and is highest in the distal regions of the large bowel.

See Also the Following Articles

Colonic Motility • Duodenal Motility • Electrogastrography • Gastric Motility • Interstitial Cells of Cajal • Small Intestinal Motility

Further Reading

Bass, P., Code, C., and Lambert, E. (1961). Motor and electric activity of the duodenum. *Am. J. Physiol.* **201**, 287–291.

Schuster, M., Crowell, M., and Koch, K. (eds.) (2002). "Schuster Atlas of Gastrointestinal Motility," 2nd Ed. B. C. Decker, Inc., Hamilton.

Szurszewski, J. (1997). The enigmatic electrical slow wave. *Gastroenterology* **113**, 1785–1787.

Szurszewski, J. (1998). A 100-year perspective on gastrointestinal motility. *Am. J. Physiol.* **274**, G447–G453.

Behçet's Disease

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

aphthosis Occurrence of numerous mucosal ulcers, or apthae. Commonly occurs as an idiopathic disorder or can occur as a part of disorders such as Behçet's disease or Crohn's disease. In Behçet's disease, the aphthae are typically numerous, large, and deep.

pathergy Excessive inflammatory reaction to tissue trauma. Characteristic of Behçet's disease, but also seen in pyoderma gangrenosum.

uveitis Inflammation of the uveal tract of the eye, which consists of the iris, cyclic body (anteriorly; anterior uveitis, iritis, and iridocyclitis), and the choroid plexus and retinal layers (posteriorly; posterior uveitis). Inflammation of the entire uveal tract is referred to as panuveitis.

Behçet's disease is an idiopathic, intermittent inflammatory clinical syndrome characterized by recurrent genital and oral aphthosis accompanied by other systemic inflammatory manifestations that commonly include arthritis, uveitis, and skin lesions. Loss of vision from recurrent ocular disease is the principal morbidity of this disease.

EPIDEMIOLOGY

Behçet's disease is a worldwide disease but occurs more frequently in countries along the ancient Silk Road, which linked China to the Roman Empire. The highest prevalence is in Turkey (80–370 cases per 100,000). It has intermediate prevalence in Iran, Saudi Arabia, and the Far East and has a much lower prevalence in North America and Northern Europe. The disease is more common in men, and characteristically affects young adults in the third to fifth decades.

ETIOLOGY AND PATHOGENESIS

Like other autoimmune diseases, the precise cause of Behçet's disease is unknown. The increased prevalence of Behçet's disease in geographic regions along the Silk Road correlates with the prevalence of the human leukocyte antigen (HLA)-B51 allele in populations endemic to these regions; HLA-B51 appears to contribute to the risk of Behçet's disease, but the association is much less prominent in Western countries. This allele also appears to affect the severity of disease. The major histocompatibility complex class I chain-related A gene (MICA gene), which is in linkage disequilibrium with the HLA-B51 allele, is also a candidate gene for Behçet's disease. Hypotheses examining the relationship between microbial infection and the development of Behçet's disease have suggested that shared ubiquitous antigens, including the heat-shock proteins of microorganisms, may trigger cross-reactive autoimmune responses in patients with Behçet's disease.

Pathologically, the lesions in Behçet's disease are predominantly seen in the vasculature. Histopathologic analysis reveals a neutrophilic vascular reaction (leukocytoclastic vasculitis). Hypercoagulability, likely secondary to endothelial activation, is also characteristic of Behçet's disease.

CLINICAL FEATURES

Oral and Genital Lesions

Recurrent, painful, aphthous oral ulcers are frequently the first and most persistent manifestation of Behçet's disease. They resemble common oral aphthae, but when they are large, multiple, or seen on the soft palate or pharynx, they should arouse suspicion for Behçet's disease. Recurrent genital ulceration is seen in the majority of patients. In men and women, such ulcers are typically seen, respectively, on the scrotum and on the vulva, but can be seen on any mucosal surface.

Cutaneous Lesions

Common skin lesions include acneiform nodules, pseudofolliculitis, and papulopustular lesions. Occasionally, erythema nodosum and pyoderma gangrenosum occur. A positive pathergy test is considered highly specific for Behçet's disease. A 20-gauge needle is inserted in the skin of the volar surface of the forearm to a depth of 0.5 cm and is then rotated and withdrawn. The appearance of a pustule >2 mm at the puncture site 48 hours later constitutes a positive

 Copyright 2004, Elsevier (USA). All rights reserved.

pathergy test. Pathergy is distinctly less common in North American and Northern European patients.

Ocular Lesions

The classic ocular lesion seen in Behçet's disease is severe anterior uveitis resulting in visible pus in the anterior chamber (hypopyon). Anterior uveitis is typically relapsing and remitting in character. Posterior uveitis, which involves the posterior uveal tract and retinal vasculitis with its associated complications, is the main cause of permanent visual impairment in Behçet's disease. Ocular inflammation is less common in North American and Northern European patients.

Musculoskeletal

A nondestructive oligoarthritis is the extent of joint involvement in Behçet's disease and is not a dominant feature of this disease, occurring in less than half of the patients. It is often reflective of systemic disease activity. Overlap myositis with proximal muscle weakness occasionally occurs.

Nervous System

Neurological disease occurs in one-third of patients with Behçet's disease, often several years after mucocutaneous manifestations have been present. Aseptic meningitis, stroke syndromes, and cranial neuropathies can occur. Magnetic resonance imaging reveals the predilection for small brain stem vessels and periventricular sites of involvement.

Vascular

Behçet's disease can involve arterial and venous blood vessels of any size. Large-vessel involvement is relatively more common. Clinical syndromes include inferior vena cava obstruction and Budd–Chiari syndrome. Deep venous thrombosis and superficial thrombophebitis occur quite commonly in Behçet's disease.

Gastrointestinal

Ulceration can occur in any part of the gastrointestinal tract but is most frequently seen in the terminal ileum and cecum. Perforation and bowel infarction from mesenteric vasculitis have also been reported in Behçet's disease. The ulcerations can be longitudinal, fissured, or aphthoid. Symptoms include abdominal pain, melena, or hematochezia. The clinical and endoscopic appearance may be indistinguishable from that of Crohn's disease, creating difficulty in differentiating these disorders, because patients with Crohn's disease frequently exhibit oral aphthae as well.

Other

Renal disease is relatively uncommon in Behçet's disease. Amyloidosis has been reported. Epididymitis is also rare. A variety of rare cardiac lesions have been reported. An overlap syndrome with relapsing polychondritis, the "mouth and genital ulcers with inflamed cartilage" (MAGIC) syndrome, can occur.

DIAGNOSIS AND DIFFERENTIATION FROM OTHER CONDITIONS

Recurrent aphthous stomatitis is a common disorder. Ulcers in Behçet's disease are typically larger, deeper, take longer to heal, and may cause scarring. The diagnosis of Behçet's disease should not be assigned unless other clinical features are present. In 1990, the International Study Group for Behçet's disease proposed criteria for the diagnosis of the disease (Table I). Other disorders characterized by recurrent mucocutaneous

TABLE I Criteria for Diagnosis of Behçet's Disease[a]

Criterion	Observation
Primary	
Recurrent oral ulceration	Minor aphthous, major aphthous, or herpetiform ulceration observed by physician or patient, recurring at least three times in one 12-month period
Plus two of the following symptoms	
Recurrent genital ulceration	Aphthous ulceration or scarring observed by physician or patient
Eye lesions	Anterior uveitis, posterior uveitis, or cells in the vitreous (on slit-lamp examination) or retinal vasculitis observed by ophthalmologist
Skin lesions	Erythema nodosum observed by physician or patient, pseudofolliculitis, or papulopustular lesions; or acneiform nodules observed by physician in postadolescent patients not receiving corticosteroid treatment
Positive pathergy test	Read by physician at 24–48 hours

[a] Modified from International Study Group for Behçet's Disease (1990).

lesions include Crohn's disease, human immunodeficiency virus (HIV) infection, recurrent orogenital aphthosis, herpes simplex infection, and cyclic neutropenia. Differentiation from Crohn's disease may be particularly challenging because many other clinical features are also shared, i.e., anterior uveitis, terminal ileal ulcers, and arthritis. Posterior uveitis is rare in Crohn's disease and bowel histology is often distinctly different.

THERAPY AND PROGNOSIS

Mucocutaneous disease is often treated with topical and intralesional steroids. In refractory cases, systemic steroids are employed. Colchicine may also be effective. Steroid-sparing medications include azathioprine, cyclosporine, methotrexate, and thalidomide. Ocular disease is treated very aggressively, often using topical steroid drops or steroid injections into the episcleral space (subtenon injection); with steroid-resistant disease, cyclosporine or methotrexate is often used. Interferon α is emerging as a useful agent for both mucocutaneous and ocular disease. Small open trials have also shown beneficial effects of antitumor necrosis factor (anti-TNF) therapy.

Behçet's disease is a progressive disease with an undulating course. Morbidity is higher in the young, in males, and in patients from the Middle or Far East. Blindness and vascular disease are the cause of the greatest morbidity in Behçet's disease.

See Also the Following Articles

Budd–Chiari Syndrome • Crohn's Disease • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Bayraktar, Y., Balkanci, F., Bayraktar, M., *et al.* (1997). Budd–Chiari syndrome: A common complication of Behçet's disease. *Am. J. Gastroenterol.* **92**, 858.

Hamuryudan, V., Mat, C., Saip, S., *et al.* (1998). Thalidomide in the treatment of the mucocutaneous lesions of the Behçet syndrome. A randomized, double-blind, placebo-controlled trial. *Ann. Intern. Med.* **128**, 443.

International Study Group for Behçet's Disease (1990). Criteria for diagnosis of Behçet's disease. *Lancet* **335**, 1078–1080.

Kaklamani, V. G., and Kaklamanis, P. G. (2001). Treatment of Behçet's disease—An update. *Semin. Arthritis Rheum.* **30**, 299.

Lee, R. G. (1986). The colitis of Behçet's syndrome. *Am. J. Surg. Pathol.* **10**(12), 888–893.

Rozenbaum, M., Rosner, I., and Portnoy, E. (2002). Remission of Behçet's syndrome with TNF-α blocking treatment. *Ann. Rheum. Dis.* **61**, 283.

Sakane, T., Takeno, M., Suzuki, N., and Inaba, G. (1999). Behçet's disease. *N. Engl. J. Med.* **341**, 1284.

Yazici, H., Basaran, G., Hamuryudan, V., *et al.* (1996). The ten-year mortality in Behçet's syndrome. *Br. J. Rheumatol* **35**, 139.

Belching

DANIEL C. BUCKLES AND RICHARD W. MCCALLUM
Kansas University Medical Center

achalasia Disorder in which the lower esophageal sphincter fails to relax properly.
aerophagia Swallowing air.
gastroparesis Disorder in which the muscular contractions of the stomach are lacking and the stomach does not empty its contents.
Helicobacter pylori Gram-negative bacteria that can infect stomach mucosa.
hypochlorhydria Condition of having decreased amounts of stomach acid.
proton pump inhibitor Class of medication that blocks acid production in the stomach.
vagal Pertaining to the vagus nerve of the parasympathetic nervous system.

Belching, also known as eructation or burping, is defined as the involuntary and sometimes noisy regurgitation of air from the stomach and mouth. It is a normal physiological activity that commonly occurs after eating.

PHYSIOLOGY

The sequence of events that lead to belching has become fairly well understood through detailed manometric and fluoroscopic studies. A rise in intragastric pressure will lead to transient lower esophageal sphincter relaxations (TLESRs). These then result in pressure equalizations between the esophageal body and the stomach, and intragastric gas is refluxed into the esophagus. This occurs commonly; during the majority of times, the intraesophageal gas is not perceived by the patient, and the gas is returned to the stomach by secondary esophageal peristalsis. If the volume of intraesophageal gas is sufficient, relaxation of the upper esophageal sphincter is triggered. The intraesophageal air passes out of the oropharyngeal cavity and a belch is produced if the intrathoracic pressure generated is adequate. This sequence is thought to be under vagal control.

CLINICAL SYNDROMES

As a normal physiological event, there is no definition as to what constitutes an abnormal volume or frequency of belching. In general, belching comes to the attention of the clinician only when the patient perceives that it is abnormal or problematic. A number of organic gastrointestinal disorders may give rise to excessive belching. These include esophageal motility disorders, such as achalasia, which may lead to retained esophageal contents. Gastric outlet obstruction and gastroparesis may cause increased gastric pressure resulting in belching. Gastroesophageal reflux disease or a hiatal hernia may also lead to excessive belching. *Helicobacter pylori* infection, in the presence or absence of peptic ulcer disease, may lead to excessive production of ammonia and bicarbonate through urea splitting. Small bowel bacterial overgrowth, seen in patients with surgically altered small intestinal anatomy, motility disorders, and hypochlorhydria from autoimmune causes or chronic proton pump inhibitor use, may lead to excessive hydrogen and methane production. Often such disorders will have accompanying symptoms, such as, pain, vomiting, regurgitation, change in bowel habits, or weight loss, and a detailed history and physical examination can guide further necessary evaluation. Also, for poorly understood reasons, patients may unconsciously attempt to lessen epigastric or thoracic pain not of gastrointestinal origin by purposely swallowing air and belching. This could be part of an anxiety response. It is therefore important to think of angina pectoris in appropriate risk patients who present with the new onset of excessive belching.

After organic conditions that lead to excessive belching have been excluded, the majority of patients will have aerophagia, or air swallowing, as the cause of their complaints. Sometimes easily correctable behavioral causes of aerophagia can be identified. These would include cigarette smoking, sucking hard candy, ill-fitting dentures, or drinking through a straw, for example. Other patients will reflexively swallow air during stressful or anxiety-provoking situations. The remainder will habitually swallow air as part of a functional gastrointestinal syndrome. Aerophagia rarely leads to any significant adverse outcomes except in children or mentally retarded persons, who may swallow such large amounts of air that massive

 Copyright 2004, Elsevier (USA). All rights reserved.

abdominal distension occurs; a rare case of perforation has been reported in such individuals.

DIAGNOSTIC TESTING

A detailed history will usually guide the need for specific testing. In most situations, a plain film to check the distribution of bowel gas is appropriate. Barium esophagography can assess esophageal diameter and determine whether a hiatal hernia is present. Esophagogastroduodenoscopy (EGD) may be helpful in ruling out peptic ulcer disease, sequelae of reflux, or other structural lesions. Urea breath testing, examination of gastric biopsy, stool antigen tests, and serum antibody detection are all available to identify *H. pylori* infection. Hydrogen breath testing can identify abnormal small bowel bacterial overgrowth through the detection of excess hydrogen production after glucose or lactulose challenge. Finally, cardiac stress testing can help identify those with angina as a cause of excessive belching.

TREATMENT

If a specific organic disorder is identified as the underlying cause of excessive belching, then treatment aimed at alleviating that cause is usually sufficient to correct the complaint. In patients with aerophagia, a number of measures may be tried. Sometimes behavioral modification, such as chewing and eating more slowly, quitting tobacco, and avoiding hard candy and carbonated beverages, may be helpful. In patients with anxiety as an etiology, anxiolytics and antidepressants may be appropriate. Other patients with habitual aerophagia may benefit from referral to a psychologist for behavioral therapy and counseling to help them control the urge to belch. Such patients should be reminded that, although their habit may be personally troubling and socially distasteful, it is generally a benign disorder and the prognosis is excellent.

Pharmacologic treatments aimed at directly modifying belching have been largely disappointing. Simethicone is a defrothicant that acts to coalesce small bubbles of gas. Many simethicone-containing products are available and are widely used, but convincing studies regarding their effectiveness are lacking. Other agents designed to neutralize or suppress acid production, such as proton pump inhibitors, histamine-2 (H2) blockers, or antacids, would not be expected to be useful in the absence of a specific acid-related disorder.

See Also the Following Articles

Achalasia • Flatulence • Gastric Outlet Obstruction • Gastroesophageal Reflux Disease (GERD) • *Helicobacter pylori* • Hiatal Hernia

Further Reading

Berk, J. E. (1995). Gas. *In* "Bockus Gastroenterology" (W. S. Haubrich and F. Schaffner, eds.), 5th Ed., pp. 115–121. W. B. Saunders, Philadelphia.

McNalley, E. F., Kelly, J. E., and Ingelfinger, F. J. (1964). Mechanism of belching: Effects of gastric distension with air. *Gastroenterology* **64**, 254–259.

Rao, S. S. C. (1997). Belching, bloating, and flatulence: How to help patients who have troublesome abdominal gas. *Postgrad. Med.* **101**, 263–269.

Bezoars

HAGITH NAGAR
Tel Aviv Sourasky Medical Center and Tel Aviv University

bezoar Any of various calculi found chiefly in the gastrointestinal organs and formerly believed to possess magical properties. The prefixes phyto-, tricho-, and lacto- are used to define the major bezoar constituents: vegetable matter, hair, and milk products, respectively.

Bezoars are intralumenal masses consisting of undigested matter that may lodge at any level within the alimentary canal. These objects consist of plant material, hair, dairy products, or foreign bodies. Therapy consists of enzyme dissolution, endoscopic extraction, or surgery.

HISTORY

The word "bezoar" is derived from an Arabic corruption of the Persian word "*panzahar*," meaning "against poison," a reference to curative properties that were formerly ascribed to these objects. Bezoars have been described for more than 3000 years, having been noted in the intestines of sheep and goats. In ancient times, people used bezoars as cameos for good luck, health, and youth. In some cases, the objects were ingested as treatment for dysentery, leprosy, snake poisoning, vertigo, and epilepsy.

Bezoars were highly valued; indeed, Debakey and Ochsner describe one case in which a bezoar was given in exchange for a castle, and it was even possible to "rent" a bezoar. As forgeries became prevalent, tests were developed to examine their authenticity.

The first trichobezoar in a human was described as a postmortem finding in 1779 and the first report of surgical removal of a bezoar was reported by Schonborn in 1883. Diagnosis through palpation of an abdominal mass was reported in 1896 by Stelzner. The first gastroscopic removal of a bezoar was performed by McKechne in 1972.

CLINICAL FEATURES

Bezoars have been reported in patients ranging in age from a few months to 90 years and at all levels of the gastrointestinal tract from the esophagus to the rectum. Areas of physiological or anatomical narrowing, or those that result from surgery or peritoneal adhesion, are particularly common sites. Although they can occur at any portion of the gastrointestinal tract, most are found in the stomach. A single bezoar may completely fill the stomach and even extend into the small bowel (Figs. 1 and 2).

Bezoars are classified into four types, depending on their contents. Phytobezoars consist of undigested portions of fruits or vegetables, trichobezoars consist of hair, and lactobezoars consist of congealed milk products. The fourth category may contain any of a wide range of foreign objects, such as coins, gallstones, and ingested medications ("pharmacobezoar").

Trichobezoar usually presents as a gastric mass, occasionally extending into the duodenum and even into the jejunum. As many as 90% of trichobezoars are diagnosed in adolescent girls. Often, underlying stress is present and the patient suffers from trichophagia and trichotillomania. Cases in which the hair mass continued as far as the cecum have been referred to as "Rapunzel syndrome," after the famed fairy tale character who used her long hair to escape from a castle. The clinical presentation of trichobezoar depends on the location and size of the mass and may include vomiting due to partial or complete obstruction, weight loss, or an abdominal mass that may be thought to be malignancy. In

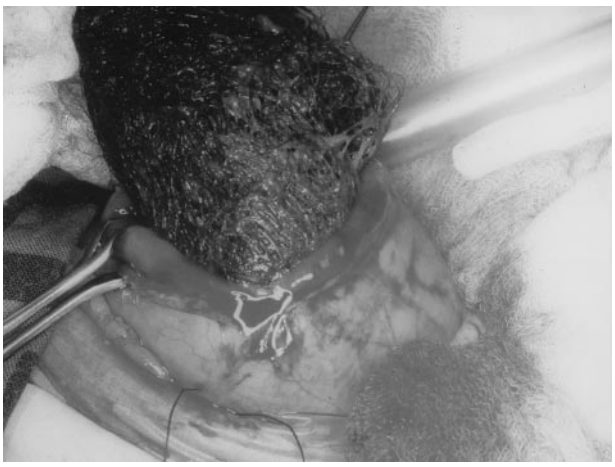


FIGURE 1 Extraction of a trichobezoar through a gastrotomy.

 Copyright 2004, Elsevier (USA). All rights reserved.

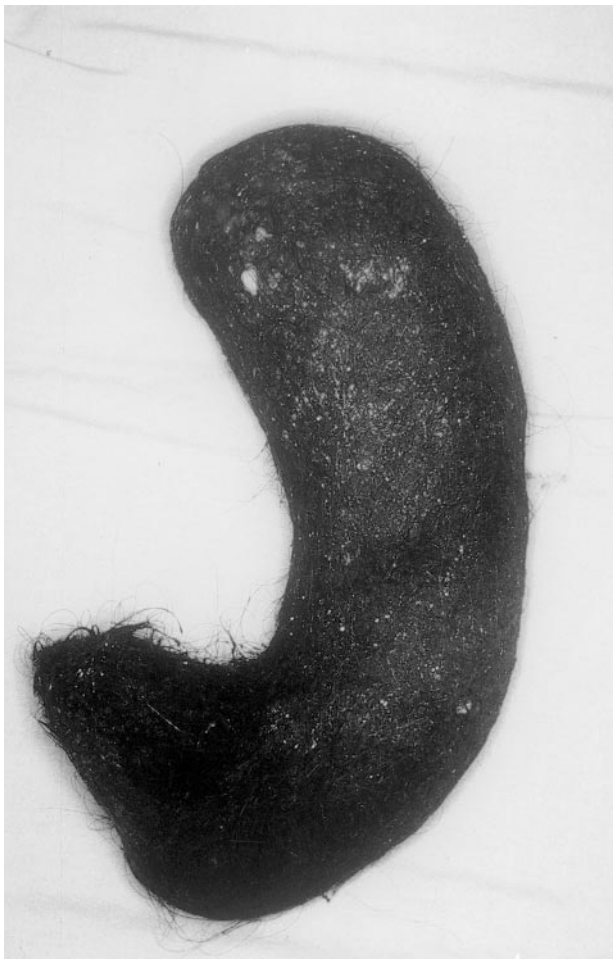


FIGURE 2 Trichobezoar. Note that the specimen has retained the shape of the stomach.

addition to mechanical obstruction, bezoars may be complicated by anemia, gastrointestinal ulcers, or perforation.

Lactobezoar consists of a compact mass of undigested milk concretions located within the gastrointestinal tract. Most are reported in infants, in particular preterm infants on caloric-dense formulas. Lactobezoars may precipitate gastric outlet obstruction, which mimics a variety of medical and surgical conditions.

The diagnosis of bezoar is confirmed using ultrasonography (US), computerized tomography (CT), or barium swallow. Endoscopy may be useful in both the diagnosis and the management of gastric bezoar. CT is considered more accurate than ultrasonography and has been used to diagnose bezoars at all levels of the gastrointestinal tract. Findings on US consist of hyperechoic band-like lesions and acoustic shadows, whereas CT demonstrates a mass that contains air pouches, without postcontrast enhancement.

Therapeutic options include medical dissolution, endoscopy extraction, or surgery. Gastroscopic extraction is the treatment of choice for gastric phytobezoars. Endoscopic suction removal of gastric phytobezoars using a large-channel endoscope is efficacious and safe. Endoscopically guided electrohydraulic lithotripsy has been advocated as a safe and effective modality for gastric bezoars. Alternative endoscopic methods have included the use of a water jet, forceps, or snare and basket. Lactobezoars may be removed in this way or dissolved using oral enzymes. Surgery is the treatment of choice for trichobezoar, with removal of the entire mass, including the "tail," via gastrotomy.

Esophageal bezoars are rare and tend to occur in patients with structural or functional abnormalities of the esophagus. Additionally, enteric feeding formulas that include sucralfate and casein have also been implicated in the formation of esophageal bezoars, particularly in the setting of decreased esophageal pH.

See Also the Following Articles

Computed Tomography (CT) • Endoscopic Ultrasonography • Foreign Bodies • Pica • Ultrasonography

Further Reading

Chen, M. K., and Beierle, E. A. (2001). Gastrointestinal foreign bodies. *Pediatr. Ann.* **30**, 736–742.

Debakey, M., and Ochsner, A. (1939). Bezoars and concretions: A comprehensive review of the literature with an analysis of 303 collected cases and presentation of 8 additional cases. *Surgery* **5**, 132–160.

DuBose, T. M., V, Southgate, W. M., and Hill, J. G. (2001). Lactobezoars: A patient series and literature review. *Clin. Pediatr.* **40**, 603–606.

Lee, J. (1996). Bezoars and foreign bodies of the stomach. *Gastrointest. Endosc. Clin. N. Am.* **6**, 605–619.

Nagar, H., and Resnick, I. (2000). Trichobezoar of the stomach. *Is. Med. Assoc. J.* **2**, 641–642.

Ripolles, T., Garcia-Aguayo, J., Martinez, M. J., and Gil, P. (2001). Gastrointestinal bezoars: Sonographic and CT characteristics. *Am. J. Roentgenol.* **177**, 65–69.

Bile Composition

ALAN F. HOFMANN
University of California, San Diego

ABC transporters ATP-stimulated membrane transporters that are responsible for transporting molecules "uphill" (against their concentration gradient); such molecules have an ATP-binding cassette, hence the designation "ABC."

amphipathic Denoting molecules that have hydrophobic and hydrophilic domains such that they are surface active and often self-associate to form aggregates (micelles).

bile acids Molecules that are present in high concentrations in bile and are responsible for its physiological properties. Chemically, bile acids are compounds having the cholane or cholestane nucleus with an acidic group on the terminal carbon of the side chain. One to three hydroxyl groups are present on the nucleus, and there may be an additional hydroxyl group on the side chain. Bile acids are formed in hepatocytes from cholesterol. They occur in bile conjugated in an amide linkage with glycine or taurine.

bile alcohols Molecules that are present in high concentrations in bile in certain species (cartilaginous fish, herbivorous fish, ancient mammals). Chemically, they are compounds with a cholestane nucleus, having a hydroxyl group on the terminal carbon of the side chain. One to three hydroxyl groups are present on the nucleus, and there may be one to three additional hydroxyl groups on the side chain.

bile salts Collective term denoting both bile acids and bile salts.

bilirubin Tetrapyrrole (brown in color); the major bile pigment present in the bile of most mammals; the end product of heme metabolism. When bilirubin is oxidized, it becomes biliverdin.

biliverdin Oxidized bilirubin (green in color); the major bile pigment in the bile of some fish and mammals. When biliverdin is reduced, it becomes bilirubin.

canaliculus Space formed between the apical membranes of hepatocytes, into which canalicular bile is secreted; the smallest unit of the biliary tract.

cholesterol Sterol that is present in cell membranes. Cholesterol is white and insoluble in water. It has the cholestene nucleus with a double bond at C_5-C_6 nucleus and a hydroxy group on the third carbon atom.

cholic acid Primary bile acid; very common in many mammals. Chemically, it is a C_{24} bile acid with hydroxyl groups at carbons 3, 7, and 12.

conjugation In metabolism, the addition of a molecule that renders a lipophilic molecule more hydrophilic and usually makes it more water soluble. Molecules commonly used for conjugation include sulfate, glucuronic acid, glutathione, glycine, and taurine.

enterohepatic circulation Movement of molecules from the liver, to the biliary tract, to the intestine, to the portal venous system, and to the liver. The only molecules that are known to have an efficient enterohepatic circulation are bile salts.

flavinoid Collective term for a variety of polycyclic compounds that are present in vegetables, especially soy products. Many of the molecules have antiestrogen properties (these are termed "phytoestrogens").

flippase Transporter that "flips" a molecule in the membrane from one side to the other. In the canaliculus, mdr2 flips phosphatidylcholine from the cytosolic side to the luminal side.

gallstones Concretions formed in the biliary tract; composed of molecules that are insoluble in water, usually cholesterol and/or calcium bilirubinate.

Gibbs–Donnan equilibrium Distribution of anions and cations when an impermeable anion is present on one side of the membrane. Electrostatic effects lead to enrichment of divalent cations on the side of the impermeable anion. In bile, this results in the concentration (activity) of Ca^{2+} ions being higher in gallbladder bile than in plasma.

γ-glutamyl transpeptidase Enzyme located on the luminal face of the canaliculus (in some species) and the bile duct epithelium; cleaves the amide bonds of glutathione to constituent amino acids. The concentration of the enzyme increases in liver disease.

glutathione Tripeptide consisting of glutamic acid, cysteine, and glycine. The sulfhydryl group on the cysteine is used for conjugation of a number of organic anions, especially those containing halogen groups.

immunoglobulins A, M, and G Molecules with potent bacteriostatic properties; IgA is the dominant immunoglobulin of intestinal secretions and is found also in saliva and bile.

micelles Polymolecular aggregates formed by amphipathic molecules, such as detergents and bile salts; their presence above a certain concentration is termed the "critical micellization concentration" (CMC).

mucin Protein that has many carbohydrate molecules covering its surface; the dominant protein of mucous.

 Copyright 2004, Elsevier (USA). All rights reserved.

phosphatidylcholine Zwitterionic phospholipid present in cell membranes and in bile; chemically, has a glycerol backbone. Fatty acids are esterified to the first two hydroxyl groups; a phosphate group is attached to the third hydroxyl group, yielding phosphatidic acid; the phosphate group is esterified to choline (hence the name, phosphatidylcholine).

primary bile acid Synthesized in the hepatocyte from cholesterol.

saturation, or cholesterol saturation Number, expressed as a fraction or a percentage, describing the cholesterol concentration of a bile sample relative to its equilibrium solubility (at saturation). Mathematically, it is the cholesterol concentration divided by the equilibrium solubility in the sample, or that predicted to hold for a sample based on studies of model systems simulating bile.

secondary bile acid Formed by modification of the hydroxyl groups of a primary bile acid. Common changes involve removal of the hydroxyl group at C-7 (7-dehydroxylation) or oxidation of any hydroxyl group to an oxo group, or epimerization of any hydroxyl group (a change of an α-hydroxyl group to a β-hydroxyl group or vice versa).

solubility product Mathematical product of the concentration of the anion times the concentration of the cation of a salt that has a low solubility in water. When the solubility product of a salt is exceeded, the solution is supersaturated, and the salt can potentially precipitate from solution.

sphincter of Oddi Valve at the end of the biliary tract where the tract empties into the duodenum. In humans, the common bile duct and pancreatic ducts merge at the sphincter of Oddi.

triangular coordinates Method for showing the relative proportions of three constituents that together add up to 1 (or 100%). Because there are only two degrees of freedom (the third is equal to the total minus the first and second), it is possible to represent compositions of such mixtures in two dimensions.

vesicles Lipid aggregates composed of a bilayer of amphipathic lipids. Vesicles may consist of a single bilayer (unilamellar vesicles) or multiple bilayers (multilamellar vesicles). In bile, vesicles are composed mostly of phosphatidylcholine and cholesterol. Bile salt molecules adsorb to vesicles and, if present in sufficiently high concentration, change the vesicles into mixed micelles.

Bile is a digestive and excretory fluid formed by secretion of solutes (and accompanying water) into the biliary tract. Most bile originates from hepatocytes. These secrete bile into the biliary canaliculi, i.e., the spaces between the apical membranes of hepatocytes. Canalicular bile drains into the bile ducts, which merge to form the common bile duct that in turn drains into the small intestine. Bile is modified be secretion and/or absorption as it passes down the bile ductules and ducts or when it is stored in the gallbladder. Biliary constituents are organic or inorganic. The lipid component of the organic fraction consists of three major classes—bile salts (bile acids), phospholipids (mostly phosphatidylcholine), and cholesterol. Three major classes of bile salts are present in the bile of vertebrates—C_{27} bile alcohols (conjugated with sulfate), C_{27} bile acids (conjugated with taurine), and C_{27} bile acids (conjugated with glycine or taurine). In most mammals, bile salts are composed predominantly of C_{24} bile acids.

INTRODUCTION

Bile is an aqueous fluid secreted by the liver into the biliary tract. Bile is distinguished from other digestive secretions by its color—brown or green—and by its high concentration of bile salts. Bile is colored because it contains bile pigments, tetrapyrroles formed from heme. Biliverdin is green; when reduced, it becomes bilirubin, which is brown. Bilirubin is the bile pigment in most vertebrates, but in some species (certain fish and the rabbit), biliverdin is the dominant bile pigment. The emphasis herein is on bile in mammals, because relatively little information is available on bile composition in amphibians, reptiles, fish, and birds.

Bile salts are surface-active molecules formed in the hepatocyte from cholesterol. Bile salts are present in bile as anions that self-associate to form polymolecular aggregates termed "micelles." The presence of such micelles gives bile potent solubilizing properties, and in mammals, the micelles in bile also contain phospholipids and cholesterol. Bile is bitter in taste because of its high concentration of bile salts. Bile is isosmotic and slightly alkaline.

In most vertebrates, the biliary tract contains a globular reservoir, the gallbladder, in which bile is stored and concentrated between meals. When a meal is eaten, the gallbladder contracts and the sphincter at the end of the biliary tract (sphincter of Oddi) relaxes. As a result, bile enters the small intestine, where it promotes the digestion and absorption of dietary lipids.

Bile is also an excretory fluid in that it serves as a vehicle for the excretion of molecules that cannot be eliminated in urine. In contrast to renal excretion, in which the driving force is filtration by the glomerulus, biliary excretion relies on transporters in the canalicular membrane. Biliary excretion serves to eliminate ions or molecules that are bound to plasma proteins. Such substances include a variety of lipophilic organic molecules and polyvalent metal cations. When these molecules enter the intestine via bile, they are poorly absorbed, thereby leading to their elimination from the body in feces.

Historical Aspects

Because the gallbladder is readily detected at autopsy, examination of the appearance of bile for clues to the nature of disease has been practiced since antiquity. Bile was one of the four "humors" proposed by Aristotle; the term "melancholia" was used to suggest that when bile was black, disease was present, causing a state of depression. Because gallbladder bile is readily obtained when livestock are slaughtered for meat, studies on its composition began nearly two centuries ago. Isolation of bile salts—the major constituents of bile—was performed early in the nineteenth century. Pure compounds were isolated and named before any idea of their chemical structure was available. The Greek word *chole* for "bile" gave rise to the name "cholic acid," from which the names of many other bile salts are derived.

Bile contains a greater concentration of cholesterol in humans than in any other vertebrate. When cholesterol molecules crystallize and the crystals aggregate, a gallstone is formed. In the late eighteenth century, gallstones were dissolved in hot alcohol and cholesterol crystals were obtained. The crystals were named "cholesterin" (later changed to cholesterol), the name denoting the solids in bile.

The major bile salts in bovids are the two bile acids, cholic acid and deoxycholic acid. They occur in bile as taurine or glycine conjugates. When bile is heated with alkali, the unconjugated bile acids, cholic acid and deoxycholic acid, are formed. These are readily isolated and have been available as laboratory chemicals in relatively pure form for at least half a century. As a result of their availability, a great deal is known about their properties.

TYPES OF BILE

Bile is formed by secretion of solutes into the canaliculus, a space formed between the apical membranes of adjacent hepatocytes. Such bile, termed canalicular bile, cannot be sampled, because the canaliculus is too small to be sampled by micropuncture techniques. Bile enters the biliary ductules, where it is modified by solute absorption and secretion. The biliary ductules have an arboreal architecture. The ductules progressively anastomose, ultimately forming the right and left hepatic ducts; these in turn merge to form the common hepatic duct. Bile sampled at this point is termed "ductular bile." Ductular bile may either enter the gallbladder or bypass the gallbladder to enter the small intestine. Bile in the gallbladder may be sampled *in vivo* by gallbladder puncture during surgery or *ex vivo* after gallbladder removal; such bile is termed "gallbladder bile." When studies seek to ascertain the composition of gallbladder in the healthy subject, a tube is placed into the duodenum and gallbladder contraction is induced by intravenous injection of a peptide. Such bile that is collected in the duodenum is termed "duodenal bile" because it contains a mixture of gallbladder and hepatic bile as well as pancreatic secretions.

CLASSIFICATION OF BILIARY CONSTITUENTS

Biliary constituents may be divided into two broad classes—organic and inorganic. The organic fraction, in turn, is divided into a lipid fraction and a nonlipid fraction. The lipid fraction is defined as bile salts (or bile acids), phospholipids, and cholesterol. The nonlipid fraction contains other organic molecules, most of which are of endogenous origin. The major organic nonlipid is bile pigment (bilirubin or biliverdin). Other trace constituents include conjugates of steroids, fat-soluble vitamins, and proteins and peptides. Exogenous molecules may also be present in bile (for example, drugs, ingested bile acids, and dietary constituents such as flavinoids). Table I describes a classification of biliary constituents.

Organic molecules that are larger than sucrose can enter bile only by carrier-mediated transport. For a substance to be present in bile, it must be a substrate for one or more of the canalicular transporters. These transporters, which have been cloned in the past decade, have an ATP-binding cassette (ABC).

Hepatic bile is formed by the flow of water and filterable solutes across the paracellular junctions between the hepatocytes. Such flow occurs in response to the osmotic stimulus of molecules that are actively secreted into the canalicular space. In most species, the osmotically active agents are bile salts and probably anions such as bicarbonate and chloride. Glutathione also contributes. Because bile salts and most organic molecules in bile are anions, an equivalent concentration of cations must be present. The dominant cations are those of plasma (Na^+, K^+, Ca^{2+}, and Mg^{2+}) and have concentrations that are similar to those of an ultrafiltrate of plasma. Canalicular secretion of copper, iron, and cations of heavy metals promotes their biliary excretion. A summary of the composition of hepatic and gallbladder bile is given in Table II.

TABLE I Classification of Biliary Constituents

Organic constituents	Inorganic constituents
Lipid fraction	**Cations**
Bile acids or alcohols	*Plasma electrolytes*
Conjugated	Sodium
Unconjugated	Potassium
Phospholipids	Calcium
Phospholipid classes	Magnesium
Individual species	*Plasma metals*
Cholesterol	Copper
Nonlipid fraction	Iron
Bile pigments	*Trace metals*
Bilirubin/biliverdin	Many polyvalent cations
Trace constituents	**Anions**
Steroids (conjugated)	*Plasma electrolytes*
Oxysterols	Bicarbonate
Fatty acids (unesterified)	Chloride
Vitamin B_{12}	Phosphate
Fat-soluble vitamins (A, D, E, K—conjugates)	Sulfate
Filtrable solutes	
Sugars, amino acids, organic acids, purines, pyrimidines	
Proteins and peptides; autocoids	
Plasma proteins: albumin, IgM, IgA, apoproteins A-I, A-II, B, and C	
Bile-specific proteins: IgA, mucins, APF/CPB	
Peptides	
Glutathione	
Autocoids; signaling molecules	
Prostanoids	
Cyclic AMP	
Xenobiotics; dietary constituents	
Vitamin B_{12} congeners	
Dietary lipids (e.g., plant sterols)	
Drugs and their metabolites	
Flavinoids	

BILIARY LIPIDS

The following descriptions of biliary lipids apply to mammals. In some reptiles and fish, no biliary lipids other than bile salts are present.

Bile Salts

Bile salts are the dominant anion in gallbladder bile. Active secretion of bile salt anions by a canalicular transporter into the canaliculus generates an osmotic force that pulls water and filterable solutes into the canaliculus. Bile salts are derived from cholesterol by a multienzyme process. Three major classes of bile salts are present in vertebrates—bile alcohols with 27 carbon atoms (C_{27} bile alcohols), C_{27} bile acids, and C_{24} bile acids. (In most mammals, bile salts are composed predominantly of bile acids, as distinguished from bile alcohols, and much of the clinical literature uses the term "bile acids").

Bile salts are secreted in conjugated form. The terminal carbon of bile salts contains a functional group—a hydroxy group in bile alcohols and a carboxyl group in bile acids. The functional group serves to couple bile acids to the conjugating moiety. Bile alcohols are esterified with sulfate. The carboxyl group of the C_{27} bile acids is coupled to the amino group of taurine, with the result that such conjugated bile acids are termed "N-acyl amidates" or "N-acyl aminoamidates." The C_{24} bile acids are conjugated with taurine in some species, with glycine in other species, and with both taurine and glycine in some species, including humans. Conjugated bile salts are strong acids and are fully ionized at the pH conditions prevailing in bile and the small intestine. Therefore, conjugation of bile acids increases their aqueous solubility. Conjugation also prevents precipitation by Ca^{2+} ions.

Bile salts are hydroxylated on the nucleus and sometimes on the side chain. All bile salts have a 3-hydroxy group because they are derived from cholesterol, which is a 3β-hydroxy sterol. Cholesterol 7α-hydroxylase is a microsomal enzyme that is rate limiting for bile salt synthesis, and bile salts with hydroxy groups at carbons 3 and 7 can be considered to be the building block on which an additional nuclear hydroxyl group is added. All hydroxyl groups are present on one face (the α face) of the bile acid molecule, causing bile acids to be amphipathic. Trivial names of bile acids are based on the pattern of nuclear substituents as well as on the animal from which the compound was first isolated.

During enterohepatic cycling of bile acids, bacteria in the distal intestine remove the 7-hydroxy group of bile acids, giving rise to one or more 7-deoxy bile acids. The 7-deoxy bile acids are termed "secondary" bile acids because they are formed by bacterial enzymes, and thus should be distinguished from "primary" bile acids that are synthesized in the liver.

Bile salts have a metabolism that differs from that of other biliary lipids. They are efficiently absorbed from the distal small intestine and are returned to the liver in portal venous blood. They are transported efficiently into the hepatocyte and again secreted into bile, thus undergoing an enterohepatic circulation. Efficient intestinal absorption leads to a recycling pool of bile salts, with the result that the transhepatocyte flux of bile acids greatly exceeds bile acid biosynthesis. A small fraction of bile acids is not absorbed in the distal intestine and is

TABLE II Composition of Bile in the Adult Human: Lipids and Ions

	Composition[a]				
	Hepatic bile		Gallbladder bile		
Component	mmol/liter	Mole fraction	mmol/liter	Mole fraction	Comment
Biliary lipids					
Conjugated bile acids	10–40	~0.7	70–150	~0.7	Mole fraction of biliary lipids, i.e., bile acids, phospholipids, cholesterol
Cholic acid conjugates	4–16	0.4	28–60	0.4	Mole fraction of conjugated bile acids
CDCA[b] conjugates	4–16	0.4	28–60	0.4	
DCA[c] conjugates	2–8	0.2	14–30	0.2	Conjugates of lithocholic acid and UDCA are also present, but at <5%
Phospholipids	3–12	~0.3	21–45	~0.3	Mole fraction of biliary lipids
Cholesterol	1–4	0.05–0.1	7–15	0.05–0.1	Mole fraction of biliary lipids; cholesterol/phospholipid ratio higher in gallstone patients
pH	7.1–7.5		6.6–7.2		
Electrolytes					
Na^+	146 ± 8		210 ± 12		Na^+ concentration depends on extent of concentration in gallbladder
K^+	4.8 ± 0.5		12.7 ± 3.4		
Cl^-	105 ± 11		66 ± 33		
HCO_3^-	22 ± 3		14 ± 8		
Ca^{2+} (total)	1–2		2–12		
Ca^{2+} (ionized)	0.5–1.2		0.5–2.0		
Mg^{2+}	—		6.9 ± 2.2		
Fe	—		0.016 ± 0.01		
Cu	—		0.096 ± 0.05		
PO_4^{2-}	0.2–0.9		Not reported		

[a] Mean ± SD or range.
[b] CDCA, Chenodeoxycholic acid.
[c] DCA, Deoxycholic acid.

excreted in the feces. Fecal loss of bile salts is balanced by *de novo* biosynthesis.

In humans, biliary bile acids consist of the two primary bile acids, cholic acid and chenodeoxycholic acid, each comprising about 40% of biliary bile acids. The remainder is made up of deoxycholic acid (formed by bacterial dehydroxylation of cholic acid), lithocholic acid (formed by bacterial dehydroxylation of chenodeoxycholic acid), and ursodeoxycholic acid (formed by bacterial epimerization of chenodeoxycholic acid). Each bile acid is conjugated with glycine or taurine.

Phospholipids

In most mammals, bile contains phospholipids that are predominantly phosphatidylcholine (PC). (The term "lecithin" is sometimes used for biliary phospholipids.) The phospholipids in human bile have been analyzed by high-performance liquid chromatography (HPLC) in order to quantify individual PC species. A large number (up to 30) of PC molecules are present in human bile. In general, these tend to have a C_{16} saturated or monounsaturated fatty acid moiety in the first position and a C_{18} mono- or diunsaturated molecule in the second position.

PC is synthesized in the hepatocyte and travels to the canalicular membrane by a phospholipid-binding protein. The canalicular membrane contains the usual variety of membrane phospholipids, and a "flippase" selectively flips PC molecules across the canalicular membrane. The accumulation of phospholipid on the luminal face of the canalicular membrane is evidenced by vesicular evaginations that can be visualized by electron microscopy.

Biliary phospholipid serves to protect the biliary epithelium from the cytotoxic effects of bile salts. In the presence of phospholipids, bile salts form mixed

micelles at a lower concentration than is observed in the absence of phospholipids. Biliary phospholipid also serves to enhance greatly the solubility of cholesterol in bile, a property that is important in humans because the ratio of cholesterol (the solute) to the solvent (bile salts and PC) is much higher than in other mammalian species.

Cholesterol

Cholesterol is present in bile in unesterified form. Cholesterol in bile has no function and its presence in bile indicates that not all cholesterol in the liver was converted to bile acids, or, as a corollary, that cholesterol excretion is mediated by biliary excretion of cholesterol as well as bile acids.

Cholesterol has a low aqueous solubility and is maintained in solution by being incorporated into the mixed bile salt phospholipid micelles present in bile. Because the solubility of cholesterol depends on the concentration of the mixed micelles, it has been common practice to express cholesterol concentration as a mole fraction of either total biliary lipids or bile salts plus PC. It is also common practice to express the relative proportions of the biliary lipids using triangular coordinates. From these proportions and from the equilibrium solubility of cholesterol in model systems composed of the three biliary lipids, it is possible to calculate cholesterol "saturation." Gallbladder bile that is supersaturated in cholesterol is a risk factor for the formation of cholesterol gallstones.

When bile is supersaturated with cholesterol, additional phases besides mixed micelles are present in bile. These include cholesterol-rich vesicles. The vesicular phase may be isolated by centrifugation or exclusion chromatography. Besides cholesterol, bile also contains trace amounts of other sterols. These include endogenous sterols such as oxysterols, which are intermediates in bile acid biosynthesis. Bile also contains plant sterols such as sitosterol, campesterol, and methosterol.

NONLIPID ORGANIC MOLECULES

Bile Pigments

The major biliary nonlipid is bilirubin, which is formed in the hepatocyte by reduction of biliverdin. Bilirubin is classified as a nonlipid because it is present in bile as a glucuronide conjugate, and in this chemical form is not appreciably incorporated into the mixed micelles present in bile.

Bilirubin is formed in the reticuloendothelial cells from heme and enters the hepatocyte by uptake from sinusoidal plasma. In the hepatocyte, bilirubin enters the smooth endoplasmic reticulum, where one or both of its two carboxylic groups are esterified to the C-1 hydroxyl group of glucuronic acid. In humans, bilirubin is present predominantly in the form of the diglucuronide, but in other animals, the monoglucuronide may be the dominant bile pigment. Bilirubin, in contrast to bile acids, does not undergo enterohepatic cycling, so that biliary secretion is approximately equal to synthesis from heme.

Organic Anions Present in Trace Amounts

Other lipophilic molecules are present in bile in trace amounts. These include glucuronides or sulfates of steroid hormones, prostanoids, flavinoids (of dietary origin), and conjugates of fat-soluble vitamins and/or their metabolites. Biliary secretion is not considered to play a major role in the excretion of these compounds and they are not known to have any functional effects on the biliary tract or small intestine. Their presence in bile results from their serving as substrates for the rather nonspecific anion transporters of the canalicular membrane.

Proteins and Peptides

Only three proteins have appreciable concentrations in bile—immunoglobluin A (IgA), mucin, and an anionic polypeptide fraction/Ca^{2+}-binding protein (APF/CPB), which is amphipathic and has phospholipid and Ca^{2+}-binding properties. IgA is secreted into bile from cholangiocytes in humans; in other mammals, it is thought to enter bile from hepatocytes. IgA serves to inhibit bacterial growth in the biliary tract and may also bind to cholesterol crystals, inhibiting deposition of additional cholesterol molecules. Mucin is present in hepatic bile and gallbladder bile. Biliary mucin is predominantly type 3 and 5B. Biliary mucin also enters bile from peribiliary "glands" that secrete mucin into the larger bile ducts. In gallbladder bile, mucin (type 2) originates in part as a secretion by goblet cells that are present in the gallbladder epithelium. The presumed role of mucin is to prevent bacterial adhesion to epithelial cell membranes. The anionic, amphipathic protein(s) APF/CBP bind Ca^{2+} and phospholipids, but their biological functions have not as yet been elucidated.

A variety of plasma proteins are present in bile in trace amounts, including several apoproteins (of lipoproteins), albumin, transferrin, α-2-macroglobulin, fibrinogen, α-1-antitrypsin, and haptoglobin. These proteins are considered to enter bile in part by

exocytosis of vesicles from hepatocytes, cholangiocytes, or cholecystocytes. Other modes of entry of proteins are desorption from the canalicular membrane and leakage across patulous paracellular junctions. IgG is considered to play a role in nucleation of cholesterol gallstones. Antigen–antibody complexes have also been reported to be secreted in bile. Concentrations of some proteins in gallbladder bile are summarized in Table III.

Bile also contains enzymes, presumably arising from desorption of enzymes on the luminal face of the canaliculus, from exocytosis into bile of vesicles from the hepatocyte and the biliary tract epithelium, and from sloughing of biliary tract epithelial cells. Enzymes that have been measured in bile include lactate dehydrogenase, 3-hydroxybutyrate dehydrogenase, malate dehydrogenase, glucose-6-phosphate dehydrogenase, ornithine carbamoyltransferase, aspartate aminotransferase, alanine aminotransferase, creatine kinase, cholinesterase, alkaline phosphatase, lysozyme, β-glucuronidase, aminopeptidase, γ-glutamylpeptidase, and fructose biphosphate adolase. A variety of lysosomal enzymes have also been measured in bile.

The tripeptide, glutathione is secreted into bile, mostly in the form of its disulfide, its oxidation product. During passage along the biliary tract, it is hydrolyzed to its constituent amino acids (glycine, cysteine, and glutamine) which are in part absorbed.

TABLE III Composition of Bile in the Adult Human: Proteins and Peptides

Component	Hepatic bile (g/ml)	Gallbladder bile (g/ml)
Total proteins	50–1000	1000–7000
Mucin		100–400
Immunoglobulins		
IgA		100–900
IgM	2–160	
IgG	30–500	
Albumin	100–1000	
Transferrin	10–160	
α_2-Macroglobulin	2–100	
APF/CBP	1	
Apoproteins		
ApoA-I	1–3	15–25
ApoA-II	1–2	7–15
C-I	8–18	5–12
C-II	2–4	2–5
ApoB	7–15	25–40
Glutathione		
Cyclic AMP		

Xenobiotics

Bile serves as the major excretory route for a number of lipophilic drugs and/or their metabolites. Anionic drugs and their glucuronides or sulfates are secreted into bile by a nonspecific organic anion transporter, the multidrug-resistance-associated protein 2 (Mrp2), and occasionally by a bile acid transporter, the bile salt export pump (BSEP). Cationic and some uncharged drugs are transported by the multidrug-resistance 1 (MDR1) P-glycoprotein. Because these compounds do not have an enterohepatic circulation, their biliary excretion does not exceed the daily dosage, and their concentration is usually quite low when compared to that of bile salts.

The bile acid ursodeoxycholic acid (ursodiol) is used to treat cholestatic liver disease and cholesterol gallstone disease. It is conjugated in the liver and secreted in bile. With continued administration, its pool size increases and it may become the dominant biliary bile acid.

INORGANIC CONSTITUENTS

Cations

As previously noted, biliary cations are those that are present in an ultrafiltrate of plasma. Polyvalent cations are secreted into bile by one (and possibly more) canalicular metal transporter(s). During concentration of bile in the gallbladder, the activity of Ca^{2+} increases because of Gibbs–Donnan equilibrium effects. The result is that precipitation of calcium salts in the gallbladder is not uncommon and may occur whenever the solubility product of a calcium salt is exceeded. Calcium salts of unconjugated bilirubin, carbonate, phosphate, uncommon bile acids, a number of anionic drugs, and oxalate have been observed. The most common calcium salt in gallstones is calcium bilirubinate.

Anions

The concentration of bicarbonate in hepatic bile varies widely between species, and the concentration of chloride varies reciprocally. The guinea pig is unique among mammals in having a bicarbonate concentration of 80 m*M*, and bile flow in this species appears to be driven mainly by bicarbonate secretion. In other species, bicarbonate concentration may exceed plasma concentration because of bicarbonate secretion by cholangiocytes.

CHANGES IN COMPOSITION DURING TRANSIT IN THE BILIARY TRACT

Bile Ducts

Biliary Lipids

The apical sodium dependent bile salt transporter (ASBT), which is a membrane bile salt–sodium cotransporter, is present in cholangiocytes, and some reabsorption of conjugated bile salts is likely to occur; however, this flux is small. No information exists on phospholipid or cholesterol absorption in the bile ducts, but it has been speculated that the low cholesterol content of bile in some species may result from cholesterol absorption in the biliary tract.

Despite the lack of major changes in concentration of biliary lipids, the physical state of bile may change considerably during passage along the bile ducts. During biliary lipid secretion, vesicles are desorbed from the canalicular membrane. Such vesicles may transform to micelles during the flow of bile in the ducts. Such a physicochemical transformation is simply the result of time, and the biliary ductules are unlikely to play a role in this process.

Nonlipid Organic Molecules

Unconjugated bilirubin is transported to only a negligible extent across the canalicular membrane. However, it is likely that if unconjugated bilirubin enters canalicular bile, it will be absorbed partly during transit in the biliary ductules.

Canalicular bile may be hypertonic, and becomes isosmotic during transit through the biliary ductules, because their paracellular junctions are permeable to water. Canalicular bile contains filterable solutes such as short-chain fatty acids, glucose, and amino acids. These are absorbed in part. Glutathione is hydrolyzed to its constituent amino acids by γ-glutamyl transpeptidase, which is located on the apical membrane of cholangiocytes; in humans, it is also present on the outer surface of the canalicular membrane.

Molecules present in conjugated form (glucuronides, sulfates, glutathione conjugates) are not believed to be absorbed in the biliary ductules. In some species, glutathione conjugates of xenobiotics undergo cleavage of the outer amide bonds of glutathione, leaving a cysteine conjugate (termed a "mercapturic acid derivative") of the molecule. If molecules are transported into bile in unconjugated form, and if they are sufficiently lipophilic, they will be absorbed in part in the biliary ductules to undergo a cholehepatic pathway. As noted previously, mucin and IgA are secreted into ductular bile by the biliary epithelium.

Inorganic Constituents

Bile ducts contain the cystic fibrosis transmembrane (chloride) regulator (CFTR) and chloride and possibly bicarbonate are secreted by cholangiocytes in response to the hormone secretin. Chloride undergoes chloride/bicarbonate exchange, resulting in an apparent bicarbonate secretion. Sodium ions and accompanying water molecules enter via the paracellular junctions, with the result that an isotonic solution of sodium bicarbonate is added to ductal bile.

Gallbladder

Biliary Lipids

Bile is concentrated in the gallbladder by removal of chloride, bicarbonate, and accompanying cations. Removal of biliary lipids occurs to a much smaller extent. In humans, with prolonged fasting, bile becomes less saturated in cholesterol, indicating absorption of cholesterol. Phospholipids and bile salts are also absorbed but to a smaller extent. The apical bile salt sodium transporter is present in the gallbladder mucosa. During gallbladder storage, there is little change in the phospholipid/bile salt ratio of gallbladder bile.

Nonlipid Organic Molecules

Conjugated bilirubin is not absorbed by the gallbladder wall. During prolonged storage in the gallbladder, bilirubin glucuronides may undergo spontaneous hydrolysis, leading to the formation of unconjugated bilirubin. This may be absorbed, solubilized in micelles, and/or precipitated as the insoluble calcium salt.

The nonspecific anion transporters MRP2 and MRP3 have been recently identified in gallbladder epithelium, indicating that the gallbladder mucosa might secrete organic anions into bile or absorb organic anions from bile and transport them into blood.

The protein composition and content change during gallbladder storage. In humans, haptoglobin, IgM, and IgA, as well as α1-glycoprotein, are reported to be absorbed by the gallbladder. At the same time, mucin, albumin, and IgG are secreted into bile.

Inorganic Constituents

The major change in gallbladder composition during storage is removal of chloride and bicarbonate anions as well as accompanying cations. Water is removed at the same time, with the result that bile remains isotonic. Removal of electrolytes and water is considered to occur by a double-ion exchange mechanism of the type that is responsible for water removal in the ileum. A Na^+/H^+ antiporter transports hydrogen ions into the

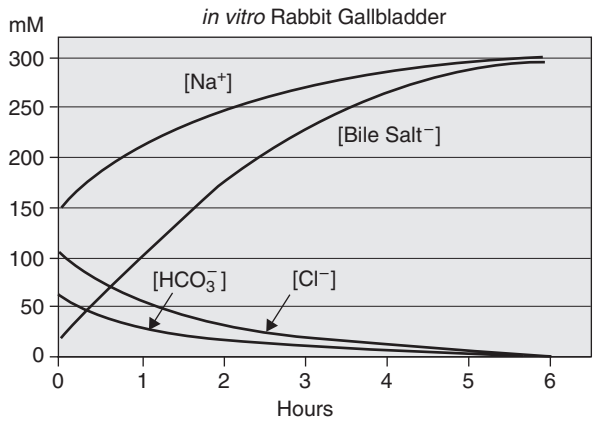


FIGURE 1 Changes in bile ion concentrations during concentration of bile in the gallbladder. Data were obtained using the rabbit gallbladder *in vitro*, but similar concentration changes are likely to occur during *in vivo* concentration of bile in the human gallbladder. However, hepatic bile bicarbonate concentration is lower in humans (about 25 m*M*) than in the rabbit (in which it is about 60 m*M*. From Moore and Hofmann (1993). Reproduced with permission from the publisher.

lumen, and sodium ions into the cell. At the same time, a chloride/bicarbonate exchanger transports chloride into the cell and bicarbonate into the lumen. In the lumen, the hydrogen ions and bicarbonate ions combine to form carbon dioxide and water, thus destroying osmoles. The sodium and potassium ions that have entered the cell are pumped out at the basolateral membrane, and this efflux of cations in turn pulls water and chloride ions through the paracellular junctions.

During this concentration process, bile is acidified. With prolonged storage, gallbladder pH may decrease from pH 7.5 to pH 5.5. Bile remains isotonic, even though total sodium concentrations and bile salt concentrations may be as high as 300 m*M*. Ca^{2+} increases modestly because of a Gibbs–Donnan equilibrium effect. A schematic depiction of the changes in concentration of the major ions in bile is shown in Fig. 1.

See Also the Following Articles

Barrier Function in Lipid Absorption • Bile Flow • Biliary Tract Anatomy • Bilirubin and Jaundice • Cholesterol Absorption • Gallstones, Pathophysiology of

Further Reading

Afdhal, N. H. (2000). "Gallbladder and Biliary Tract Diseases." Marcel Dekker, New York.

Cooper, A. D., (ed.) (1999). Bile salts: Metabolic, pathologic, and therapeutic considerations. *Clin.Gastroenterol.* **28**, 1–254.

Gleeson, D., Hood, K. A., Murphy, G. M., and Dowling, R. H. (1993). Calcium and carbonate ion concentrations and gallbladder and hepatic bile. *Gastroenterology* **102**, 1707–1716.

Gustafsson, U., Sahlin, S., and Einarsson, C. (2000). Biliary lipid composition in patients with cholesterol and pigment gallstones and gallstone-free subjects: Deoxycholic acid does not contribute to formation of cholesterol gallstones. *Eur. J. Clin. Invest.* **30**, 1099–1106.

Haigh, W. G., and Lee, S. P. (2001). Identification of oxysterols in human bile and pigment gallstones. *Gastroenterology* **121**, 118–23.

He, C., Fischer, S., Meyer, G., Muller, I., and Jungst, D. (1997). Two-dimensional electrophoretic analysis of vesicular and micellar proteins of gallbladder bile. *J. Chromatogr. A* **776**, 109–115.

Hofmann, A. F. (ed.). (1984). Physical chemistry of bile. *Hepatology* **4**(suppl.), 1S–252S.

Hofmann, A. F. (1989). Overview of bile secretion. *In* "Handbook of Physiology. Section on the Gastrointestinal System" (S. G. Schultz, ed.), pp. 549–566. American Physiological Society, Bethesda, Maryland.

Hofmann, A. F. (1994). Bile acids. *In* "The Liver: Biology and Pathobiology" (I. M Arias, J. L. Boyer, N. Fausto, W. B. Jakoby, D. A. Schachter, and D. A. Shafritz, eds.), 3rd Ed., pp. 677–718. Raven Press, New York.

Hofmann, A. F., and Strasberg, S. M. (eds.). (1990). Biliary cholesterol transport and precipitation. *Hepatology* **12**(suppl.), 1S–244S.

Keulemans, Y. C., Mok, K. S., deWit, L. T., Gouma, D. J., Groen, A. K. (1998). Hepatic bile versus gallbladder bile: A comparison of protein and lipid concentration and composition in cholesterol gallstone patients. *Hepatology* **28**, 11–16.

Klaassen, C. D. (1976). Biliary excretion of metals. *Drug Metab. Rev.* **5**, 165–196.

Lentner, C., (ed.) (1981). Bile. *In* "Ciba Scientific Tables," 8th Ed., Vol. I. Units of Measurement, Body Fluids, Composition of the Body, Nutrition. Ciba-Geigy Corporation, West Caldwell, New Jersey.

Miettinen, T. E., Iesaniemi, Y. A., Gylling, H., Jartvinen, H., Silvennoinen, E., and Meittinen, T. A. (1996). Noncholesterol sterols in bile and stones of patients with cholesterol and pigment stones. *Hepatology* **23**, 274–280.

Moore, E. W., and Hofmann, A. F. (1993). The biliary tract: Bile formation and gallbladder function (slides and text). Project of the American Gastroenterological Association. Milner-Fenwick, Timonium, Maryland.

Ostrow, J. D. (ed.) (1986). "Bile Pigments and Jaundice." Marcel Dekker, New York.

Shiffman, M. I., Sugerman, H. J., Kellum, J. M., and Moore, E. W. (1992). Changes in gallbladder bile composition following gallstone formation and weight reduction. *Gastroenterology* **103**, 214–221.

Bile Duct Injuries and Fistulas

MUFTI SAADAT A. FAROOQI
Keck School of Medicine, University of Southern California

biliary fistula An abnormal passage or communication from the biliary system to another location.
cholangiography Injection of a radiocontrast dye into the biliary system to study the anatomy; it can be performed percutaneously via the liver or endoscopically via the ampulla of Vater, located in the duodenum.
cholecystitis Inflammation of the gallbladder.
cholelithiasis The presence of stones in the bile duct system.

A fistula is defined as an abnormal communication between the lumen of a hollow viscus and another hollow organ or the integument. Abdominal fistulas generally are classified by their sites of origin and termination, by the volume and composition of drainage, and by their etiology. Fistulous connections can occur between the biliary tract and various structures, including the enteric tract, bronchial tree, skin, and blood vessels. They can develop as a complication of chronic cholelithiasis (i.e., biliary enteric fistula) or infection such as liver abscess or hydatid disease (i.e., bronchobiliary, biliary cutaneous, or bilioportal fistulas). They can also result from blunt abdominal or operative trauma (i.e., biliary cutaneous fistulas). Internal fistulas connect the biliary system to other abdominal and thoracic organs and cavities. External fistulas are connections between the biliary tract and the skin.

INTRODUCTION

Biliary fistulas were first mentioned by Bartholini in 1654. Thilesus described the first external biliary fistula in 1670. A thorough analysis of biliary fistula was first presented in 1890 by Courvoisier, who reported a large series of approximately 500 cases, including 131 patients with gallstone ileus and 169 patients with spontaneous external biliary fistula.

Biliary fistulas are usually the result of acute suppurative cholecystitis associated with cholelithiasis. The suppurative process leads to necrosis and perforation. The site of spontaneous perforation and subsequent fistula is in the gastroenteric system, in the thoracic cavity, or more rarely through the abdominal wall. Biliary fistulas are commonly reported as a complication of cholecystectomy, common bile duct exploration, inadvertent operative injury of the bile duct, or penetrating trauma.

ETIOLOGY

Biliary fistulas are incidental findings at cholecystectomy in approximately 5% of patients. Cholecystocutaneous fistulas to the abdominal wall occur but are exceedingly rare. More than 90% of internal biliary fistulas occur as a result of cholelithiasis and acute or chronic cholecystitis. During repeated episodes of inflammation, adhesions form between the gallbladder and adjacent structures, and eventually a stone (or stones) erodes into adjacent viscera. Fistulas to the duodenum account for 70% of gallstone-related internal fistulas and most of the remainder are to the stomach or colon. Although fistulas not associated with stones more often involve the common bile duct rather the gallbladder, the gallbladder may be affected.

With more widespread use of laparoscopic cholecystectomy, the incidence of bile duct injury, including biliary fistulas, has increased (compared to the incidence associated with open cholecystectomy). Bile leakage from the cystic duct remnant is among the most common injuries reported as a complication of laparoscopic cholecystectomy. The most common cause of cystic duct leaks involves imprecise application of clips on the duct or their subsequent dislodgement during the procedure. Other potential mechanisms of injury include trauma from cannulation during cholangiography, inadvertent thermal injury, necrosis of the duct from acute inflammation or dissection, induced devascularization, and ischemia.

Biliary fistulas may also arise from intrahepatic ducts and the common duct. Small anomalous bile ducts (of Luschka), if present, drain directly from the liver bed into the gallbladder. These are severed during cholecystectomy and, if not closed operatively, result in a bile leak into the gallbladder fossa.

 Copyright 2004, Elsevier (USA). All rights reserved.

CLINICAL PRESENTATION

Internal Biliary Fistulas

Most cholecystoduodenal fistulas are asymptomatic or result in nonspecific digestive complaints. Gallstones may occasionally pass through the fistula tract and cause gallstone ileus. A preoperative diagnosis of an uncomplicated cholecystoduodenal fistula is rarely made.

Patients with spontaneous development of cholecystocholic fistulas present with fever, chills, and abdominal pain due to the influx of bacteria into the biliary tract. Because the ileum, the main site of bile resorption, is bypassed, a choleric enteropathy may develop as a sign of bile acid loss. The patients present with nausea, weight loss, and steatorrhea.

Choledochoduodenal fistulas (due to duodenal ulcer disease or biliary neoplasms) are usually diagnosed incidentally, because they are rarely associated with biliary symptoms. In rare instances, cholangitis, jaundice, and abnormal liver tests may occur as indicators of a concomitant biliary tract infection or obstruction.

The hallmark of thoracobiliary or bronchobiliary fistulas (due to trauma, malignancies, liver abscess, parasitic liver disease, choledocholithiasis, postoperative biliary stenosis, or rare congenital disorders) is the presence of bile pigments in the sputum (biliptysis). Bronchiolitis is usually present. Other symptoms include right upper abdominal and pleuritic chest pain.

Fever, chills, or leukocytosis is seen in only half the cases. A right pleural effusion is almost always present.

External Biliary Fistulas

Spontaneous external biliary fistulas are usually a complication of acute cholacystitis with underlying cholelitiasis. They are now extremely rare due to better diagnosis.

An increase in iatrogenic bile duct injury has been seen with the advent of laparoscopic surgery, especially laparoscopic cholecystectomy. These injuries may become apparent in the immediate postoperative period or may develop over months. Patients may present with jaundice, fever, bile peritonitis, biliary fistula, or abdominal pain.

DIAGNOSIS

Plain Films

Gas in the biliary tree is visible on plain films in most patients with a biliary fistula. Previous surgery (choledochoduodenostomy, cholecystojejunostomy, or sphincterotomy), patulous sphincter of Oddi, and ascending cholangitis with a gas forming organism should be considered in the differential diagnosis.

Barium Contrast Examination

An upper gastrointestinal series or a barium enema is frequently useful in detecting and evaluating the cause of a biliary–enteric fistula when the diagnosis is unclear after clinical exam and a plain film. It may be especially useful in the diagnosis of biliary–enteric fistulas due to peptic ulcer disease, carcinoma, diverticulitis, or Crohn's disease.

Ultrasound

The role of sonography is limited, since the fistula is often too small to be imaged. The disappearance of previously documented gallstones in the appropriate clinical context maybe suggestive of a fistula. Chronic inflammatory changes in the gallbladder and associated mass lesion or biliary tract obstruction may be demonstrated.

CT Scan

CT is most helpful when the diagnosis of pneumobilia is unclear from the plain film or when biliary tract obstruction or a mass lesion is present.

Nuclear Medicine

99mTc-IDA scintigraphy is most useful in delineating bronchobiliary fistulas (congenital or traumatic), postoperative or traumatic fistulas, and biliary–enteric anastomotic leaks, often obviating direct cholangiography.

Cholangiography

Direct cholangiography is often necessary to opacify the fistula, demonstrate obstruction, and delineate the pathologic anatomy precisely. However, small fistulas may not be filled.

TREATMENT

Internal Biliary Fistulas

Biliary–Enteric Fistulas

Surgery is unnecessary in asymptomatic or high-risk patients with cholecystoduodenal fistula. If a cholecystoduodenal fistula is discovered incidentally during abdominal surgery, a cholecystectomy and closure of the duodenal defect are sufficient in low-risk patients.

Patients with cholecystocolic fistulas should undergo immediate surgical treatment, including cholecystectomy and closure of the fistulous communication. A laparoscopic approach has been recently described. Segmental resection of the colon may be necessary. Treatment of uncomplicated choledochoduodenal fistula is usually not indicated. Most authors agree that the therapy is dictated by the symptoms of the underlying peptic ulcer disease and not by presence of the fistula. Closure of the fistula may be achieved by conservative management of the ulcer disease. Surgical treatment of the fistula may be necessary in the presence of severe ulcer disease with perforation, bleeding, or obstruction or upon the onset of acute cholangitis.

Thoracobiliary and Bronchobiliary Fistulas

Surgery is the mainstay of treatment in congenital bronchobiliary fistulas. Excision of the fistula through a right thoracotomy is usually performed. Interventional radiology and endoscopic techniques, including stenting to reduce distal biliary obstruction, may provide a safe treatment of acquired fistulas. Currently, operative approaches should be considered only if percutaneous or endoscopic interventions have failed.

External Biliary Fistulas

Definitive treatment of a spontaneous external biliary fistula requires removal of the gallbladder and excision of the fistulous tract. Alternatively, the fistula may be curetted and left to heal spontaneously. An iatrogenic external biliary fistula is a serious and difficult complication of biliary surgery. Delayed surgical repair with a biliodigestive anastomosis after resolution of the inflammation is often preferable. However, this surgery is associated with a high morbidity. When the fistula is well contained and the patient does not present with any signs of sepsis or bile peritonitis, a nonsurgical treatment can advised. A conservative treatment with endoscopic sphincterotomy, stenting, or placing of a nasobiliary catheter to reduce intraductal pressure can facilitate spontaneous healing.

PROGNOSIS

The prognosis in patients with biliary fistula is generally fair, depending in large part on the specific type of fistula, the clinical presentation, and the presence of comorbid conditions.

See Also the Following Articles

Biliary Tract, Anatomy • Cholecystectomy • Cholelithiasis, Complications of • Fistula • Gallstones, Pathophysiology of • Sphincterotomy

Further Reading

Morrissey, K., and McSherry, C. (1994). Internal biliary fistula and gallstone ileus. *In* "Surgery of the Liver and Biliary Tract," pp. 909–922. Churchill-Livingstone, Philadelphia.

Rose, D. M., Rose, A. T., Chapman, W. C., *et al.* (1998). Management of bronchobiliary fistula as a late complication of hepatic resection. *Am. Surg.* 64, 873–876.

Ruderman, R. L., Laird, W., Reingold, M. M., and Rosen, I. B. (1975). External biliary fistula. *Can. Med. Assoc. J.* **113**, 875–878.

Wagner, G. R., and Passaro, E., Jr. (1971). Choledochoduodenal fistula secondary to duodenal ulcer. *Arch. Surg.* **103**, 21–24.

Walker, G., and Large, A. (1954). Choledochoduodenal fistula: Its surgical management. *Ann. Surg.* **139**, 510.

Bile Flow

RICHARD M. GREEN
Northwestern University Feinberg School of Medicine

bile Fluid secreted by the liver into the biliary system and subsequently into the gut. Bile is composed of water, organic anionic and cationic substances, electrolytes, and several proteins and lipids. In addition, bile is the route of excretion for many toxic substances. It contains endogenous or exogenous substances that are excreted by the liver, frequently after they undergo hepatic biotransformation.

bile acid Organic acid that is derived from cholesterol and possesses a steroid nucleus ring. Bile acids are amphiphilic substances, with both hydrophilic and hydrophobic regions of the molecule, and can thus serve as detergents to solubilize lipids.

bile canaliculus Domain of polar hepatocytes that connects with the biliary system. It contains many transport proteins that are required for bile formation.

bile salt Organic anion that is the ionized form of a bile acid.

bile salt-dependent bile flow Component of bile flow that is attributed to the osmotic effect of bile salts secreted into the bile canaliculus.

bile salt-independent bile flow Component of bile flow that is not attributed to the osmotic effect of bile salts, and is commonly attributed to the active secretion of electrolytes and other substances.

biliary system Anatomic area communicating from the liver to the gut. It includes small bile ductules, bile ducts, and the gallbladder.

bilirubin Porphyrin molecule that is a chemical breakdown product of heme metabolism. Bilirubin is very hydrophobic (insoluble in water) and must be metabolized into its water-soluble glucuronide form prior to secretion into bile. Bilirubin is yellow, thus, when serum bilirubin levels are elevated, jaundice becomes evident.

choleretic Physiologic increase in bile flow. A choleretic substance will stimulate the formation and secretion of bile.

cholestasis Pathophysiologic state characterized by an impairment of bile flow. This can be caused by a physical obstruction of the biliary system or by an impairment of bile formation.

gallbladder Organ of the biliary system that serves the functions of storing and concentrating bile. When the gallbladder is stimulated, such as following a meal, its contents are emptied into the biliary system and subsequently flow into the gut to enhance the digestive process. However, many animals lack gallbladders and humans usually have no symptoms after removal of the gallbladder.

Bile, composed of water, organic anionic and cationic substances, electrolytes, and several proteins and lipids, is the route of excretion for many toxic substances. It contains endogenous or exogenous substances excreted by the liver, frequently after they undergo hepatic biotransformation. Bile formation is an essential function of the liver. Although other organs such as the kidney and intestines are essential for the absorption and secretion of multiple xenobiotic and endobiotic substances, the liver has an important role in the metabolism of many of these substances prior to their excretion into bile.

INTRODUCTION

Many of the physiologic and molecular mechanisms responsible for bile formation have recently been elucidated, as well as the regulation of these transport proteins in both normal physiologic and pathophysiologic states. The transport function of bile formation by the liver is vectoral, occurring in a unidirectional manner. Substances are taken up from the blood of the portal circulation at the hepatic sinusoidal (basolateral) surface, are transported across the hepatocyte to the bile canaliculus, and are secreted into the canalicular space. Once the bile has left the canaliculus it may be further modified by the cholangiocytes, which are the epithelial cells lining the biliary system. Bile is often stored and concentrated in the gallbladder and is ultimately secreted via the bile ducts into the intestine. It is also important to know that bile salts, which are a major component driving the formation and flow of bile, are reabsorbed in the ileum (the distal portion of the small intestine). *De novo* bile salt synthesis quantitatively accounts for only a small amount of the bile salts that are secreted by the liver each day. The overwhelming majority of bile salts secreted into the bile each day comes from cycling in the enterohepatic circulation, in which the bile salts cycle from the liver via the bile duct into the intestine, are absorbed in the ileum and thus enter the portal circulation, return to the liver, and are taken up by the sinusoidal (basolateral) bile salt transporters. These processes may be significantly altered in many pathophysiologic disease states.

 Copyright 2004, Elsevier (USA). All rights reserved.

BILE FORMATION

Mechanisms of Bile Formation

The anatomical structure of the liver is unique in many aspects. The liver parenchyma is perfused by two distinct vascular systems, from the portal vein and from the hepatic artery. The ultrastructure of the liver also consists of a periportal and pericentral area where blood flows unidirectionally from the periportal to the pericentral end of the hepatic sinusoid. Bile formation occurs in the liver, and bile flow occurs in the opposite direction, arising in the pericentral area, crossing the midzone, and continuing to the periportal hepatocytes before it enters through the bile ductular system. There is a zonal distribution of many liver-specific proteins, including transport and metabolic proteins, which also occur in this unique anatomical and physiologic distribution.

Hepatocytes are polar epithelial cells, which have basolateral (sinusoidal) and canalicular (apical) domains, separated by tight junctions. The sinusoidal membrane is actively involved in the uptake of numerous substances into the hepatocyte from the portal circulation. Portal blood can flow within the space of Disse (between the endothelium of the portal venule and the sinusoidal membrane), and multiple substances in the blood may be taken up by one of the numerous sinusoidal uptake mechanisms. There have been several putative and definitively defined bile salt uptake mechanisms identified on the basolateral membrane of the hepatocyte. The majority of bile salt uptake is via sodium-dependent bile salt cotransport, and a minority of bile salt uptake occurs via a sodium-independent pathway. The sodium-dependent taurocholate cotransporter peptide (NTCP, in humans; Ntcp, in rodents) is the major bile salt uptake mechanism in humans and rodents. There have been several other putative human bile salt transport proteins defined, including epoxide hydrolase, although this remains controversial. A large class of organic anion transport proteins (OATPs) have also been identified in the basolateral membrane of both rodent and human liver cells; these proteins transport numerous organic anionic substances, including bile salts. In addition, there are several other transport proteins and transport mechanisms for electrolytes, including the Na^+,K^+-ATPase, the Ca^{2+}-ATPase, as well as numerous ion channels.

The Na^+,K^+-ATPase is instrumental in creating the electrochemical sodium gradient that drives sodium-dependent cotransport. Although several other channels and transporters are important in maintaining the transcellular sodium gradient, the Na^+,K^+-ATPase is an ATP-dependent protein complex that creates the electrochemical gradient. Activity of this pump has been correlated with many hepatocellular transport functions, although more recent data have also emphasized the import on bile formation of specific transport proteins that have been identified and cloned over the past decade.

Over the past decade, numerous ATP-binding cassette (ABC) proteins have been identified in the liver canalicular membrane. Members of this transporter family include several P-glycoproteins, including multidrug-resistance (MDR) proteins and MDR-associated proteins (MRPs). In addition, numerous other ABC proteins have been demonstrated to be important for the secretion of phospholipid, glutathione, and other organic anionic and cationic substances. Finally, the liver secretes numerous other organic and inorganic anions and cations as well as water (involving aquaporins or water channels). Many of these transport proteins are highly regulated and are significantly modified by numerous hormones, intracellular second messengers, pharmacologic agents, and in pathophysiologic disease states.

Functions of Bile Formation

Several physiologic functions of bile formation have been identified to be important for normal function. Bile formation provides an essential root of excretion for many xenobiotic and endobiotic substances, including bilirubin. The diminished biliary secretion of bilirubin can cause jaundice, which is often the presenting sign of hepatic disease. Clinical jaundice is often the presenting sign of liver disease because bilirubin is "easily seen" due to its characteristic yellow color. Patients with clinical jaundice also frequently have other concomitant hepatic impairments of metabolism and secretion. The liver, due to its unique microvasculature, is also frequently involved in removing from the circulation large molecules that may be bound to albumin or other transport proteins. Similarly, many protein-bound drugs are cleared from the body via biliary excretion. The liver possesses numerous metabolic pathways that frequently will metabolize substrates in the circulation prior to their hepatobiliary secretion. Many hydrophobic substances, which cannot be cleared by the kidney, are cleared from the body only after hepatic biotransformation into hydrophilic substances, which can be cleared in the bile and/or urine.

A second important function of bile formation and/or bile flow into the intestine is the enhanced absorption of hydrophobic or lipid-soluble substances. These fat-soluble substances include numerous lipids

(e.g., cholesterol) and essential fat-soluble vitamins. Bile salts serve as detergents for the solubilization of lipids and fat-soluble nutrients so they can be absorbed in the small intestine. Many cholestatic conditions cause an impaired ability of the body to absorb fat-soluble vitamins, including vitamin K. Many of the clotting factors produced by the liver are vitamin K dependent, and cholestasis can frequently cause severe coagulopathy due to impaired solubilization and the resultant malabsorption of vitamin K. Thus, bile formation and bile flow are essential for both the excretion of numerous substances from the body and the absorption of lipids and other essential fat-soluble nutrients. An impairment of bile flow from the liver into the intestine may be caused by an anatomic obstruction of the biliary system, as well as by an impairment of bile formation.

Bile Acid-Dependent and Bile-Independent Bile Formation

Because bile acids are the major organic solute in bile, it is not surprising that bile acid secretion is a major driving force for bile flow. Formation of canalicular bile has often been functionally divided into two components: bile acid-dependent bile formation and bile acid-independent bile formation. The bile acid-dependent component of bile formation is typically attributed to the osmotic effect of the bile acids. In support of this, there are data from humans, dogs, rats, and rabbits that there is a linear relationship between canalicular bile flow and bile acid output. In addition, many non-micelle-forming bile acids exhibit a larger and more significant choleretic potency than do many micelle-forming bile acids. The bile acid-independent component of bile formation is likely due to the effects of ion secretion as well as to the secretion of bicarbonate, glutathione, and bilirubin. The bile also contains numerous biliary lipids, organic anions, organic cations, and some biliary proteins.

EXTRAHEPATIC CHOLESTASIS

Extrahepatic obstruction of bile flow may be caused by any benign or malignant condition obstructing the bile ducts or adjacent anatomical structures. The initial management of a jaundiced patient or a patient with evidence of cholestasis typically involves a radiologic evaluation of the biliary system. Ultrasonography can reveal both intrahepatic and extrahepatic biliary dilatation. It is a sensitive imaging modality for the detection of cholelithiasis (gallstones) and is often the initial procedure of choice for imaging jaundiced patients. It may also identify the presence of choledocholithiasis (common bile duct stones), obstructing masses in the head of the pancreas or in the liver, as well as biliary obstruction caused by large perihepatic lymph nodes (which can cause extrinsic compression of the bile ducts).

Abdominal computer tomography (CT) scanning may also be useful for the evaluation of extrahepatic obstruction. Biliary dilatation may be evident, and this imaging modality may be more effective than ultrasonography in identifying and characterizing masses of the liver or pancreas, or lymphadenopathy. Endoscopic retrograde cholangiopancreatography (ERCP) has been the gold standard for the evaluation of biliary dilatation. In addition, it is useful not only diagnostically, but also therapeutically when biliary obstruction must be endoscopically treated. Percutaneous transhepatic cholangiography (PTC), intraoperative, or T-tube cholangiography can also be employed to define the biliary anatomy. Magnetic resonance cholangiography (MRC) is also highly effective in evaluating the anatomy of the biliary system. Recent evidence indicates that it is almost as sensitive as ERCP, and may be more cost effective due to the lower complication rate. However, it does not offer the option of therapeutic intervention at the time of the procedure.

Extrahepatic obstruction of the biliary system may commonly be caused by several benign or malignant conditions. Common bile duct stones (choledocholithiasis), bile duct strictures from malignant or benign causes, and extrahepatic compression of the bile duct from pancreatic or hepatic masses may all cause obstruction. These may include primary adenocarcinomas of the pancreas or hepatocellular or metastatic carcinoma, or may potentially be due to metastases and lymphadenopathy of the porta-hepatis lymph nodes. Lymphadenopathy due to lymphoma, neoplasia, or many infectious diseases may potentially all cause biliary obstruction.

INTRAHEPATIC CHOLESTASIS

An impairment of bile flow can also occur in the absence of an anatomic obstruction of the biliary system. This condition is termed "intrahepatic cholestasis," and is due to an impairment of bile formation and/or impaired biliary secretion, often with resultant pruritis or jaundice. Over the past decade, many of the molecular mechanisms responsible for bile formation have been defined, as well as their regulation in pathophysiologic diseases that result in cholestasis. In addition, many genetic diseases responsible for impaired bile formation

have also been elucidated, which often present in the neonatal or pediatric population.

Intrahepatic cholestasis may be caused by many factors, including medication, toxins, inflammation, hepatic regeneration, or inborn errors of metabolism. As previously discussed, distinct hepatocellular transport systems are responsible for the biliary secretion of the various organic anionic, organic cationic, or nonionic substances secreted into the bile. Many sinusoidal (basolateral) transporters have been identified that are responsible for the uptake of xenobiotics or endobiotics from the portal blood into the hepatocytes. Following the hepatocellular uptake of these substances, they may be biotransformed by one of many metabolic pathways found in the hepatocyte. Many of these pathways are responsible for the conversion of hydrophobic (fat-soluble) substances into hydrophilic (water-soluble) substances, so that they may be secreted into bile. These substances are subsequently transported to the canalicular membranes of the hepatocytes for biliary secretion. Intracellular transport mechanisms of many of these substances remain to be elucidated, but likely involve diffusion, binding to intracellular binding proteins, and potentially membrane-to-membrane transport. These substances are subsequently secreted into the biliary system by distinct canalicular transport proteins. Diminished expression or function of canalicular membrane transporters or other hepatic transporters and/or binding proteins may be responsible for intrahepatic cholestasis.

Medications are a common cause of intrahepatic cholestasis. Although, virtually any drug can cause abnormal liver function, certain medications are more frequently implicated in causing cholestasis. Anabolic steroids, estrogens, protease inhibitors such as indinavir, trimethoprim/sulfamethoxazole, phenothiazines, and many antihypertensive medications are frequent causes of an elevated serum bilirubin or alkaline phosphatase. Total parenteral hyperalimentation (TPN) also often causes cholestasis. Severe inflammation due to infection and sepsis may result in cholestasis, a clinical entity often termed "the jaundice of sepsis." Similarly, "benign postoperative jaundice" is a multifactorial entity due to cholestasis from inflammation and medications, often with concomitant hepatic ischemia and increased bilirubin production due to the breakdown of red blood cells from blood transfusions or hematomas. Other less common causes of intrahepatic cholestasis include the intrahepatic cholestasis of pregnancy or benign recurrent intrahepatic cholestasis (BRIC), which are likely due to mutations of canalicular transport proteins, potentially in combination with hyperestrogenemic or other stressor states.

Finally, acute or chronic liver disease due to hepatitis, primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), the vanishing bile duct syndromes, or other cholestatic diseases may also present as intrahepatic cholestasis. In addition, cirrhosis due to almost any cause may cause cholestasis. Certain types of hepatitis, such as alcohol-induced liver injury or certain viral infections (hepatitis A, hepatitis B, hepatitis C, Ebstein–Barr virus, cytomegalovirus), may also present with particularly cholestatic forms of hepatitis. The cholestasis caused by these etiologies typically resolves when the inciting agent or acute disease process resolves. However, in certain forms of genetic or acquired diseases that cause progressive liver disease, the cholestasis may continue to worsen and cause significant morbidity and mortality.

SUMMARY

In summary, bile formation, biliary secretion, and the resultant bile flow into the intestine are essential functions of the liver and hepatobiliary system. Recent scientific advances have helped elucidate the mechanisms of bile formation and the pathophysiology in many disease states. Hepatobiliary secretion is an essential route of elimination for many xenobiotic and endobiotic substances, and significant symptoms such as jaundice and pruritis can develop when bile flow is diminished. Extrahepatic obstruction can be caused by intrinsic tumors of the biliary system or by external compression of the bile duct. Intrahepatic cholestasis can be due to a host of pathophysiologic states and is typically characterized by impairment of the molecular mechanisms of bile formation. Elimination of extrahepatic obstruction and treatment of the primary causes of intrahepatic cholestasis will often result in normalization of bile flow, and symptomatic improvement of the patient.

Acknowledgments

This work was supported by grants RO1 DK59580 and RO1 HD40027 and Veteran's Administration Merit Review Funding.

See Also the Following Articles

Bile Composition • Biliary Tract, Development • Bilirubin and Jaundice • Cholelithiasis, Complications of • Gallstones, Pathophysiology of • Vitamin K: Absorption, Metabolism, and Deficiency

Further Reading

Elferink, R. O., and Groen, A. K. (2002). Genetic defects in hepatobiliary transport. *Biochim. Biophys. Acta* **1586**(2), 129–145.

Erlinger, S. (1996). Mechanisms of hepatic transport and bile secretion. *Acta Gastroenterol. Belg.* **59**(2), 159–162.

Hofmann, A. F. (1989). Current concepts of biliary secretion. *Dig. Dis. Sci.* **34**(Suppl. 12), 16S–20S.

Meier, P. J., and Stieger, B. (2002). Bile salt transporters. *Annu. Rev. Physiol.* **64**, 635–661.

Thompson, R., and Jansen, P. L. (2000). Genetic defects in hepatocanalicular transport. *Semin. Liver Dis.* **20**(3), 365–372.

Bile Formation

GERALD SALEN AND ASHOK K. BATTA

University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, and Veteran's Administration Medical Center, East Orange, New Jersey

amphiphilic Functioning in both oil and water; bile salts are amphiphilic in nature because they have both a polar, hydrophilic surface, which contains the hydroxyl groups and a carboxyl group, and a nonpolar, cyclopentanophenanthrene hydrophobic surface.

apical sodium-dependent bile acid transporter Brush border glycoprotein involved in the reabsorption of conjugated bile acids in the ileum.

bile salt export pump ATP-binding cassette type of membrane transporter located at canalicular microvilli of the hepatocytes; preferentially transports conjugated tri- and dihydroxy bile acids from the sinusoids into the canaliculi.

cholehepatic shunt Mechanism to explain hypercholeresis (greatly increased bile flow) resulting from hepatic secretion of unconjugated bile acids. According to this mechanism, the unconjugated bile acid is secreted by the hepatocyte in anionic form, protonated by carbonic acid via carbonic anhydrase, passively reabsorbed in the protonated form by the biliary ductular cells, and carried back to the hepatocytes via the periductular capillary plexus, thus generating HCO_3^- within the biliary tree. The bile acid is thus available at the sinusoidal membrane for another enterohepatic cycle, which may go on until the bile acid is biotransformed.

farnesoid X receptor Orphan nuclear receptor expressed predominantly in the liver, kidney, adrenals, and intestine; a very potent ligand for hydrophobic bile acids. This receptor binds to bile acids as a heterodimer with the retinoid X receptor α, and the heterodimer thus formed regulates the transcription of several genes involved in bile acid homeostasis.

micelle Colloidal aggregates of bile salts with molecular weights of 16,000 to 40,000. Above their critical micellar concentrations, bile salts form micelles with their lipid-soluble (hydrophobic) surfaces facing each other and the hydrophilic polar hydroxyl groups exposed to the water phase. These micelles are able to take up other amphiphilic solutes, such as cholesterol and lecithin, to form mixed micelles.

Na^+-dependent hepatocellular cotransporting protein Sodium/taurocholate cotransporting protein present in basolateral rat liver plasma membranes; accounts for most of the physiological properties of Na^+-dependent bile acid uptake in intact liver. This glycoprotein in rats has 362 amino acids and an apparent molecular weight of 51,000. The human analogue consists of 349 amino acids with 77% amino acid homology to the rat liver protein.

organic anion transport protein Polypeptide in rat liver; a glycoprotein with 670 amino acids and a native molecular weight of approximately 80,000 in sinusoidal plasma membranes. The human analogue also consists of 670 amino acids and shows a 67% amino acid homology to the rat liver protein.

Bile formation is essential for normal intestinal lipid digestion and absorption, excretion of lipid-soluble xenobiotics and endogenous toxins, and cholesterol homeostasis. Interference in regular bile formation can

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Elferink, R. O., and Groen, A. K. (2002). Genetic defects in hepatobiliary transport. *Biochim. Biophys. Acta* **1586**(2), 129–145.

Erlinger, S. (1996). Mechanisms of hepatic transport and bile secretion. *Acta Gastroenterol. Belg.* **59**(2), 159–162.

Hofmann, A. F. (1989). Current concepts of biliary secretion. *Dig. Dis. Sci.* **34**(Suppl. 12), 16S–20S.

Meier, P. J., and Stieger, B. (2002). Bile salt transporters. *Annu. Rev. Physiol.* **64**, 635–661.

Thompson, R., and Jansen, P. L. (2000). Genetic defects in hepatocanalicular transport. *Semin. Liver Dis.* **20**(3), 365–372.

Bile Formation

GERALD SALEN AND ASHOK K. BATTA

University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, and Veteran's Administration Medical Center, East Orange, New Jersey

amphiphilic Functioning in both oil and water; bile salts are amphiphilic in nature because they have both a polar, hydrophilic surface, which contains the hydroxyl groups and a carboxyl group, and a nonpolar, cyclopentanophenanthrene hydrophobic surface.

apical sodium-dependent bile acid transporter Brush border glycoprotein involved in the reabsorption of conjugated bile acids in the ileum.

bile salt export pump ATP-binding cassette type of membrane transporter located at canalicular microvilli of the hepatocytes; preferentially transports conjugated tri- and dihydroxy bile acids from the sinusoids into the canaliculi.

cholehepatic shunt Mechanism to explain hypercholeresis (greatly increased bile flow) resulting from hepatic secretion of unconjugated bile acids. According to this mechanism, the unconjugated bile acid is secreted by the hepatocyte in anionic form, protonated by carbonic acid via carbonic anhydrase, passively reabsorbed in the protonated form by the biliary ductular cells, and carried back to the hepatocytes via the periductular capillary plexus, thus generating HCO_3^- within the biliary tree. The bile acid is thus available at the sinusoidal membrane for another enterohepatic cycle, which may go on until the bile acid is biotransformed.

farnesoid X receptor Orphan nuclear receptor expressed predominantly in the liver, kidney, adrenals, and intestine; a very potent ligand for hydrophobic bile acids. This receptor binds to bile acids as a heterodimer with the retinoid X receptor α, and the heterodimer thus formed regulates the transcription of several genes involved in bile acid homeostasis.

micelle Colloidal aggregates of bile salts with molecular weights of 16,000 to 40,000. Above their critical micellar concentrations, bile salts form micelles with their lipid-soluble (hydrophobic) surfaces facing each other and the hydrophilic polar hydroxyl groups exposed to the water phase. These micelles are able to take up other amphiphilic solutes, such as cholesterol and lecithin, to form mixed micelles.

Na^+-dependent hepatocellular cotransporting protein Sodium/taurocholate cotransporting protein present in basolateral rat liver plasma membranes; accounts for most of the physiological properties of Na^+-dependent bile acid uptake in intact liver. This glycoprotein in rats has 362 amino acids and an apparent molecular weight of 51,000. The human analogue consists of 349 amino acids with 77% amino acid homology to the rat liver protein.

organic anion transport protein Polypeptide in rat liver; a glycoprotein with 670 amino acids and a native molecular weight of approximately 80,000 in sinusoidal plasma membranes. The human analogue also consists of 670 amino acids and shows a 67% amino acid homology to the rat liver protein.

Bile formation is essential for normal intestinal lipid digestion and absorption, excretion of lipid-soluble xenobiotics and endogenous toxins, and cholesterol homeostasis. Interference in regular bile formation can

Copyright 2004, Elsevier (USA). All rights reserved.

lead to life-threatening complications. Knowledge of the intricacies involved in bile formation and its flow in the enterohepatic circulation is integral to developing therapeutic treatments of hepatobiliary diseases.

WHAT IS BILE?

Bile is a complex fluid; it is isosmotic with plasma and is composed primarily of water, inorganic electrolytes, and organic solutes such as bile acids, phospholipids, cholesterol, and bile pigments. Bile is secreted by the hepatocytes, which transport a wide variety of endogenous and exogenous substances from blood into the bile capillaries. Bile formation begins at the level of hepatocytes by the net movement of water and solutes into the microvilli of the bile capillaries, which are formed by the apposition of the cell membranes of the hepatocytes. The hepatocytes are typically arranged in single-cell-thick plates and are joined via the bile canaliculi, approximately 1 μm in diameter from the basolateral, or sinusoidal, domain. Bile salts, cholesterol, phospholipids, and conjugated bilirubin enter the lumens of the bile capillaries together with water and inorganic electrolytes by a process of active transport. Neighboring canaliculi join together and empty into small terminal ductules (canals of Hering), which in turn convey the bile to larger ductules, the intralobular and interlobular ducts, and eventually to the extrahepatic bile ducts. Bile flows through these extrahepatic bile ducts into the gallbladder (when present), where it is stored, and into the intestine, where cholesterol and the lipids from the phospholipids are absorbed in the jejunum and conjugated bile acids are absorbed in the ileum. In the gallbladder, the sodium, calcium, chloride, and bicarbonate ions and water are reabsorbed, resulting in a 5- to 10-fold concentration of the remaining solid constituents of bile. The daily volume of bile secretion in humans averages 500–800 ml.

After food is ingested and the gastric contents, in particular fat, reach the duodenum, secretin is released from the duodenal mucosa, which stimulates the bile duct epithelium to release water and bicarbonate. Simultaneously, the hormone cholecystokinin is released from the duodenal mucosa, which stimulates the gallbladder to contract and the sphincter of Oddi to relax, with the result that the bile flows into the duodenum. The bile salts are more than 95% reabsorbed in the terminal ileum and return to the liver by a process of enterohepatic cycling, to be excreted again into the bile. With the completion of eating and gastric emptying, the neural and hormonal stimuli cease and the sphincter of Oddi closes so that the gallbladder resumes collecting, concentrating, and storing the bile. The relative proportions of the major organic solutes in hepatic bile are illustrated in Table I.

TABLE I Composition of Hepatic and Gallbladder Bile

Bile component	Hepatic bile (%)	Gallbladder bile (%)
Water	97	89
Solids	3	11
Bile salts	0.2–2.0	6
Bilirubin	0.02–0.07	2.5
Cholesterol	0.06–0.16	0.2–0.4
Phospholipids	0.04	0.1–0.4
Fat	0.12	0.3–1.2
Inorganic salts	1.0	0.8

BILE COMPOSITION

Bile acids or bile salts are the major organic solutes in bile and are derived from the liver as well as the intestine. Primary bile acids (cholic acid and chenodeoxycholic acid, in humans) are synthesized from cholesterol in the liver, and secondary bile acids (deoxycholic acid and lithocholic acid, in humans) are produced from primary bile acids by bacteria during their intestinal transit (Fig. 1). Bile acids and cholesterol have the same steroid skeleton (the cyclopentanophenanthrene nucleus with methyl groups between the rings at positions 18 and 19 and a side chain at C-17). Most bile acids have a C_5 side chain with the terminal carbon as the carboxyl group. However, the major bile acids in alligators, 5β-cholestanoic acids, have a C_8 side chain with the terminal carbon as a carboxyl group (Fig. 1). In addition, bile acids have hydroxyl groups in the nucleus, at C-3, C-7, and/or C-12, and the majority of bile acids contain a cis-A/B ring fusion. After synthesis, bile acids are converted into their glycine and taurine conjugates by hepatic enzymes, forming bile salts that are then secreted into the bile. Conjugated bile acids are more hydrophilic than the free (unconjugated) bile acids and possess lower pK_a values, factors that enhance their water solubility and decrease their ability in the small intestine to traverse cell membranes by passive diffusion . Further, glycine and taurine conjugates of bile acids are resistant to hydrolysis by pancreatic enzymes, so that they can accumulate in the small intestine at a high enough intraluminal concentration to facilitate fat digestion and absorption. Secondary bile acids are formed in the intestinal lumen by bacterial deconjugation and dehydroxylation of primary bile salts and are passively absorbed and account for 20% of the human biliary bile salt pool.

Bile salts are amphiphilic [literally, "loving" (mixing with) both oil and water] because they have both a polar,

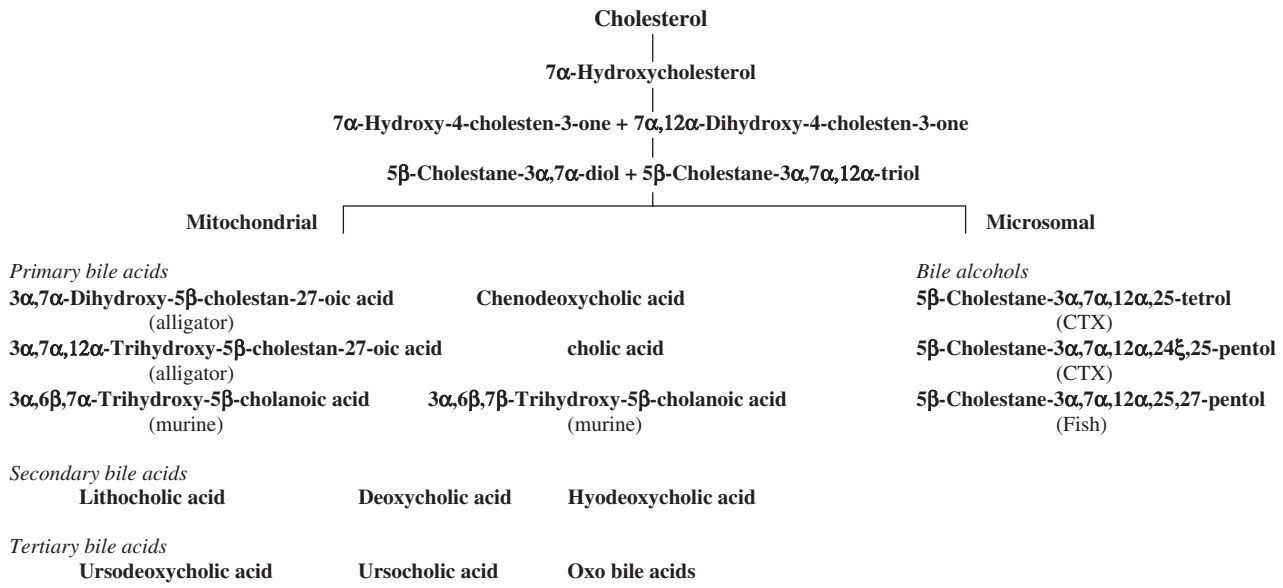


FIGURE 1 Metabolism of cholesterol in humans and animal species. In the classic bile acid synthetic pathway, a series of ring modifications precede side chain oxidation to yield 5β-cholestane-3α,7α-diol and 5β-cholestane-3α,7α,12α-triol, which are then converted into bile acids either via the mitochondrial 27-hydroxylation pathway or via the microsomal 25-hydroxylation pathway. Several bile alcohols play a role in cerebrotendinous xanthomatosis (CTX).

hydrophilic surface and a nonpolar, hydrophobic surface. Above their critical micellar concentrations, bile salts form micelles (colloidal aggregates with molecular weights of 16,000–40,000), with their lipid-soluble (hydrophobic) surfaces facing each other and the hydrophilic polar groups exposed to the water phase. These micelles are able to take up other amphiphilic solutes, such as cholesterol and lecithin, to form mixed micelles. The aggregation of bile salts in micelles reduces their osmotic effectiveness, and the micelles also bind sodium and potassium and thereby remove them from the ionic, osmotically active state. The sodium ion activity in gallbladder bile is measured to be 148–186 nmol/liter, whereas the sodium content is 220–340 mmol/liter. Thus, even though bile is isosmolar with plasma, it contains very high concentrations of bile salts and sodium. The major ionic constituents of bile are listed in Table II.

GALLBLADDER FUNCTION

The primary functions of the gallbladder are collecting, concentrating, and storing bile during interdigestive periods; emptying the bile by smooth muscle contraction in response to cholecystokinin; bile acidification; moderating hydrostatic pressure within the biliary tract; and absorbing lipid-soluble organic components of bile, such as unconjugated dihydroxy bile acids, radiocontrast agents, unsaturated fatty acids, lecithin, lysolecithin, and even free cholesterol. The greatly increased concentration of bile salts and lecithin in the gallbladder bile helps keep relatively large concentrations of cholesterol in solution; this greatly reduces the risk of cholesterol gallstone formation due to cholesterol supersaturation in the gallbladder, where the bile remains stored for hours to days. With the 6- to 10-fold concentration of bile salts in the gallbladder bile during fasting, which accounts for up to 60% of the total bile acid pool, the hepatic bile acid secretory rate and bile salt concentration within the hepatic bile ducts

TABLE II Ionic Constituents of Hepatic and Gallbladder Bile

Ionic component	Hepatic bile (mmol/liter)	Gallbladder bile (mmol/liter)
Na^+	174	220–340
K^+	6.6	6–10
Cl^-	55–107	1–10
HCO_3^-	34–65	0–17
Bile salts	28–42	290–340
Ca^{2+}	6	25–32
Mg^{2+}	0.5	
Osmolality	299 mOsm/liter	299 mOsm/liter

are markedly reduced. However, the secretion of biliary lecithin and cholesterol is reduced, with the result that lecithin-rich vesicles are formed in the bile acid-depleted bile; these vesicles carry increased amounts of cholesterol that may precipitate.

The concentrated gallbladder bile is largely the result of electroneutral Na^+-coupled Cl^- transport and passive water movement. In addition, HCO_3^- transport also takes place via Cl^-/HCO_3^- exchange and perhaps substitution by short-chain fatty acids, in particular butyrate. The net result is that the gallbladder bile is isotonic to plasma and is composed of high concentrations of Na^+, bile salts, and Ca^{2+} and lower concentrations of chloride and bicarbonate, compared to hepatic bile.

BILE SALT-DEPENDENT AND -INDEPENDENT BILE FORMATION

Bile formation can be considered to be a result of active secretion of electrolytes and bile acids followed by passive water flow. Three different modes of bile acid uptake by the liver have been identified: nonsaturable, Na^+-dependent saturable, and Na^+-independent saturable. The Na^+-dependent transport of bile acids is quantitatively the most common mode of uptake and seems physiologically to be the most important mechanism. The Na^+-independent component has been detected in a few studies and is nonspecific and shared by other organic anions (e.g., sulfobromophthalein). Most studies have shown that there is a Na^+-dependent uptake of conjugated bile acids, in particular taurocholate, and a Na^+-independent uptake of all unconjugated bile acids and certain nonbile acid anions, such as sulfobromophthalein, bilirubin, and many other xenobiotic substrates. The solute uptake is mediated by certain carrier proteins that have been recently characterized by photoaffinity labeling with fluorescent bile salt derivatives and by cloning techniques. The apparent molecular weight of the bilirubin and sulfobromophthalein carrier is 60,000 or 55,000; of the nonesterified fatty acid transporter, 40,000; and of the bile acid transporter, 48,000–50,000. Most of these transport proteins are located in the basolateral membrane of the liver.

Sinusoidal Transport

The sinusoidal uptake of conjugated bile acids, such as taurocholate, is primarily mediated by a secondary active transport process driven by the inwardly directed Na^+ gradient maintained by Na^+,K^+-ATPase located in the basolateral membrane. Studies on hepatocellular membrane vesicles show that Na^+/taurocholate cotransport is responsible for the uptake across the basolateral membrane. The transport rates of cholate and taurocholate are decreased in the absence of extracellular sodium, but are not completely suppressed, suggesting that the inward transport of bile acids is composed of both Na^+-dependent and Na^+-independent components. In contrast to the Na^+-dependent bile salt uptake, the physiological and biochemical properties of the largely nonspecific Na^+-independent hepatocellular bile salt uptake pathways are less well defined. The driving force for the Na^+-independent uptake of bile acids is not known, but most studies suggest the role of a Na^+-dependent bile acid transporter identified as the Na^+/taurocholate cotransporting protein (NTCP).

The amphipathic bile salts and nonbile salt organic anions reach the sinusoidal or basolateral surface of hepatocytes largely as albumin-bound complexes and pass across the fenestrae of sinusoidal endothelial cells. Dissociation of the bile salts, sulfobromophthalein, and fatty acids is facilitated by the microenvironment of the space of Disse or by the sinusoidal plasma membrane. However, the hepatocellular uptake of these organic anions is a saturable function of the unbound ligand concentrations at physiological albumin levels and involves Na^+-dependent and Na^+-independent transport systems.

The *Xenopus laevis* sinusoidal Na^+/taurocholate cotransporting protein (Ntcp) was the first Na^+-dependent hepatocellular cotransporting protein identified and characterized by expression cloning of *X. laevis* oocytes. It is a glycoprotein with 362 amino acids and an apparent molecular weight of 51,000 in isolated basolateral rat liver plasma membranes. The human analogue Ntcp (human NTCP) has also been identified and consists of 349 amino acids with 77% amino acid homology with the rat liver Ntcp. A number of recent studies have shown that Na^+/taurocholate cotransporting protein can account for most, if not all, physiological properties of Na^+-dependent bile acid uptake in intact liver. In addition, expression cloning in *X. laevis* oocytes also identified the rat organic anion-transporting polypeptide 1 (oatp1) from rat liver as well as the human organic anion-transporting polypeptide (OATP) from human liver. Rat oatp1 is a glycoprotein with 670 amino acids and a native molecular weight of approximately 80,000 in sinusoidal plasma membranes isolated from rat liver. Human OATP also consists of 670 amino acids and shows a 67% amino acid homology with the rat liver oatp1, thereby suggesting that both transporters may belong to the same transporter gene family, although significant differences between the two exist on both structural and functional levels. Cloning

Ntcp/NTCP and oatp1/OATP has increased our understanding of the molecular physiology of sinusoidal bile acid, organic anion, and drug uptake into rat and human livers. Both Ntcp and NTCP are physiologically relevant hepatocellular Na^+-dependent bile acid carriers, whereas the cloning of oatp1 and OATP has confirmed the existence of Na^+-independent multispecific bile acid carriers at the basolateral plasma membrane domain of mammalian hepatocytes. The transport properties of oatp1 can account at the same time for the multispecific bile acid carrier, the neutral steroid carrier, and the charge-independent organic cation carrier.

Canalicular Transport

Transport of bile acids across the canalicular membrane of the hepatocyte represents the rate-limiting step in overall transport of bile acids from blood into bile. This canalicular excretion of monoanionic bile acids, which preferentially transports conjugated tri- and dihydroxy bile acids, occurs largely via an ATP-dependent processor. The transporter, the bile salt export pump (BSEP), has been cloned and is a member of the ATP-binding cassette superfamily of proteins. The BSEP is related to the multidrug resistance gene products, or P-glycoproteins, is expressed exclusively in the liver, and has been localized to the canalicular domain of hepatocytes. The expression of this transporter has been shown to be sensitive to bile acid flux through the hepatocyte, possibly at the level of transcription of the BSEP gene. The promoter of the BSEP gene was very recently cloned and was shown to contain an inverted repeat-1 (IR-1) element, which serves as a binding site for the farnesoid X receptor (FXR), a nuclear receptor for bile acids. The FXR heterodimerizes with the retinoid X receptor α (RXRα), and when bound to bile acids, the complex FXR/RXRα regulates the transcription of several genes involved in bile acid homeostasis. These studies demonstrate a critical role of FXR as a physiological bile acid sensor integrating the expression of BSEP in bile acid homeostasis.

ELECTROLYTE TRANSPORT AND BILE FORMATION

Sodium and potassium are the major extra- and intracellular cations in all living cells. The mechanism of the bile salt-independent fraction of canalicular bile flow is considered to involve the Na^+, K^+-activated adenosine triphosphatase sodium pump in the canalicular membrane, whereby three sodium ions are pumped out of the cells for every two potassium ions pumped in, resulting in a net intracellular negative potential that amounts to approximately −35 mV in intact hepatocytes. In cultured hepatocytes, it has been demonstrated that substrate-induced (alanine, taurocholate) sodium influx is associated with a rapid increase in Na^+, K^+-ATPase cation pumping, indicating that pump activity is directly modulated by the intracellular sodium concentration. The basolateral Na^+, K^+-ATPase-mediated electrogenic cation pumping represents a major driving force for both Na^+-coupled solute transport across the sinusoidal membrane and potential dependent anion secretion across the canalicular membrane.

Plasma membrane Na^+/H^+ exchange (or antiport) that involves a sequential uptake of Na^+ by Na^+, K^+-ATPase followed by uptake of protons via Na^+/H^+ exchange is selectively located in the basolateral membrane in the hepatocyte and serves a variety of interrelated metabolic and transport functions, including regulation of intracellular pH and cell volume control. This Na^+/H^+ exchange plays an important role in the generation of bile acid-induced choleresis, e.g., the choleresis induced by physiological bile acids, such as taurocholate and the ursodeoxycholic acid-stimulated bicarbonate secretion. Although a Na^+/Cl^- or a $Na^+/K^+/Cl^-$ cotransport has not been found in rat liver cells, the presence of an electroneutral Cl^-/HCO_3^- antiport is demonstrated at the canalicular membrane and provides an explanation for the role of Na^+/H^+ exchange in the canalicular secretion of bicarbonate stimulated by ursodeoxycholic acid. Also, a Na^+/HCO_3^- cotransport has been demonstrated in the basolateral, but not canalicular, membrane vesicles; it mediates Na^+-coupled HCO_3^- influx into rat hepatocytes under physiological conditions and helps explain the Na^+ dependency of the ursodeoxycholic acid-induced hypercholeresis.

Indirect evidence suggests that transport of glutathione across the canalicular plasma membrane into bile contributes to the formation of the bile acid-independent fraction of bile flow and to an increase in biliary glutathione that is associated with a corresponding increase in bile flow. This tripeptide is secreted into canalicular bile in relatively high concentrations via a specific carrier-mediated mechanism. Quantitative measurements show that bile formation with glutathione is of the same order as that with bile acids when measured at low rates of bile acid excretion, i.e., in the absence of micelles.

ENTEROHEPATIC CIRCULATION

The enterohepatic circulation of bile acids involves hepatic bile acid formation and secretion into the bile, accumulation of bile in the gallbladder, bile emptying into the small intestine, bile acid reabsorption from the

small intestine, and bile acid transport into the liver via the portal blood. The intestinal conservation of bile acids is approximately 95% efficient and reflects both passive and active reabsorption processes. The total bile acid pool in humans is about 3–5 g and cycles 6 to 10 times daily, depending on the number of times food is ingested. In this way, 20–30 g of bile acids enter the proximal small intestine in 24 hours. The daily 0.5-g fecal loss of bile acids is balanced by hepatic *de novo* synthesis of bile acids from cholesterol. The bile acids exert two major actions on bile formation in mammals. First, they stimulate bile flow, primarily as a result of osmotic effects of their active secretion into the biliary canaliculi, and second, they stimulate the biliary secretion of the water-insoluble amphipaths lecithin and cholesterol, with which they form mixed micelles. Based on the observation that biliary bile acid output and bile flow are linearly related in all animal species studied so far, bile acid-stimulated bile flow has been attributed to active secretion of bile acids followed by osmotically driven water flow. However, the choleretic efficiency may vary greatly for the same bile acid given to different animal species, or for different bile acids given to the same animal species, and these differences are not always readily explained by differing micelle-forming properties of the bile acids.

Hypercholeresis and Cholehepatic Shunt

Unconjugated bile acids produce choleresis in excess of that predicted by osmotic action of the secreted bile acid, and it is accompanied by a selective increase in biliary HCO_3^- concentration. This phenomenon is called hypercholeresis, which is bile acid-dependent bile flow that appears to be greater than what can be explained by the osmotic effects of the bile acid recovered in bile. Hypercholeresis is observed on feeding a dose of a bile acid that exceeds the capacity of the liver to conjugate fully the infused bile acid with glycine and taurine; the result is that a proportion of the unconjugated bile acid is secreted into the bile. This phenomenon is also observed with infusion of nor-bile acids (which have four carbon atoms in the side chain, instead of the five carbons in the normal bile acids). The hepatic enzymes are unable to conjugate nor-bile acids. Two mechanisms have been proposed to explain hypercholeresis. In the first, the unconjugated bile acid causes intracellular alkalinization by increasing H^+ excretion across the sinusoidal membrane; the increased intracellular pH then induces HCO_3^- excretion across the canalicular membrane. The second mechanism is explained by what is called the "cholehepatic shunt pathway," according to which the unconjugated bile acid is secreted by the hepatocyte in anionic form, protonated by carbonic acid in the biliary tree via carbonic anhydrase, passively reabsorbed in the protonated form by the biliary ductular cells, and carried back to the hepatocytes via the periductular capillary plexus, thus generating HCO_3^- within the biliary tree. The bile acid is thus available at the sinusoidal membrane for another enterohepatic cycle, which may go on until the bile acid is biotransformed.

Intestinal Absorption

An important component of the enterohepatic circulation is the active reabsorption of conjugated bile acids in the ileum, via the ileal apical sodium-dependent bile acid transporter (ASBT), an integral brush border membrane glycoprotein that cotransports sodium and bile acids to ileocytes. Conjugated trihydroxy bile acids (taurocholate) are preferred over dihydroxy (taurochenodeoxycholate) bile acids, and unconjugated bile acids are passively absorbed throughout the small intestine. Several recent studies suggest that the ASBT is regulated by intraluminal bile acids, and conditions that reduce intestinal bile acid concentrations, such as fasting, biliary diversion, and extrahepatic cholestasis, decrease bile acid transport into brush border membrane vesicles. However, others have found a feedback inhibition of ASBT expression after taurocholate feeding in the rat and guinea pig.

Bile acids are transported across the ileal brush border membrane aided by the ASBT, and the accumulation of bile acids within the enterocyte against its electrochemical gradient is driven by the inwardly directed Na^+ gradient, maintained by a basolateral Na^+, K^+-ATPase. Intracellular transport of bile acids in the enterocyte is mediated by several cytosolic and microsomal proteins. At the basolateral membrane of the ileum, bile acids leave the enterocyte by a Na^+-independent anion exchange process. Absorption of bile acids into the portal blood occurs mainly by way of hydrophobic binding to albumin, although a small fraction is bound to lipoproteins. Bile acid uptake from the portal blood, completing the enterohepatic circulation, is typically expressed as the first-pass extraction (percent bile acid removed during a single passage through the hepatic acinus). The fractional extraction is greater for the hydrophilic bile acids compared to the hydrophobic bile acids (80–90% for taurocholate vs. approximately 50% for unconjugated ursodeoxycholic acid). The small fraction of bile acids that escapes absorption from the small intestine undergoes bacterial modification in the colon: deconjugation, 7-dehydroxylation, epimerization of hydroxyl group(s), and oxidation of

hydroxyl groups to oxo-bile acids. The predominant secondary bile acids formed in the colon are the 7-dehydroxylated bile acids, lithocholic acid and deoxycholic acid, formed by 7α-dehydroxylation of chenodeoxycholic acid and cholic acid, respectively. These secondary bile acids are passively absorbed and brought into enterohepatic circulation.

CLINICAL IMPLICATIONS

Cholestasis

Cholestasis is a functional defect in bile formation at the level of the hepatocyte (i.e., intrahepatic cholestasis, an example of which is primary biliary cirrhosis) or an obstruction to bile flow within the biliary tract (i.e., extrahepatic cholestasis, which includes primary sclerosing cholangitis). Extrahepatic cholestasis may be caused by choledocholithiasis, pancreatic and periampullary carcinoma, biliary strictures, and pancreatitis. On the other hand, several mechanisms may play a role in the pathogenesis of intrahepatic cholestasis, including alterations in sinusoidal membrane function and composition, alterations in cytoskeletal organization and function, alterations in tight junction permeability, and impairment of canalicular membrane structure and function.

Regardless of the mechanism of cholestasis, the net effect is retention of solutes, including bile acids, bilirubin, and cholesterol, which are preferentially excreted into bile under normal conditions. It has been recently shown that the sinusoidal sodium-dependent bile acid cotransporter (NTCP), the canalicular Na^+, K^+-ATP-dependent-bile acid transport activities, and the expression of the multidrug resistance protein 2 (Mrp2), which is responsible for the biliary excretion of amphiphilic anions, are all down-regulated in experimental models of intra- and extrahepatic cholestasis. Hepatic retention of bile acids results in decreased biliary excretion into the intestine, so that intestinal bile acid concentrations are reduced to levels below those required for normal fat digestion and absorption, and steatorrhea and deficiencies of fat-soluble vitamins may result. Reduced hepatic uptake also results in greater systemic spillover, so that serum bile acids increase over normal levels and are generally used as an indicator for liver disease, including cholestasis. Jaundice results from retention of bilirubin. Defective cholesterol excretion leads to the formation of xanthelasma, xanthomas, and alteration in the erythrocyte membrane leading to target and spur cell formation. The retained bile acids in the hepatocyte are subject to further enzymatic modification, including 6-hydroxylation, glucuronidation, and sulfation, which can further increase cholestasis, and the renal route of elimination becomes predominant.

Defects in Intestinal Transport and Metabolism

Crohn's disease, ileal resection, and bypass are characterized by defective bile salt absorption and may give rise to cholerrheic or a steatogenic enteropathy. In disorders involving more than 100 cm of ileum, hepatic synthesis of new bile salts cannot compete with intestinal loss to maintain a critical concentration of bile salts in the proximal small intestine for normal fat digestion, and steatorrhea with or without diarrhea occurs. In disorders with lesser degrees of involvement, increased concentrations of bile salts in the colon result in net fluid and electrolyte secretion, and bile acid-induced diarrhea with minimum fat malabsorption occurs.

In steatogenic enteropathy, hydroxy fatty acids formed in the colon by bacterial enzymatic hydroxylation of dietary lipids cause net fluid secretion and diarrhea. Idiopathic bile acid catharsis, a chronic diarrheal illness characterized by bile acid malabsorption, results from a relative lack of ileal bile acid transporter proteins. Disorders that are associated with acidic duodenal pH, such as exocrine pancreatic insufficiency, result in greater passive absorption of bile acids or their intraluminal precipitation. The intraluminal bile acid concentration becomes low, contributing to impaired fat digestion and steatorrhea. In small bowel bacterial overgrowth, bile acids undergo bacterial deconjugation and dehydroxylation to form unconjugated deoxycholic acid and lithocholic acid, which are absorbed passively by nonionic diffusion. Intraluminal bile salt concentration is again compromised and steatorrhea may occur.

Defects in Enterohepatic Circulation Dynamics

The bile acid pool cycles 6–10 times daily through the enterohepatic circulation. This is accomplished by the coordinated action of two physical pumps (the gallbladder and the small intestine), two chemical pumps (the transport systems of the hepatocyte and the terminal ileal erythrocyte), and two sphincters (the sphincter of Oddi and the ileocecal valve). The dynamics of the enterohepatic circulation are influenced by alterations in the gallbladder or the small intestine. Rapid intestinal transit increases cycling frequency whereas delayed intestinal transit decreases cycling frequency. In celiac

disease, reduced release of cholecystokinin in the intestinal mucosa results in impaired gallbladder emptying and stagnation of bile acids in the biliary tree. Cholecystectomy results in increased 7α-dehydroxylation of cholic acid due to increased exposure to colonic bacteria, and an increased deoxycholic acid pool is reported.

Defects in Synthesis

Continuous bile acid synthesis from cholesterol is required to maintain the bile acid pool in the enterohepatic circulation and compensate for intestinal loss. Reduced bile acid synthesis would result in serious complications, e.g., fecal loss will not be replenished, absorption of cholesterol and fat-soluble substances would decrease so that cholesterol excretion would be reduced, and the bile acid-dependent bile flow would stop. In cirrhosis, there is a reduced total bile acid pool, resulting from a reduced microsomal 12α-hydroxylation that results in a reduction of the cholate pool. Intestinal 7α-dehydroxylation of cholic acid is reduced so that the pool of deoxycholic acid is greatly reduced. Further, the increased sulfation/glucuronidation of hepatic bile acids that takes place results in increased renal clearance.

In the rare inherited lipid storage disease, cerebrotendinous xanthomatosis, which is characterized by progressive neurologic disturbances, premature atherosclerosis, cataracts, and tendon xanthomas, mitochondrial 27-hydroxylation is deficient and bile acid synthesis is decreased. Cholesterol synthesis is increased due to loss of feedback inhibition by bile acids, resulting in cholestanol formation. An alternative pathway for cholic acid synthesis becomes active, operating via microsomal bile alcohol formation (see Fig. 1). Large amounts of 5β-cholestane-3α,7α, 12α, 25-tetrols and of 5β-cholestane-3α,7α,12α,24, 25-pentols accumulate in the hepatocyte. The bile alcohols that are secreted in the bile after glucuronidation are poorly absorbed and largely excreted in the stool and urine. Treatment with chenodeoxycholic acid suppresses increased cholesterol synthesis, and in turn, cholestanol and bile alcohol levels, and may improve neurologic symptoms.

Zellweger's syndrome is another rare and fatal autosomal recessive disorder associated with multiple craniofacial dystoses, central nervous system abnormalities, generalized hypotonia, hepatomegaly, and renal cysts. Lack of peroxisomes in the liver and kidney results in defective β-oxidation of long-chain fatty acids and bile acids in this disease, and abnormal C_{27} bile acids accumulate. Smith–Lemly–Opitz syndrome is another devastating, often fatal, autosomal recessive disease characterized by dysmorphic facial features, limb abnormalities, genital disorders, and widespread defects in many endocrine glands and the liver, kidneys, urinary system, heart, lungs, and skeleton. The affected children are mentally retarded, often show severe failure to thrive, and most often die prematurely. Defective 7-dehydrocholesterol-Δ^7-reductase results in highly reduced plasma and tissue cholesterol with concomitant accumulation of the precursor, 7-dehydrocholesterol (5,7-cholestadien-3β-ol), and an isomer, 8-dehydrocholesterol (5,8-cholestadien-3β-ol). Reduced cholesterol levels result in highly reduced circulating bile acid levels, and intestinal fat absorption is impaired. Prenatal diagnosis is possible and supplementation with dietary cholesterol and bile acid seems to show limited promise.

SUMMARY

It is clear that bile flow is a very complex process that involves not only bile acids but also a large number of other components. The Na^+, K^+-ATPase is central to bile formation and current research has shown the involvement of a variety of transport proteins, including NTCP, OATP, and BSEP. The orphan nuclear receptors, liver X receptor α (LXRα), FXR, and RXR, have been shown to be powerful regulators of bile acid synthesis and homeostasis. Identification of novel nuclear receptor ligands and target genes may allow for the design of receptor-specific therapeutic agents to assist in the control of many hepatobiliary diseases.

See Also the Following Articles

Bile Composition • Bile Flow • Biliary Tract, Anatomy • Cholestatic Diseases, Chronic • Gallstones, Pathophysiology of

Further Reading

Bahar, R. J., and Stolz, A. (1999). Bile acid transport. *Gastroenterol. Clin. North Am.* **28**, 27–58.

Dawson, P. A., and Oelkers, P. (1995). Bile acid transporters. *Curr. Opin. Lipidol.* **6**, 109–114.

Gurantz, D., Schteingart, C. D., Hagey, L. R., Steinbach, J. H., Grotmol, T., and Hofmann, A. F. (1991). Hypercholeresis induced by unconjugated bile acid infusion correlates with recovery in bile of unconjugated bile acids. *Hepatology* **13**, 540–550.

Hofmann, A. F. (1994). Bile acids. *In* "The Liver: Biology and Pathobiology" (I. Arias, H. Popper, D. Schachter, and D. A. Shafritz, eds.), pp. 677–718. Raven, New York.

Meier, P. J. (1988). Transport polarity of hepatocytes. *Semin. Liver Dis.* **8**, 293–307.

Meier, P. J., Eckhardt, U., Schroeder, A., Hagenbuch, B., and Stieger, B. (1997). Substrate specificity of sinusoidal bile acid and organic anion uptake in rat and human liver. *Hepatology* **26**, 1667–1677.

Nathanson, M. H., and Boyer, J. L. (1991). Mechanisms and regulation of bile secretion. *Hepatology* **14**, 551–565.

Repa, J. J., and Mangelsdorf, D. J. (2000). The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. *Annu. Rev. Cell Div. Biol.* **16**, 459–481.

Reuben, A. (1984). Bile formation: Site and mechanisms. *Hepatology* **4**, 15S–24S.

Biliary Stricture

EUGENE R. PANTANGCO AND AMANDEEP SAHOTA
Keck School of Medicine, University of Southern California

cholecystectomy Removal of the gallbladder either by laparoscopic technique or open abdominal surgical incision.

endoscopic retrograde cholangiopancreatogram Procedure performed by cannulation of the common bile duct and pancreatic duct by means of a flexible endoscope and retrograde injection of radio-opaque contrast media in order to demonstrate all portions of the entire biliary tree and pancreatic ducts.

magnetic resonance cholangiopancreatogram Noninvasive radiologic test (magnetic resonance imaging) that enables detailed resolution of the biliary tree.

percutaneous transhepatic cholangiogram Procedure that permits direct visualization of the biliary tree via percutaneous placement of a fine needle through the lower right chest wall through the hepatic parenchyma and into the right or left bile duct. Injection of radio-opaque contrast media enables the visualization of the proximal biliary tree and common bile duct.

primary sclerosing cholangitis Inflammatory process involving the biliary tree, causing obstruction. It is often seen in men and is associated with inflammatory bowel disease. Diagnosis is by cholangiography, which often shows diffuse stricturing, beaded appearance.

stent Hollow tubular device made of plastic or metal; can be placed endoscopically or percutaneously to aid in draining the biliary tree.

A biliary stricture is often grouped with gallstones as a frequent cause of biliary duct obstruction. Although strictures caused by malignancy do occur (e.g., in pancreatic cancer or cholangiocarcinoma), the majority of biliary strictures stem from benign processes such as an operative procedure or an inflammatory response (e.g., pancreatitis). Postoperative bile duct strictures tend to be a frustration for surgeons because more than three-quarters of the iatrogenic bile duct injuries go unrecognized at the time of surgery. Strictures located distal to the cystic duct (i.e., closer to the pancreas and duodenum) typify most of the benign postoperative strictures, whereas strictures proximal to the cystic duct tend to be malignant in origin.

ETIOLOGY

Biliary Strictures Due to Operative Trauma

Bile duct injuries have been reported following procedures involving the liver, pancreas, and stomach. (For example, strictures can occur at the site of a biliary–enteric anastomosis done for reconstruction.) More commonly, however, a majority of postoperative biliary strictures occur following cholecystectomy with or without exploration of the common bile duct (CBD) by intraoperative cholangiogram (IOC). The site of bile duct injury is often located within 1 cm of the confluence of the right and left hepatic ducts. Common errors involving excessive cauterization, misplacement of surgical clips, and underrecognition of anatomic variants have been noted as possible reasons for injury in this area. The incidence of postoperative biliary stricture following open cholecystectomy has been reported to be 0.1–0.5% in a large series of patients who underwent laparoscopic cholecystectomy. Bile duct injury as reported from a group at the University of California, Los Angeles, had an incidence of 0.2–0.4%. Many researchers feel that the incidence of injury following

Copyright 2004, Elsevier (USA). All rights reserved.

Meier, P. J., Eckhardt, U., Schroeder, A., Hagenbuch, B., and Stieger, B. (1997). Substrate specificity of sinusoidal bile acid and organic anion uptake in rat and human liver. *Hepatology* **26**, 1667–1677.

Nathanson, M. H., and Boyer, J. L. (1991). Mechanisms and regulation of bile secretion. *Hepatology* **14**, 551–565.

Repa, J. J., and Mangelsdorf, D. J. (2000). The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. *Annu. Rev. Cell Div. Biol.* **16**, 459–481.

Reuben, A. (1984). Bile formation: Site and mechanisms. *Hepatology* **4**, 15S–24S.

Biliary Stricture

EUGENE R. PANTANGCO AND AMANDEEP SAHOTA
Keck School of Medicine, University of Southern California

cholecystectomy Removal of the gallbladder either by laparoscopic technique or open abdominal surgical incision.

endoscopic retrograde cholangiopancreatogram Procedure performed by cannulation of the common bile duct and pancreatic duct by means of a flexible endoscope and retrograde injection of radio-opaque contrast media in order to demonstrate all portions of the entire biliary tree and pancreatic ducts.

magnetic resonance cholangiopancreatogram Noninvasive radiologic test (magnetic resonance imaging) that enables detailed resolution of the biliary tree.

percutaneous transhepatic cholangiogram Procedure that permits direct visualization of the biliary tree via percutaneous placement of a fine needle through the lower right chest wall through the hepatic parenchyma and into the right or left bile duct. Injection of radio-opaque contrast media enables the visualization of the proximal biliary tree and common bile duct.

primary sclerosing cholangitis Inflammatory process involving the biliary tree, causing obstruction. It is often seen in men and is associated with inflammatory bowel disease. Diagnosis is by cholangiography, which often shows diffuse stricturing, beaded appearance.

stent Hollow tubular device made of plastic or metal; can be placed endoscopically or percutaneously to aid in draining the biliary tree.

A biliary stricture is often grouped with gallstones as a frequent cause of biliary duct obstruction. Although strictures caused by malignancy do occur (e.g., in pancreatic cancer or cholangiocarcinoma), the majority of biliary strictures stem from benign processes such as an operative procedure or an inflammatory response (e.g., pancreatitis). Postoperative bile duct strictures tend to be a frustration for surgeons because more than three-quarters of the iatrogenic bile duct injuries go unrecognized at the time of surgery. Strictures located distal to the cystic duct (i.e., closer to the pancreas and duodenum) typify most of the benign postoperative strictures, whereas strictures proximal to the cystic duct tend to be malignant in origin.

ETIOLOGY

Biliary Strictures Due to Operative Trauma

Bile duct injuries have been reported following procedures involving the liver, pancreas, and stomach. (For example, strictures can occur at the site of a biliary–enteric anastomosis done for reconstruction.) More commonly, however, a majority of postoperative biliary strictures occur following cholecystectomy with or without exploration of the common bile duct (CBD) by intraoperative cholangiogram (IOC). The site of bile duct injury is often located within 1 cm of the confluence of the right and left hepatic ducts. Common errors involving excessive cauterization, misplacement of surgical clips, and underrecognition of anatomic variants have been noted as possible reasons for injury in this area. The incidence of postoperative biliary stricture following open cholecystectomy has been reported to be 0.1–0.5% in a large series of patients who underwent laparoscopic cholecystectomy. Bile duct injury as reported from a group at the University of California, Los Angeles, had an incidence of 0.2–0.4%. Many researchers feel that the incidence of injury following

Copyright 2004, Elsevier (USA). All rights reserved.

laparoscopic cholecystectomy is initially high because of surgical inexperience.

Biliary Stricture Following Liver Transplantation

Biliary tract complications are a major cause of morbidity following liver transplantation, with several series reporting incidences from 13 to 25%. Strictures have been attributed to hepatic artery or portal vein thrombosis, graft rejection, or tissue ischemia reperfusion injury. In addition, reflux cholangitis secondary to Roux-limb reconstruction and infection by cytomegalovirus (CMV) have been reported as stricture etiologies following liver transplantation.

Biliary Stricture Secondary to Pancreatic Processes

In 1886, Riedel described bile duct obstruction resulting from benign pancreatic disease. Pancreatic processes can often cause distal CBD pathology because that part of the biliary tree runs in a groove on the posterior surface of the pancreas before joining the pancreatic duct in the ampulla and emptying into the duodenum. Tumors, cysts, or extensive edema within the surrounding pancreatic tissue can cause extrinsic compression of the distal CBD. In one study, the prevalence of CBD stenosis approached 100% in the clinical setting of chronic alcoholic pancreatitis, with a corresponding elevated alkaline phosphatase level twice the normal limit for at least 4 weeks. Overall, a prevalence of chronic CBD stenosis caused by chronic alcoholic pancreatitis ranges from 4 to 10%.

Biliary Stricture Associated with Primary Sclerosing Cholangitis

Patients with primary sclerosing cholangitis (PSC) and diffuse ductal disease have been found to have a dominant stricture 15–20% of the time. Severe disease with development of cirrhosis is often treated surgically (e.g., liver transplantation); however, endoscopic stenting has been helpful as a temporizing or palliative measure for those patients who are not surgical candidates. It is important to note that malignancy must be excluded prior to endoscopic treatment. An Amsterdam group has reported successful outcomes after multiple endoscopic stenting of dominant strictures in PSC. In this 10-year retrospective study, endoscopic therapy was technically successful in 84% of patients.

Biliary Stricture Due to Other Causes

Although less common than benign causes, primary malignancy (pancreatic cancer, cholangiocarcinoma) or metastatic disease can cause biliary strictures. In the literature, other benign causes of stricture include parasitic infestation, pericholedochal abscess, choledochal cysts, or vascular rings. External trauma has been reported to cause stricture 1.7% of the time. Tuberculosis and cystic fibrosis are additional processes that have been reported to cause biliary stricture.

CLINICAL PRESENTATION

A majority (80%) of the patients with postoperative biliary stricture present with symptoms within a year after the initial operation (10% within the first week, 70% within first 6 months). Patients may present with jaundice, an elevated alkaline phosphatase, or they may have signs of bile peritonitis (e.g., abdominal pain and fever) secondary to a bile leak or fistula. A clinical diagnosis can be delayed in patients with advanced biliary cirrhosis. They may present with signs and symptoms of portal hypertension, including increased abdominal girth or mass, confusion, or bleeding from varices. Strictures secondary to pancreatic causes can present with elevated transaminases, amylase, lipase or a persistent low-grade hyperbilirubinemia.

DIAGNOSIS

Preliminary imaging studies such as ultrasound and computed tomography (CT) scan are helpful in detecting intrahepatic and extrahepatic ductal dilatation. The most useful diagnostic test is cholangiography by an endoscopic retrograde cholangiopancreatogram (ERCP), percutaneous transhepatic cholangiogram (PTC), and, more recently, a magnetic resonance cholangiopancreatogram (MRCP). PTC may be more effective in patients with proximal strictures because it allows for therapeutic drainage and delineates anatomy. On the other hand, ERCP is advantageous for distal strictures due to pancreatic causes by enabling visualization of the pancreatic ductal system. Prophylactic antibiotic coverage prior to cholangiography is frequently recommended. In cases of bile leakage and recognized postsurgical complications, radionuclide hepatobiliary scintigraphy has been a helpful diagnostic tool. In cases with equivocal results or for patients who are clinically unfit for invasive procedures, MRCP offers improved imaging resolution, which can aid in diagnosis and treatment. The weaknesses of MRCP include

patient claustrophobia, the high cost of the procedure, and the variability of a radiologist's interpretation.

TREATMENT

The management of a stricture is often based on its location. Bismuth, a surgeon, developed a classification of bile duct strictures determined by the anatomic level of involvement. Bismuth I is a type of stricture more than 2 cm from the common hepatic duct (CHD), Bismuth II is a type of stricture less than 2 cm from the CHD, Bismuth III is a high ductal (proximal) stricture where confluence is preserved, Bismuth IV is a high ductal stricture where confluence is destroyed, and Bismuth V is an anomalous right duct. The Bismuth classification continues to be used as a tool in the surgical or medical management of benign biliary stricture.

In the majority of cases in which biliary stricture results from operative trauma, the goal of reoperative surgery is to reestablish bile flow into the proximal small intestine. Possible surgical procedures include choledochojejunostomy, choledochoduodenostomy, or a Roux-en-Y hepaticojeju-nostomy. Most of these surgeries have produced good long-term results with success rates as high as 94%. Recurrent strictures after surgical repair have occurred in up to 10% of cases over a period of 7 months. Biliary stents and endoscopic dilations have been used as adjuncts both before and after reoperative surgery, but this practice is still often debated.

Treatment for biliary strictures resulting from chronic pancreatitis is controversial. If patients present with persistent jaundice and/or cholangitis, surgical decompression is indicated. Sphincteroplasty has been unsuccessful in most series, primarily due to the length of the distal common bile duct stricture. However, the approach to asymptomatic patients with an isolated increase in the level of alkaline phosphatase is debatable. Some clinicians advocate surgical decompression to avoid the possible complications of persistent cholestasis, including secondary biliary cirrhosis. Others advocate a conservative approach, with periodic liver function testing, especially when the alkaline phosphatase is normal or only mildly elevated and the proximal ducts are not dilated.

Because the majority of strictures are in the postoperative setting, most notably cholecystectomy, measures can be taken to help decrease the incidence. According to the surgical literature, there are some main points to remember during surgery: mobilizing the gallbladder by dissecting in a plane next to the surface of a gallbladder, always identify the junction between the common bile duct and the cystic duct; and perform an operative cholangiography, know the major variations of common hepatic duct, and never use cautery or apply clips blindly.

See Also the Following Articles

Cholangitis, Sclerosing • Cholecystectomy • Liver Transplantion • Percutaneous Transhepatic Choliangiography (PTC)

Further Reading

Bergman, J. J., van den Brink, G. R., Rauws, E. A., *et al.* (1996). Treatment of bile duct lesions after laparoscopic cholecystectomy. *Gut* **38**, 141–147.

Deviere, J., Cremer, M., Baize, M., Love, J., Sugai, B., and Vandermeeren, A. (1994). Management of common bile duct stricture caused by chronic pancreatitis with metal mesh self-expandable stents. *Gut* **35**, 122–126.

Richardson, M. C., Bell, G., and Fullarton, G. M. (1996). Incidence and nature of bile duct injuries following laparoscopic cholecystectomy: An audit of 5913 cases. West of Scotland laparoscopic cholecystectomy audit group [see comments]. *Br. J. Surg.* **83**, 1356–1360.

Roslyn, J. J., and Tompkins, R. K. (1991). Reoperation for biliary strictures. *Surg. Clin. North Am.* **71**, 109.

Way, L. W., Bernhoft, R. A., and Thomas, M. J. (1981). Biliary stricture. *Surg. Clin. North Am.* **61**, 963.

Biliary Tract, Anatomy

SHOBHA SHARMA
Emory University Hospital

liver segment Portion of the liver defined by an independent blood supply.
sectoral bile duct Bile ducts formed by the merging of segmental bile ducts.
segmental bile duct Bile ducts draining segments of the liver.

The anatomy of the biliary system may be divided into that of the intrahepatic ducts and the extrahepatic ducts. The segmental intrahepatic bile ducts join to form the left and right hepatic ducts. At the hilum of the liver, these two ducts merge to form the common hepatic duct. The common hepatic duct is joined by the cystic duct (that drains the gall bladder) to form the common bile duct. The extrahepatic biliary system is composed of the common hepatic duct, the cystic duct and the common bile duct.

LIVER SEGMENTS

The liver is divided into eight segments based on the fact that the portal venous and hepatic arterial blood supply to and hepatic venous and biliary drainage from these segments are independent of each other and there are no significant anastomoses between segments. This permits segmental resections of the liver. Segment I is the posteriorly located caudate lobe. The primary division of the rest of the liver into left and right lobes is along a line that extends from the center of the gallbladder bed to the left side of the inferior vena cava (Cantlie's line). The left lobe is divided into anterior and posterior sectors. The posterior sector is segment II. The anterior sector consists of segment III and segment IV (the quadrate lobe). In 67% of individuals, the left hepatic duct is formed after the posterior sectoral duct bile ducts (draining segment IV) join the anterior sectoral duct (formed by the merging of the ducts draining segments II and III). In 25% of individuals, the segment IV duct drains into the segment III duct and in 4% there is drainage into both the segment III duct and the right sectoral duct. Drainage of the segment IV duct(s) into the common hepatic duct, into the segment II duct, and into the right sectoral duct occurs in 1% of individuals, for each variation.

SECTORAL DUCTS

The right lobe is divided into four segments. The first division separates the right lobe into antero-medial and postero-lateral sectors, each of which is further subdivided into anterior and posterior segments, i.e., segments V, VI, VII, and VIII. In 91% of individuals, the right anterior sectoral duct drains segments V and VIII and the right posterior sectoral duct drains segments VI and VII. The two sectoral ducts join to form the right hepatic duct. Variations in this anatomy are noted in approximately 10% of individuals. The segment V duct drains directly into the right hepatic duct in 5% of individuals and into the posterior sectoral duct in 4%. The segment VIII bile duct drains into the posterior sectoral duct in 20% of individuals. Variations in segment VI bile duct anatomy include direct drainage into the anterior sectoral duct (10%), the right hepatic duct (2%), and the common hepatic duct (2%). The caudate lobe (segment I) drains into both the left and right biliary systems.

The left and right hepatic ducts join to form the common hepatic duct in 57 to 72% of individuals. Variations in this anatomy primarily involve the right sectoral ducts. They include the right anterior and posterior sectoral ducts separately joining the left hepatic duct in 12% (no right hepatic duct) of individuals. A right sectoral duct drains into the left hepatic duct in 6% of individuals (right anterior sectoral duct in 1% and right posterior sectoral duct in 5%). In 20% of individuals, a right sectoral duct drains directly into the common hepatic duct (the right anterior sectoral duct in 16% and the right posterior sectoral duct in 4%). Finally, all the sectoral ducts may join to form the common hepatic duct without forming the right and left hepatic ducts (3%).

COMMON BILE DUCT

The common bile duct measures approximately 10 cm in length and 0.5 cm in diameter and lies to the right of the hepatic artery and anterior to the portal vein. It descends behind the posterior aspect of the superior duodenum along the free edge of the lesser omentum

 Copyright 2004, Elsevier (USA). All rights reserved.

and deviates toward the right in a groove on the posterior aspect of the head of the pancreas. Here it lies anterior to the inferior vena cava and enters the descending duodenum on its posterior-medial aspect. The sphincteric muscle of Oddi derived from the duodenal smooth muscle surrounds the common bile duct and the opening of the duct into the duodenal lumen is marked by a mucosal protrusion called the ampulla or papilla of Vater. The pancreatic duct accompanies the common bile duct within the sphincter of Oddi in 70–85% of individuals.

See Also the Following Articles

Biliary Tract, Development • Duodenum, Anatomy • Hepatic Circulation • Gastrointestinal Tract Anatomy, Overview • Liver, Anatomy

Further Reading

Couinaud, C. (1957). Les voies biliaires intra-hepatiques. *In* "Le Foie: Etudes anatomiques et chirurgicales" pp. 119–145. Masson et Cie, Libraires de l'academie de Medecine, Paris.

Healey, J. E., and Schroy, P. C. (1953). Anatomy of the biliary ducts within the human liver: Analysis of the prevailing pattern of branchings and the major variations of the biliary ducts. *Am. Med. Assoc. Arch. Surg.* **66**, 599–616.

Romanes, G. J. (1986). Ducts of the liver. *In* "Cunningham's Manual of Practical Anatomy", 15th ed., Vol. 2, pp. 152–153. English Language Book Society/Oxford University Press, Hong Kong.

Biliary Tract, Benign Tumors of

NAHID HAMOUI AND PETER F. CROOKES
Keck School of Medicine, University of Southern California

choledochojejunostomy Surgical procedure in which the common bile duct is connected to the jejunum.

choledochoscopy Diagnostic aid during surgery in which a flexible endoscope is placed directly into the common bile duct.

endoscopic retrograde cholangiopancreatography Diagnostic procedure for imaging the biliary tree. A flexible fiberoptic endoscope is placed in the duodenum and a fine flexible tube is inserted into the common bile duct through the ampulla of Vater. A radio-opaque substance is instilled directly into the duct and serial X-ray films are taken.

hemobilia Blood in the bile ducts, which then passes into the duodenum and is either vomited or passed per rectum.

hepaticojejunostomy Surgical procedure in which the common hepatic duct is connected to the jejunum.

magnetic resonance cholangiopancreatography Method of imaging the biliary tree and pancreatic duct without using radiation.

stent Hollow plastic tube inserted into a narrowing in a hollow tract of the body as a nonoperative method of keeping the tube patent.

Whipple operation (pancreaticoduodenectomy) Major surgical procedure in which the head of the pancreas, the duodenum, and the lower end of the common bile duct are all removed, usually because of cancer of the head of the pancreas.

Benign biliary tumors are nonmalignant masses in the biliary tree. These masses may be neoplastic or congenital in origin and they may also result from inflammatory processes. It is necessary to distinguish these tumors from reactions to surgical trauma and from other systemic diseases known to cause strictures in the biliary system, such as sclerosing cholangitis.

INTRODUCTION

Benign biliary tumors are extremely uncommon. They have been reported in 0.1% of all biliary tract surgeries and constitute only 6% of all extrahepatic bile duct

Copyright 2004, Elsevier (USA). All rights reserved.

and deviates toward the right in a groove on the posterior aspect of the head of the pancreas. Here it lies anterior to the inferior vena cava and enters the descending duodenum on its posterior-medial aspect. The sphincteric muscle of Oddi derived from the duodenal smooth muscle surrounds the common bile duct and the opening of the duct into the duodenal lumen is marked by a mucosal protrusion called the ampulla or papilla of Vater. The pancreatic duct accompanies the common bile duct within the sphincter of Oddi in 70–85% of individuals.

See Also the Following Articles

Biliary Tract, Development • Duodenum, Anatomy • Hepatic Circulation • Gastrointestinal Tract Anatomy, Overview • Liver, Anatomy

Further Reading

Couinaud, C. (1957). Les voies biliaires intra-hepatiques. *In* "Le Foie: Etudes anatomiques et chirurgicales" pp. 119–145. Masson et Cie, Libraires de l'academie de Medecine, Paris.

Healey, J. E., and Schroy, P. C. (1953). Anatomy of the biliary ducts within the human liver: Analysis of the prevailing pattern of branchings and the major variations of the biliary ducts. *Am. Med. Assoc. Arch. Surg.* **66**, 599–616.

Romanes, G. J. (1986). Ducts of the liver. *In* "Cunningham's Manual of Practical Anatomy", 15th ed., Vol. 2, pp. 152–153. English Language Book Society/Oxford University Press, Hong Kong.

Biliary Tract, Benign Tumors of

NAHID HAMOUI AND PETER F. CROOKES
Keck School of Medicine, University of Southern California

choledochojejunostomy Surgical procedure in which the common bile duct is connected to the jejunum.

choledochoscopy Diagnostic aid during surgery in which a flexible endoscope is placed directly into the common bile duct.

endoscopic retrograde cholangiopancreatography Diagnostic procedure for imaging the biliary tree. A flexible fiberoptic endoscope is placed in the duodenum and a fine flexible tube is inserted into the common bile duct through the ampulla of Vater. A radio-opaque substance is instilled directly into the duct and serial X-ray films are taken.

hemobilia Blood in the bile ducts, which then passes into the duodenum and is either vomited or passed per rectum.

hepaticojejunostomy Surgical procedure in which the common hepatic duct is connected to the jejunum.

magnetic resonance cholangiopancreatography Method of imaging the biliary tree and pancreatic duct without using radiation.

stent Hollow plastic tube inserted into a narrowing in a hollow tract of the body as a nonoperative method of keeping the tube patent.

Whipple operation (pancreaticoduodenectomy) Major surgical procedure in which the head of the pancreas, the duodenum, and the lower end of the common bile duct are all removed, usually because of cancer of the head of the pancreas.

Benign biliary tumors are nonmalignant masses in the biliary tree. These masses may be neoplastic or congenital in origin and they may also result from inflammatory processes. It is necessary to distinguish these tumors from reactions to surgical trauma and from other systemic diseases known to cause strictures in the biliary system, such as sclerosing cholangitis.

INTRODUCTION

Benign biliary tumors are extremely uncommon. They have been reported in 0.1% of all biliary tract surgeries and constitute only 6% of all extrahepatic bile duct

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Benign Bile Duct Tumors

Epithelial tumors
Adenoma
Cystadenoma
Papilloma
Multiple biliary papillomatosis
Adenomyoma
Granular cell tumor
Neural tumors
Neurinoma
Paraganglioma
Amputation neuroma
Leiomyoma
Inflammatory tumor
Heterotopic tissue
Heterotopic gastric tissue
Heterotopic pancreas tissue

neoplasms. However, in the interest of optimal management, it is important to consider these neoplasms in the differential diagnosis of obstructive jaundice. Benign tumors of the bile ducts, despite their varied origins and histological appearance, are similar in many aspects of their clinical presentation, methods of diagnosis, and treatment. Pathologically, there are at least seven subtypes, as indicated in Table I.

CLINICAL PRESENTATION

Symptomatic benign biliary tumors generally present with obstructive jaundice and/or right upper quadrant pain, most often of long duration. The clinical and radiological presentation of these patients may mimic biliary calculi, cancer, or inflammatory conditions of the bile ducts. Depending on tumor location, it is also possible for patients to be asymptomatic. Very rarely, these patients present with other symptoms, such as bleeding. A single case of death due to hemorrhage from an adenoma has been reported, and biliary papillomatosis has also been associated with hemobilia and may present with anemia. Biliary calculi are reported in only 20% of patients with adenomatous tumors and are not reported in the majority of patients with other types of benign bile duct lesions. The first study usually performed on a jaundiced patient is an ultrasound, which will confirm the diagnosis of obstructive jaundice but will give no indication of the pathology. This is especially true when no biliary calculi are present. In this circumstance, further imaging of the biliary tree is necessary, usually by endoscopic retrograde cholangiopancreatography (ERCP), although magnetic resonance cholangiopancreatography (MRCP) is increasingly being used. Even ERCP is usually not diagnostic of the lesion, although it is helpful in defining tumor location, extension, and size, as well as the status of the intrahepatic ductal system (Fig. 1). These features are useful in operative planning, even if the exact histological diagnosis is most often made postoperatively.

TUMOR SUBTYPES

Epithelial Tumors

Adenoma

The most common benign biliary tumors are those arising from the epithelial tissue lining the ducts. Chu's review of benign biliary neoplasms, published in 1950, found that 26 out of 30 cases reviewed were epithelial tumors. The anatomical distribution of these tumors is illustrated in Fig. 2. Most are found in the ampulla, with the common bile duct being the next most common site. Adenomas vary grossly from a few millimeters to several centimeters, although tumors as large as 15 cm have been reported. Grossly, they are firm, gray–white, and nonencapsulated. Histologically, they are composed of small, round, tubules resembling bile ducts in a fibrous stroma. The ducts may have a tortuous configuration and secrete mucin, but not bile. The epithelial lining consists of cuboidal, lightly basophilic cells that have regular nuclei. The supporting stroma is scanty and variably inflamed and becomes progressively hyalinized with time. It has been suggested that papillary cancers of the extrahepatic ducts may arise from preexisting adenomas, but because of the rarity of these lesions, it is difficult to prove. The origin of biliary duct adenoma is debatable. Plausible hypotheses

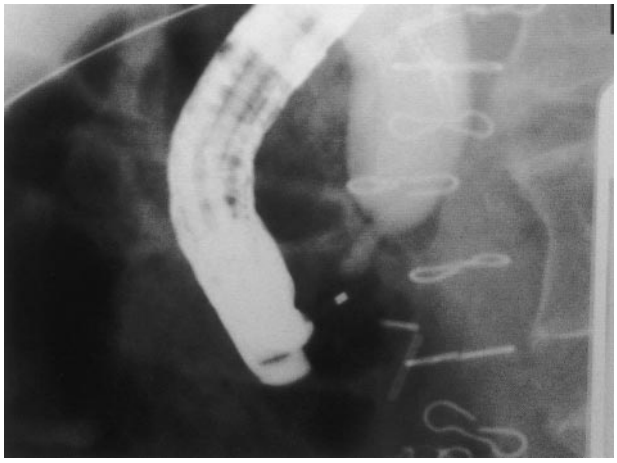


FIGURE 1 Endoscopic retrograde cholangiopancreatography of ampullary adenomyoma.

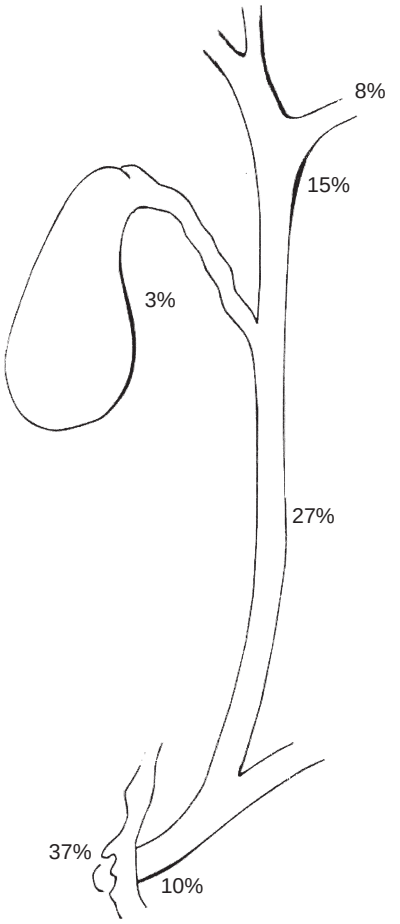


FIGURE 2 Location and frequency of epithelial tumors in the extrahepatic biliary system. Reprinted from Beazley and Blumgart, "Benign tumours and pseudotumours of the biliary tract." *In* "Surgery of the Liver and Biliary Tract" (L. H. Blumgart, Ed.), 2nd Ed., p. 944, Copyright 1994, by permission of the publisher Churchill Livingstone.

include developmental abnormality, hamartoma, a true neoplasm, or a reactive process to a focal injury.

Cystadenomas

Cystadenomas are rare cystic tumors of epithelial origin that arise in the liver, the majority in the right lobe, or less commonly in the extrahepatic biliary system. There are two histological variants, a mucinous type and a serous type. The more common mucinous type histologically resembles the mucinous cystadenoma of the pancreas. The origin of cystadenomas is unclear. They may be derived from embryonic foregut rests or from hamartomas. The cystadenoma (CA) is predominantly a tumor of middle-aged women; females account for 90% of reported cases. In contrast to other benign biliary tumors, abdominal swelling and pain are the most frequent presenting symptoms, although biliary obstruction, hemorrhage, rupture, and vena caval obstruction have been reported. Also, unlike other benign tumors of the bile ducts, imaging studies are reasonably specific and may permit preoperative diagnosis. The classic ultrasound appearance is an anechoic mass with internal septations that are highly echogenic, whereas computer tomography (CT) usually shows a smooth, thick-walled cyst with fine internal septae. The appearance of this tumor on magnetic resonance imagery (MRI) has also been described. Elevated CA 19-9 tumor marker levels have been reported. Grossly, the tumor usually consists of multilocular cysts ranging in diameter from 2.5 to 28 cm and containing a mucinous or gelatinous fluid. The inner surface may be smooth or trabeculated. Microscopically, the tumor is lined by mucin-secreting columnar to cuboidal epithelium with pale eosinophilic cytoplasm and basally oriented nuclei. The tumor is generally single layered but may form small papillary foldings. Focal intestinal metaplasia is occasionally observed. The stromal cells are spindle shaped or rarely oval, resemble ovarian stroma, and are immunoreactive with vimentin, actin, and desmin. Mucinous cystadenoma behaves as a slow-growing tumor but has a tendency to become malignant over a period of years. In one large study series, 6 of 18 cystadenocarcinomas had areas of preexisting benign cystadenoma and 7 of 51 cystadenomas had foci of dysplasia.

The serous variety of cystadenoma histologically resembles serous cystadenoma of the pancreas. It consists of numerous small cystic spaces lined by a single layer of cuboidal cells with clear cytoplasm containing glycogen. The serous cystadenomas rest on a basement membrane but, in contrast to the mucinous variety, are not surrounded by a mesenchymal stroma. Serous cystadenomas are not known to undergo malignant transformation.

Papillomas

Papillomas typically measure a few centimeters, although Leriche, in 1934, reported a case of a papillomatous tumor weighing 750 g in the common bile duct of a 4-year-old child. Grossly, the tumors may be firm, elevated masses or soft, vascular, sessile, or pedunculated growths. Histologically, the tumors are composed of thick or delicate papillae covered by a layer of tall columnar epithelium with a loose connective tissue core.

Biliary papillomatosis is a premalignant disorder consisting of multicentric papillary adenomas in the biliary tract. It is a rare disorder, with only about 50 cases described in the literature. The disease is seen in middle-aged to older men and women at a male : female

ratio of about 2 : 1. Pathologically, the lesion consists of soft, friable, papillary masses filling dilated intrahepatic and extrahepatic ducts. The gallbladder and major pancreatic duct may also be involved. The intervening liver parenchyma may be green or fibrotic. Histologically, the dilated ducts contain multiple papillary adenomas, which are composed of branching papillary fronds covered by mucus-secreting columnar epithelium covering fibrovascular stalks. The epithelial layer is adenomatous, with varying degrees of atypia. Paneth cells and endocrine cells may also be present. A point mutation of the K-*ras* oncogene has been found and may indicate a high probability of progression to carcinoma.

Preoperative diagnosis may be made by cholangiography, although the filling defects seen with this study have been confused in the past with air bubbles or intrabiliary blood clots. MRCP has also been used to image this lesion. Published reports have recommended intraoperative choledochoscopy of all accessible intrahepatic ducts to define the extent of the disease.

Adenomyoma

The majority of adenomyomas are located in the stomach, duodenum, and jejunum, and they also occur with some frequency in the gallbladder. Adenomyoma of the biliary tract is much rarer and was first reported in 1942; a review of world literature conducted in 2000 yielded 25 additional cases. Of these, 84% were located in the common bile duct or ampulla of Vater. A typical example is shown in Fig. 3. Histologically, they consist of a dense mass of intertwined smooth muscle bundles, collagen fibers, and epithelial structures that create distorted lumina.

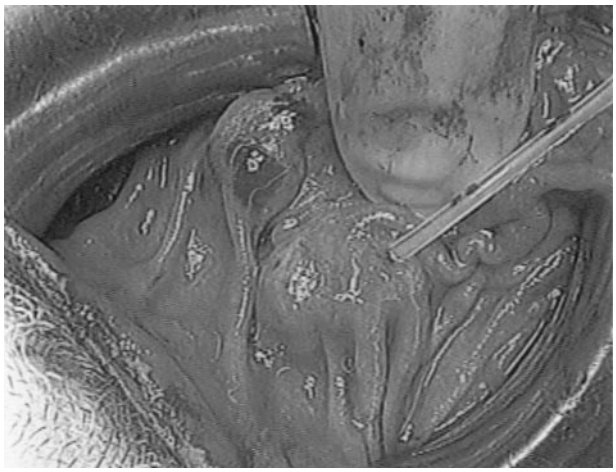


FIGURE 3 Ampullary adenomyoma.

Granular Cell Tumor

Granular cell tumor is a rare benign tumor occasionally found in the biliary tree, although it is more commonly reported in other locations; the first description of a granular cell tumor in 1926 was in the skin. A granular cell tumor of the biliary tree was first reported in 1952; there have been about 50 reports in the literature since that time. It was originally believed that these tumors originated in muscle, and they were consequently classified as myoblastomas. The currently accepted theory is that they originate from Schwann cells, and can be seen by light microscopy, electron microscopy, and immunohistochemical staining to be very similar to nerve cells. Of the cases so far reported, 62% are in black women with a median age of 31 years. The most common location in which these tumors are found is in the common bile duct, although they have been reported in the gallbladder, cystic duct, and common hepatic and hepatic ducts. In five published reports, biliary granular cell tumors have presented concomitantly with granular cell tumors in other locations, three out of five being in the skin and two in the stomach wall. Grossly, the tumors are yellow–white and usually less than 3 cm (Fig. 4). Microscopically, they consist of polygonal granular eosinophilic cells that react with periodic acid–Schiff staining and have centrally located small vesicular nuclei (Fig. 5). Although preoperative radiological studies are not specific for this lesion, intraoperative frozen section does allow diagnosis. There have been no malignant granular cell tumors reported in the biliary system, although there have been several reported in the skin.

FIGURE 4 Granular cell tumor of the intrapancreatic bile duct. Reproduced with permission from *Cancer* 53(10), 2179. Copyright 1984, American Cancer Society. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

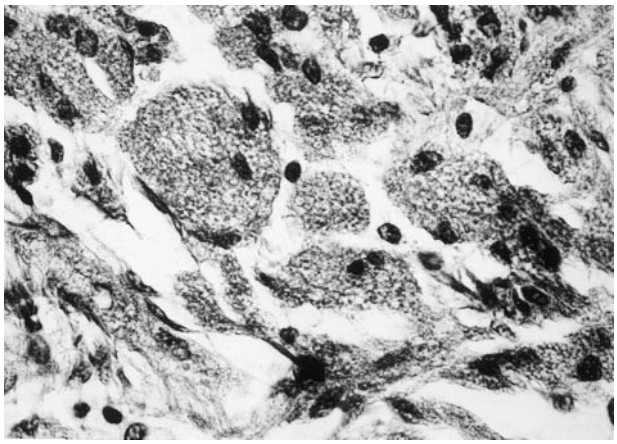


FIGURE 5 Microscopic appearance of tumor cells, showing small nuclei and abundant granular cytoplasm. Reproduced with permission from *Cancer* **53**(10), 2180. Copyright 1984, American Cancer Society. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

Neural Tumors

Neural tumors of the biliary system are infrequently reported and are thought to arise from the network of neural tissue surrounding the extrahepatic ducts. In 1955, a neurinoma of the common bile duct was reported in a 40-year-old woman who presented with jaundice and pain. The lesion was locally excised, leaving the duct intact, and the patient was reported well 18 months later. Von Recklinghausen's neurofibromatosis has also been recorded in one instance to have caused obstructive jaundice due to a periampullary nodular tumor. There have been several reported cases of paraganglioma of the extrahepatic ducts that presented with jaundice. These tumors, also known as nonchromaffin tumors of the sympathetic nervous system, most commonly occur in the adrenal medulla. Additional locations reported include the mediastinum, small intestine, retroperitoneum, urinary bladder, kidney, stomach, trachea, and tongue, as well as in the gallbladder. Microscopically, paraganglioma is described as nests of large, uniformly polygonal, round cells with abundant finely granular eosinophilic cytoplasm and round nuclei surrounded by a delicate fibrovascular septa. Metastasis has not been reported for biliary paragangliomas but is known to occur for these tumors in other locations. Their malignant potential cannot be predicted on histologic grounds.

Amputation neuroma is not a true neoplasm but a disorganized proliferation of a severed nerve. It most commonly arises after cholecystectomy from the cystic bile duct stump and grows extraluminally, in which location it rarely causes obstructive jaundice. However, it can occur in the main hepatic bile ducts and cause jaundice and pain. This usually follows cholecystectomy, although in one case it occurred in a patient who had a lymph node dissection of the hepatoduodenal ligament for gastric cancer. Amputation neuroma has also been reported as a rare cause of biliary obstruction following liver transplantation.

Leiomyomas

Although leiomyomas are the most common benign tumors of the esophagus, stomach, and small intestine, they are among the rarest in the extrahepatic biliary tree, with only five reported cases. This may be explained by the paucity of smooth muscle fibers in the extrahepatic biliary system. Of the five reported lesions, three were found in the common bile duct, one at the ampulla of Vater, and one at the hepatic duct bifurcation.

Inflammatory Tumors

Inflammatory tumors are nonmalignant lesions that may resemble neoplasms during preoperative investigation and even on gross inspection during laparotomy. One large study series from the Hammersmith Hospital reported 104 patients who had laparotomy for presumed malignant biliary obstruction, seven of which were found to have benign disease. In all seven patients, the obstructing lesion was removed and biliary reconstruction was performed by hepaticojejunostomy. Of these seven patients, six were alive and five were asymptomatic, from 19 to 49 months postoperatively. One patient died of cholangitis with cholangiographic evidence of progression of sclerosing cholangitic lesions.

Histologic examination of these lesions shows an extensive increase of fibrous tissue with subepithelial mucous glandular proliferation. Glandular cells are well differentiated, with elongated to rounded nuclei showing normal polarity. Clumps of lymphocytes are present in both perivascular and perineural locations, but they do not infiltrate the walls of the bile ducts. It is emphasized here that it is important to have tissue diagnosis of bile duct stricture before treating potentially benign disease with a therapy, such as stenting, normally reserved for short-term palliation of malignant disease, because this is associated with a higher complication rate than definitive surgical therapy. Attempts have been made to use mutation analysis of oncogenes such as K-*ras* to distinguish benign from malignant biliary strictures preoperatively, but this is not yet practical with current technology.

Heterotopic Tissue

Symptomatic heterotopic tissue arising in the biliary tree has been rarely reported. Whitaker *et al*. first observed heterotopic gastric mucosa arising in a cystic duct that had obstructed the gallbladder. Gastric mucosa has also been reported in the common bile duct, the common hepatic duct, and the ampulla of Vater. There are two hypotheses for development of this tissue: metaplasia with heterotopic differentiation and congenital development from multipotential endodermal tissue. Seven cases of heterotopic pancreatic tissue have also been reported in the common bile duct and ampulla of Vater.

TREATMENT AND PROGNOSIS

Treatment for benign bile duct tumors depends more on location of the tumor than on specific tumor type. Because of the rarity of these lesions, there is no consensus on standard of care, and only general principles can be advanced. These patients should be carefully investigated preoperatively. If the patient goes to surgery without a definitive diagnosis, it is advisable to obtain an intraoperative frozen section, because histological diagnosis permits better planning of the extent of surgical excision. Operations reported in the literature have included local excision, transduodenal papillotomy, curettage, local duct resection with reconstruction, hepatectomy, and occasionally a Whipple operation for periampullary tumors originally thought to be malignant. Tumors such as cystadenoma and granular cell tumor are known to recur following incomplete excision, although complete resection is usually curative.

Treatment of biliary papillomatosis deserves special note in that it is more complex, and patients for the most part have a less favorable course compared to other benign diseases of the biliary tree. In the past, many patients were treated with palliative techniques, such as cholecystectomy, curettage, and internal or external drainage. These treatments may relieve obstruction temporarily but are usually followed by recurrence. If the lesions are limited to one liver lobe, hepatectomy should be performed, although even with this treatment the disease may recur. In the five reported cases in which hepatectomy was performed, two patients appeared to be cured at the 6-month and 4-year followup exams. One patient died 6 years after resection and had diffuse malignant tumors of the right lobe of the liver, and two patients had recurrences 6 months and 3 years after left hepatic lobectomy. Hepatic transplantation has recently been proposed as an alternative therapy. The role of chemotherapy is unproved because of the rarity of the disease.

SUMMARY

Benign biliary tumors, although rare, are important considerations in the differential diagnosis of obstructive jaundice. These tumors vary considerably in their natural history and malignant potential, although their clinical presentations are similar and also mimic the presentation of cholelithiasis and cholangiocarcinoma, which are much more common. With many of these tumors, it is difficult to arrive at the correct diagnosis preoperatively, although there are a few, such as cystadenoma and intrahepatic biliary papillomatosis, that can be identified using cholangiography. In many cases, correct diagnosis can be accomplished only by intraoperative frozen section. This becomes important in cases in which more radical surgery is being contemplated than may be required, such as a pancreaticoduodenectomy. It is also important not to consign the patient to palliative therapy, such as insertion of a stent, when the patient may be cured by surgical excision. It may be possible that future advances in molecular biology will permit more precise preoperative diagnoses and treatment planning.

Acknowledgments

We thank Dr. Para Chandrasoma and Dr. Rod Mateo for their assistance in providing illustrations for this article.

See Also the Following Articles

Bile Duct Injuries and Fistulas • Biliary Tract, Development • Bilirubin and Jaundice

Further Reading

Beazley, R. M., and Blumgart, L. H. (1994). Benign tumours and pseudotumours of the biliary tract. *In* "Surgery of the Liver and Biliary Tract" (L. H. Blumgart, ed.), 2nd Ed., pp. 941–953.

Cattell, R. B., Braasch, J. W., and Kahn, F. (1962). Polypoid epithelial tumors of the bile ducts. *N. Engl. J. Med.* **266**, 57–61.

Chu, P. T. (1950). Benign neoplasms of the extrahepatic bile ducts. *Arch. Pathol.* **54**, 84–97.

Colombari, R., and Tsui, W. M. S. (1995). Biliary tumors of the liver. *Semin. Liver Dis.* **15**, 402–412.

Gouma, D. J., Singh Mutum, S., Benjamin, I. S., and Blumgart, L. H. (1984). Intrahepatic biliary papillomatosis. *Br. J. Surg.* **71**, 72–74.

Hadjis, N. S., Collier, N. A., and Blumgart, L. H. (1985). Malignant masquerade at the hilum of the liver. *Br. J. Surg.* **72**, 659–661.

Biliary Tract, Development

SHOBHA SHARMA
Emory University Hospital

biliary tract Branching tubular network draining bile from the liver into the intestine.
duodenum Proximal small intestine into which biliary tract drains.
hepatocytes Individual cells constituting the liver.

The biliary tract consists of the organs and ducts that produce, transport, store, and secrete bile into the duodenum in digestive processes. The biliary system includes the liver, gallbladder, and bile ducts.

FETAL ORIGIN

The biliary tract and the liver arise from the endoderm of the distal foregut. On day 22 of gestation, a thickening called the hepatic plate develops in the ventral wall of the duodenum. This plate continues to proliferate, forming the hepatic diverticulum or liver bud that grows into the septum transversum (the plate of mesenchyme that divides the coelomic space into the thoracic and abdominal cavities and subsequently becomes the diaphragm). Within the septum transversum, the hepatic diverticulum proliferates into hepatic cords that give rise to the hepatocytes, biliary canaliculi, and intrahepatic bile ducts. Simultaneously, blood lakes form between these hepatic cords, followed by the development of capillaries and portal and hepatic veins. This vasculature is derived from the vitelline and umbilical blood vessels that traverse the septum transversum. Though the hepatocytes are intimately related to the sinusoids, they are separated from the portal vein radicals by connective tissue derived from the septum transversum.

DEVELOPMENT

Formation of intrahepatic bile ducts begins at approximately 9–10 weeks of gestation. The first ducts to appear are the left and right hepatic ducts at the hilum of the liver, and development progresses centrifugally toward the periphery of the liver. Biliary epithelium is derived from the hepatocytes immediately adjacent to the mesenchyme surrounding the portal veins. These hepatocytes change their phenotype to form the ductal plate. In contrast to the larger hepatocytes, the ductal plate cells have a low cuboidal shape and acquire cytokeratin 19 and 7 expression that is not present in the hepatocytes. Partial reduplication of the ductal plate occurs and the initial bile ducts are formed by fusion of these two layers at the portal mesenchymal interface. Between 12 and 14 weeks of gestation, these ducts migrate into the portal mesenchyme that now separates them from hepatocytes. Throughout the gestational period, this process extends more peripherally and development of the smallest bile ducts continues in the first couple of months after birth. The extrahepatic biliary system, which includes the common hepatic, cystic and common bile ducts, develops from the portion of the hepatic diverticulum between the duodenum and the septum transversum. On day 26, 4 days after the initiation of the hepatic diverticulum, another diverticulum develops from the ventral duodenal wall; this is destined to develop into the cystic duct and gall bladder. The tissues at the junction of the hepatic and cystic duct diverticuli proliferate to form the common bile duct, which elongates as the duodenum grows away from the septum transversum and undergoes rotation.

Even though studies in rodent models demonstrate the relevance of hepatocyte growth factor (HGF) in mesenchymal epithelial interactions that may be relevant to biliary epithelial proliferation, there is evidence to suggest that direct hepatocyte–biliary epithelial interactions are involved in orderly bile ductal development in the human.

See Also the Following Articles

Biliary Tract, Anatomy • Development, Overview • Duodenum, Anatomy • Hepatic Circulation • Hepatocytes • Liver, Anatomy • Liver, Development

Further Reading

Auth, M. K. H., Joplin, R. E., Okamoto, M., Ishida, Y., McMaster, P., Neuberger, J. M., Blaheta, R. A., Voit, T., and Strain, A. J. (2001). Morphogenesis of primary human biliary epithelial cells: Induction in high-density culture or coculture with autologous human hepatocytes. **33**, 519–529.

 Copyright 2004, Elsevier (USA). All rights reserved.

Eyken, P. V., Sciot, R., Callea, F., Van Der Steen, K., Moerman, P., and Desmet, V. J. (1989). The development of the intrahepatic bile ducts in man: A keratin-immunohistochemical study. *Hepatology* **8**, 1586–1595.

Larsen, W. J. (2001). Development of the gastrointestinal tract. *In* "Human Embryology" (L. S. Sherman, S. S. Potter, and W. J. Scott, eds.), 3rd Ed., pp. 240–241. Churchill Livingstone Harcourt Heath Sciences, Philadelphia.

Biliary Tract, Developmental Anomalies of the

ARDATH K. YAMAGA AND FRANK R. SINATRA

Children's Hospital Los Angeles, Women's and Children's Hospital, and Keck School of Medicine, University of Southern California

apoptosis Death of a cell in a programmed manner.
atresia Absence of a normal opening or lumen.
cholestasis State in which there is a reduction or inhibition of the flow of bile.
hypersplenism Condition, usually associated with portal hypertension, in which the spleen is enlarged, sequestering platelets and red and white blood cells.
jaundice Visible yellow staining of the skin and sclera due to an increase in bile pigments in the serum.
lithiasis Formation of stones of any kind (e.g., gallstones).
polysplenia Syndrome consisting of multiple spleens, usually bilateral, with rudimentary and accessory splenic tissue.
porta hepatis (hilum) Location in the liver between the caudate and quadrate lobes that contains the portal vein, hepatic artery, hepatic nerves, hepatic ducts, and lymphatic vessels.
portoenterostomy (Kasai operation) Surgery performed for extrahepatic biliary atresia whereby a Roux-en-Y loop of jejunum is anastomosed to the porta hepatis.
pruritus State of itching.
situs inversus viscerum Condition in which the viscera are transposed in the abdominal cavity, with the liver on the left side and the stomach and spleen on the right.
TORCH infection Toxoplasmosis, rubella, cytomegalovirus, and herpesvirus congenital infections associated with fetal malformations; syphilis is also associated with congenital infections.

Developmental anomalies of the biliary tract are common causes of obstructive jaundice in infants and children. In most cases, these disorders appear to be due to a defect in biliary embryogenesis or to acquired obliteration of the biliary ductal system. Although some of these disorders are amenable to surgical correction, many will continue to impart some degree of impaired bile flow and subsequent development of hepatic fibrosis or cirrhosis. The diagnostic and therapeutic challenge for the clinician is to identify, as rapidly as possible, those patients for whom surgical correction is beneficial. For the remainder of the patients for whom there are no surgical options, treatment is aimed at the prevention and treatment of the complications of cholestasis.

INTRODUCTION

Congenital abnormalities of the liver and biliary tract are usually diagnosed in early infancy and childhood (Table I). Most affected infants and children present with conjugated hyperbilirubinemia. Any infant with jaundice beyond 2 weeks of age (the acceptable range

TABLE I Common Developmental Anomalies of the Biliary Tract

Extrahepatic biliary atresia
Choledochal cyst
Intrahepatic bile duct paucity
Infantile polycystic disease/congenital hepatic fibrosis
Cystic dilatation of intrahepatic ducts

Copyright 2004, Elsevier (USA). All rights reserved.

Eyken, P. V., Sciot, R., Callea, F., Van Der Steen, K., Moerman, P., and Desmet, V. J. (1989). The development of the intrahepatic bile ducts in man: A keratin-immunohistochemical study. *Hepatology* **8**, 1586–1595.

Larsen, W. J. (2001). Development of the gastrointestinal tract. *In* "Human Embryology" (L. S. Sherman, S. S. Potter, and W. J. Scott, eds.), 3rd Ed., pp. 240–241. Churchill Livingstone Harcourt Heath Sciences, Philadelphia.

Biliary Tract, Developmental Anomalies of the

ARDATH K. YAMAGA AND FRANK R. SINATRA

Children's Hospital Los Angeles, Women's and Children's Hospital, and Keck School of Medicine, University of Southern California

apoptosis Death of a cell in a programmed manner.
atresia Absence of a normal opening or lumen.
cholestasis State in which there is a reduction or inhibition of the flow of bile.
hypersplenism Condition, usually associated with portal hypertension, in which the spleen is enlarged, sequestering platelets and red and white blood cells.
jaundice Visible yellow staining of the skin and sclera due to an increase in bile pigments in the serum.
lithiasis Formation of stones of any kind (e.g., gallstones).
polysplenia Syndrome consisting of multiple spleens, usually bilateral, with rudimentary and accessory splenic tissue.
porta hepatis (hilum) Location in the liver between the caudate and quadrate lobes that contains the portal vein, hepatic artery, hepatic nerves, hepatic ducts, and lymphatic vessels.
portoenterostomy (Kasai operation) Surgery performed for extrahepatic biliary atresia whereby a Roux-en-Y loop of jejunum is anastomosed to the porta hepatis.
pruritus State of itching.
situs inversus viscerum Condition in which the viscera are transposed in the abdominal cavity, with the liver on the left side and the stomach and spleen on the right.
TORCH infection Toxoplasmosis, rubella, cytomegalovirus, and herpesvirus congenital infections associated with fetal malformations; syphilis is also associated with congenital infections.

Developmental anomalies of the biliary tract are common causes of obstructive jaundice in infants and children. In most cases, these disorders appear to be due to a defect in biliary embryogenesis or to acquired obliteration of the biliary ductal system. Although some of these disorders are amenable to surgical correction, many will continue to impart some degree of impaired bile flow and subsequent development of hepatic fibrosis or cirrhosis. The diagnostic and therapeutic challenge for the clinician is to identify, as rapidly as possible, those patients for whom surgical correction is beneficial. For the remainder of the patients for whom there are no surgical options, treatment is aimed at the prevention and treatment of the complications of cholestasis.

INTRODUCTION

Congenital abnormalities of the liver and biliary tract are usually diagnosed in early infancy and childhood (Table I). Most affected infants and children present with conjugated hyperbilirubinemia. Any infant with jaundice beyond 2 weeks of age (the acceptable range

TABLE I Common Developmental Anomalies of the Biliary Tract

Extrahepatic biliary atresia
Choledochal cyst
Intrahepatic bile duct paucity
Infantile polycystic disease/congenital hepatic fibrosis
Cystic dilatation of intrahepatic ducts

Copyright 2004, Elsevier (USA). All rights reserved.

for physiologic jaundice) should be evaluated by a fractionated bilirubin assay. If the conjugated fraction of bilirubin is elevated, further evaluation should be performed. Although a large number of genetic, metabolic, and infectious disorders may produce neonatal cholestasis, extrahepatic biliary atresia remains among the most common etiologies. Other developmental anomalies of the biliary tract, although less common, represent important potentially correctable causes of infantile cholestasis.

NORMAL DEVELOPMENT OF THE BILIARY TRACT

In the human fetus, the liver begins formation in the third to fourth week of gestation. Immature liver cells begin as outgrowths of the foregut endodermal epithelium; the outgrowths merge with a mesodermal region called the septum transversum. As the endodermal cells proliferate, the liver parenchyma and intrahepatic bile ducts are formed. The liver cells grow in thick sheets with surrounding vasculature, which become the hepatic sinusoids. The sinusoids become the template for the architectural pattern that is assumed by the mature liver.

The bile duct forms as a connection between the fetal liver and the duodenum. The gallbladder and cystic duct connecting the gallbladder to the hepatic duct are formed as outgrowths of the common bile duct. Both are initially hollow, then become solid and once again recanalized by week 12 of gestation, when bile is formed and drains into the gastrointestinal tract. It is the presence of bile that gives the meconium its characteristic dark color. Subsequently, the duodenum rotates and the bile duct assumes its final location, entering the duodenum posteriorly.

The Kupffer cells, connective tissue cells, and hematopoietic cells arise from the mesoderm. The hematopoietic cells reside in the liver until the last 2 months of gestation, when the source of precursor cells is shifted to the bone marrow. At birth, only small regions of hematopoietic cells remain in the liver.

Normal development of the liver and the biliary tract requires a highly coordinated combination of cell proliferation and apoptosis. Developmental anomalies of the liver are presumably related to a defect in this interaction. As a result, there is a large spectrum of abnormalities, including malformation of the entire liver or one of its lobes, positional anomalies (situs inversus viscerum), and hepatic vascular malformations. The most common neonatal anomaly of the biliary tract is extrahepatic biliary atresia.

EXTRAHEPATIC BILIARY ATRESIA

Extrahepatic biliary atresia (EHBA), or bile duct obstruction, is a result of an idiopathic inflammatory process that results in the destruction of the common bile duct, at any point between the porta hepatis and the duodenum. A fibrous cord replaces the normal bile duct and the outflow tract of the liver is therefore obstructed. When persistent obstruction occurs, biliary cirrhosis and its sequelae ensue. EHBA is the most common structural abnormality causing chronic cholestasis in infancy and is the most frequent indication for liver transplantation in the pediatric age group. Although this disease usually presents in the first 2 months of life, in most cases it is unlikely to be a true congenital malformation but is more likely acquired in late pregnancy or after birth.

Characteristics of Extrahepatic Biliary Atresia

EHBA occurs worldwide with an incidence ranging from 1 in 8000 to 1 in 25,000 live births. The female: male ratio is 1.4 : 1 and, in the United States, the incidence is higher in African-Americans and Asians compared to Caucasians. Caucasians have significantly lower survival rates compared to African-Americans, Hispanics, and Asians, but the reasons for this remain unknown. There is no known genetic basis for this disorder.

Kasai has classified extrahepatic biliary atresia into three types. Type I, occurring in 10–15% of cases, consists of an obstruction at the common bile duct and is considered "correctable." Type II is obstruction of the common hepatic duct and includes both correctable and uncorrectable forms. Type III is an "uncorrectable" obstruction of the hepatic ducts at the porta hepatis and occurs in 75–80% of cases.

There are at least two apparent phenotypes of biliary atresia. The embryonic (fetal) type accounts for less than one-third of the cases, and these children have onset of cholestasis immediately after birth. These cases are often associated with other anomalies, including polysplenia, congenital heart disease, situs inversus viscerum, intestinal malrotation and atresias, bilobed right lung, preduodenal portal vein, and azygous continuation of the inferior vena cava. Therefore, the embryonic type is thought to occur early in fetal life and represents a true congenital anomaly. The second type is the perinatal type, in which there is a jaundice-free period followed by later onset of jaundice. This insult most likely occurs after birth, because bile duct remnants are present in the porta hepatis. Landing has suggested that neonatal hepatitis, biliary atresia, and choledochal cyst

TABLE II Characteristics of Extrahepatic Biliary Atresia

Symptoms	Physical signs	Laboratory data[a]
Initial	Jaundice	Conjugated hyperbilirubinemia
Jaundice	Firm liver	Elevated aminotransferases
Acholic or light-colored stools	± Enlarged liver	Elevated alkaline phosphatase
Dark urine	± Enlarged spleen	Elevated γ-glutamyl transpeptidase
Late	Wasting of extremities	Prolonged prothrombin time
Pruritus	± Ascites	Normal to low albumin
Fussiness; crying more than usual		Normal thyroid studies
Abdominal distension		Normal α-1-antitrypsin phenotype
Hematemesis		Negative blood and urine cultures
		Negative for TORCH, VDRL, and HIV infections
		Ultrasound: enlarged liver, enlarged spleen, small or absent gallbladder
		Iminodiacetic acid scan: normal or reduced uptake; absent biliary excretion

[a] TORCH infections include toxoplasmosis, rubella, cytomegalovirus, and herpesvirus; the Venereal Disease Research Laboratory (VDRL) test is for syphilis.

represent a spectrum of disorders initiated by a perinatal insult. A perinatal viral infection has been suggested as the most likely initiating event. Other studies have also suggested an association between viral infections and biliary atresia. Potential etiologic agents include cytomegalovirus, rotavirus (groups A and C), and reovirus type 3.

Diagnosis of Extrahepatic Biliary Atresia

Because EHBA presents similarly to other causes of neonatal cholestasis, differentiating the etiology can be difficult (Table II). The clinical presentation of biliary atresia usually consists of a term infant with the onset of jaundice between 2 and 5 weeks of age. Stools without pigmentation (acholic) are common. On physical examination, the jaundiced infant has an enlarged and firm liver and may have splenomegaly. Laboratory studies reveal mildly to moderately elevated total serum bilirubin levels (usually less than 8 mg/dl, with a conjugated fraction greater than 2 mg/dl, or more than 20% of the total), normal to mildly elevated alanine and aspartate transaminases (ALT and AST), and markedly elevated alkaline phosphatase and γ-glutamyl transpeptidase (GGT). As the disease progresses, biliary cirrhosis develops, with resultant portal hypertension.

The two radiologic studies that are helpful in the diagnosis of biliary atresia are hepatobiliary ultrasound and nuclear scintigraphy. The ultrasound can identify a choledochal cyst, or other etiologies of extrahepatic obstruction, as well as associated anomalies, such as polysplenia, situs inversus viscerum, preduodenal portal vein, and discontinuous inferior vena cava. The gallbladder is often absent or small, even after fasting. Biliary dilatation is not usually visualized. Intravenous technetium-99 m-labeled iminodiacetic acid (IDA) is normally taken up by hepatocytes and excreted into the biliary system. In the case of biliary atresia, the uptake is normal, or delayed in the presence of hepatic dysfunction, and excretion into the duodenum is not visualized. The sensitivity of the test is 97−100%, although the specificity ranges from 43 to 97%. Phenobarbital improves sensitivity by increasing the excretion of the imaging agent. The recommended dose of phenobarbital is 5 mg/kg/day for 3 to 5 days prior to the hepatobiliary scan. Computer tomography (CT) and magnetic resonance imaging (MRI) add little additional diagnostic information.

Evaluation of hepatic histopathology is the most reliable method for diagnosing EHBA. Percutaneous liver biopsy is often obtained for tissue diagnosis prior to surgical exploration. The histologic features of biliary atresia are those of extrahepatic obstruction. There are variable degrees of bile stasis, and bile plugs within the portal ducts. Although specific for biliary obstruction, these findings are present in less than half of the cases. Bile duct proliferation is present with or without lymphocytic inflammation in the portal tracts. Additionally, giant cell transformation of the hepatocytes may be present. The portal triads are enlarged with variable degrees of fibrosis (Fig. 1). With continued cholestasis, portal and periportal fibrosis may progress to bridging fibrosis and cirrhosis.

If the diagnosis in EHBA is suspected, an intraoperative cholangiogram is recommended. The

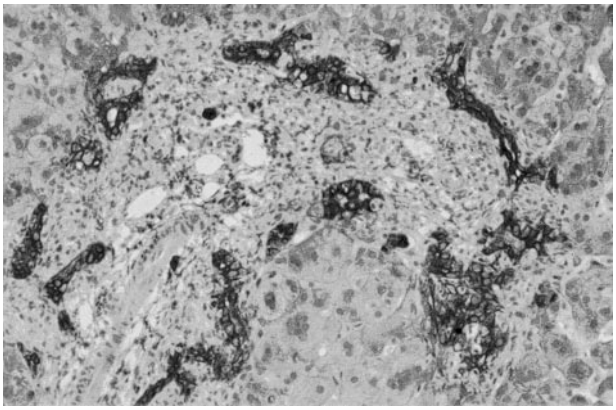


FIGURE 1 Extrahepatic biliary atresia. Bile duct proliferation, mixed inflammatory infiltrate, and fibrosis expand the portal tract. The bile duct cells are prominent throughout the portal zone (darkly stained cells) (cytokeratin stain, original magnification ×40). Courtesy of Hector Monforte, Children's Hospital Los Angeles, California.

usual approach, through the gallbladder, may be difficult due to the smallness of the gallbladder. Meticulous evaluation of the entire biliary tree is necessary because intrahepatic bile duct paucity can be mistaken for EHBA and is not amenable to surgery.

Extrahepatic Complications of EHBA

Extrahepatic complications of biliary atresia include nutritional deficiencies. Poor weight gain and wasting of the extremities are common and are due to the combination of cachexia and fat malabsorption. Intestinal absorption of fat-soluble vitamins is dependent on bile salts, and when bile flow is reduced, as in the case of biliary atresia, vitamin deficiencies can occur. The most common problem is vitamin D deficiency, resulting in rickets or osteopenia. Vitamin E deficiency can lead to progressive neuromuscular symptoms and can be diagnosed early by the loss of deep tendon reflexes. Vitamin A deficiency can lead to retinopathy. Additionally, vitamin K is important for synthesis of clotting factors and can be monitored with a serum prothrombin time. Oral supplementation is recommended at two to four times the recommended daily requirement (Table III). Vitamin D can be supplemented with cholecalciferol (D_3) or ergocalciferol (D_2). Vitamin E is available in two forms, but the micellized form, D-α-tocopherol polyethylene glycol succinate (TPGS), is recommended. Fat malabsorption from the diet also occurs. Medium-chain triglycerides (MCTs) are absorbed by the intestine without need for lipolysis. Therefore, diets containing MCTs are recommended in the form of MCT-containing formulas or as a dietary supplement. However, even those infants that consume adequate calories often suffer from protein-calorie malnutrition, as evidenced by wasting of their extremities.

The most life threatening complication of biliary atresia is portal hypertension. Despite surgical relief of the biliary obstruction, the intrahepatic disease often progresses. One-fifth to one-third of children will not have significant excretion of bile after portoenterostomy and up to three-fourths of children who initially had good bile flow will develop chronic liver disease. This often progresses to biliary cirrhosis and resultant portal hypertension. Almost half of children who survive for 5 or more years after operation have hepatomegaly and/or splenomegaly. Therefore, they are at risk for complications of cirrhosis, including jaundice, hypersplenism, ascites, and bleeding from esophageal varices. Ascites is managed by sodium restriction and diuretics (spironolactone). Persistent ascites despite sodium restriction and oral diuretics may require intravenous colloid, such as albumin, followed by diuretics or paracentesis if the child is having symptoms of respiratory distress or feeding difficulties due to compression of the stomach. Aggressive management is important because ascites increases the risk for spontaneous bacterial peritonitis and is an indicator of progressive hepatic decompensation.

The presence of significant esophageal varices is bimodal in the presence of successful surgical intervention. The first peak occurs in young survivors, between 9 and 23 months, and the second peak occurs in those over 5 years. The larger and more numerous the varices, the higher the risk of bleeding. Varices will develop in about half of those children with splenomegaly. Of

TABLE III Commonly Used Medications for Extrahepatic Biliary Atresia

Medications	Dose
Poly-vi-Sol	2 ml daily
Vitamin K	2.5 mg three times per week
Vitamin E	D-α-Tocopherol polyethylene glycol, 20–25 mg/kg/day
	Aquasol E, 50 units (0.1 ml) daily
Vitamin A	5000 units (0.1 ml) daily
Vitamin D	Ergocalciferol (D_2), 8000 units (1 ml) daily
	Calcitriol, 0.01–0.04 μg/kg/day
Ursodeoxycholic Acid	10–25 mg/kg/day (divided into two daily doses)
Bactrim	2 ml, twice daily (5–10 mg trimethoprim/kg/day)

those, half will develop variceal bleeding. Endoscopic therapy with sclerotherapy or band ligation is the initial treatment of choice.

Treatment of EHBA

Timely diagnosis of biliary atresia is important because early surgical intervention is associated with improved outcome. The Kasai portoenterostomy has altered the outcome of infants with "uncorrectable" biliary atresia, a previously uniformly fatal disease. If the atresia is in the distal common bile duct (type I), the proximal portion can be directly connected to the intestine via Roux-en-Y anastomosis. However, if the proximal common hepatic ducts demonstrate atresia throughout the porta hepatis, the primary therapy is the Kasai portoenterostomy. The gallbladder and extrahepatic biliary tree are excised. Subsequently, the porta is extensively dissected to the surface of the liver, and a segment of jejunum is anastomosed around the small intrahepatic bile ducts. Due to this alteration of the anatomy, the most common complication of the portoenterostomy is ascending cholangitis. Prognosis is directly related to establishing bile flow and resolution of jaundice.

In 82% of Kasai's patients, surgical correction within the first 60 days of life produced satisfactory bile excretion. If the surgery was performed after 90 days of age, the success rate declined to less than 38%. The prognosis of the portoenterostomy depends upon the initial success of the surgery, the extent of the fibrosis, and postoperative complications. In the event of abrupt reduction in bile outflow after previously satisfactory bile excretion, suggesting scar tissue within the porta hepatis, reoperation to reestablish bile flow may be recommended. However, reoperation due to poor initial bile drainage has little advantage at establishing bile flow.

After satisfactory bile flow has been surgically established, laboratory values do not always completely return to normal. Serum bilirubin falls to normal early postoperatively, but ALT and GGT fluctuate between normal to moderately elevated. Alkaline phosphatase often remains elevated but may return to normal over time. Despite normal bilirubin levels, some patients have progressive liver disease. By age 5 years, 40–80% have esophageal varices at endoscopy. Figure 2 summarizes the prognosis in children with EHBA undergoing Kasai portoenterostomy.

Ascending cholangitis is the most significant postoperative complication following portoenterostomy, initially described in 68% of patients. Therefore, surgical modifications of the intestinal anastomosis have

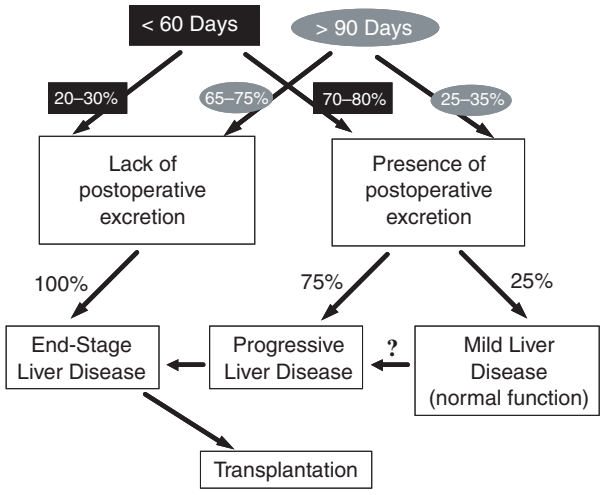


FIGURE 2 Prognosis following Kasai portoenterostomy for extrahepatic biliary atresia. Modified from Sinatra (2001).

been performed and the incidence of cholangitis has been reduced. The symptoms of cholangitis include fever (greater than 38 °C), elevated white blood cell count and erythrocyte sedimentation rate, positive bacterial culture or evidence of sepsis, and simultaneous reduction in bile drainage as evidenced by elevated total and conjugated bilirubin. Bacterial cholangitis prior to the age of 2 years can cause obstruction of the previously patent biliary tree. This complication is almost always limited to children who have established good bile flow, suggesting the direct ascension of intestinal bacteria into the biliary system from the intestinal limb. Treatment includes intravenous antibiotics with both anaerobic and gram-negative coverage. Aggressive management is imperative because cholangitis is associated with progressive fibrosis of the intrahepatic biliary tree. Prophylactic antibiotics remain controversial but many medical centers use a sulfamethoxazole–trimethoprim combination (Septa, Bactrim) for up to 1 year postoperatively.

Prior to the introduction of the Kasai portoenterostomy, the mean life expectancy of an infant with extrahepatic biliary atresia was 11 months. According to Kasai's series, the 10-year survival rate postoperatively was 33% and currently has increased to 54%, with those undergoing surgery prior to 2 months of age having a better prognosis than those undergoing surgery after 3 months of age. Over 50% of pediatric liver transplantations are performed for biliary atresia, and approximately 80% of children undergoing portoenterostomy for biliary atresia will require liver transplantation sometime during their lifetime. The patient survival after liver transplantation is high, 87% at 1 year and 84% at 3 years, with graft survival of 77% at 1 year

and 69% at 3 years, according to the United Network for Organ Sharing.

CHOLEDOCHAL CYST

A congenital segmental cystic dilatation of the biliary system, or choledochal cyst, can be located in any portion of the extrahepatic biliary tract. Symptoms may occur at any age and choledochal cyst must be included in the differential diagnosis of a neonate with cholestasis. Choledochal cysts are 2 to 4 times more common in females, have the highest incidence in Asia, and occur in 1 in 13,000 to 1 in 2,000,000 live births. There is an uncommon association with other abnormalities of the biliary tract, such as biliary atresia, double common duct, double gallbladder, and polycystic or hypoplastic kidneys. The classic clinical presentation includes a triad of pain, abdominal mass, and jaundice, but occurs in only 15% of patients. Neonates usually present with persistent or intermittent jaundice. Other symptoms include fever and vomiting. The laboratory results are nonspecific and demonstrate an elevation of conjugated bilirubin and amino-transferases.

The diagnosis is usually made by pre- or postnatal ultrasound examination that demonstrates extrahepatic and/or intrahepatic bile duct dilatation. Magnetic resonance cholangiopancreatography is another diagnostic tool to evaluate the biliary tree. Todani has classified congenital bile duct cysts into five types, based on location. Type I consists of a fusiform, cystic dilatation of the common bile duct (Fig. 3). Type II is a single outpouching of the common bile duct and/or the gallbladder. Type III is a cystic lesion at the entry into the duodenum, also known as a choledochocoele. In type IV there are multiple cysts in the extrahepatic bile ducts, with or without involvement of the intrahepatic bile ducts. Type V consists of intrahepatic dilatations, similar to those seen in Caroli's disease.

The diagnosis requires an exploratory laparotomy with operative cholangiogram, endoscopic retrograde cholangiography, or transhepatic cholangiography to visualize the entire biliary tree. Once the diagnosis has been confirmed, excision of the entire cyst(s) and gallbladder, followed by Roux-en-Y choledochojejunostomy, is recommended. If untreated, obstructive biliary disease may occur, including bile stasis, cholelithiasis, and pancreatitis, with ultimate development of biliary cirrhosis and portal hypertension. Additionally, there is a high rate (20-fold) of malignant transformation in the residual tissue. The most common malignancy is adenocarcinoma, located in the bile duct, gallbladder, liver, or pancreas. Even after resection, an increased incidence of biliary malignancy exists.

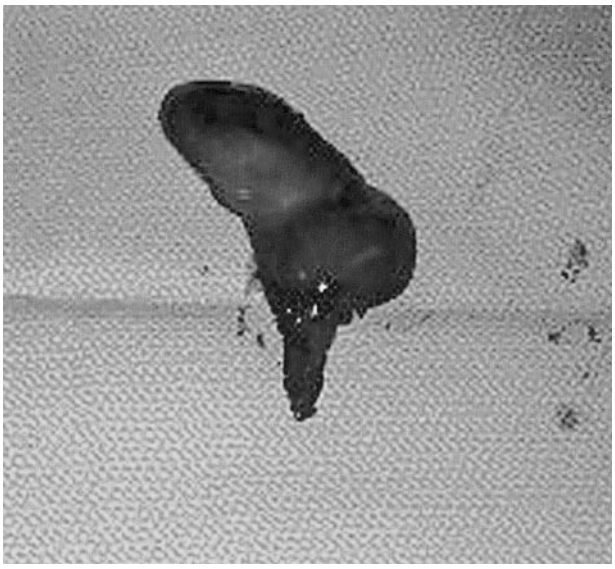


FIGURE 3 Type I choledochal cyst. The fusiform dilatation is located in the proximal region of the extrahepatic biliary tree. This operative specimen is from a 3-year-old female with conjugated hyperbilirubinemia.

INTRAHEPATIC BILE DUCT PAUCITY

Bile duct paucity is defined as absence or reduction of normal intralobular hepatic bile ducts, usually less than 0.5 bile ducts per portal triad. The normal number of bile ducts in term infants and adults is 0.9–1.8 bile ducts per portal triad. The pathogenesis of bile duct paucity is unclear but may be the result of congenital absence, partial absence, atrophy, or response to injury.

The most common condition associated with intrahepatic bile duct hypoplasia in infants and children is arteriohepatic dysplasia or syndromic paucity of intrahepatic bile ducts (Alagille's syndrome). Described by Daniel Alagille in 1975, this group of patients has a variable combination of hepatic ductular hypoplasia and characteristic facies, vertebral malformations, eye findings, and cardiac anomalies. The extrahepatic biliary tree is patent but may be hypoplastic.

Alagille's syndrome occurs worldwide with an incidence of approximately 1 in 100,000. Genetic studies have demonstrated an autosomal dominant inheritance pattern with variable penetrance and expression. Deletion or translocation within the region of chromosome 20p12 has been linked to Alagille's syndrome. This region has been mapped to the human *Jagged 1* gene, which encodes a ligand for the Notch receptor that mediates cell–cell interaction. Further studies are

underway to evaluate how this genetic deletion results in paucity of intrahepatic bile ducts in addition to the other features of this syndrome.

Persistent jaundice is usually the first sign in infants and is often followed by the onset of intense pruritus. Laboratory findings include elevation of conjugated bilirubin, alkaline phosphatase, GGT, cholesterol, triglycerides, and serum bile acids. On physical examination, the children have a recognizable facial pattern, including broad forehead and pointed chin which produce the triangular shape, in addition to mid-face hypoplasia, deeply set eyes and an elongated nose with a flattened tip. Hepatomegaly is usually present. A cardiac murmur can suggest peripheral pulmonic stenosis (in approximately 90% of patients), tetralogy of Fallot, ventricular or atrial septal defects, aortic stenosis, or coarctation. The vertebral anomalies include butterfly vertebrae (hemivertebrae) in the thoracic area, and the ophthalmologic exam may demonstrate the presence of posterior embryotoxon, a prominence of Schwalbe's line within the anterior chamber of the eye. Posterior embryotoxon occurs in up to 15% of the normal population and can result in glaucoma. Short stature and delayed pubertal development are also common. Cutaneous xanthomas are noted in the presence of extreme elevations of serum phospholipids, with serum cholesterol concentrations often in excess of 500 mg/dl. Other affected organs include the kidneys, pancreas, and vasculature of the central nervous system.

Pruritus and vitamin deficiencies are common complications of prolonged cholestasis. Initial speculation suggested that pruritus was due to the elevation of serum bile acids, but the cause is currently unknown. Oral therapies include ursodeoxycholic acid, cholestyramine, phenobarbital, and rifampin. However, when these are ineffective, phototherapy, partial external biliary diversion, and plasmapheresis have been suggested. Fat-soluble vitamins should be monitored, and supplementation as needed should be provided.

In conjunction with associated features, hepatic histopathology provides the most reliable information in diagnosing intrahepatic biliary duct paucity. However, in young infants, liver biopsy specimens may not demonstrate absence or paucity of intrahepatic interlobular ducts, even in the presence of cholestasis. In fact, bile duct proliferation and giant cell transformation with reduced portal tract size and number may be present in the first months of life. Therefore, the importance of a cholangiogram when EHBA is suspected cannot be emphasized enough, because intrahepatic bile duct disease may be exacerbated by performance of a portoenterostomy and is associated with higher mortality when performed. As a child ages, there is a progressive loss of intrahepatic bile ducts with a variable progression to fibrosis or cirrhosis.

The prognosis of Alagille's syndrome is generally good and is directly related to the severity of liver and/or cardiac involvement. The severity of liver disease does not correlate with the severity of extrahepatic manifestations. The morbidity due to pruritus, hyperlipidemia, and vitamin deficiencies may be significant, and treatment is directed at these complications. Patients with Alagille's syndrome may develop hepatic malignancy, most commonly hepatocellular carcinoma. Approximately 15–20% of patients will develop progressive liver disease and require liver transplantation. Transplantation is also indicated to improve the child's quality of life, usually when associated with intractable pruritus, in the absence of cirrhosis.

Other causes of intrahepatic bile duct paucity include progressive familial intrahepatic cholestasis, α1-antitrypsin deficiency, hypopituitarism, Down's syndrome, congenital infections, graft-versus-host disease, and organ rejection.

INFANTILE POLYCYSTIC DISEASE/ CONGENITAL HEPATIC FIBROSIS

The intrahepatic biliary ducts are formed by immature hepatocytes and branches of the portal vein, which comprise, in the ductal plate, a double-walled tube with a narrow lumen. Remodeling of this plate by cell proliferation in combination with apoptosis is important for development of the mature biliary tract. Lack of remodeling of these immature structures results in ductal plate malformations (Fig. 4). This appears to be the basic

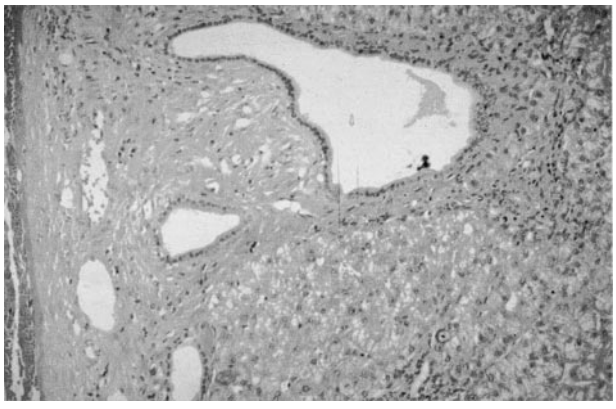


FIGURE 4 Ductal plate malformation in congenital hepatic fibrosis. The portal tract has abnormal and dilated intrahepatic bile duct structures with significant amounts of portal fibrosis (hematoxylin and eosin stain, original magnification ×40).

mechanism in the development of congenital diseases of the intrahepatic bile ducts. Ductal plate malformations appear to be associated with destruction of intrahepatic bile ducts by an inflammatory process. The cysts may be diffuse or localized, causing variable amounts of biliary dilatation and increased fibrous tissue. Multiple syndromes of childhood have been associated with cystic lesions of the liver.

The presence of multiple hepatic ductal cysts is associated with cystic lesions in other organs, especially the kidneys. Cystic disease in the kidney affects the tubular structures of the nephron or collecting ducts. Autosomal recessive polycystic kidney disease (ARPKD) is a rare disorder with an incidence of 1 in 6000 to 1 in 40,000 live births. Genetic studies have revealed an autosomal recessive inheritance pattern, and the gene has been mapped to chromosome 6p. The liver is usually enlarged but liver function is normal. Cystic dilatation of the ductules at the periphery of the portal zone and portal fibrosis are often seen only microscopically. These findings often increase with age. These cystic bile ducts are in communication with the rest of the biliary system and increase the risk of cholangitis, lithiasis, and neoplasm.

Autosomal dominant polycystic kidney disease (ADPKD) is more common, with an incidence of 1 in l000 live births, and the genes have been mapped to chromosomes 16p (85% of cases) and 4q (15% of cases). This disorder is often diagnosed in the adult, not during childhood, because the intrahepatic cystic dilatations may not communicate with the biliary tree and are usually asymptomatic.

Congenital hepatic fibrosis (CHF) is an autosomal recessive disease in which presinusoidal fibrosis is associated with distorted bile duct structures that do not communicate with the biliary tree. The hepatocytes are preserved but portal hypertension develops. Treatment of CHF is directed at controlling variceal bleeding. CHF may occur alone but is also associated with cystic changes of the larger biliary ducts, as in Caroli's syndrome, and 20% of cases have associated renal abnormalities, most commonly ARPKD (less commonly, ADPKD). Other associations include vascular abnormalities, including congenital heart disease and aneurysms, and malformation syndromes such as trisomies 9 and 13.

CYSTIC DILATATION OF INTRAHEPATIC DUCTS (CAROLI'S DISEASE)

Cystic dilatation of the larger, segmental intrahepatic bile ducts, or Caroli's disease, involves multiple segments of the biliary tract. These macroscopic, nonobstructive saccular dilatations alternate with areas of stenosis and are contiguous with the unaffected biliary tract. There is an absence of other histologic abnormalities. These ductal abnormalities are associated with an increased risk of repetitive attacks of cholangitis, intrahepatic lithiasis, and progression to cirrhosis, hepatic failure, and cholangiocarcinoma. Renal abnormalities, most commonly ARPKD, are associated with Caroli's disease.

See Also the Following Articles

Alagille Syndrome • Ascites • Bilirubin and Jaundice • Cirrhosis • Liver Cysts • Neonatal Cholestasis and Biliary Atresia • Portal Hypertension and Esophageal Varices • Pruritus of Cholestasis

Further Reading

Balistreri, W. F. (1999). Liver disease in infancy and childhood. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 1357–1512. Lippincott Williams & Wilkins, Pennsylvania.

Balistreri, W. F., Grand, R., Hoofnagle, J. H., Suchy, F. J., Rychman, F. C., Perlmutter, D. H., and Sokol R. J. (1996). Biliary atresia: Current concepts and research directions–Summary of a symposium. *Hepatology* **23**(6), 1682–1692.

Bates, M. D., Bucuvalas, J. C., Alonso, M. H., and Ryckman, F. C. (1998). Biliary atresia: Pathogenesis and treatment. *Semin. Liver Dis.* **18**(3), 281–293.

Birnbaum, A., and Suchy F. J. (1998). The intrahepatic cholangiopathies. *Semin. Liver Dis.* **18**(3), 263–269.

Chiba, T., Hanamatsu, M., Mochizuki I., Ohi, R., and Kasai, M.. (1983). Follow-up studies on the long-term survivors of biliary atresia—Clinical aspects. *In* "Biliary Atresia and Its Related Disorders" (M. Kasai, ed.), pp. 239–248. Elsevier Science Publ., New York.

Ohi, R. (2001). Surgery for biliary atresia. *Liver* **21**, 175–182.

Piccoli, D. A., and Spinner, N. B. (2001). Alagille syndrome and the *Jagged1* gene. *Semin. Liver Dis.* **21**(4), 525–534.

Sadler, T. W. (1990). Digestive system. *In* "Langman's Medical Embryology" (T. W. Sadler, ed.), 6th Ed., pp. 237–259. Williams & Wilkins, Maryland.

Sinatra, F. R. (2001). Liver transplantation in biliary atresia. *Pediatr. Rev.* 22(5), 166–168.

Bilirubin and Jaundice

WEIZHENG W. WANG AND STEPHEN D. ZUCKER
University of Cincinnati College of Medicine

cholestasis Systemic accumulation of biliary constituents resulting from impaired bile flow.
Crigler–Najjar syndrome Inherited deficiency of bilirubin UDP-glucuronosyltransferase activity.
Dubin–Johnson syndrome Conjugated hyperbilirubinemic disorder caused by an inherited deficiency of the MRP2 canalicular transporter.
Gilbert's syndrome Genetic polymorphism resulting in impaired bilirubin conjugation.
Rotor syndrome Genetic disorder of unclear etiology characterized by conjugated hyperbilirubinemia.

Bilirubin (from the Latin "*bilis*," meaning bile, and "*rubor*," meaning red) is a bile pigment formed during the catabolism of heme-containing compounds, primarily hemoglobin. Excessive accumulation of bilirubin, as a result of enhanced production or impaired elimination, results in yellow discoloration of the skin, sclera, and mucus membranes, which is termed jaundice.

CHEMICAL PROPERTIES OF BILIRUBIN

Bilirubin is a nearly symmetric tetrapyrrole consisting of two rigid planar dipyrrole units joined by a methylene bridge, which is stabilized in a ridge-tile conformation by two trios of internal hydrogen bonds (Fig. 1). Bilirubin exhibits poor aqueous solubility because it is predominantly in the un-ionized diacid form at physiologic pH. For this reason, plasma bilirubin is bound primarily to albumin and, to a much lesser degree, high-density lipoproteins, with only a trivial fraction present as the free monomer.

BILIRUBIN PRODUCTION AND ELIMINATION

Eighty percent of daily bilirubin production (250–400 mg) is derived from erythrocyte hemoglobin, with the remaining 20% resulting from the degradation of other hemoproteins (e.g., cytochromes, myoglobin, catalases). Bilirubin is generated by the sequential activity of heme oxygenase and biliverdin reductase enzymes, which are present at high levels in

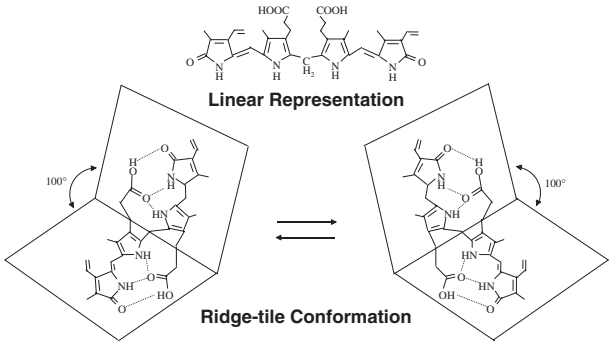


FIGURE 1 The chemical structure of bilirubin. The bilirubin molecule consists of two slightly asymmetrical, rigid, planar dipyrrinones connected by a central CH_2 bridge. The bottom panel depicts the two optical enantiomers of bilirubin in their folded, ridge-tile conformation that is stabilized by intramolecular hydrogen bonds (dotted lines).

reticuloendothelial cells of the spleen, Kupffer cells of the liver, macrophages, and intestinal epithelium. Heme oxygenase is an NADPH-dependent enzyme that initiates the opening of the porphyrin ring by catalyzing the oxidation of the α-carbon bridge of heme, leading to the release of iron and the formation of carbon monoxide and the green pigment, biliverdin. Biliverdin reductase subsequently mediates the reduction of biliverdin to bilirubin, also via an NADPH-dependent mechanism. Bilirubin is cleared from the systemic circulation almost exclusively by the liver. Hepatocellular uptake of bilirubin occurs both by passive diffusion and by a facilitated transport mechanism. In the hepatocyte, bilirubin is conjugated with glucuronic acid to form water-soluble bilirubin mono- and diglucuronides. This reaction is catalyzed by the microsomal enzyme UDP-glucuronosyltransferase (UGT). To date, 15 human UGT isoforms have been identified, each with distinctive substrate specificity. Of these, 8 are encoded by the UGT1A locus, including the bilirubin-specific isoform (UGT1A1). Bilirubin conjugates are actively secreted across the canalicular membrane of the hepatocyte into bile by multidrug-resistance-associated protein 2 (MRP2). Less than 4% of bilirubin in normal human bile is in the unconjugated form.

 Copyright 2004, Elsevier (USA). All rights reserved.

ENTEROHEPATIC CYCLING OF BILIRUBIN

In the colon, bilirubin conjugates are hydrolyzed by bacterial β-glucuronidases. The resultant unconjugated bilirubin either precipitates as insoluble calcium salts or is further reduced by colonic bacteria to a series of molecules termed urobilinogens, which undergo enterohepatic recycling and are eventually excreted into the urine and feces. Under conditions in which bile salt malabsorption occurs (e.g., ileal resection), unconjugated bilirubin is solublized in the intestinal lumen and passively diffuses across the mucosal epithelium, resulting in enhanced bilirubin flux and increased rates of pigment gallstone formation.

BILIRUBIN TOXICITY

Bilirubin is neurotoxic at high concentrations, a phenomenon known as kernicterus (bilirubin encephalopathy). Infants are at particularly high risk for kernicterus when serum bilirubin levels exceed 20–25 mg/dl (340–425 μ*M*). The most affected regions of the brain include the basal ganglia and the brainstem nuclei for oculomotor and auditory function. Phototherapy can reduce serum bilirubin levels by inducing geometric and structural isomerization, thereby enhancing excretion into bile.

JAUNDICE

Jaundice can be observed clinically when hyperbilirubinemia reaches levels that exceed 3 mg/dl (50 μ*M*), approximately three times the upper limit of normal. Hyperbilirubinemia can be classified into two main categories: (1) unconjugated hyperbilirubinemia as a result of bilirubin overproduction or impaired bilirubin conjugation and (2) conjugated hyperbilirubinemia due to impaired canalicular excretion or biliary obstruction (Table I).

TABLE I Classification of Hyperbilirubinemia

- **Unconjugated hyperbilirubinemias**
  - **Bilirubin overproduction**
    - Hemolysis (e.g., sickle cell disease, glucose-6-phosphate dehydrogenase deficiency)
    - Ineffective erythropoiesis (e.g., thalassemia, megaloblastic anemia)
    - Hematoma
  - **Impaired hepatic conjugation**
    - Inherited disorders
      - Crigler–Najjar syndrome
      - Gilbert's syndrome
    - Drugs (e.g., chlorpromazine, rifampin, irinotecan, HIV protease inhibitors)
  - **Multifactorial**
    - Neonatal (physiologic) jaundice
    - Chronic hepatic congestion (e.g., congestive heart failure)
    - Jaundice of sepsis
    - Portosystemic shunts
    - Wilson's disease
    - Zieve's syndrome (associated with alcoholic hepatitis)
- **Conjugated hyperbilirubinemias**
  - **Intrahepatic cholestasis**
    - Inherited disorders
      - Dubin–Johnson syndrome
      - Rotor syndrome
      - Progressive familiar intrahepatic cholestasis
    - Primary biliary cirrhosis
    - Primary sclerosing cholangitis
    - Drugs and toxins (e.g., arsenic, sulfonylureas)
    - Infiltrative process (e.g., sarcoidosis, amyloidosis)
    - Total parental nutrition
    - Cholestasis of pregnancy
    - Liver transplant rejection
    - Hepatocellular injury (e.g., hepatitis A, fibrosing cholestatic hepatitis)
    - Paraneoplastic syndrome (e.g., lymphoma)
  - **Extrahepatic cholestasis (biliary obstruction)**
    - Choledocholithiasis
    - Malignancy
      - Intrinsic (e.g., cholangiocarcinoma, ampullary carcinoma)
      - Extrinsic (e.g., pancreatic carcinoma, hepatocellular carcinoma, metastatic disease)
    - Biliary stricture (e.g., postsurgical, sclerosing cholangitis)
    - Parasitic infections (e.g., *Ascaris lumbricoides*, liver flukes)

Unconjugated Hyperbilirubinemias

Bilirubin Overproduction

Extravascular hemolysis, intravascular hemolysis, ineffective erythropoiesis, or large hematomas may cause increased bilirubin production through enhanced heme catabolism. Because hepatic bilirubin clearance is extremely efficient and the bone marrow cannot increase erythrocyte production more than eightfold, ongoing steady-state hemolysis generally does not produce plasma unconjugated hyperbilirubinemia beyond 4–5 mg/dl. Bilirubin concentrations that exceed this level are indicative of coincident hepatic dysfunction.

Crigler–Najjar Syndrome

Crigler–Najjar syndrome is a rare, autosomal recessive disorder of bilirubin metabolism. It has been classified into two distinct forms (types I and II) based upon the severity of disease. Crigler–Najjar type I manifests as extreme jaundice and kernicterus as a result of absent hepatic UGT1A1 activity and is nearly always fatal unless liver transplantation is performed. In contrast, type II disease, which is caused by small point mutations that

reduce but do not abrogate enzyme activity, is associated with lower serum bilirubin concentrations and affected individuals typically survive into adulthood without neurologic impairment. Patients with Crigler–Najjar type II frequently respond to phenobarbital treatment, which reduces serum bilirubin levels by approximately 25%.

Gilbert's Syndrome

Gilbert's syndrome is the most common inherited disorder of bilirubin glucuronidation. Affected individuals exhibit isolated unconjugated hyperbilirubinemia, with levels as high as 6 mg/dl occurring in the setting of fasting, febrile illness, or physical stress. This benign condition is caused by a polymorphism in the promoter TATA element of the gene encoding UGT1A1, leading to a TA insertion into the wild-type $A(TA)_6TAA$ sequence. Liver homogenates from individuals homozygous for the Gilbert's $A(TA)_7TAA$ genotype exhibit a 50% reduction in bilirubin-conjugating activity. Approximately 9% of the general population is homozygous and 30% heterozygous for the Gilbert's polymorphism.

"Physiologic Jaundice" of the Neonate

Mild unconjugated hyperbilirubinemia ($\leq$6 mg/dl) is commonly observed in the neonatal period, generally peaking at 3 to 4 days of age. This condition arises from the combination of increased production, decreased conjugation, and enhanced enterohepatic cycling of bilirubin. Bilirubin production is two to three times higher in neonates than in adults due to increased turnover of erythrocytes. Furthermore, bilirubin UGT1A1 activity is very low at birth, rising to adult levels over the initial 2 weeks of life. Newborns also have a relative paucity of enteric flora, resulting in decreased conversion of bilirubin to urobilinogen and high intestinal activity of β-glucuronidase, leading to enhanced deconjugation of bilirubin, which is then absorbed across the intestinal mucosa.

Conjugated Hyperbilirubinemias

Conjugated hyperbilirubinemia results either from impaired canalicular secretion of conjugated bilirubin into bile (intrahepatic) or from obstruction of the biliary system (extrahepatic).

Dubin–Johnson Syndrome

Dubin–Johnson syndrome is a rare, benign, autosomal recessive disorder of the liver characterized by chronic conjugated hyperbilirubinemia. It is caused by mutations in the MRP2 gene. MRP2 is responsible for the canalicular secretion of non-bile-acid organic anions, such as conjugated bilirubin and other glucuronide- or glutathione-conjugated compounds. The liver is histologically normal except for the presence of dense pigment within hepatocytes (lysosomes containing polymers of epinephrine metabolites), which produces a grossly black appearance.

Rotor Syndrome

Individuals with the autosomal recessive Rotor syndrome also exhibit chronic conjugated hyperbilirubinemia. However, in contrast to Dubin–Johnson syndrome, in which total urinary coproporphyrin excretion is normal with 80% excreted as coproporphyrin I, Rotor syndrome patients exhibit elevated levels of urinary coproporphyrins with approximately 65% coproporphyrin I. The liver exhibits a normal appearance and the underlying molecular defect remains to be elucidated.

Progressive Familial Intrahepatic Cholestasis

Progressive familial intrahepatic cholestasis is a heterogeneous group of disorders, characterized by defective secretion of biliary components. Mutations in several canalicular transport proteins have been described, including the aminophospholipid transferase (FIC1), the bile salt export pump (BSEP), and the multidrug resistance protein 3 P-glycoprotein (MDR3), which serves as a phosphatidylcholine flipase. These disorders typically present in infancy or childhood and are associated with growth retardation and progressive liver failure.

Cholestasis

Any process that impairs canalicular secretion of biliary conjugates or obstructs the flow of bile can cause cholestasis and concomitant conjugated hyperbilirubinemia. As canalicular transport processes are typically impaired earlier in the course of liver disease than is hepatic conjugation, hepatocellular injury tends to manifest as conjugated hyperbilirubinemia. In extrahepatic cholestasis (biliary obstruction), conjugated bile pigments and other components of bile, such as bile salts and alkaline phosphatase, spill over into the blood.

See Also the Following Articles

Bile Composition • Bile Flow • Kernicterus • Liver Cysts • Neonatal Hyperbilirubinemia

Further Reading

Bosma, P. J., Chowdhury, J. R., Bakker, C., Gantla, S., de Boer, A., Oostra, B. A., Lindhout, D., Tytgat, G. N., Jansen, P. L., Elferink, O., *et al.* (1995). The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *N. Engl. J. Med.* **333**, 1171–1175.

Brink, M. A., Slors, J. F., Keulemans, Y. C., Mok, K. S., De Waart, D. R., Carey, M. C., Groen, A. K., and Tytgat, G. N. (1999). Enterohepatic cycling of bilirubin: A putative mechanism for pigment gallstone formation in ileal Crohn's disease. *Gastroenterology* **116**, 1420–1427.

Dennery, P. A., Seidman, D. S., and Stevenson, D. K. (2001). Neonatal hyperbilirubinemia. *N. Engl. J. Med.* **344**, 581–590.

Goessling, W., and Zucker, S. D. (2000). Role of apolipoprotein D in the transport of bilirubin in plasma. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G356–G365.

Gourley, G. R. (1997). Bilirubin metabolism and kernicterus. *Adv. Pediatr.* **44**, 173–346.

Kadakol, A., Ghosh, S. S., Sappal, B. S., Sharma, G., Chowdhury, J. R., and Chowdhury, N. R. (2000). Genetic lesions of bilirubin uridine-diphosphoglucuronate glucuronosyltransferase (UGT1A1) causing Crigler–Najjar and Gilbert syndromes: Correlation of genotype to phenotype. *Hum. Mutat.* **16**, 297–306.

Lee, J., and Boyer, J. L. (2000). Molecular alterations in hepatocyte transport mechanisms in acquired cholestatic liver disorders. *Semin. Liver Dis.* **20**, 373–384.

McDonagh, A. F., and Lightner, D. A. (1985). Like a "shrivelled blood orange"—Bilirubin, jaundice, and phototherapy. *Pediatrics* **75**, 443–455.

Nogales, D., and Lightner, D. A. (1995). On the structure of bilirubin in solution: $^{13}C[^{1}H]$ heteronuclear Overhauser effect NMR analyses in aqueous buffer and organic solvents. *J. Biol. Chem.* **270**, 73–77.

Ostrow, J. D., Mukerjee, P., and Tiribelli, C. (1994). Structure and binding of unconjugated bilirubin: Relevance for physiological and pathophysiological function. *J. Lipid Res.* **35**, 1715–1737.

Thompson, R., and Jansen, P. L. (2000). Genetic defects in hepatocanalicular transport. *Semin. Liver Dis.* **20**, 365–372.

Zucker, S. D., Goessling, W., and Hoppin, A. G. (1999). Unconjugated bilirubin exhibits spontaneous diffusion through model lipid bilayers and native hepatocyte membranes. *J. Biol. Chem.* **274**, 10852–10862.

Zucker, S. D., Qin, X., Rouster, S. D., Yu, F., Green, R. M., Keshavan, P., Feinberg, J., and Sherman, K. E. (2001). Mechanism of indinavir-induced hyperbilirubinemia. *Proc. Natl. Acad. Sci. USA* **98**, 12671–12676.

Boerhaave's Syndrome

MAKAU LEE
University of Mississippi Medical Center, Jackson

Boerhaave's syndrome Spontaneous, transmural tear of the esophagus, with free perforation.
Hamman sign Mediastinal crunching sound with heartbeat.
mediastinitis Infection involving the mediastinum.
odynophagia Pain on swallowing foods or liquids.

Herman Boerhaave, in 1724, described a spontaneous, transmural tear of the esophagus, with free perforation resulting from very forceful retching and vomiting. Boerhaave's syndrome is seen frequently in alcoholism and is highly lethal if diagnosis and treatment are not prompt.

CLINICAL FEATURES

Although most patients report antecedent retching and vomiting, Boerhaave's syndrome can result from any maneuver that causes a sudden increase in intraabdominal pressure. A majority of the esophageal ruptures occur in the distal esophagus on the left side. Boerhaave's syndrome tends to occur outside of the hospital setting and has a high mortality rate from mediastinitis, abscess formation in the mediastinum and pleural space, tissue necrosis, and septicemia, due to a delay in diagnosis. Delays in the diagnosis of Boerhaave's syndrome are common because the initial trauma leading to the perforation frequently occurs during an alcoholic stupor, and the clinical manifestations of this condition are generally nonspecific.

DIAGNOSIS

Clinically, patients with Boerhaave's syndrome may present with chest, neck, and abdominal pain and

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Bosma, P. J., Chowdhury, J. R., Bakker, C., Gantla, S., de Boer, A., Oostra, B. A., Lindhout, D., Tytgat, G. N., Jansen, P. L., Elferink, O., *et al.* (1995). The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *N. Engl. J. Med.* **333**, 1171–1175.

Brink, M. A., Slors, J. F., Keulemans, Y. C., Mok, K. S., De Waart, D. R., Carey, M. C., Groen, A. K., and Tytgat, G. N. (1999). Enterohepatic cycling of bilirubin: A putative mechanism for pigment gallstone formation in ileal Crohn's disease. *Gastroenterology* **116**, 1420–1427.

Dennery, P. A., Seidman, D. S., and Stevenson, D. K. (2001). Neonatal hyperbilirubinemia. *N. Engl. J. Med.* **344**, 581–590.

Goessling, W., and Zucker, S. D. (2000). Role of apolipoprotein D in the transport of bilirubin in plasma. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G356–G365.

Gourley, G. R. (1997). Bilirubin metabolism and kernicterus. *Adv. Pediatr.* **44**, 173–346.

Kadakol, A., Ghosh, S. S., Sappal, B. S., Sharma, G., Chowdhury, J. R., and Chowdhury, N. R. (2000). Genetic lesions of bilirubin uridine-diphosphoglucuronate glucuronosyltransferase (UGT1A1) causing Crigler–Najjar and Gilbert syndromes: Correlation of genotype to phenotype. *Hum. Mutat.* **16**, 297–306.

Lee, J., and Boyer, J. L. (2000). Molecular alterations in hepatocyte transport mechanisms in acquired cholestatic liver disorders. *Semin. Liver Dis.* **20**, 373–384.

McDonagh, A. F., and Lightner, D. A. (1985). Like a "shrivelled blood orange"—Bilirubin, jaundice, and phototherapy. *Pediatrics* **75**, 443–455.

Nogales, D., and Lightner, D. A. (1995). On the structure of bilirubin in solution: $^{13}C[^{1}H]$ heteronuclear Overhauser effect NMR analyses in aqueous buffer and organic solvents. *J. Biol. Chem.* **270**, 73–77.

Ostrow, J. D., Mukerjee, P., and Tiribelli, C. (1994). Structure and binding of unconjugated bilirubin: Relevance for physiological and pathophysiological function. *J. Lipid Res.* **35**, 1715–1737.

Thompson, R., and Jansen, P. L. (2000). Genetic defects in hepatocanalicular transport. *Semin. Liver Dis.* **20**, 365–372.

Zucker, S. D., Goessling, W., and Hoppin, A. G. (1999). Unconjugated bilirubin exhibits spontaneous diffusion through model lipid bilayers and native hepatocyte membranes. *J. Biol. Chem.* **274**, 10852–10862.

Zucker, S. D., Qin, X., Rouster, S. D., Yu, F., Green, R. M., Keshavan, P., Feinberg, J., and Sherman, K. E. (2001). Mechanism of indinavir-induced hyperbilirubinemia. *Proc. Natl. Acad. Sci. USA* **98**, 12671–12676.

Boerhaave's Syndrome

MAKAU LEE
University of Mississippi Medical Center, Jackson

Boerhaave's syndrome Spontaneous, transmural tear of the esophagus, with free perforation.
Hamman sign Mediastinal crunching sound with heartbeat.
mediastinitis Infection involving the mediastinum.
odynophagia Pain on swallowing foods or liquids.

Herman Boerhaave, in 1724, described a spontaneous, transmural tear of the esophagus, with free perforation resulting from very forceful retching and vomiting. Boerhaave's syndrome is seen frequently in alcoholism and is highly lethal if diagnosis and treatment are not prompt.

CLINICAL FEATURES

Although most patients report antecedent retching and vomiting, Boerhaave's syndrome can result from any maneuver that causes a sudden increase in intraabdominal pressure. A majority of the esophageal ruptures occur in the distal esophagus on the left side. Boerhaave's syndrome tends to occur outside of the hospital setting and has a high mortality rate from mediastinitis, abscess formation in the mediastinum and pleural space, tissue necrosis, and septicemia, due to a delay in diagnosis. Delays in the diagnosis of Boerhaave's syndrome are common because the initial trauma leading to the perforation frequently occurs during an alcoholic stupor, and the clinical manifestations of this condition are generally nonspecific.

DIAGNOSIS

Clinically, patients with Boerhaave's syndrome may present with chest, neck, and abdominal pain and

Copyright 2004, Elsevier (USA). All rights reserved.

odynophagia, dysphagia, hoarseness, aphonia, vomiting, hematemesis, and respiratory distress. Physical examination may reveal subcutaneous crepitation, mediastinal crunching sound with the heartbeat (also called Hamman sign), fever, and shock. Of note, only one-third of patients with Boerhaave's syndrome present with the classic Mackler triad of chest pain, vomiting, and subcutaneous emphysema. Leukocytosis is common. Radiographic findings include pleural effusion, mediastinal widening, hydropneumothorax, and pneumomediastinum. Esophagrams performed with water-soluble contrast medium (such as gastrograffin) are generally diagnostic. The differential diagnosis should include Mallory–Weiss tear, esophageal intramural hematoma, peptic ulcer disease and its complications (such as bleeding and perforation), aortic dissection, myocardial infarction, pericarditis, pulmonary embolism, spontaneous pneumothorax, and pancreatitis.

MANAGEMENT

Boerhaave's syndrome is a highly lethal condition that demands early surgical consultation. Treatment is primarily surgical, involving repair of the perforation, with or without resection of the affected segment of the esophagus. Diagnosed patients should not be managed nonoperatively because diagnosis typically occurs at a time when gross contamination of the surrounding tissues is already advanced and surgical intervention is required.

See Also the Following Articles

Dysphagia • Emesis • Esophageal Surgery • Esophageal Trauma • Esophagus, Anatomy • Mallory–Weiss Tear

Further Reading

Baehr, P. H., and McDonald, G. B. (1998). Esophageal disorders caused by infection, systemic illness, medications, radiation, and trauma. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., pp. 519–539. W. B. Saunders, Philadelphia.

Faigel, D. O., and Fennerty, M. B. (1999). Miscellaneous diseases of the esophagus. *In* "Textbook of Gastroenterology" (T. Yamada., ed.), 3rd Ed., pp. 1304–1325. Lippincott Williams & Wilkins, Wilkins, Philadelphia.

Younes, Z., and Johnson, D. A. (1999). The spectrum of spontaneous and iatrogenic esophageal injury. *J. Clin. Gastroenterol.* **29**, 306–317.

Borborygmus

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

borborygmi Sounds emanating from the digestive tract.
intestinal power propulsion One of the patterns of motility occurring in the small and large intestine.
migrating motor complex One of the patterns of motility in the stomach and small intestine.

Sounds originating from the digestive tract are termed borborygmus in its singular form and borborygmi in the plural. The term was derived from the Latin word for "rumbling." Borborygmi can be heard during physical examination by placing the listening device on the abdominal surface. Exaggerated borborygmi can often be heard without the aid of listening devices by someone positioned within approximately 1 m of the individual. Movements of gas and liquid in the stomach and intestines cause borborygmi. The muscles in the walls of the stomach and intestines contract in a variety of organized patterns that propel the liquid and gaseous contents within the lumen. Rapid propulsion generates turbulence in the mixed medium of liquid and gas bubbles. The passage of turbulence, especially through orifices of reduced diameter, produces sounds of sufficient volume to be heard away from the abdomen. These sounds may be variously described as rumbling, "growling," "bubbling," "gurgling," etc. There are two kinds of bowel sounds. Normal sounds are those associated with the normal digestion of a meal and the clearance of undigested residue after digestion of a meal is complete. Pathologic bowel sounds are those that reflect a disordered condition. Knowledge of the relations of bowel sounds and specific pathologic conditions can be helpful in clinical diagnosis of patient complaints.

NORMAL BOWEL SOUNDS

The stomach, small intestine, and large intestine can generate borborygmi. Each of these specialized organs differs in its normal physiology. Each differs in its patterns of muscle contractions and propulsive motility. The properties of the solid, liquid, and gaseous contents also differ for the three organs. Together, these factors determine the nature of the sounds that can be heard coming from the abdomen.

Stomach

The stomach is divided into two functional compartments. The proximal compartment is for storage. Its musculature relaxes to increase compartmental volume as food is ingested and slowly contracts to decrease the volume as emptying of the stomach proceeds. The distal compartment has powerful musculature and a contractile pattern that are specialized for grinding food into the small-sized particles required for emptying and for propulsion of the contents through the sphincter that separates the stomach from the small intestine. Propulsive contractile waves are ring-like in the distal stomach and occur at a frequency of 3 per minute in humans. They start in the midstomach and travel to the junction of the stomach with the small intestine. The ring-like contractions vary in strength from weak contractions that can barely be seen in the lumen to very strong contractions that obliterate the lumen. Agitation of the contents by the forces of these contractions accounts for the sounds emanating from the stomach.

The sounds of the stomach are most audible when the subject is reclining in such a way as to bring the gas bubble into the distal compartment of the stomach. In humans, the sounds may continue in lock-step fashion for 2 h or more after a large meal.

"Growling" sounds from the stomach and the sensation of "hunger pangs" several hours after a meal reflect a pattern of motility termed the migrating motor complex (MMC). The MMC is a specialized pattern of gastrointestinal motility that starts in the stomach and slowly migrates to the small intestine and onward toward the small intestinal junction with the large intestine. The first MMC starts in the stomach when digestion and absorption of a meal have reached completion. A second MMC starts in the stomach as the first one reaches the end of the small intestine. Eighty to 120 min are required for the MMC to travel the length of the small intestine. This cycling of the MMC is repeated at approximately hourly intervals until such time as the next meal is ingested.

Ring-like propulsive contractions in the distal stomach are strongest during the occurrence of an MMC.

 Copyright 2004, Elsevier (USA). All rights reserved.

Secretions and gas bubbles accumulate in the stomach between occurrences of the MMC and their movements during an MMC account for the gastric noises. Hunger pangs are believed to be the sensations an individual experiences while the strong contractions of the MMC are occurring in the distal compartment of the stomach.

Small Intestine—Large Intestine

The rhythmic sounds of the small intestine are different in quality from the sounds of the stomach when monitored with a listening device on the abdominal surface. Walter C. Alvarez, of the Mayo Clinic, described the small intestinal sounds as "either soft crepitations, rattling explosive discharges or sometimes slow rumbles" that differ from the "gushing, explosive sounds of the stomach." The sounds from the small intestine are best heard in the lower left quadrant of the abdomen.

Walter B. Cannon (1871–1945), the much respected physiologist at Harvard University, first reported gurgling sounds that were interpreted to be due to contractions of the colon. After the subject was given an enema, the sounds came at regular intervals and were associated with cramp-like lower abdominal pain.

Contractions of the wall musculature and the movements of the lumenal contents cause the borborygmi that emanate from the small intestine or colon. A specialized pattern of intestinal motility, termed power propulsion, most likely accounts for the exaggerated sounds from the small or large bowel. Power propulsion involves strong, long-lasting contractions that propagate for extended distances along the intestine. The contractions are part of an efficient propulsive mechanism that rapidly strips the lumen clean as it travel at speeds of approximately 1 cm/s over long lengths of intestine. Power propulsion accomplishes mass movement of intralumenal material in normal and disordered states. Its occurrence in disordered states (e.g., stimulant laxatives or intestinal infections) is associated with abdominal cramp-like pain.

PATHOLOGIC BOWEL SOUNDS

Ability to hear increased bowel sounds is among the most common physical findings in mechanical obstruction of the bowel. In suspected cases of intestinal obstruction, auscultation is generally included in the physical examination. Borborygmi are often the earliest physical evidence of obstruction. The sounds are described as loud and bubbling, like those made by the emptying of a bottle of water. They can be heard in large rushes and be succeeded by periods of relative quiescence. Frequently, fine high-pitched metallic-like "tinkles" that are often localized to a small abdominal region are heard.

Borborygmi heard in mechanical obstruction are accentuated during the waves of abdominal pain experienced by the patient. The paroxysms of pain and the rushes of sound occur simultaneously because both are produced by powerful contractions of the intestinal musculature. Increased strength of contraction in this case presumably reflects adjustment by the intestinal nervous system for propulsive forces to overcome the obstruction. Whereas audible borborygmi are useful diagnostic signs of mechanical blockage of passage through the intestine, the absence of bowel sounds is a characteristic feature of paralytic ileus, which is the pathologic absence of intestinal motility.

See Also the Following Articles

Colonic Motility • Colonic Obstruction • Duodenal Motility • Gastric Motility • Migrating Motor Complex • Pathologic and Paralytic Ileus • Power Propulsion • Small Intestinal Motility

Further Reading

Alvarez, W. (1928). "The Mechanics of the Digestive Tract." Paul B. Hoeber, New York.

Cannon, W. (1905). Auscultation of the rhythmic sounds produced by the stomach and intestines. *Am. J. Physiol.* **14**, 339–353.

Summers, R., and Liu, C. (1995). Approach to the patient with ileus and obstruction. *In* "Textbook of Gastroenterology" (T. Yamada, D. Alpers, C. Owyang, D. Powell, and F. Silverstein, eds.), 2nd Ed., Vol. 1, pp. 796–798. Lippincott, Philadelphia.

Brain–Gut Axis

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

afferent trafficking Inflow of sensory information from the gut to the brain.
efferent trafficking Outflow of information to the gut from the brain and spinal cord.
homeostasis Maintenance of a constant internal state in an organ or the whole body.

The brain–gut axis is a physiologic system comprising all of the neural centers and the incorporated neurophysiologic mechanisms involved between the head and gut. Two-way trafficking of neural information in the moment-to-moment control of activities in the digestive tract is the rule in the brain–gut axis.

EFFERENT–AFFERENT TRAFFICKING

Efferent trafficking is the outflow of information (i.e., efferent information) to the gut from the brain and spinal cord. Outflow information is generally command information for gut behavior determined by the brain to be necessary for bodily homeostasis at a given instant in time. Afferent trafficking is the inflow of sensory information from the gut to the brain. A steady stream of sensory information informs the brain of the ongoing state of the gut. The integrative microcircuits in the brain require accurate information on the state of the gut in order to issue commands for effective adjustments of state. State adjustments include (1) amount of acid being secreted in the stomach, (2) the size of the stomach reservoir during ingestion of food, (3) rate of emptying of the stomach into the small intestine, and (4) reversal of propulsive motility during vomiting.

Afferent information in the brain–gut axis underlies the sensations that an individual refers to the gut. This aspect of information flow to the brain stands aside from the necessity of afferent information for the functions of the automatic feedback control loops that exist between the central nervous system and gut. Most gut feelings would not be felt without the transmission of afferent information and processing of the information in the brain. A majority of the sensations referred to the gut are unpleasant. Often, when extreme, the sensations are motivation for consultation with a physician. Some well-known examples are heartburn, bloating, cramping abdominal pain, and urgency to defecate.

EMOTIONAL GUT

The underpinnings of the "emotional gut" are centered in the brain part of the brain–gut axis. Connections between the neural networks in the higher centers of the brain (e.g., cerebral cortex) and the gut account for the projections to the digestive tract of emotions such as fright, stress, and anger. Projections to the gut of emotions of psychogenic origin are often profound. A repugnant experience may be repeated as nausea and vomiting either at the thought of or on a repetition of the experience. Recently appreciated psychosocial aspects of functional bowel disorders direct additional attention to the brain–gut axis. The evidence suggests that symptoms of the irritable bowel syndrome may arise later in life as a consequence of physical or emotional abuse in earlier life.

See Also the Following Articles

Autonomic Innervation • Chest Pain, Non-Cardiac • Emesis • Enteric Nervous System • Irritable Bowel Syndrome • Nausea

Further Reading

Drossman, D., Creed, F., Olden, K., Svedlund, J., Toner, B., and Whitehead, W. (2000). Psychosocial aspects of the functional gastrointestinal disorders. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 372–391. Degnon Associates, McLean, Virginia.

Salt, W. (1997). "Irritable Bowel Syndrome and the Mind–Body Brain–Gut Connection." Parkview Publ., Columbus, Ohio.

Wood, J., Alpers D., and Andrews, P. (2000). Fundamentals of neurogastroenterology: Basic science. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 31–90. Degnon Associates, McLean, Virginia.

 Copyright 2004, Elsevier (USA). All rights reserved.

Breath Tests

ALAN F. HOFMANN
University of California, San Diego

alveolar air Expired breath that is representative in composition of the gas present in the alveoli, the smallest divisions of the respiratory tree. Alveolar air is obtained by having the subject expire normally and then having the subject forcefully expire the residual air in his or her lungs.

endolytic breath test A breath test based on metabolism of a substrate labeled with isotopic carbon by tissue enzymes.

stable isotope An isotope containing more or fewer neutrons in the nucleus than those observed in the most abundant form of the element. The stable isotope used in breath tests is ^{13}C, which can be measured by mass spectrometry or infrared spectroscopy.

ureolysis Cleavage of urea to carbon dioxide and ammonia by bacterial enzymes.

xenolytic breath test A breath test based on formation of hydrogen or labeled carbon dioxide from an ingested substrate by bacterial enzymes.

This article summarizes the principles of breath tests and describes their clinical purposes and utility. Breath tests are diagnostic tests in which the concentration of labeled carbon dioxide or hydrogen in breath is measured in order to assess the rate of a physiological or pathophysiological process.

INTRODUCTION

Breath tests may be divided into two types according to what is measured in breath—labeled carbon dioxide or hydrogen. They may also be divided into endolytic tests, in which the rate to be measured is mediated by tissue enzymes, or xenolytic tests, in which the substrate is metabolized by bacterial enzymes. Since tissue enzymes do not form hydrogen, all hydrogen breath tests are xenolytic tests. Labeled carbon dioxide tests may be either endolytic or xenolytic.

Other substances may be measured in breath in the research laboratory to provide metabolic information, but discussion of the use of such substances is beyond the scope of this article. Examples are breath ethane to assess cytochrome P450 function, breath carbon monoxide to assess the rate of bilirubin biosynthesis or airway inflammation, hydrogen peroxide to detect airway inflammation, mercaptans in breath as an indicator of halitosis, or ammonia to detect ureolysis of ingested urea or to infer impaired urea synthesis caused by impaired hepatic function. However, such measurements are not usually considered to be "breath tests." In addition, there are other studies examining specific substrates to detect abnormalities in uncommon diseases. These include measurement of $^{13}CO_2$ excretion after administration of cholesterol-1-[^{13}C]oleate incorporated into chylomicrons or measurement of $^{13}CO_2$ production after administration of [^{13}C]galactose in inborn errors of galactose metabolism.

LABELED CARBON DIOXIDE BREATH TESTS

Overview

Labeled carbon dioxide breath tests use either ^{13}C- or ^{14}C-labeled substrates. ^{13}C is a stable isotope that offers the advantage of avoiding radioactivity exposure to the subject; however, ^{13}C has the disadvantage of being more difficult to measure than ^{14}C as well as being a more costly isotope. ^{14}C is a radioactive isotope of carbon that gives off a weak beta particle. The half-life is several thousand years. Because the mass of breath CO_2 that is collected is large (1–2 mmol) and because of the high efficiency of liquid scintillation spectrometric measurement of ^{14}C, only very small quantities of radioactivity need to be administered (5–10 μCi). The radiation hazard to the patient has been calculated to be negligible and the [^{14}C]urea breath test has been approved for marketing in the United States.

Measurement of Labeled Carbon Dioxide in Breath

^{13}C is a stable isotope that can be measured by using a dedicated mass spectrometer or by infrared spectroscopy. When mass spectroscopy is used, breath is collected in an impermeable bag or in an evacuated test tube and mailed to a mass spectrometry facility for processing. Mass spectrometric measurements have been automated and are rapid, accurate, and precise. Measurement of ^{13}C by infrared spectroscopy has improved

 Copyright 2004, Elsevier (USA). All rights reserved.

greatly in the past few years, and a tabletop instruments that permit measurements of ^{13}C within minutes have been developed. Such instruments permit ^{13}C breath tests to be performed in the clinic or at the bedside with results available shortly after the test is performed.

Measurement of $^{14}CO_2$ is made in two ways. In the first method, breath CO_2 is trapped in a known volume of organic base. To achieve this, alveolar air is blown into the base that contains an acid–base indicator until the color changes, indicating that the base has been fully consumed by expired carbon dioxide. Radioactivity is determined by liquid scintillation spectroscopy. In the second method, breath is trapped in a base and the content of ^{14}C is determined using a Geiger–Muller radiation detector.

Both of these methods give breath-specific activity, that is, the enrichment of breath CO_2 by the administered isotope. To convert enrichment or specific activity to percentage dose expired per unit time, a certain rate of endogenous CO_2 production is assumed, e.g., 9 μmol CO_2 per kilogram per hour.

Endolytic Labeled Carbon Dioxide Breath Tests

In endolytic breath tests, cleavage of the target bond or availability of the substrate to cleavage is rate-limiting in the conversion of an ingested or injected substrate to CO_2. It is assumed that all of the reactions after cleavage are rapid so that the rate of labeled CO_2 production is directly proportional to the rate of cleavage of the target bond. Octanoate, for example, is an excellent "leaving group," as it is rapidly metabolized to CO_2 in the hepatocyte and is not incorporated into ester lipids.

The endolytic labeled carbon dioxide breath test that appears to have some clinical utility is the [^{14}C]octanoate or [^{13}C]octanoate breath test, which is used to measure the rate of gastric emptying. Octanoate is poorly absorbed from the stomach, rapidly absorbed from the small intestine, and rapidly oxidized to CO_2. Accordingly, the rate of appearance of labeled carbon dioxide in the breath is, in principle, a valid indicator of gastric emptying. The disease to be detected is delayed gastric emptying as occurs, for example, in diabetic neuropathy. At present, the test is investigational. The test is not as sensitive as imaging studies of gastric emptying, but is much easier to conduct.

For measurement of pancreatic function, cholesterol-1-[^{14}C]octanoate has been proposed. The cholesterol ester is a substrate for the nonspecific esterase (bile acid-dependent) present in pancreatic juice. The test is not specific for pancreatic disease. It is also not sensitive when compared to direct (invasive) measurements of pancreatic enzyme secretion. The test is investigational at present.

For assessment of lactase activity, a [^{14}C]lactose test has been described and, in limited studies, was sensitive and specific. However, the hydrogen breath test is equally sensitive and specific, [^{14}C]lactose is not readily available, and the test has had little application.

For assessment of hepatic function, a large number of substrates have been improved. These include substrates for microsomal demethylating enzymes such as aminopyrine, methacetin, and erythromycin. Amino acids containing a labeled carboxyl group have also been proposed; the carboxyl carbon is rapidly converted to CO_2 by mitochondrial enzymes. One limitation of microsomal substrates is that the subsequent metabolism of the liberated methyl group is complex and may be impaired in liver disease.

Most of these tests detect severe liver disease, but they have not proven to have clinical utility when compared to other tests of hepatic function such as plasma levels of bilirubin or albumin or measurement of prothrombin time. However, a number of these tests are under active investigation and it appears premature to dismiss them as lacking clinical utility.

Xenolytic Labeled Carbon Dioxide Breath Tests

During the past decade, the treatment of peptic ulcer disease has undergone a revolution because of the astonishing finding that the majority of peptic ulcer disease is caused by *Helicobacter pylori* colonization of the gastroduodenal mucosa. Recognition that *H. pylori* had urease activity and that urease is lacking in all tissues led to the introduction of the urea-labeled carbon breath test (urea breath test, UBT). In this test, urea labeled with ^{13}C or ^{14}C is administered orally and breath CO_2 is collected during the first half hour or hour. The UBT was approved for marketing using urea labeled with either ^{13}C or ^{14}C and has been used widely throughout the world. Because of its excellent specificity and sensitivity, it has been used to diagnose the presence of *H. pylori* as well as to test efficacy of therapy.

Several substrates tagged with labeled carbon have been used to detect bacterial overgrowth in the small intestine. The [^{13}C]xylose or [^{14}C]xylose breath test is based on the rapid oxidation of xylose by bacteria compared to minimal oxidation by endogenous enzymes. The test is investigational and has not been widely used in part because of the rarity of the clinical condition as well as the limited sensitivity of the test. Moreover, rapid small intestinal transit results in the xylose being transported to the colon where it is rapidly oxidized by bacteria. A combined hydrogen-labeled carbon dioxide

breath test in which a positive test is signaled by the appearance of labeled CO_2 before a hydrogen peak has been proposed.

The cholyl-[^{13}C]glycine or cholyl-[^{14}C]glycine breath test is based on the ability of intestinal bacteria to deconjugate the substrate and the rapid oxidation of the liberated glycine (by bacterial and/or tissue enzymes) to CO_2. Only the first carbon of glycine is tagged with labeled carbon, as conversion of this atom to CO_2 is greater than and incorporation into tissue proteins is less than that for the second carbon atom. The test is investigational and has not been widely used because of the rarity of the clinical condition and because of the incomplete sensitivity of the test. Incomplete sensitivity for the detection of bacterial overgrowth is explained by overgrowth with bacteria lacking the enzymes to deconjugate cholylglycine.

Clinical Utility of Labeled Carbon Dioxide Breath Tests

Adaptation of labeled carbon dioxide breath tests in clinical practice has been slow for multiple reasons. Only the UBT has been widely used. Breath tests other than the UBT require prolonged collection of breath and the taking of multiple samples. For substrates given orally, delayed gastric emptying may give false-negative results. Intravenous injection solves this problem, but transforms a noninvasive procedure into an invasive procedure. In many disease conditions, there are competing diagnostic procedures. The technical problems of analyzing rapid and accurate measurement of $^{13}CO_2$ or $^{14}CO_2$ in breath have been solved, but the complexity of reproducible substrate delivery and the subsequent oxidation of the released moiety remain. It is possible that breath tests could be valuable in veterinary medicine where their noninvasive aspects make them attractive for diagnostic procedures.

HYDROGEN BREATH TESTS

Overview

In the healthy person, hydrogen is formed by anaerobic bacteria in the colon from dietary carbohydrate that escapes absorption in the small intestine. Some individuals do not form hydrogen, presumably because such individuals have a different colonic flora. Hydrogen breath tests measure the interaction of a substrate that generates hydrogen when metabolized by anaerobic bacteria. Rapid hydrogen production occurs under three conditions: (1) a substrate that is normally efficiently absorbed is malabsorbed and passes into the colon, for example, lactose in lactase-deficient individuals; (2) there is overgrowth of anaerobic bacteria in the small intestine; or (3) there is rapid small intestinal transit of a nonabsorbed substance, e.g., lactulose.

Measurement of Hydrogen in Breath

Expired alveolar air is collected in an impermeable bag or into the intake of a measuring device. Breath hydrogen is currently detected by specific hydrogen detectors. These are not expensive and are accurate and precise. They are based on the use of a metallic membrane that is permeable to hydrogen gas. Breath hydrogen may also be determined accurately by gas chromatography. Because dietary carbohydrate is not completely absorbed, there is always some hydrogen present in expired air. What is recorded is a peak of hydrogen above the baseline level.

Types of Hydrogen Breath Tests and Their Clinical Utility

Hydrogen breath tests are used to diagnose lactase deficiency. Increased breath hydrogen after a lactose load is a specific and rather sensitive test for lactase deficiency. Measurement of breath hydrogen after a glucose or xylose load appears to be a useful test as a screening test for overgrowth of anaerobic bacteria in the small intestine. Finally, hydrogen breath tests are used in research studies to measure small intestinal transit time by observing the time for breath hydrogen to appear after an oral load of a nonabsorbed carbohydrate such as lactulose, even though the lactulose itself causes more rapid small intestinal transit.

See Also the Following Articles

Bacterial Overgrowth • Carbohydrate and Lactose Malabsorption • Gastric Emptying • Halitosis • *Helicobacter pylori* • Small Intestinal Motility

Further Reading

Klein, P. D. (2001). 13C breath tests: Visions and realities. *J. Nutr.* **131**, 1637S–1642S.

Lee, J. S., Camilleri, M., Zinsmeister, A. R., Burton, D. D., Kost, L. J., and Klein, P. D. (2000). A valid, accurate, office based nonradioactive test for gastric emptying of solids. *Gut* **46**, 768–773.

Mion, F., Ecochard, R., Guitton, J., and Ponchon, T. (2001). 13CO2 breath tests: Comparison of isotope ratio mass spectrometry and nondispersive infrared spectrometry. *Gastroenterol. Clin. Biol.* **25**, 375–379.

Riordan, S. M., McIver, C. J., and Duncombe, V. M., *et al.* (1995). Factors influencing the 1-g 14C-xylose breath test for bacterial overgrowth. *Am. J. Gastroenterol.* **90**, 1455–1460.

Riordan, S. M., McIver, C. J., Walker, B. M., *et al.* (1996). The lactulose breath hydrogen test and small intestinal bacterial overgrowth. *Am. J. Gastroenterol.* **91**, 1795–1803.

Budd–Chiari Syndrome

THOMAS W. FAUST* AND JEFFREY A. TUVLIN†

** University of Pennsylvania Health System, Philadelphia and † University of Chicago Hospitals and Clinics*

mesocaval shunt Surgically created shunt promoting flow from the superior mesenteric vein into the inferior vena cava, bypassing an occluded hepatic outflow tract.

side-to-side portacaval shunt Surgically created shunt promoting flow from the portal vein into the inferior vena cava, bypassing an occluded hepatic outflow tract.

transjugular intrahepatic portosystemic shunt Internal jugular vein is used to create communication between the systemic and portal circulations in order to decompress the portal circulation.

venoocclusive disease Occlusion of the terminal hepatic venules caused by nonthrombotic fibroobliterative endophlebitis.

Hepatic venous outflow obstruction may lead to a variety of disorders known as the Budd–Chiari syndrome. In this syndrome, patients present with occlusion of the terminal hepatic venules, major hepatic veins, or the inferior vena cava. Patients may be asymptomatic, or they may present with fulminant hepatic failure or cirrhosis with portal hypertensive complications. Although many options are available to diagnose and treat Budd–Chiari syndrome, patients are typically refractory to medical therapy alone. Interventional radiologic techniques may be appropriate and effective in selected cases. Decompressive surgical shunts are recommended for patients with acute or subacute liver injury, whereas orthotopic liver transplantation is standard care for patients with fulminant hepatic failure or decompensated cirrhosis.

ETIOLOGY

Occlusion of the Inferior Vena Cava and Major Hepatic Veins

Membranous and nonmembranous occlusions of the hepatic veins and inferior vena cava (IVC) are responsible for hepatic outflow obstruction and the Budd–Chiari syndrome. Hematologic disorders and myeloproliferative diseases are the most frequent causes of nonmembranous venous obstruction in the Western world (Table I). Paroxysmal nocturnal hemoglobinuria, antithrombin III deficiency, factor V Leiden mutation, and deficiencies of protein C and protein S are additional causes of Budd–Chiari syndrome.

Solid tumors (primary hepatocellular, renal, adrenal, pulmonary, pancreatic, and gastric carcinomas) and vascular neoplasms arising within the hepatic veins or the vena cava can also produce Budd–Chiari syndrome and hepatic failure. Rare causes of nonmembranous hepatic venous outflow obstruction include infections, collagen vascular diseases, inflammatory bowel disease, and many systemic diseases. Whereas nonmembranous venous occlusion is common in the West, membranous occlusion is more common in the Far East, South Africa, and India (Table I). Membranous obstruction can occur in children, but is more prevalent in adults.

Occlusion of the Terminal Hepatic Venules

Nonthrombotic, fibroobliterative endophlebitis of the terminal hepatic venules within the liver is referred to as venoocclusive disease. As with occlusion of the

TABLE I Etiologies of Budd–Chiari Syndrome

Nonmembranous	Membranous
Myeloproliferative disorders	Webs
Paroxysmal nocturnal hemoglobinuria	Membranes
Antithrombin III deficiency	
Proteins C and S deficiency	
Factor V Leiden mutation	
Antiphospholipid antibodies	
Lupus anticoagulant	
G20210A (factor II gene mutation)	
Neoplasms	
Infections	
Collagen vascular diseases	
Behcet's disease	
Sarcoidosis	
Oral contraceptives	
Pregnancy	
Inflammatory bowel disease	
Cirrhosis	
Polycystic liver disease	
Idiopathic	

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Etiologies of Venoocclusive Disease

Pyrrolizidine alkaloids	Cirrhosis
Medications and toxins	Cryptogenic
Cytosine arabinoside	Alcoholic
6-Thioguanine	Hepatitis B
Cyclophosphamide	Systemic lupus erythematosus
Nitrosoureas	Allergic granulomatous arteritis
Busulfan	Thorotrast
Vitamin A	Oral contraceptives
Arsenic	Insecticides
Total body irradiation	

major hepatic veins and the IVC, occlusion of these smaller veins may lead to the Budd–Chiari syndrome. Multiple etiologies have been reported (Table II). Patients at risk for developing venoocclusive lesions include recipients of bone marrow and renal allografts. Other conditions associated with small vessel occlusion include cryptogenic cirrhosis or cirrhosis secondary to alcohol or hepatitis B. Patients with alcoholic hepatitis may also exhibit focal obliteration of intrahepatic veins. Vitamin A toxicity, arsenic poisoning, insecticide exposure, 6-thioguanine, intraarterial 5-fluoro-2′-deoxyuridine, oral contraceptives, and thorotrast have rarely been associated with venoocclusive disease.

PATHOLOGY

Pathologically, there are distinct differences between acute and chronic hepatic venous outflow obstruction. Grossly, acute outflow obstruction is associated with an enlarged, smooth, red–purple liver. Histologically, centrilobular congestion and sinusoidal dilatation are appreciated. Atrophy, necrosis, and centrizonal hepatocyte dropout with extension to periportal regions are associated with severe injury (Fig. 1). In chronic Budd–Chiari syndrome, the inferior vena cava receives direct outflow from the caudate lobe of the liver, which compensates for venous outflow obstruction of the major hepatic veins. The result is caudate lobe hypertrophy with atrophy and cirrhosis of the remaining hepatic segments. Histologically, there is complete obliteration of the central veins associated with midzonal and centrilobular fibrosis with or without cirrhosis.

CLINICAL PRESENTATION

The clinical presentation of patients with Budd–Chiari syndrome is highly variable, depending on the extent and rate of outflow obstruction. Patients may present in many ways, ranging from an asymptomatic state to fulminant hepatic failure or cirrhosis with portal hypertensive complications. Hepatomegaly and ascites are present in greater than 85% of patients, whereas esophagogastric varices, splenomegaly, and prominent collaterals are less frequently seen. Right upper quadrant pain, nausea, vomiting, hepatomegaly, and ascites are characteristic of acute obstruction. Jaundice and splenomegaly can be present but are usually mild. Most patients present with a subacute course (less than 6 months) characterized by vague right upper quadrant discomfort, hepatomegaly, mild to moderate ascites, and splenomegaly; jaundice may be either absent or mild. Chronic venous outflow obstruction is likely if symptoms have been present for greater than 6 months and are associated with fatigue, bleeding varices, encephalopathy, coagulopathy, hepatorenal syndrome, and/or malnutrition. Fulminant hepatic failure is a

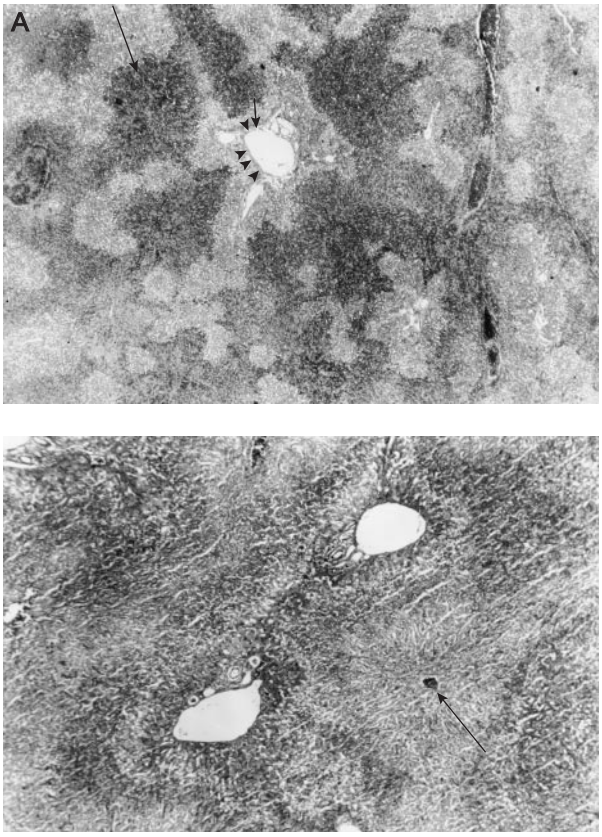


FIGURE 1 Acute Budd–Chiari syndrome. (A) Low-power view of centrilobular congestion, atrophy, necrosis, and dropout of centrizonal hepatocytes (arrow) with periportal sparing (arrowheads) (hematoxylin and eosin stain, ×40). (B) Higher magnification of acute injury, with thrombus identified within the hepatic vein (arrow) (hematoxylin and eosin stain, ×100). Courtesy of John Hart, University of Chicago Hospitals, Chicago, Illinois.

rare manifestation of Budd–Chiari syndrome and is typically the result of complete occlusion of all major hepatic veins. Without treatment, progressive encephalopathy, coagulopathy, and death are inevitable within 8 weeks of venous outflow obstruction.

DIAGNOSTIC EVALUATION

Laboratory Investigation

Standard laboratory investigation is rarely useful in the assessment of patients with suspected Budd–Chiari syndrome. Aminotransferases are typically normal or mildly abnormal; however, elevated enzyme values are not specific for venous outflow obstruction. Likewise, the serum bilirubin, alkaline phosphatase, and prothrombin time can be either normal or mildly abnormal but are not useful in diagnosing Budd–Chiari syndrome. Ascitic fluid analysis adds little specificity to the diagnosis, although it may be useful in ruling out other causes of abdominal fluid accumulation. The workup for Budd–Chiari syndrome cannot be considered complete without a comprehensive workup for an underlying hypercoagulable state.

Medical Imaging

Abdominal ultrasound is the best screening test for the evaluation of the hepatic veins, vena cava, and portal vein. In experienced hands, the sensitivity of ultrasound approaches 85–95%. Enlarged, stenotic, and tortuous hepatic veins identify acute venous occlusion, whereas the veins of patients with chronic disease may be more difficult to visualize. Ultrasound may reveal caval compression by a hypertrophic caudate lobe, or obstruction by thrombus, tumor, or membranes. Hepatic venous-to-venous "spiderweb" collaterals are highly suggestive of Budd–Chiari syndrome. Doppler technology added to conventional ultrasound greatly increases sensitivity.

Computer tomography (CT) and magnetic resonance imaging (MRI) may reveal nonvisualization of vessels or obstruction by thrombus. In addition, these studies readily detect hepatic parenchymal disease, ascites, and splenomegaly.

Angiography remains the gold standard for the diagnosis of Budd–Chiari syndrome. Thrombi commonly form at the junction of the major hepatic veins with the cava or just distal to the venous orifices. Injection of contrast medium classically reveals intrahepatic collateral vessels or recanalized veins, giving the classic spiderweb appearance. Pressure measurements obtained during angiography provide useful information prior to surgical decompression.

Liver Biopsy

Liver biopsy is useful in establishing the diagnosis of Budd–Chiari syndrome and in evaluating the severity of the disease. Based on the severity, physicians can select the most appropriate therapy for individual patients. Angiography and liver biopsy should be obtained from all patients considered to be surgical candidates. Biopsies should be bilobar, because injury may be variable from one lobe to the other.

MANAGEMENT

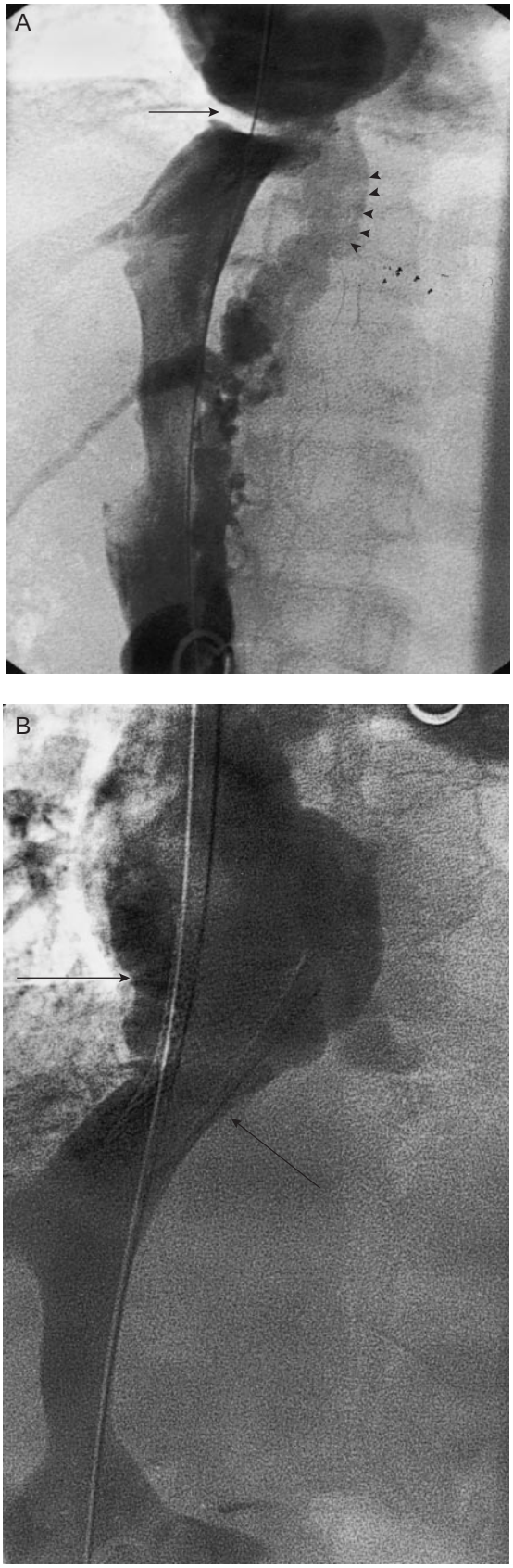
The goals of treatment for patients with Budd–Chiari syndrome include alleviating venous obstruction and preserving hepatic function through eradication of centrilobular congestion. Despite medical therapy and interventional radiologic techniques, the majority of patients will require surgical decompression.

Medical Therapy

Although frequently used in patients with Budd–Chiari syndrome, low-sodium diets, diuretics, and therapeutic paracentesis are generally ineffective because these methods do little to reverse the underlying pathophysiology. The use of anticoagulants and thrombolytics for patients with acute incomplete thrombotic occlusion must be balanced against the typically short window of opportunity to decompress the liver surgically.

Interventional Radiology

Percutaneous transluminal balloon angioplasty is an exciting therapy for hepatic venous outflow obstruction secondary to caval webs or hepatic venous stenoses. Although excellent short-term results are achievable, sustained patency rates have been disappointing. Wire, laser, or needle-assisted angioplasty and intraluminal stenting are emerging techniques (Fig. 2). Transjugular intrahepatic portosystemic shunts (TIPSs) are frequently used as a bridge to transplantation for patients with end-stage liver disease and ascites refractory to medical therapy, or for patients presenting with fulminant hepatic failure. Because patients can develop TIPS thrombosis or pseudointimal hyperplasia after placement, surveillance of TIPS patency with Doppler ultrasonography and/or angiography is recommended. Additional studies will be required before the TIPS can be routinely recommended as treatment for patients with either acute or chronic outflow obstruction.



Surgical Procedures

Surgical options for patients with Budd–Chiari syndrome include decompressive procedures and orthotopic liver transplantation (OLT). A variety of surgical shunts are available for decompression to relieve centrilobular congestion and necrosis. Decompression should be considered the standard of care for patients with acute or subacute venous occlusion. After surgery, long-term anticoagulation is recommended. Prior to surgery, liver biopsy and pressure measurements of the portal vein, IVC, and right atrium should be obtained to determine the best method of decompression. The procedure of choice for patients with either acute or subacute Budd–Chiari syndrome is the creation of a side-to-side portacaval shunt, which converts the portal vein into an effective outflow tract into the IVC. Greater than 85% of patients who receive portacaval shunts enjoy long-term survival, with improvement in hepatosplenomegaly, ascites, liver function tests, and histology; however, decompensation may be observed in patients with marginal hepatic reserve. Mesocaval shunts provide effective portal decompression for patients unable to receive side-to-side shunts. With successfully placed mesocaval shunts, long-term survival is possible for 75–95% of patients. Unfortunately, shunt thrombosis may develop in 20–55% of patients who receive mesocaval shunts. Depending on the venous anatomy, other decompressive shunts may need to be considered. Other surgical techniques can also provide adequate decompression in appropriately selected patients. Surgical reconstruction of the vena cava and hepatic venous ostia, transatrial membranotomy with finger fracture or excision, cavoplasty with autologous pericardial patch, and dorsocranial resection of the liver with hepatoatrial anastamosis have shown efficacy in limited studies.

OLT is the preferred option for patients with Budd–Chiari syndrome and fulminant liver failure or decompensated cirrhosis. Transplantation may also be appropriate for patients with significant liver disease who decompensate after receiving decompressive shunts or for patients with shunt failure. In patients without significant preoperative hemodynamic instability or multiorgan failure, survival after transplantation

FIGURE 2 (A) Angiographic view of an occluded inferior vena cava at its junction with the right atrium. There is a stenosis due to thrombus (arrow), with subsequent filling of the lumbar venous plexus (arrowheads). (B) Postinterventional dilatation revealing good flow of contrast into the right atrium (arrow) and lack of filling of the lumbar venous plexus. Courtesy of Brian Funaki, University of Chicago Hospitals, Chicago, Illinois.

is similar to that of decompressive shunts. Survival of patients transplanted for Budd–Chiari syndrome is also comparable to that of patients receiving allografts for other noncholestatic liver diseases, and is better than that of patients receiving transplants for malignancy or chronic hepatitis B. Selected patients, such as those with underlying hypercoagulable states not cured with OLT, require long-term anticoagulation.

SUMMARY

Budd–Chiari syndrome is an uncommon disorder associated with obstruction of the terminal hepatic venules, hepatic veins, IVC, and/or right atrium. Sinusoidal congestion, centrilobular necrosis, hepatic fibrosis, and cirrhosis can develop as a consequence of hepatic venous outflow obstruction. Patients typically present with symptoms and signs of subacute disease or cirrhosis, although an occasional patient may present without any evidence of significant liver injury. Routine laboratory and ascitic fluid analyses are rarely helpful in establishing the diagnosis. Doppler sonography is an excellent screening test, and CT and MRI complement sonography as noninvasive studies. Angiography remains the best technique for the diagnosis of venous occlusion. Pressure measurements at the time of angiography and bilobar liver biopsy are also helpful in guiding the clinician to the most appropriate therapy.

Low-sodium diets, diuretics, and paracentesis are of limited benefit in patients with Budd–Chiari syndrome. Thrombolytic therapy may be useful for early disease associated with fresh thrombosis. Selected patients may also benefit from interventional radiologic techniques such as percutaneous angioplasty with or without stenting. The TIPS can serve as a bridge to transplantation, but additional studies are warranted to assess the impact of TIPS placement on long-term survival. Surgical decompression is appropriate for patients with acute or subacute venous occlusion; alleviation of sinusoidal congestion and hepatic necrosis is possible. OLT is reserved for patients with Budd–Chiari syndrome associated with fulminant hepatic failure, decompensated cirrhosis, or failed surgical shunts. With careful patient selection and improvements in surgical technique, perioperative care, and immunosuppression, long-term patient survival after OLT is expected.

See Also the Following Articles

Cirrhosis • Fulminant Hepatic Failure • Hepatic Circulation • Liver Biopsy • Liver Transplantation • Portal Hypertension and Esophageal Varices • Ultrasonography

Further Reading

Denniger, M. H., Chait, Y., Casadevall, N., *et al.* (2000). Cause of portal or hepatic venous thrombosis in adults: The role of multiple concurrent factors. *Hepatology* **31**, 587–591.

Dilawari, J. B., Bambery, P., Chawla, Y., *et al.* (1994). Hepatic outflow obstruction (Budd–Chiari syndrome): Experience with 177 patients and review of the literature. *Medicine* **73**, 21–36.

Emre, A., Kalayci, G., Ozden, I., *et al.* (2000). Mesoatrial shunt in Budd–Chiari syndrome. *Am. J. Surg.* **179**, 304–308.

Ganger, D. R., Klapman, J. B., McDonald, V., *et al.* (1999). Transjugular intrahepatic portosystemic shunt (TIPS) for Budd–Chiari syndrome and portal vein thrombosis. *Am. J. Gastroenterol.* **94**, 603–608.

Ganguli, S. C., Ramzan, N. N., McKusick, M. A., *et al.* (1998). Budd–Chiari syndrome in patients with hematological disease: A therapeutic challenge. *Hepatology* **27**, 1157–1161.

Hemming, A. W., Langer, B., Greig, P., Taylor, B. R., Adams, R., and Heathcote, J. (1996). Treatment of Budd–Chiari syndrome with portosystemic shunt or liver transplantation. *Am. J. Surg.* **171**, 176–181.

Henderson, J. M., Warren, D., Millikan, W. J., *et al.* (1990). Surgical options, hematologic evaluation, and pathologic changes in Budd–Chiari syndrome. *Am. J. Surg.* **159**, 41–50.

Klein, A. S., and Cameron, J. L. (1990). Diagnosis and management of the Budd–Chiari syndrome. *Am. J. Surg.* **160**, 128–133.

Klein, A. S., Sitzman, J. V., Coleman, J., Herlong, F. H., and Cameron, J. L. (1990). Current management of the Budd–Chiari syndrome. *Ann. Surg.* **212**, 144–149.

Langnas, A. N., and Sorrell, M. F. (1993). The Budd–Chiari syndrome: A therapeutic gordian knot? *Semin. Liver Dis.* **13**, 352–359.

Mahmoud, A. E., Mendoza, A., Meshikhes, A. N., *et al.* (1996). Clinical spectrum, investigations and treatment of Budd–Chiari syndrome. *Q. J. Med.* **89**, 37–43.

Millikan, W. J., Henderson, J. M., Sewell, C. W., *et al.* (1985). Approach to the spectrum of Budd–Chiari syndrome: Which patients require portal decompression? *Am. J. Surg.* **149**, 167–176.

Mitchell, M. C., Boitnott, J. K., Kaufman, S., Cameron, J. L., and Maddrey, W. C. (1982). Budd–Chiari syndrome: Etiology, diagnosis and management. *Medicine* **61**, 199–218.

Olzinski, A. T., and Sanyal, A. J. (2000). Treating Budd–Chiari syndrome: Making rational choices from a myriad of options. *J. Clin. Gastroenterol.* **30**(2), 155–161.

Orloff, M. J., Daily, P. O., Orloff, S. L., *et al.* (2000). A 27-year experience with surgical treatment of Budd–Chiari syndrome. *Ann. Surg.* **232**, 340–352.

Pisanni-Ceretti, A., Indra, M., Prestipino, F., *et al.* (1998). Surgical and radiologic treatment of primary Budd–Chiari syndrome. *World J. Surg.* **22**, 48–54.

Ringe, B., Lange, H., Oldhafer, K., *et al.* (1995). Which is the best surgery for Budd–Chiari syndrome: Venous decompression or liver transplantation?—Single center experience with 50 patients. *Hepatology* **21**, 1337–1344.

Tilanus, H. W. (1995). Budd–Chiari syndrome. *Br. Surg.* **82**, 1023–1030.

Wang, Z. G., and Jones, R. S. (1996). Budd–Chiari syndrome. *Curr. Probl. Surg.* **33**, 83–211.

Bulimia Nervosa

JULIE E. B. NOLAN AND THOMAS D. GERACIOTI, JR.
University of Cincinnati and Veterans Affairs Medical Center

arrhythmia Abnormal rhythm of the heartbeat.
binge (binge eating) Consumption in a discrete time frame (usually $\leq$2 h) of an amount of food that is definitely excessive in comparison with the amount of food consumed by most others in similar circumstances. The consumption of large quantities of food at certain celebratory or holiday feasts, for example, is not usually considered to constitute pathological bingeing. A single binge-eating episode may take place at more than one location—such as following dinner at a friend's with eating at home—but day-long snacking is not a binge.
hyperamylasemia Abnormally high concentrations of the digestive enzyme amylase in the blood.
hypokalemia Abnormally low potassium concentrations in the blood.
osteoporosis Diminished mineralization of bone.
ostopenia Reduced bone formation.
parotid gland hypertrophy Enlargement of the salivary glands.
postprandial satiety Sense of satiation or fullness after eating.

The possibility of eating large quantities of food in one feeding episode is contained in the human behavioral repertoire, having developed under prehistorical evolutionary pressure to take advantage of transiently available food during periods of overall food scarcity. In recent decades, binge eating followed by vomiting—or other methods of purging—has been used as a means of body weight control, wherein the gustatory, self-soothing, and orally gratifying aspects of eating can be repeatedly obtained without an increase in body mass. If used routinely by vulnerable individuals, a pattern of binge eating followed by vomiting can become entrained and, in a sense, addictive.

CORE SYMPTOMS OF BULIMIA NERVOSA

The modern syndrome of bulimia nervosa was first described in London by Russell in 1979, in a 30-patient case series, as an eating disorder characterized by recurrent, uncontrolled binge-eating episodes in combination with compensatory behavior aimed at preventing weight gain. Individuals with bulimia nervosa commonly ingest up to 10,000 calories per day. The most common postbinge compensatory behaviors in individuals with bulimia nervosa are self-induced vomiting and laxative, diuretic, or enema misuse. Excessive exercise, fasting, and misuse of medications—including appetite suppressants (for example, amphetamines), cathartics (for example, syrup of ipecac), and metabolism-stimulating agents (for example, thyroid hormone)—are also commonly manifested. A formal diagnosis of bulimia nervosa, according to the American Psychiatric Association, requires that both binge eating and purging (or other compensatory behaviors) occur at least twice per week for at least 3 months (see Table I).

Excessive concern about or preoccupation with body shape, size, or weight is also a hallmark of bulimia nervosa. These individuals are focused upon maintaining or achieving a desirable body weight and shape. Therefore, pathological binge eating that exists in the absence of intense body consciousness (and in the absence of attempts to control body weight or shape) is not bulimia nervosa. Although a substantial minority of individuals with bulimia nervosa have histories of anorexia nervosa, most people with bulimia nervosa have a body weight within the normal range.

Patients with bulimia nervosa show an awareness of having an eating disorder, in contradistinction to patients with anorexia nervosa, in whom denial of an eating problem is intrinsic to the illness. However, individuals with bulimia nervosa are frequently ashamed about their eating pattern and usually—but certainly not always—maintain strict secrecy about it. Since the majority of people with bulimia nervosa are within a normal weight range, or are only slightly under- or overweight, it is difficult to identify someone with bulimia nervosa by sight (whereas, of course, the wasted, skeletal appearance of individuals with anorexia nervosa is quite obvious). Thus, individuals with bulimia nervosa are not frequently recognized by others to have an eating disorder.

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I DSM-IV Criteria for Bulimia Nervosa

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   1. Eating, in a discrete period of time (e.g., within any 2-h period), an amount of food that is definitely larger than most people would eat during a similar period of time and in similar circumstances.
   2. A sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating).

B. Recurrent inappropriate compensatory behavior in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise.

C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for 3 months.

D. Self-evaluation is unduly influenced by body shape and weight.

E. The disturbance does not occur exclusively during episodes of anorexia nervosa.

Purging type

During the current episode of bulimia nervosa, the person has regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

Nonpurging type

During the current episode of bulimia nervosa, the person has used other inappropriate compensatory behaviors, such as fasting or excessive exercise, but has not regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

BULIMIA NERVOSA SUBTYPES

Two subtypes of bulimia nervosa have been characterized, the purging type and the nonpurging type. Individuals with bulimia nervosa who purposefully vomit after binging or who misuse laxatives or enemas to expel undigested or incompletely digested food from the gastrointestinal tract have the purging type of bulimia nervosa. Misuse of diuretics is also regarded as a purging behavior. Those individuals with bulimia nervosa who do not purge themselves of undigested food or create a diuresis after binge eating have the nonpurging type of the disorder. Those individuals with the nonpurging type nevertheless engage in behaviors aimed at eliminating the potentially adverse effects of binge eating on body weight and shape by excessive exercise and/or fasting. In many cases, women with bulimia nervosa are intense, obsessive exercisers or "obligatory athletes."

Purging is present in approximately 90% of patients with bulimia nervosa, with forced vomiting present in the overwhelming majority of these individuals. As noted by Russell is his original description of the syndrome in 1979, "with repeated practice, the act of vomiting becomes effortless." Some individuals with bulimia nervosa do not have to stimulate the gag reflex at all in order to vomit; forceful contraction of the abdominal musculature suffices. Laxative abuse is seen in more than half of patients.

VARIANTS OF BULIMIA NERVOSA

In clinical practice, both "subclinical" (low-grade) forms of bulimia nervosa and variants of bulimia nervosa are seen about as frequently as the bulimia nervosa syndrome per se. These variants include, but are not limited to, those that (1) involve binge eating fewer than two times per week; (2) involve binge eating without compensatory, weight-control behavior (so-called binge eating disorder, see below); or (3) involve purging, such as self-induced vomiting, without binge eating. For example, the individual who engages in binge eating followed by self-induced vomiting an average of once per week for months or years has an eating disorder that does not meet the full diagnostic criteria for bulimia nervosa due to a lower frequency of bingeing. The normal-weight or slightly underweight woman who shows overconcern about body weight, intense interest in food, frequently diets, and routinely chews food and spits it out before swallowing it does not meet diagnostic criteria for bulimia nervosa due to the lack of binge eating. Other individuals, again mostly women, induce vomiting after intake of regular or modest amounts of food as a diet strategy; they do not binge eat. In these cases, perhaps one or more meals are digested but most other bouts of eating end in self-induced emesis. These women are at an increased risk of eventually developing full-blown bulimia nervosa. If one of these individuals does not show a compensatory increase in eating, fasts, or vomits after every meal, then she moves on the continuum of disordered eating toward anorexia nervosa, bulimic subtype. The unwieldy, catch-all diagnosis of "eating disorder, not otherwise specified" has been adopted by the American Psychiatric Association to capture these and other eating disorder variants.

Binge eating disorder is a common condition (2–5% in community studies in the United States and 30% among obese patients seeking weight loss) that is

TABLE II DSM-IV Criteria for Binge-Eating Disorder

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   1. Eating, in a discrete period of time (e.g., within any 2-h period), an amount of food that is definitely larger than most people would eat in a similar period of time under similar circumstances.
   2. A sense of lack of control over eating during the episodes (e.g., a feeling that one cannot stop eating or control what or how much one is eating).

B. The binge-eating episodes are associated with three (or more) of the following:
   1. eating much more rapidly than normal;
   2. eating until feeling uncomfortably full;
   3. eating large amounts of food when not feeling physically hungry;
   4. eating alone because of being embarrassed by how much one is eating;
   5. feeling disgusted with oneself, depressed, or very guilty after overeating.

C. Marked distress regarding binge eating is present.

D. The binge eating occurs, on average, at least 2 days a week for 6 months.
   Note: The method of determining frequency differs from that used for bulimia nervosa; future research should address whether the preferred method of setting a frequency threshold is counting the number of days on which binges occur or counting the number of episodes of binge eating.

E. The binge eating is not associated with the regular use of inappropriate compensatory behaviors (e.g., purging, fasting, or excessive exercise) and does not occur exclusively during the course of anorexia nervosa or, bulimia nervosa.

characterized by chronic binge eating without compensatory weight-restricting behavior. Given the lack of weight-controlling compensatory behaviors, individuals with binge eating disorder are frequently obese. Binge eating disorder is seen in both men and women across the life span. Although an accepted diagnosis by practicing clinicians, the most recent Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-IV), published in 1994, proposed binge eating disorder as a new diagnostic entity about which there was insufficient information to include as an official category. It is likely that the diagnosis will be officially recognized by the American Psychiatric Association with the next DSM edition. See Table II for the complete 1994 DSM criteria.

Bulimia nervosa and anorexia nervosa can occur simultaneously, such as constitutes the binge-eating/purging subtype of anorexia nervosa, commonly called bulimarexia. Bulimia nervosa and anorexia nervosa can also occur alternately within the same individual over time.

Yet another eating disorder in which hyperphagia is an essential component is night eating syndrome, described by Stunkard and colleagues in 1955 and consisting of morning anorexia and evening or nighttime hyperphagia with agitation and insomnia. At a minimum, this syndrome appears to reflect phase-delayed circadian rhythms.

EPIDEMIOLOGY

Bulimia nervosa is generally a syndrome seen in young women, afflicting approximately 1–4% of college-aged women in community samples, although it is also seen in older women (mostly with chronic or chronic, intermittent bulimia nervosa that had its onset in youth, but occasionally as a new onset syndrome) and in men. Indeed, the overwhelming majority of people with bulimia nervosa, over 90% in clinical samples, are female. This gender distribution is often regarded to be a result of a societal or cultural premium on feminine beauty that is transmitted to the young woman through either conscious or unconscious channels (most conspicuously via mass media). That eating disorders are more common in industrialized nations and in affluent individuals is consistent with the notion of environmental influences on the development of bulimia nervosa. That white women much more frequently develop bulimia nervosa and anorexia nervosa than black women is also taken to reflect the relatively higher value placed on thinness in the former culture. Black and white women are at roughly equal risk for binge eating disorder—that is, unrestrained eating—but white women are more likely to show compensatory fasting or purging behavior and concerns about body weight and shape. However, if a female—regardless of race—internalizes the view that slender beauty is ideal, body dissatisfaction, severe fasting, and even subsequent binge eating and full-blown bulimia nervosa may develop.

Impaired satiety, secretive eating, binge eating, postprandial vomiting, and other symptoms of abnormal eating are also surprisingly common in preschool children, although much more work is needed to understand and characterize these phenomena and their frequency.

RISK FACTORS, COMORBIDITIES, AND ASSOCIATED FEATURES

Dieting, especially severe and/or repeated dieting, is a major risk factor for development of bulimia nervosa and other eating disorders. Although most individuals who diet do not develop an eating disorder, even moderate dieting by adolescent girls increases the risk of later developing an eating disorder by several-fold. Many bulimic patients have had histories of anorexia nervosa.

Genetic factors are increasingly appreciated as a factor in the development of eating disorders. Studies of the differences in rates of bulimia nervosa between dizygotic and monozygotic twins suggest that approximately 50% of the risk of developing bulimia nervosa is genetic. However, the interplay between inherited vulnerabilities and intrafamilial relationships is hard to definitively tease out. Maternal body dissatisfaction increases the likelihood that an eating disorder will develop in her progeny. The majority of patients with bulimia nervosa have a first-degree relative with a mood disorder, substance abuse disorder (most often alcoholism), or an eating disorder. Indeed, the majority of patients who present for treatment of bulimia nervosa have suffered from another psychiatric syndrome either prior to the onset of the eating disorder or currently with the eating disorder. Most commonly, a mood (affective) disorder, such as major depression, atypical depression, or bipolar affective disorder (manic depression and its low-grade variants), is seen in patients with bulimia nervosa, but a range of other psychiatric syndromes, including anxiety disorders (such as posttraumatic stress disorder, panic disorder, and obsessive–compulsive disorder), substance abuse (especially of tobacco and alcohol), and personality disorders, are also seen.

COMPLICATIONS, PHYSICAL STIGMATA, AND MORBIDITY

Despite the normal or even vigorously healthy appearance of most individuals with bulimia nervosa, physical stigmata do exist as a result of complications from the chronically disordered eating pattern (see Table III). Calluses on the dorsum of the hand or knuckles (Russel's sign) are often seen. These calluses eventually develop after the repeated thrusting of the fingers into the throat in order to induce vomiting, due to the pressure of the teeth on the skin. Loss of enamel from the teeth—even a moth-eaten appearance of the dentition—can be seen following years of very frequent contact with acidic vomitus. Bilateral parotid (salivary) gland hypertrophy (swollen cheeks) can develop from the hyperstimulation of excessive eating and frequent vomiting and confers a characteristic "chipmunk" facies. The breath may be foul from recent or frequent passage of partly digested food during emesis. Rectal prolapse, most likely from forced vomiting-related increases in intra-abdominal pressure or from straining during defecation, has been reported in a number of patients, invariably preceded by complaints of rectal bleeding.

TABLE III Potential Complications of Bulimia Nervosa

Dental
Erosion of dental enamel
Chipped and/or ragged teeth
Frequent caries
Cardiac arrhythmias
Cardiac myopathies
Calluses or scars over knuckles (Russell's sign)
Constipation/delayed gut transit time
Esophageal reflux, spasms, rarely tears
Foul breath odor
Gastric distension (rarely, gastric perforation, rupture)
Hypokalemia
Hyperamylasemia
Hypothermia
Lowered seizure threshold
Osteopenia
Osteoporosis
Parotid gland hypertrophy ("chipmunk" appearance)
Rectal bleeding (rarely, rectal prolapse)

An important, frequently encountered metabolic complication is hypokalemia due to frequent vomiting and/or frequent use of laxatives or enemas; cardiac arrhythmias sometimes follow. Surprisingly, some patients show substantial hypokalemia without cardiac arrhythmia or other apparent adverse effects, perhaps due to long-term organismic adaptations to the situation. For example, a young woman with long-standing bulimia nervosa had a potassium level of 2.1 mEq/L but still performed with great success in a demanding exercise competition. Other patients with similarly low potassium levels have irregular heartbeats and require emergency cardiac monitoring and electrolyte normalization.

Use of syrup of ipecac to induce vomiting can weaken or damage the cardiac myocardium. Circulating pancreas-derived amylase concentrations increase following a binge–purge episode and, clinically, elevated blood amylase levels are a reliable indicator of such behavior.

Despite the numerous complications that can develop (see Table III for a partial list), the morbidity

associated with bulimia nervosa is surprisingly low—certainly much lower than was feared when the first 30 cases were described by Russell in 1979 (although some of those cases would now be classified as "eating disorders not otherwise specified"). As opposed to the situation in anorexia nervosa, wherein growth, genital sexuality, and fertility essentially cease or fail to develop, these parameters are often quite adequate in the person with bulimia nervosa. Moreover, unlike the case with anorexia nervosa, death from a complication of bulimia nervosa is very rare. Gross mortality rates in patients with bulimia nervosa have been found to be at least 0.3%, including deaths from traffic accidents, suicide, malnutrition, and drug interactions. The crude mortality rate in anorexic patients is perhaps 20 times higher.

PATHOPHYSIOLOGY

Patients with bulimia nervosa almost always show prominent impairments in their experience of postprandial satiety. Although the degree of hunger experienced by patients varies—with some showing great ability to suppress hunger and fast between binge eating episodes—once the ingestion of food occurs they are at high risk for uncontrollable food intake. In part, the inability or impaired ability to experience postprandial satiety reflects a general impairment in the communication of peripherally generated satiety signals to the brain. Thus, most patients with bulimia nervosa have impaired postprandial secretion of the satiety-inducing hormone cholecystokinin (CCK) from the proximal intestine. Normally, postprandial CCK-mediated signals promote closure of the gastric pylorus (increasing the feeling of fullness) and travel via the vagus nerve to brainstem and hypothalamic feeding centers that mediate the satiety response. Abnormalities in this signaling system could be genetic and/or acquired (perhaps as a result of frequent binge eating). Part of the mechanism of antidepressant drug treatment of bulimia nervosa involves increasing the postprandial elaboration of CCK.

Abnormally low integrated function of the indoleamine serotonin has also been identified in patients with bulimia nervosa. In this regard, serotonin in intestinal enterochromaffin cells has been found to be necessary for the contractile effect of CCK. Serotonin also appears to have a satiety-inducing effect in the central nervous system. Low brain serotonin function also underlies obsessive–compulsive disorder, elements of which are routinely encountered in patients with bulimia nervosa, and other eating disorders, in whom seemingly nonsensical behaviors are irresistible (such as the eating of 5000 calories within an hour in a person who is trying to remain trim). High doses of pharmacologic agents that increase the whole-body availability of serotonin are efficacious in both bulimia nervosa and obsessive–compulsive disorder.

If patients with bulimia nervosa maintain a normal weight, the hypothalamic–pituitary–thyroid, hypothalamic–pituitary–adrenal, and hypothalamic–pituitary–gonadal neuroendocrine axes remain surprisingly normal.

NATURAL HISTORY AND TREATMENT

The natural history of bulimia nervosa is highly variable, with perhaps 50% undergoing spontaneous remission without treatment after an active, syndromic period of months to years. Other individuals—whether eaters of normal quantities of food, overeaters, or binge eaters—immediately give up on self-induced emesis as a method of weight control after attempting it. However, a substantial minority of those who develop bulimia nervosa, up to 20%, go on to chronically demonstrate the full-blown syndrome at 10-year follow-up, whereas others continue to show subsyndromal elements over the long term. Treatment studies show significantly higher rates of short-term remission than naturalistic follow-up studies, but the long-term benefits of treatment interventions have not been established. Approximately 30% of well-improved patients eventually experience a bulimic relapse. Thus, although bulimia nervosa is generally much more benign than anorexia nervosa, some individuals suffer from chronic bulimia nervosa, lasting even decades. Little is known regarding the risks for developing chronic bulimia nervosa, although the comorbid presence of severe borderline personality disorder is a negative prognostic sign.

The standard approach to the treatment of bulimic patients is a combination of antidepressant medication and psychotherapy. Usually, treatment is instituted and accomplished on an outpatient basis, as opposed to the situation in anorexia nervosa or bulimarexia where inpatient correction of malnutrition and severe wasting is usually necessary. Brief inpatient admissions are reserved for the emergency control of severe electrolyte imbalances and other severe bulimia-related complications (including severe depression and acute suicidality).

A variety of antidepressants, including selective serotonin reuptake inhibitors (usually in high

doses), monoamine oxidase inhibitors, and tricyclic antidepressants, are successful in ameliorating or reducing bulimic behavior in the majority of patients. However, antidepressant treatment alone is inadequate in most cases; complementary intensive psychotherapy is also indicated.

The most well-studied psychotherapy for bulimia nervosa is cognitive-behavioral psychotherapy, with results suggesting that this intervention is efficacious. Cognitive therapy focuses upon the morbid fears of obesity and worries about body size and shape, with cognitive and behavioral exercises aimed at correcting cognitive distortions and low self-esteem and at reducing perfectionism and the hyperfocus on the body. However, intensive, exploratory, psychodynamic psychotherapy with an experienced clinician—in combination with antidepressant medication—is also helpful in most cases and has the advantage of not focusing strictly on eating disorder symptoms (and other superficial, behavioral manifestations of the psyche), but also provides the context for the uncovering of occult or unconscious psychological problems or conflicts.

Psychologically, patients with bulimia nervosa often, but not always, show substantial impulsivity and a great deal of interpersonal neediness and object hunger (hunger for caring and empathic relatedness with others), for which oral gratification often substitutes. Analogous impairments to feeding-related satiety are often seen clinically in the form of seemingly insatiable desire for close contact or merger with another person. The stable presence of a caring, empathically attuned psychotherapist is in itself highly therapeutic for such patients. Nevertheless, long-established binge–purge (or binge–fasting) cycles can become deeply and organismically entrained, and pointed interventions, such as antidepressant medications and nutritional oversight, aimed at normalizing the appetitive and satiety signal processing mechanisms (optimally together with exploratory psychotherapy) are sometimes therapeutically necessary.

SUMMARY

In fewer than 25 years, major strides have occurred in the understanding of the clinical course, pathophysiology, psychopathology, and treatment of bulimia nervosa. Currently, the efficacy of both psychotherapy and psychopharmacologic interventions in the treatment of this common eating disorder has been well established. Ideally, these treatment modalities should be combined.

See Also the Following Articles

Anorexia Nervosa • Appetite • Emesis • Nausea • Nutritional Assessment

Further Reading

Bulik, C. M., Sullivan, P. F., Wade, T. D., and Kendler, K. S. (2000). Twin studies of eating disorders: A review. *Int. J. Eating Dis.* **27**, 1–20.

Geracioti, T. D., Jr., and Liddle, R. A. (1988). Impaired cholecystokinin secretion in bulimia nervosa. *N. Engl. J. Med.* **319**, 683–688.

Heatherington, M. M. (2000). Eating disorders: Diagnosis, etiology, and prevention. *Nutrition* **7/8**, 547–551.

Keel, P. K., and Mitchell, J. E. (1997). Outcome in bulimia nervosa. *Am. J. Psychiatry* **154**, 313–321.

Mitchell, J. E., Peterson, C. B., Myers, T., and Wonderlich, S. (2001). Combining pharmacotherapy and psychotherapy in the treatment of patients with eating disorders. *Psychiatric Clin. North Am.* **24**, 315–323.

Russell, G. (1979). Bulimia nervosa: An ominous variant of anorexia nervosa. *Psychol. Med.* **9**, 429–448.

Stunkard, A. (1997). Eating disorders: The last 25 years. *Appetite* **29**, 181–190.

Stunkard, A. J., Grace, W. J., and Wolf, H. G. (1955). The night-eating syndrome: A pattern of food intake among certain obese patients. *Am. J. Med.* **19**, 78–86.

Yager, J., *et al.* (2000). American Psychiatric Association Practice Guidelines: Practice Guidelines for the Treatment of Patients with Eating Disorders (Revision). *Am. J. Psychiatry* **157**(Suppl. 1), 1–39.

Calcitonin Gene-Related Peptide (CGRP)

PETER HOLZER
University of Graz, Austria

calcitonin gene-related peptide First neuropeptide discovered by molecular analysis of the calcitonin gene, in the absence of any information about its functional role.

calcitonin receptor-like receptor One of the three proteins that form the calcitonin gene-related peptide receptor complex (the other two proteins are the receptor-associated membrane protein-1 and the receptor component protein). The calcitonin receptor-like receptor consists of seven transmembrane domains, is coupled to G proteins, and serves as the receptor protein that recognizes calcitonin gene-related peptide as ligand.

extrinsic afferent nerve fibers Originate from the nodose ganglia (vagal afferents) and dorsal root ganglia (spinal afferents) and connect the gut with the brain stem and spinal cord, respectively.

intrinsic enteric neurons Originate from the myenteric and submucosal nerve plexuses within the wall of the gastrointestinal tract and supply all effector tissues of the gut.

mucosal homeostasis Complex array of protective mechanisms coordinated by neural and other control systems to ensure that the mucosa survives undamaged the onslaughts that occur during digestion; the gastrointestinal mucosa is endangered by toxic, antigenic, and pathogenic food ingredients as well as by potentially harmful secretory products such as pepsin and acid.

neuropeptide Oligo- or polypeptide that is expressed by specific neurons and serves as an extracellular messenger of those neurons.

receptor-associated membrane protein-1 Single-membrane-spanning protein (also termed receptor activity-modifying protein); required to guide the intracellular trafficking of calcitonin receptor-like receptor to the cell membrane and to endow calcitonin receptor-like receptor with high affinity for calcitonin gene-related peptide. Association of calcitonin receptor-like receptor with receptor-associated membrane protein-1 is a prerequisite for the formation of functional calcitonin gene-related peptide receptors.

receptor component protein Intracellular peptide important for coupling calcitonin receptor-like receptor to the G protein G_s/adenylate cyclase signal transduction pathway.

Calcitonin gene-related peptide is a 37-amino-acid neuropeptide; within the gastrointestinal tract, this peptide occurs primarily in extrinsic afferent nerve fibers and intrinsic enteric neurons. Following its release, calcitonin gene-related peptide acts on specific receptors that are expressed by enteric neurons and gastrointestinal effector systems in a tissue-specific manner. In so doing, the peptide plays a role in the regulation of blood flow, exocrine and endocrine secretory activity, mucosal homeostasis, motor activity, and nociception. It appears that the peptide-specific receptors represent a significant target for therapeutic intervention in various gastroenterological disease entities. Although the molecular pharmacology of calcitonin gene-related peptide actions is not yet completely understood, new information on receptor structure and function suggests that it may be feasible to design selective ligands for the calcitonin gene-related peptide receptor and to evaluate their utility as therapeutics.

INTRODUCTION

Calcitonin gene-related peptide (CGRP) was the first biologically active peptide to be discovered by recombinant DNA technology, instead of the traditional approach of tissue extraction, purification, and bioassay. In 1983, M. G. Rosenfeld and co-workers discovered not only this neuropeptide but also a new regulatory mechanism of cell-specific gene expression. Expression of calcitonin gene-related peptide results from the alternative splicing of RNA transcribed from the calcitonin gene. Primary RNA transcripts are processed to messenger RNA (mRNA) for calcitonin in the thyroid C cells, whereas the mRNA for CGRP is formed predominantly in neurons of the central and peripheral nervous system.

Since the publication of Rosenfeld's work, CGRP has been localized to distinct neurons as well as some endocrine and immune cells in the gastrointestinal (GI) tract. The complex structure of CGRP receptors has only recently been disclosed; functional CGRP receptors are now known to consist of three proteins, calcitonin receptor-like receptor (CRLR), receptor-associated membrane protein-1 (RAMP-1), and receptor component protein (RCP). The cell-selective expression of CGRP receptors on GI effector systems and the peptide's biological actions in the gut suggest

 Copyright 2004, Elsevier (USA). All rights reserved.

that CGRP plays multiple roles in the neural control of digestive activity. These functions include mucosal-blood flow, mucosal homeostasis, exocrine and endocrine secretory processes, motor activity, and GI nociception. Experimental work indicates that a disturbance of the CGRP system may contribute to a number of GI disorders and that a correction of these perturbations may be of therapeutic potential.

THE CGRP FAMILY

CGRP-α (CGRP-I) is encoded by the calcitonin/CGRP-α gene on chromosome 11 of the human genome and represents a 37-amino-acid polypeptide, the sequence of which varies slightly among different vertebrate species (Fig. 1), although mouse and rat CGRP-α are identical. Common to all of the CGRPs is a disulfide bridge between Cys-2 and Cys-7 and an amidated C terminus. Mouse, rat, and human tissues contain an additional peptide, CGRP-β (CGRP-II), which has a high sequence homology to CGRP-α (Fig. 1). Importantly, though, the gene from which CGRP-β is derived does not encode calcitonin. Because the biological activities of CGRP-α and CGRP-β are generally similar, it is thought that both CGRP homologues act on a similar receptor population.

CGRP belongs to a family of bioactive peptides, often termed the calcitonin peptide family, which comprise calcitonin, CGRP-α, CGRP-β, amylin (islet amyloid polypeptide), and adrenomedullin (Fig. 1). The sequence homologies range from 20 to 30% between CGRP and adrenomedullin as well as CGRP and calcitonin, to 40 to 50% between CGRP and amylin. Encoded by separate genes, amylin, like CGRP, is a 37-amino-acid polypeptide, whereas adrenomedullin is made of 52 amino acids. Although CGRP, amylin, and adrenomedullin share some biological effects, there is now good evidence that they act via different CGRP, amylin, and adrenomedullin receptors (which, however, are related to each other).

CGRP RECEPTORS

The biological effects of CGRP are brought about by interaction with specific membrane receptors. Pharmacologically, two CGRP receptor subtypes, termed $CGRP_1$ and $CGRP_2$ receptors, have been proposed to exist because certain CGRP analogues differ in their potencies in various bioassays. Characteristic of the $CGRP_1$ receptors is their sensitivity to the antagonistic effect of the human $CGRP_{8-37}$ fragment. Although known to be receptors that are coupled to G proteins and linked to adenylate cyclase, functional CGRP receptors have long resisted molecular identification because they are assembled from three different proteins (Fig. 2), CRLR, RAMP-1, and RCP. Despite the fact that CRLR is the CGRP-recognizing protein, the CGRP receptor becomes functional only if the seven-transmembrane domain-containing CRLR is associated with the single-membrane-spanning RAMP-1. This chaperone protein is important for the intracellular translocation of CRLR and its insertion into the plasma membrane and is essential for conferring a CGRP ($CGRP_1$) receptor-like binding profile on CRLR. In addition, RAMP-1 may also play a role in receptor desensitization by modulating the activation of protein kinase A. The intracellular RCP is required for efficient coupling of the receptor to the G_s/adenylate cyclase signaling pathway (Fig. 2). Following activation by an agonist, the CRLR/RAMP-1 complex is phosphorylated and internalized in a dynamin- and β-arrestin-dependent manner. The putative heterogeneity of CGRP receptors inferred from pharmacological studies has not yet been disclosed at the molecular level.

CRLR can associate not only with RAMP-1 and RCP to produce a CGRP ($CGRP_1$) receptor, but also with RAMP-2 or RAMP-3, two other chaperone proteins sharing an approximately 30% sequence homology with RAMP-1. Interestingly enough, these RAMPs dictate the pharmacological profile of CRLR because they

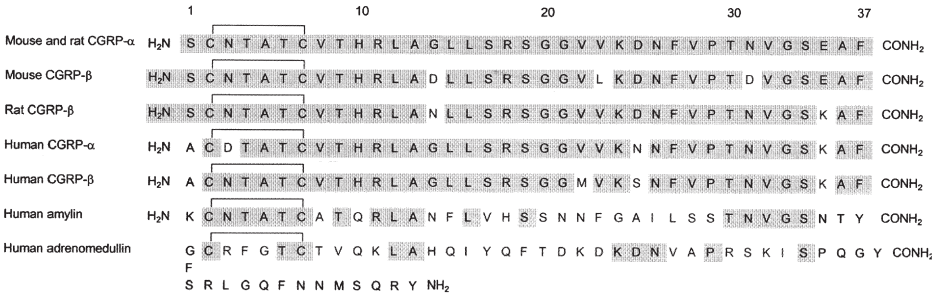


FIGURE 1 Comparison of the amino acid sequences of mouse, rat, and human CGRP-α and CGRP-β, human amylin, and human adrenomedullin.

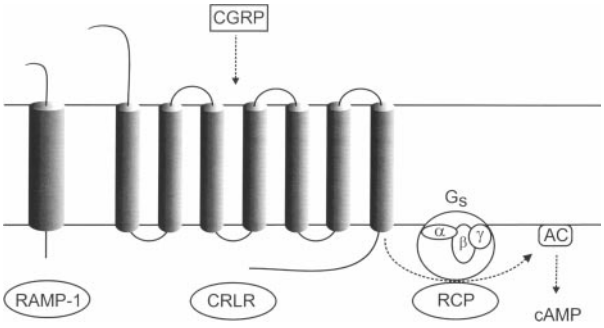


FIGURE 2 Diagram of the calcitonin gene-related peptide (CGRP) receptor complex consisting of the calcitonin receptor-like receptor (CRLR), the receptor activity-modifying protein-1 (RAMP-1), and the receptor component protein (RCP). Association of CRLR with RAMP-1 forms a high-affinity CGRP receptor that, through RCP, is effectively coupled to the G protein G_s/adenylate cyclase (AC)/cyclic adenosine monophosphate (cAMP) signaling pathway.

transfer a unique agonist selectivity profile to the receptor complex. Thus, CRLR associated with RAMP-2 behaves as an adrenomedullin receptor at which adrenomedullin-like peptides have a much greater affinity than do CGRP-like peptides. If the human calcitonin receptor isotype 2 (hCTR2) is assembled with RAMP-1 or RAMP-3, hCTR2 loses its affinity for human calcitonin but acquires high affinity for amylin, thus behaving like an amylin receptor.

Owing to the difficulties in the molecular receptor characterization, the pharmacology of CGRP receptors is still poorly understood. For a long time, $CGRP_{8-37}$ was the sole CGRP receptor antagonist available, and only recently has the first nonpeptide antagonist of CGRP receptors, BIBN4096BS, been published. This compound exhibits extremely high affinity for the human CGRP ($CGRP_1$) receptor, its K_i in the picomolar range being more than 1000-fold lower than that of human $CGRP_{8-37}$. As seen with nonpeptide antagonists for other neuropeptide receptors, the activity of BIBN4096BS is species specific, given that its potency at the rat CGRP ($CGRP_1$) receptor is 100-fold less than at the human counterpart. This variation in potency appears to be determined by species differences in the amino acid sequence of RAMP-1, notably at position 74.

EXPRESSION AND RELEASE OF CGRP IN THE GI TRACT

The principal sources of CGRP in the digestive system are extrinsic primary afferent nerve fibers and intrinsic enteric neurons (Fig. 3). These two neuronal systems contain different molecular forms of the peptide, because most of the CGRP expressed in extrinsic afferents of the rat is CGRP-α, whereas the only form of CGRP in enteric neurons is CGRP-β. The density of the GI innervation by CGRP-immunoreactive neurons varies greatly among different regions of the gut and different mammalian species, as does the relative contribution of extrinsic afferent and intrinsic enteric neurons to the overall CGRP content of the gut. Most of the CGRP found in the esophagus and stomach of small rodents is derived from extrinsic afferent neurons whereas the small and large intestine contains a sizable number of intrinsic enteric neurons expressing CGRP. The CGRP-immunoreactive neurons of the myenteric and submucosal plexus in the guinea pig intestine project primarily into the mucosa whereas those in the rat intestine issue oral and caudal projections within the plexuses as well as to the muscle layers and the mucosa.

Most of the CGRP-expressing extrinsic afferent neurons in the rodent gut originate from cell bodies in the dorsal root ganglia and reach the gut via sympathetic (splanchnic, colonic, and hypogastric) and sacral parasympathetic (pelvic) nerves while passing through prevertebral ganglia and forming collateral synapses with sympathetic ganglion cells. Within the wall of the GI tract, they supply primarily arteries and arterioles but also project to the lamina propria of the mucosa, to the submucosal and myenteric nerve plexuses, and to the circular and longitudinal muscle layers (Fig. 3). It is a characteristic of many spinal afferents in the rat, guinea pig, and canine gut that CGRP is coexpressed with the tachykinin substance P, whereas CGRP and tachykinins do not coexist in enteric neurons of these species. CGRP-positive vagal afferents originating from the nodose ganglion supply the esophagus and proximal part

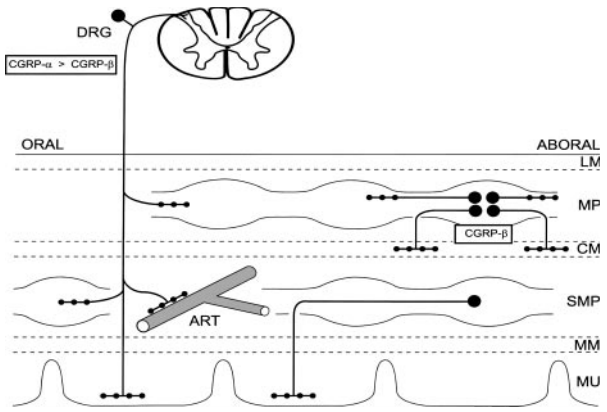


FIGURE 3 Diagram showing rat intestine extrinsic afferent and intrinsic enteric neurons that express calcitonin gene-related peptide (CGRP). DRG, dorsal root ganglion; LM, longitudinal muscle; MP, myenteric plexus; CM, circular muscle; SMP, submucosal plexus; MM, muscularis mucosae; MU, mucosa; ART, arteriole.

of the stomach but make a relatively small contribution to the content of CGRP in the gastric corpus, antrum, and intestine, given that a vast majority (80–90%) of the CGRP-containing nerve fibers in the rat stomach are derived from dorsal root ganglia.

As is expected for substances with a vesicular localization, CGRP is released from extrinsic afferent or intrinsic enteric neurons of the GI tract in a calcium-dependent manner if these cells are depolarized. CGRP release from extrinsic spinal afferents within the gut can be elicited by the vanilloid capsaicin because these nerve cells express functional vanilloid receptors of type 1. With the use of capsaicin, it has also been found that CGRP in the general circulation represents primarily an overspill of CGRP released from peri- and paravascular afferent nerve fibers. It is particularly worth noting that acidification of the mucosal tissue releases CGRP from extrinsic afferents in the stomach and duodenum.

Apart from neurons, CGRP is also found in endocrine cells of the human GI mucosa and rat pancreas and in blood-derived or resident immune cells within the lamina propria of the rat gastric mucosa. However, the quantitative and functional significance of these sources is still little known.

EFFECTS, PHYSIOLOGICAL ROLES, AND PATHOLOGICAL IMPLICATIONS OF CGRP IN THE GI TRACT

Motor Activity

The most prominent motor action of CGRP in the active gut is muscle relaxation via $CGRP_{8-37}$-sensitive CGRP receptors. This effect arises in most instances from a direct action on the muscle and leads to retardation of gastric emptying and to attenuation of motility throughout the digestive tract. However, CGRP is also able to excite enteric cholinergic motor pathways, which is in keeping with the peptide's ability to depolarize intrinsic primary afferent neurons of the myenteric plexus, to enhance the release of acetylcholine from enteric neurons, and, via $CGRP_{8-37}$-insensitive CGRP receptors on enteric neurons, to cause contraction of the resting muscle.

There is still scarce information as to whether CGRP released from intrinsic or extrinsic neurons of the gut plays a physiological role in the neural control of GI motility. The claim that CGRP released from sensory neurons contributes to distension-induced peristalsis is not universally accepted. There is, however, increasing evidence to infer that CGRP released from extrinsic afferent nerve fibers contributes to pathological disturbances of GI motility. Thus, experimental data attribute to CGRP a role in the pathological shutdown of GI motility that is associated with postoperative or peritonitis-induced ileus. This is consistent with the observation that CGRP acting via $CGRP_{8-37}$-sensitive CGRP receptors contributes to the inhibition of intestinal peristalsis that ensues after sensory neuron stimulation.

Secretory Processes

The actions of CGRP on GI ion, enzyme, mucus, and fluid secretion vary with the region and species under study and in some instances appear to depend on the experimental conditions. For instance, CGRP is able to inhibit secretagogue-evoked secretion of enzyme, bicarbonate, and fluid from the pancreas of the dog and rat *in vivo*, an effect that to a large extent is mediated by CGRP-induced release of somatostatin. In contrast, amylase secretion from isolated acini of the rat and guinea pig pancreas is enhanced by the peptide.

Electrolyte and fluid secretion in the small intestine of the dog and in the colon of the guinea pig and rat is stimulated by CGRP. Although the secretory effect of CGRP in the rat colon arises from a direct action on enterocytes, as is the case with human epithelial cell lines, the secretory action of CGRP in the guinea pig colon is mediated by enteric neurons. It remains to be shown whether CGRP plays a physiological role in the control of intestinal ion and fluid secretion. Pathologically, it appears as if the peptide contributes to the fluid secretion in the rat ileum that is evoked by *Clostridium difficile* toxin A.

There is good evidence to implicate CGRP as a transmitter by which extrinsic afferent nerve fibers contribute to the homeostatic regulation of endocrine and exocrine secretory processes in the gastroduodenal region. CGRP is highly potent in depressing basal and secretagogue-evoked output of acid and pepsin in the stomach of humans, dogs, rabbits, and rats, an action that is brought about by $CGRP_{8-37}$-sensitive CGRP receptors, that depends on somatostatin as an essential mediator, and that goes along with a depression of the release of acetylcholine, gastrin, and histamine (Fig. 4). It is of note in this context that CGRP regulates not only the release of somatostatin and gastrin but also the transcription of their genes. The physiological relevance of these effects is highlighted by the ability of the CGRP receptor antagonist $CGRP_{8-37}$ to augment basal and stimulated acid output and by the ability of acid accumulation in the gastric lumen to release CGRP from extrinsic afferent nerve fibers. By way of its effects on the release of somatostatin, gastrin,

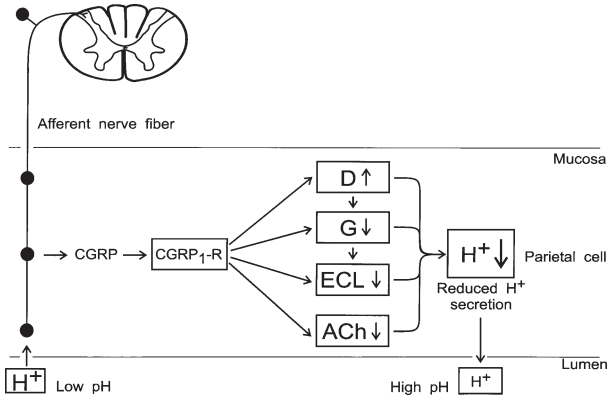


FIGURE 4 Diagram illustrating the role of calcitonin gene-related peptide (CGRP)-releasing afferent nerve fibers in the feedback control of gastric acid secretion in the rat stomach. When the acidity in the lumen rises, afferent nerve fibers release CGRP, which, via activation of $CGRP_{8-37}$-sensitive CGRP ($CGRP_1$) receptors, stimulates the release of somatostatin and inhibits the release of gastrin, histamine, and acetylcholine. D, endocrine D cells releasing somatostatin; G, endocrine G cells releasing gastrin; ECL, enterochromaffin-like cells releasing histamine; ACh, neurons releasing acetylcholine.

histamine, and acetylcholine, CGRP halts further secretion of acid and thus represents an essential transmitter in the feedback inhibition of gastric acid output (Fig. 4). The protection of the gastroduodenal mucosa from any deleterious influence of acid is supported by the peptide's ability to stimulate bicarbonate and mucus secretion.

Vascular Functions

The arteries and submucosal arterioles of the GI tract receive the densest innervation by extrinsic afferents containing CGRP. This localization of CGRP, the expression of CGRP receptors on the endothelium and smooth muscle of arteries and arterioles, and the peptide's high potency in causing vasodilatation point to a vasoregulatory function of CGRP. There is ample evidence to conclude that nonadrenergic noncholinergic dilation of the rat superior mesenteric artery is mediated by capsaicin-sensitive afferent nerve fibers releasing CGRP. In contrast, the physiological significance of CGRP in the microcirculation of the small and large intestine is little known. CGRP is able to dilate submucosal arterioles in the guinea-pig ileum, in which this neuropeptide mediates the vasodilator reaction to afferent neuron stimulation with capsaicin, whereas mucosal blood flow in the rat small and large intestine is not altered by CGRP.

The situation is different in the rat stomach, where CGRP potently enhances blood flow through the mucosa, an action that is brought about by $CGRP_{8-37}$-sensitive CGRP receptors. This CGRP-induced hyperemia arises from dilation of submucosal arterioles, but the diameter of venules remains unchanged. The dilator action of low doses of CGRP is mediated by a mechanism that involves nitric oxide; high doses of the peptide increase blood flow independently of nitric oxide. Although CGRP mediates the nitric oxide-dependent vasodilatation that afferent neuron stimulation by intragastric capsaicin elicits in the rat stomach, there is little evidence that this neuropeptide participates directly in the physiological regulation of gastric blood flow. However, pharmacological and CGRP-α gene knockout studies suggest that endogenous CGRP influences vascular function indirectly through inhibitory modulation of sympathetic nerve activity.

CGRP comes prominently into play under pathological conditions, a role that is best exemplified by the hyperemic response that ensues when the gastric mucosal barrier is disrupted by ethanol or bile salts, allowing acid to enter the tissue and damage the gastric mucosa. This acid-evoked rise of gastric mucosal blood flow involves CGRP released from extrinsic afferent nerve fibers and nitric oxide as the major vasodilator messengers (Fig. 5). The rise of gastric mucosal blood flow in response to acid influx serves a homeostatic and protective role in the gastric mucosa because it helps to neutralize and wash away intruding acid and delivers bicarbonate and other factors to defend and repair the mucosa (Fig. 5). Another example relates to the *C. difficile* toxin A-evoked inflammation in the rat ileum, which involves CGRP as a proinflammatory mediator.

Mucosal Homeostasis

There are several lines of evidence to indicate that GI mucosal integrity and repair are under the control of extrinsic afferent neurons releasing CGRP. The first hint at such a role came from the ability of CGRP to protect the mucosa in a number of experimental models of gastric injury and colonic inflammation. The action of CGRP to reduce ethanol-induced damage in the gastric mucosa is mediated by $CGRP_{8-37}$-sensitive CGRP receptors and involves both nitric oxide and K_{ATP} channels, whereas prostaglandins do not participate.

The activity of CGRP to strengthen mucosal defense appears to be of pathophysiological significance, given that the peptide mediates the gastroprotective effect of primary afferent neuron excitation by capsaicin and many other factors and drugs. Thus, blockade of CGRP receptors with $CGRP_{8-37}$ prevents the ability

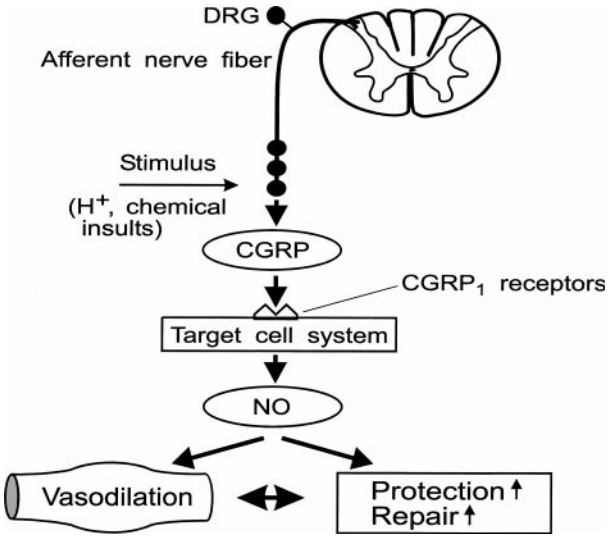


FIGURE 5 Diagram illustrating the homeostatic role of calcitonin gene-related peptide (CGRP)-releasing afferent nerve fibers in the rat gastric mucosa. Influx of acid into the mucosa through a disrupted mucosal barrier or challenge by other chemical insults stimulates these nerve fibers to release CGRP, which then activates $CGRP_{8-37}$-sensitive CGRP ($CGRP_1$) receptors to initiate local mechanisms of defense and repair. NO, nitric oxide.

of intragastric capsaicin to attenuate experimentally imposed injury, as does immunoneutralization of CGRP with polyclonal and monoclonal antibodies to the peptide. This implication of CGRP in gastric mucosal homeostasis is corroborated by the observations that $CGRP_{8-37}$ as well as active immunization of rats against CGRP exacerbates experimental injury in the stomach.

As has been pointed out by P. Holzer, CGRP released from sensory nerve fibers strengthens gastric mucosal defense by mechanisms that involve vasodilatation and hyperemia-dependent and hyperemia-independent processes. Hyperemia supports a number of gastroprotective mechanisms, including appropriate delivery of bicarbonate to the surface mucus layer, and facilitates the rapid restitution and repair of the wounded mucosa. In addition, CGRP is per se able to enhance the secretion of bicarbonate and mucus in the rat stomach and duodenum. Because CGRP-releasing afferent nerve fibers are not tonically active, it seems that they operate as a neural alarm system that, when stimulated by mucosal insults, activates mechanisms of acute defense and supports processes that aid the repair of the injured mucosa (Fig. 5). Complementary evidence for such a homeostatic role of CGRP-releasing afferent nerve fibers in the GI mucosa comes from the observation that sensory neuropathies weaken the resistance of the tissue to injury. This applies not only to the stomach, but also to the esophagus, small intestine, and colon, in which experimentally induced inflammation and damage are aggravated.

GI Sensitivity and Nociception

CGRP is a transmitter of spinal afferent neurons innervating the gut, and there is experimental evidence that CGRP can mediate GI pain and inflammatory hyperalgesia. Intraperitoneal administration of exogenous CGRP or acetic acid-induced release of endogenous CGRP triggers abdominal muscle contractions, a reaction that is indicative of pain. Of particular importance is the finding that $CGRP_{8-37}$ is able to prevent inflammation-induced hypersensitivity to colonic distension. In this respect, intrathecal $CGRP_{8-37}$ is more potent than intravenous $CGRP_{8-37}$, thus J. M. Gschossmann, E. A. Mayer, and co-workers have concluded that the site of CGRP-mediated hyperalgesia is primarily in the spinal cord.

SUMMARY

Experimental evidence indicates that, in the GI tract, CGRP is a transmitter candidate of intrinsic enteric neurons and extrinsic afferent nerve fibers. As such, this neuropeptide seems to be involved in the neural regulation of various digestive functions as deduced from the distribution of CGRP-releasing nerve fibers and CGRP receptors in the gut and the pharmacological effects of exogenous CGRP on GI motility, exocrine and endocrine secretory activity, blood flow, mucosal homeostasis, and abdominal pain. However, the physiological and pathophysiological implications of CGRP in gut function have not yet been fully characterized because potent nonpeptide CGRP receptor antagonists have not been available until recently. It would nevertheless appear, if the situation encountered in small rodents can be extrapolated to humans, that CGRP-immunoreactive afferent nerve fibers act to halt GI motility, increase gastric blood flow, inhibit gastric acid secretion, enforce GI mucosal resistance to injury, and contribute to inflammation-induced GI hyperalgesia. If so, the CGRP system in the digestive tract represents a regulatory system with considerable potential for therapeutic intervention.

Apart from acting within the gut and signaling noxious information to the spinal cord, CGRP may also play a role in the central regulation of digestive functions along the brain–gut axis. As first discovered by Y. Taché, intracerebral administration of CGRP inhibits

gastric acid secretion, ulcer formation, and motility, mostly by increasing the sympathoadrenal outflow.

See Also the Following Articles

Gastric Acid Secretion • Gastrin • Sensory Innervation • Somatostatin

Acknowledgments

Evelin Painsipp's artistry in drawing the figures is greatly appreciated. The author's work is currently supported by the Austrian Research Funds (FWF project P14295-MED) and the Austrian Federal Ministry of Education, Science and Culture (projects GZ 70.058/2-Pr/4/1999 and GZ 70.065/2-Pr/4/2000).

Further Reading

Akiba, Y., Furukawa, O., Guth, P. H., Engel, E., Nastaskin, I., and Kaunitz, J. D. (2001). Sensory pathways and cyclooxygenase regulate mucus gel thickness in rat duodenum. *Am. J. Physiol.* **280**, G470–G474.

Brain, S. D., Poyner, D. R., and Hill, R. G. (2002). CGRP receptors: A headache to study, but will antagonists prove therapeutic in migraine? *Trends Pharmacol. Sci.* **23**, 51–53.

Debas, H. T., Nelson, M. T., Bunnett, N. W., and Mulvihill, S. J. (1992). Selective release of somatostatin by calcitonin gene-related peptide and influence on pancreatic secretion. *Ann. N. Y. Acad. Sci.* **657**, 289–298.

Dockray, G. J., Forster, E. R., Louis, S. M., Sandvik, A. K., and Dimaline, R. (1992). Immunoneutralization studies with calcitonin gene-related peptide. *Ann. N. Y. Acad. Sci.* **657**, 258–267.

Doods, H., Hallermayer, G., Wu, D., Entzeroth, M., Rudolf, K., Engel, W., and Eberlein, W. (2000). Pharmacological profile of BIBN4096BS, the first selective small molecule CGRP antagonist. *Br. J. Pharmacol.* **129**, 420–423.

Evans, B. N., Rosenblatt, M. I., Mayer, L. O., Oliver, K. R., and Dickerson, I. M. (2000). CGRP-RCP, a novel protein required for signal transduction at calcitonin gene-related peptide and adrenomedullin receptors. *J. Biol. Chem.* **275**, 31438–31443.

Foord, S. M., and Marshall, F. H. (1999). RAMPs: Accessory proteins for seven transmembrane domain receptors. *Trends Pharmacol. Sci.* **20**, 184–187.

Gschossmann, J. M., Coutinho, S. V., Miller, J. C., Huebel, K., Naliboff, B., Wong, H. C., Walsh, J. H., and Mayer, E. A. (2001). Involvement of spinal calcitonin gene-related peptide in the development of acute visceral hyperalgesia in the rat. *Neurogastroenterol. Motil.* **13**, 229–236.

Hilaire, S., Bélanger, C., Bertrand, J., Laperriére, A., Foord, S. M., and Bouvier, M. (2001). Agonist-promoted internalization of a ternary complex between calcitonin receptor-like receptor, receptor activity-modifying protein 1 (RAMP1), and β-arrestin. *J. Biol. Chem.* **276**, 42182–42190.

Holzer, P. (1994). Calcitonin gene-related peptide. *In* "Gut Peptides: Biochemistry and Physiology" (J. H. Walsh and G. J. Dockray, eds.), pp. 493–523. Raven Press, New York.

Holzer, P. (1998). Implications of tachykinins and calcitonin gene-related peptide in inflammatory bowel disease. *Digestion* **59**, 269–283.

Holzer, P. (1998). Neural emergency system in the stomach. *Gastroenterology* **114**, 823–839.

Holzer, P. (2000). Calcitonin gene-related peptide in gastrointestinal homeostasis. *In* "The CGRP Family: Calcitonin Gene-Related Peptide (CGRP), Amylin, and Adrenomedullin" (D. Poyner, I. Marshall, and S. D. Brain, eds.), pp. 125–139. Landes Bioscience, Georgetown, Texas.

Holzer, P., and Barthó, L. (1996). Sensory neurons in the intestine. *In* "Neurogenic Inflammation" (P. Geppetti and P. Holzer, eds.), pp. 153–167. CRC Press, Boca Raton, FL.

Ichikawa, T., Ishihara, K., Kusakabe, T., Hiruma, H., Kawakami, T., and Hotta, K. (2000). CGRP modulates mucin synthesis in surface mucus cells of rat gastric oxyntic mucosa. *Am. J. Physiol.* **279**, G82–G89.

Juaneda, C., Dumont, Y., and Quirion, R. (2000). The molecular pharmacology of CGRP and related peptide receptor subtypes. *Trends Neurosci.* **21**, 432–438.

Oh-hashi, Y., Shindo, T., Kurihara, Y., Imai, T., Wang, Y. H., Morita, H., Imai, Y., Kayaba, Y., Nishimatsu, H., Suematsu, Y., Hirata, Y., Yazaki, Y., Nagai, R., Kuwaki, T., and Kurihara, H. (2001). Elevated sympathetic nervous activity in mice deficient in αCGRP. *Circ. Res.* **89**, 983–990.

Poyner, D. (1995). Pharmacology of receptors for calcitonin gene-related peptide and amylin. *Trends Pharmacol. Sci.* **16**, 424–428.

Rosenfeld, M. G., Emeson, R. B., Yeakley, J. M., Merillat, N., Hedjran, F., Lenz, J., and Delsert, C. (1992). Calcitonin gene-related peptide: A neuropeptide generated as a consequence of tissue-specific, developmentally regulated alternative RNA processing events. *Ann. N. Y. Acad. Sci.* **657**, 1–17.

Sternini, C. (1992). Enteric and visceral afferent CGRP neurons. Targets of innervation and differential expression patterns. *Ann. N. Y. Acad. Sci.* **657**, 170–186.

Taché, Y. (1992). Inhibition of gastric acid secretion and ulcers by calcitonin gene-related peptide. *Ann. N. Y. Acad. Sci.* **657**, 240–247.

Taché, Y., Garrick, T., and Raybould, H. (1990). Central nervous system action of peptides to influence gastrointestinal motor function. *Gastroenterology* **98**, 517–528.

Thomas, P. M., Nasonkin, I, Zhang, H. Q., Gagel, R. F., and Cote, G. J. (2001). Structure of the mouse calcitonin/calcitonin gene-related peptide α and β genes. *DNA Seq.* **12**, 131–135.

Wimalawansa, S. J. (1996). Calcitonin gene-related peptide and its receptors: Molecular genetics, physiology, pathophysiology, and therapeutic potentials. *Endocr. Rev.* **17**, 533–585.

Calcium, Magnesium, and Vitamin D Absorption, Metabolism, and Deficiency

ROBERTO CIVITELLI* AND KONSTANTINOS ZIAMBARAS†

Washington University, St. Louis, Missouri and †University of Crete, Iraklion, Greece

osteomalacia Defective mineralization of the organic matrix, resulting in excessive accumulation of osteoid in the bone tissue and increased propensity of the bones to bow or fracture under the weight of the body.

osteopenia Condition of decreased bone mass, defined as a bone mineral density between −1.0 and −2.5 standard deviations relative to the ideal peak bone mass (World Health Organization definition).

osteoporosis A skeletal disorder characterized by compromised bone strength predisposing to an increased risk of fractures. According to the World Health Organization, it is defined as a bone mineral level lower than −2.5 standard deviations from the ideal peak bone mass.

rickets A disease caused by vitamin D or phosphate deficiency during childhood; it is characterized by lack of growth plate fusion and bowing of the long bones, with defective bone matrix mineralization.

vitamin D metabolites Cholecalciferol (vitamin D_3) and ergocalciferol (vitamin D_2) are ingested with food, although ergocalciferol can also be synthesized in the skin. They both undergo two subsequent hydroxylation steps to 25-hydroxyvitamin D in the liver, and 1α,25-dihydroxyvitamin D the kidney. The former metabolite represents the storage form of the vitamin, which decreases under deficiency conditions; the latter is the hormonal form, is under the control of parathyroid hormone, and regulates intestinal calcium absorption.

In terrestrial vertebrates, mineral homeostasis involves intestine, kidney, and bone and is regulated primarily by parathyroid hormone (PTH) and 1α,25$(OH)_2$D, the biologically active, hormonal form of vitamin D. The concentration of calcium ions in the circulation is maintained within a narrow range; increased absorption of Ca^{2+} in the intestine and decreased excretion of Ca^{2+} by the kidney are usually sufficient to maintain calcium balance; however, under conditions of severe calcium deficiency, Ca^{2+} is mobilized from bone. The concentration of serum magnesium is regulated primarily by the kidney and magnesium is involved in the synthesis of PTH.

INTRODUCTION

The evolution from marine to terrestrial life required a major change in the way live organisms utilized minerals from the environment. Such a shift led to the evolution of a most exquisite biological homeostatic system that allowed organisms to absorb minerals (primarily calcium) and maintain their content in the body fluids within a narrow range. This homeostatic system, which involves intestine, kidney, and bone, is maintained primarily by two key hormones, parathyroid hormone (PTH) and 1α,25$(OH)_2$D, the hormonal form of vitamin D. Vitamin D dates back at least half a billion years. It was originally produced in ocean-dwelling phytoplankton by exposure to sunlight, probably acting as a sunscreen. With the evolution of terrestrial vertebrates, this "vitamin," which can be produced by ultraviolet (UV) irradiation of precursors in most organisms, gradually assumed its current role in the development and maintenance of the ossified skeleton.

The first priority of the integrated system that regulates mineral homeostasis is to maintain the concentration of circulating ionized calcium (Ca^{2+}) within a narrow range. Even modest deviations from the normal range must be controlled. In most cases, increased intestinal absorption and decreased renal excretion are adaptive mechanisms that are sufficient to maintain extracellular Ca^{2+} concentration within the normal range without the need to mobilize calcium from bone. However, under conditions of severe calcium deficiency and hypocalcemia, skeletal Ca^{2+} stores must be accessed to maintain circulating Ca^{2+}. This homeostatic response occurs even to the point of potentially compromising the structural integrity of the skeleton. Parathyroid hormone directs fast responses to changes in circulating calcium, whereas the vitamin D system represents a slower adaptive mechanism that maintains calcium balance, regulating primarily its input through intestinal absorption. The actions of PTH and 1α,25$(OH)_2$D are coordinated and each hormone influences the production of the other.

Magnesium, the fourth most abundant cation in the body and the major intracellular divalent cation, is not as tightly regulated as calcium, but it is directly involved in the synthesis of PTH.

 Copyright 2004, Elsevier (USA). All rights reserved.

CALCIUM

Calcium Balance

Calcium is an essential element for survival. It provides the structural integrity of the skeleton and controls vital physiologic processes, i.e., nerve excitability, muscle contraction, and blood coagulation. An adult human body contains approximately 1000 g of calcium, 99% of which is in the skeleton in the form of hydroxyapatite [$Ca_{10}(PO_4)(OH)_2$] and 1% in the extracellular fluids and soft tissues.

There are three fractions of calcium in serum: ionized (or free) calcium (~47%), protein-bound calcium (~46%), and calcium complexed to small ions, such as bicarbonate, citrate, and phosphate (~5–10%). Ionized and complexed calcium constitute the filterable calcium, but only the former is metabolically active. Albumin binds almost 90% of the protein-bound calcium, with the remainder bound to various globulins. Alterations of serum albumin concentration result in changes of total serum calcium with only minor or no changes in ionized calcium. Binding of calcium to serum proteins is pH dependent; acidosis decreases binding and increases ionized calcium, whereas alkalosis increases binding with a consequent decrease in ionized calcium. Reduced ionized calcium increases sodium permeability and enhances tissue excitability, whereas an increase in ionized calcium has the opposite effect. An example of this mechanism is the respiratory alkalosis produced in states of hyperventilation. Increased calcium binding to serum proteins causes a relative decrease of ionized calcium with consequent neuromuscular symptoms.

It is the extracellular ionized fraction of calcium that is physiologically important and it is rigidly maintained by the combined effects of PTH and $1\alpha,25(OH)_2D$. The parathyroid cells are able to sense extracellular calcium concentration via a calcium-sensing receptor and can thus respond to hypocalcemia with increased PTH secretion. As noted above, the integrated actions of PTH on distal tubular calcium reabsorption, osteoclastic bone resorption, and $1\alpha,25(OH)_2D$-mediated intestinal calcium absorption are responsible for the fine regulation of the serum calcium concentration. The sensitivity and accuracy of this integrated control allows fluctuation of ionized calcium to stay within 0.1 mg/dl in either direction from its normal setpoint. Slight decrements in the ionized calcium level elicit a prompt (within seconds) increase in the rate of PTH secretion. The increased circulating levels of PTH stimulate reabsorption of calcium from the distal kidney tubules, thus raising serum calcium. In parallel, phosphaturia is stimulated along with conversion of 25(OH)D to $1\alpha,25(OH)_2D$. The latter is responsible for a slower adaptive response that augments intestinal absorption of calcium and phosphate. If necessary, PTH and $1\alpha,25(OH)_2D$ also promote the net release of calcium and phosphate from bone. The combined calcium efflux into the extracellular fluid from the intestine and bone and the increased calcium reclamation in the distal tubules restore ionized calcium to normal, thereby inhibiting PTH secretion and closing the negative feedback loop. The most rapid changes in calcium handling in response to PTH come from the kidneys and the skeleton. Changes in renal tubular reabsorption are observed within minutes, whereas release of calcium from the skeleton occurs within 1–3 h. On the other hand, adjustments to the rate of calcium absorption in the intestine via the PTH–$1\alpha,25(OH)_2D$ axis require 24–48 h to become fully operative, so that this system does not contribute to rapid adaptive responses. Thus, only minor increases in $1\alpha,25(OH)_2D$ occur within the first few hours of a low calcium challenge, whereas prolonged hypocalcemia results in higher levels of $1\alpha,25(OH)_2D$ in the circulation. With time, an increased number of new osteoclasts appear in bone.

Calcium Deficiency

Although calcium deficiency is highly prevalent in Western societies and in many other ethnic groups, it seldom leads to clinically evident hypocalcemia. The homeostatic mechanisms described above effectively compensate for shortages of calcium intakes in mild or transitory deficiency conditions. Overt hypocalcemia develops with prolonged and severe calcium deficiency or when additional causative factors intervene to either prevent absorption or increase excretion. Therefore, hypocalcemia presents in most circumstances as an asymptomatic condition detected on routine screening and only rarely as a life-threatening medical emergency. Ionized calcium rather than total serum calcium is the primary determinant of symptoms in patients with hypocalcemia.

Perhaps the most relevant example or the metabolic consequences of subtle, but chronic calcium deficiency is represented by age-dependent (type II) osteoporosis. Aging is associated with decreased intestinal calcium absorption, decreased skin production of vitamin D, and deficient dietary intake of calcium and vitamin D. As a compensatory mechanism, PTH secretion increases, leading to increased urinary calcium reclamation and stimulation of $1\alpha,25(OH)_2D$ synthesis, but also to increased bone resorption to enhance calcium efflux from bone to the circulation. The defective calcium

intake occurs in the backdrop of an age-dependent decline in intestinal ability to adapt to a low-calcium diet. It is estimated that after the age of 40, calcium absorption decreases at a rate of 0.21% per year, and in women, an additional ~2% decline occurs after menopause. Thus, the combined effects of menopause and age lead to a 20–25% decrease in absorption efficiency in women between the ages of 40 and 60. If calcium deficiency continues for a substantial period of time, the mild secondary hyperparathyroidism developing with age may cause continued bone loss, thus effecting one of the pathophysiologic mechanisms of age-dependent osteoporosis.

The causes of calcium deficiency can be classified into three major categories, i.e., dietary deficiency, calcium malabsorption, and vitamin D-dependent deficiency (reviewed in Section IV). Although milk and other dairy products are the best source of calcium, calcium-containing supplements currently represent the preferred method of calcium intake in Western countries, especially in elderly individuals. Despite the growing evidence that an adequate calcium intake is critical not only to achieve an optimal peak bone mass but also to minimize age-dependent bone loss, a large number of Americans still fail to meet the currently recommended dietary calcium intake. The inadequacy of standard diets to provide sufficient calcium remains a problem of national relevance in most Western countries, despite food fortification. Clearly, this measure is insufficient to effectively correct the deficiency, and calcium supplementation is commonly required, especially for subjects at risk for bone loss and for elderly individuals. According to a consensus panel convened by the National Institutes of Health in 1994, the optimal daily calcium intake is 800 mg until age 10, between 1200 and 1500 mg/day for young adults, and 1000 mg for individuals in the fertile period. After menopause or in individuals older than 65, a daily intake of at least 1500 mg is required, unless estrogen replacement therapy is instituted.

As noted, most dietary calcium comes from dairy products, grains, and bony fish. Other common food constituents, such as pasta, bread, fruits, vegetables, and juices, contain much smaller amounts of calcium. They also contain large amounts of fibers, the consumption of which has increased because of their putative therapeutic role in the control of serum lipids and bowel motility, particularly in elderly people. Unfortunately, the very mechanism by which indigestible fibers facilitate intestinal transit also reduces the time for calcium absorption in the duodenum. Other mechanisms contribute to the low bioavailability of calcium in leafy green vegetables, including the relatively high content of oxalic acid in spinach and phytate in whole wheat products and other cereals, leading to the formation of insoluble and unabsorbable oxalate or phytate calcium complexes. On the other hand, the most common calcium supplement currently used is calcium carbonate, the salt with the highest concentration of calcium by weight (40%). Preparations containing calcium citrate are also very popular, as they may cause less constipation than calcium carbonate although they tend to be more difficult to ingest.

Aside from fibers, a number of food products or additives may alter calcium absorption from food. Certain dietary sugars, such as lactose, xylitol, and sorbitol, enhance intestinal calcium absorption through various mechanisms, but the potential benefits of lactose are often offset by lactose intolerance in older patients. Ethanol has negative effects on calcium balance. It decreases calcium transport and absorption in the intestinal epithelium and it can indirectly affect calcium transport through its negative effects on other organs, such as the pancreas and the liver. Likewise, there is an inverse correlation between caffeine intake and calcium balance. Caffeine, acting as mild diuretic, transiently increases the urine output of sodium that in turn drives the calcium loss. However, these changes are in most part compensated for by homeostatic responses, and an adequate calcium intake (>1200 mg) protects against any harmful effects that caffeine may have on calcium metabolism and bone mass.

Reduction of the absorptive surface area is responsible for calcium malabsorption in gastrointestinal conditions, such as Crohn's disease and celiac disease, as well as in subjects with short bowel syndrome. Malabsorption of vitamin D and loss of bile salts add to the inadequate calcium input. At the skeletal level, this results in osteopenia sometimes associated with osteomalacia, depending on whether the vitamin D or the calcium deficit prevails. Celiac disease represents a good paradigm for the combined calcium and vitamin D malabsorption in certain intestinal disorders. In celiac disease, serum levels of 25(OH)D are lower than normal, whereas $1\alpha,25(OH)_2D$ is either normal or even increased. Thus, the increase of $1\alpha,25(OH)_2D$, driven by a secondary rise of PTH, is not able to compensate for the absorption inefficiency. Correction of the malabsorption with a gluten-free diet corrects the vitamin D deficiency and prevents osteopenia or osteoporosis characteristic of untreated celiac disease.

Long-term use of anticonvulsant medications, especially phenobarbital and phenytoin, can lead to chronic, subclinical calcium deficiency with biochemical evidence of secondary hyperparathyroidism, more severe in nonambulatory or institutionalized subjects.

The most consistent abnormality is a decreased serum 25(OH)D, reduced calcium absorption, increased fecal calcium excretion with normal or even elevated concentration of circulating $1\alpha,25(OH)_2D$. Colchicine produces a dose-dependent inhibition of calcium uptake and accumulation by duodenal cells. Cytotoxic chemotherapeutic agents can also induce profound malabsorption of calcium by damaging the intestinal epithelium, which is extremely sensitive to these agents owing to the rapid turnover of the intestinal mucosa epithelium.

MAGNESIUM

Magnesium Balance

Although much less abundant than calcium, magnesium is essential in many cellular functions, especially for certain enzymatic activities. Approximately 30% of the total magnesium is protein bound (mostly to albumin), whereas 55% is ionized and 15% is complexed in salts. The kidney is primarily responsible for the regulation of the serum magnesium concentration. Approximately 5–15% of filtered magnesium is reabsorbed in the proximal convoluted tubule and at least 50% is reabsorbed in the thick ascending limb of the loop of Henle. Magnesium reabsorption follows that of sodium and water via a passive paracellular process, including solvent drag, and depends on lumenal magnesium concentration. Tubular reabsorption is normally close to saturation and any increased distal delivery of magnesium results in increased urine magnesium.

Magnesium is rather ubiquitous in food products, especially in vegetables, meats, and dairy products. As for calcium, lactose in milk can substantially enhance intestinal magnesium absorption. Since food of cellular origin is a good source of magnesium, the average dietary magnesium intake in developed countries is higher than the RDA, except in certain conditions, such as during pregnancy and lactation, as well as during chronic use of loop diuretics or after extensive burns, when the magnesium requirement is far higher than the RDA. Magnesium can be absorbed along the entire gastrointestinal tract, with the highest level of efficiency in the jejunum and ileum. Unlike calcium, net intestinal magnesium absorption is directly proportional to dietary magnesium intake, with a fractional absorption of approximately 30–40%. Although magnesium absorption is primarily passive, vitamin D may indirectly affect intestinal magnesium absorption through changes in phosphate and/or calcium absorption, as it occurs when dietary calcium is very low (200 mg/day). In such cases, the homeostatic increases in PTH and $1\alpha,25(OH)_2D$ drive active magnesium absorption, adding to the passive transport mechanism.

Magnesium Deficiency

Because magnesium is found in virtually all foods, dietary deprivation is an infrequent cause of hypomagnesemia in individuals with normal caloric intake. Magnesium deficiency is more commonly caused by excessive losses from either the gastrointestinal tract or the kidney. However, since magnesium absorption is strictly dependent on the lumenal concentration of the element, any chronic dietary restriction or deficiency results in reduced absorption. Hypomagnesemia is not uncommon in hospitalized patients in whom many factors may contribute to magnesium loss, including acidosis, continuous use of diuretics, intravenous administration of sodium-containing fluids, and drugs that increase urinary excretion of magnesium. It is estimated that up to 10% of hospitalized patients have low serum magnesium and the prevalence can be even higher in intensive care patients.

A number of dietary constituents may interfere with magnesium and decrease its bioavailability at sites of absorption. Zinc interferes with magnesium absorption probably by competing for the same absorptive sites along the intestine. Diets with high zinc content can decrease magnesium absorption as much as 10%, resulting in negative magnesium balance if dietary magnesium is low. Although the evidence that dietary calcium can interfere with magnesium absorption is tenuous at best, of more concern for the possible development of magnesium deficiency is the interference of magnesium with anions and in particular phosphate and oxalate. As occurs for calcium, insoluble magnesium phosphate or oxalate salts can form in the intestinal lumen, thus preventing the absorption of either ion. The high level of phosphate in milk formulas used to feed very-low-birth-weight babies, under the assumption that the high level of phosphate may facilitate their skeletal development, exemplifies this premise. Unfortunately, when phosphate is increased to a calcium : phosphate ratio lower than 2 : 1, fractional intestinal magnesium absorption decreases and hypomagnesemia may occur in these small babies. Leafy green vegetables are an important source of magnesium, but calcium absorbability is reduced by the presence of oxalic acid. Food processing can also modify the bioavailability of magnesium. For example, raw spinach has a poorer magnesium bioavailability than boiled or fried spinach, whereas boiling or frying does not improve spinach calcium bioavailability. Phytate, contained in many vegetables, negatively affects

magnesium absorption, as it does with calcium. A case in point is soybean, which contains large amounts of phytate, thus making diets based exclusively on soybean inadequate to provide the daily requirements of minerals.

Primary intestinal diseases associated with decreased transit time (resection or bypass of the ileum for obesity, stomach resection, chronic diarrhea/steatorrhea such as regional enteritis, or ulcerative colitis), loss of absorptive surface from intestinal mucosal damage (short bowel syndrome, nontropical sprue, radiation injury, or intestinal lymphectasia), or insufficient bile secretion (biliary fistula or atresia) can cause a magnesium deficit. It is questionable whether the vitamin D deficiency that develops in hepatobiliary disorders as a consequence of fat malabsorption has any contribution to the decreased magnesium absorption, which is largely a vitamin D-independent process. Massive large bowel resection affects magnesium kinetics more than calcium, leading to decreased magnesium absorption. Abuse of laxatives may also cause hypomagnesemia secondary to decreased transit time, whereas vomiting or nasogastric suction may contribute to magnesium depletion because of loss of the upper intestinal tract fluid, which is rich in magnesium.

Renal magnesium transport is influenced by the filtered sodium and calcium load, and administration of excessive sodium in parenteral fluids is a common factor for hypomagnesemia in hospitalized patients. Excessive urinary magnesium loss is the cause of hypomagnesemia that may occur after treatment with a wide spectrum of medications, including loop diuretics, aminoglycosides, cisplatin, cyclosporine A, amphotericin B, and pentamidine, as well as in alcoholism. Likewise, the mechanism for hypomagnesemia in poorly controlled diabetes mellitus is probably related to both increased urinary magnesium loss and dietary restrictions of grains and nuts. Intestinal magnesium absorption is also reduced in patients with hypoparathyroidism or pseudo-hypoparathyroidism, likely as a consequence of decreased serum $1\alpha,25(OH)_2D$, since magnesium deficiency corrects after therapy with vitamin D metabolites.

VITAMIN D

Vitamin D Balance

Vitamin D is a fat-soluble steroid and exists as two forms, vitamin D_2 (ergocalciferol), found in plants, and vitamin D_3 (cholecalciferol), found primarily in fatty fish. Only the latter can be synthesized in the skin from the precursor 7-dehydrocholesterol and this endogenous production can account for up to 80% of total body vitamin D. In fact, cutaneous production is the main source of vitamin D in humans worldwide. In the circulation, vitamin D is bound to vitamin D-binding protein and is taken up by the liver, where a first hydroxylation on C-25 takes place to produce 25(OH)D. This form of vitamin D is bioactive only at high concentrations, as it occurs in vitamin D intoxication. The physiologically active form is produced by further hydroxylation on C-1, which occurs in the kidney. Because of its very tightly regulated circulating levels, the presence of specific receptors in target tissues, and specific biologic actions, $1,25(OH)_2D$ is a bone fide hormone. The most important regulator is PTH, which stimulates the 1α-hydroxylase. The resulting increased $1,25(OH)_2D$ concentration represents the slow-response feedback loop that increases serum calcium by stimulating its intestinal absorption, thus correcting hypocalcemic conditions. Experimental data also suggest that phosphorus may stimulate 1α-hydroxylase activity independent of PTH, a mechanism operative in renal failure. Both 25(OH)D and $1,25(OH)_2D$ undergo hydroxylation on C-24. Although some physiologic functions have been proposed for $24,25(OH)_2D$, 24-hydroxylation renders the molecule susceptible to inactivation by side chain cleavage and oxidation.

Unlike calcium and magnesium, vitamin D is rare in food. Natural sources of appreciable amounts of vitamin D are fatty fish, such as salmon and mackerel, and in particular their livers, which account for the efficacy of cod liver oil as a cure for rickets. As for calcium, several food products, in particular cereals, bread products, and milk, are frequently fortified with vitamin D and account for a substantial share of vitamin D intake in Western populations. In the United States, federal regulations stipulate that 10 μg (400 IU) of vitamin D be added to every quart of milk. Multivitamin preparations also typically contain 10 μg. Although these doses would meet the recommended dietary allowance of vitamin D, it has been estimated that the true vitamin D requirement in the absence of sunlight exposure could be as much as 600 IU per day.

Vitamin D Deficiency

Despite the widespread use of calcium supplements containing vitamin D and food fortification, the prevalence of relative vitamin D deficiency in the Western hemisphere remains surprisingly elevated. Subclinical vitamin D deficiency, defined biochemically as low 25(OH)D with low-normal serum calcium and phosphate, increased alkaline phosphates, and

mildsecondary hyperparathyroidism, was found in approximately 10% of ambulatory women living in the midwestern United States, referred for evaluation of osteoporosis. Moreover, almost 50% of homebound, community-dwelling, elderly persons may suffer from vitamin D deficiency. Thus, the possibility of subclinical deficiency of vitamin D and calcium malabsorption should be considered in elderly subjects with poor nutrition, who spend most of their time indoors.

A variety of factors can alter the cutaneous production of vitamin D. The skin pigment melanin competes with 7-dehydrocholesterol, the precursor of vitamin D_3, for absorption of UV light. Thus, people with darker skin color require longer exposure to sunlight to produce the same amount of vitamin D_3 as those with lighter skin color. Similarly, sunscreens with a sun protection factor of 8 and above substantially reduce endogenous vitamin D synthesis. Aging significantly decreases the production of vitamin D because of lower abundance of 7-dehydrocholestrol in the aging epidermis. A person older than 70 years produces 30–35% less vitamin D_2 than a young adult, under the same sunlight exposure. Regional meteorological conditions, latitude, and seasons affect dermal vitamin D production because of the different degree of sunlight irradiation, although an average outdoor exposure during normal daily living is adequate to provide with vitamin D requirements in most circumstances.

Once vitamin D is ingested, it is incorporated into the chylomicron fraction and absorbed through the lymphatic system. Biliary cirrhosis and other chronic cholestatic syndromes with impairment of bile salt secretion result in marked reduction of serum 25(OH)D levels and calcium absorption. Excess intralumenal unabsorbed fat not only binds calcium and limits its accessibility to the transport site, it also impairs the absorption of vitamin D. On the other hand, patients with cirrhosis and marked hepatocyte loss have only modestly reduced serum 25(OH)D levels, in the absence of cholestasis, as the enzymatic reserve for 25-hydroxylation of vitamin D is adequately maintained by a relatively small number of hepatocytes. Among the many drugs that are known to interfere with the hepatic P-450 enzyme system, barbiturates are very effective at increasing catabolic hydroxylation of $1\alpha,25(OH)_2D$ and 25(OH)D to biologically inactive metabolites, leading to vitamin D deficiency and osteomalacia in the most severe cases. Intestinal malabsorption syndromes, such as Crohn's disease, cystic fibrosis, and Wipple's disease, can lead to vitamin D malabsorption, whereas diseases that affect the more distal small intestine and large intestine appear to have little effect on the intestinal absorption of fat-soluble vitamins.

In chronic renal failure, a defective production of $1\alpha,25(OH)_2D$ is the leading cause of calcium malabsorption. However, since passive calcium absorption is essentially intact, uremic patients can absorb sufficient calcium when their dietary intake is augmented to 4–10 g/day, although such doses are usually not well tolerated. Oral administration of $1\alpha,25(OH)_2D$ increases fractional calcium absorption up to 45–50% and is the mainstay of therapy in chronic renal failure. Similarly, vitamin D deficiency, hypocalcemia, and reduced intestinal calcium absorption occur in the nephrotic syndrome. These changes are attributed to hypoalbuminemia and the attendant urine loss of 25(OH)D and $1\alpha,25(OH)_2D$ bound to vitamin D-binding protein (DBP). Urinary excretion of DBP increases proportionally to the degree of proteinuria in nephrotic syndrome, whereas absorption of vitamin D is normal. In more advanced stages, 1α-hydroxylase deficiency also ensues, thus contributing to vitamin D deficiency, calcium malabsorption, hypocalcemia, and secondary hyperparathyroidism.

See Also the Following Articles

Celiac Disease • Crohn's Disease • Cystic Fibrosis • Dietary Fiber • Dietary Reference Intakes (DRI): Concepts and Implementation • Nutrition in Aging • Short Bowel Syndrome • Small Intestinal Motility • Small Intestine, Absorption and Secretion

Further Reading

Abrams, S. A. (2001). Calcium turnover and nutrition through the life cycle. *Proc. Nutr. Soc.* **60**, 283–289.

Chapuy, M.-C., and Meunier, P. J. (1997). Vitamin D insufficiency in adults and the elderly. *In* "Vitamin D" (D. Feldman, F. H. Glorieux, and L. W. Pike, eds.), p. 679. Academic Press, San Diego.

Charoenphandhu, N., Limlomwongse, L., and Krishnamra, N. (2001). Prolactin directly stimulates transcellular active calcium transport in the duodenum of female rats. *Can. J. Physiol. Pharmacol.* **79**, 430–438.

Civitelli, R., Ziambaras, K., and Leelawattana, R. (1999). Pathophysiology of calcium, phosphate, and magnesium absorption. *In* "Metabolic Diseases and Clinically Related Diseases" (L. V. Avioli and S. M. Kane, eds.), pp. 165–205. Academic Press, San Diego, CA.

Consensus Development Conference Panel (1994). Optimal calcium intake. *J. Am. Med. Assoc.* **272**, 1942–1948.

Consensus Development Conference (1994). Prophylaxis and treatment of osteoporosis. *Am. J. Med.* **90**, 107–110.

Ensrud, K. E., Duong, T., Cauley, J. A., Heaney, R. P., Wolf, R. L., Harris, E., and Cumming, S. R. (2000). Low fractional calcium absorption increases the risk for hip fracture in women with low calcium intake: Study of Osteoporotic Fractures Research Group. *Ann. Intern. Med.* **132**, 345–353.

Holick, M. F. (1994). Vitamin D: New horizons for the 21st century. *Am. J. Clin. Nutr.* **60**, 619–630.

Lo, C. W., Paris, P. W., and Clemens, T. L. (1985). Vitamin D absorption in healthy subjects and in patients with intestinal malabsorption syndromes. *Am. J. Clin. Nutr.* **42**, 644–649.

North American Menopause Society (2001). The role of calcium in peri- and postmenopausal women: Consensus opinion of the North American Menopause Society. *Menopause* **8**, 84–95.

Pattanaungkul, S., Riggs, B. L., Yergey, A. L., Vieira, N. E., O'Fallon, W. M., and Khosla, S. (2000). Relationship of intestinal calcium absorption to 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] levels in young versus elderly women: Evidence for age-related intestinal resistance to 1,25$(OH)_2$D action. *J. Clin. Endocrinol. Metab.* **85**, 4023–4027.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food, and Nutrition Board, Institute of Medicine (1997). Calcium. *In* "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride," pp. 71–145. National Academy Press, Washington, DC.

Weaver, C. M. (1994). Age related calcium requirements due to changes in absorption and utilization. *J. Nutr.* **124**, 1418S–1425S.

Webb, A. R., Kline, L., and Holick, M. F. (1988). Influence of season and latitude on the cutaneous synthesis of vitamin D3. *J. Clin. Endocrinol. Metab.* **67**, 373–378.

Campylobacter

MANFRED KIST
University Hospital Freiburg, Germany

anthropozoonosis A transmissible infection that occurs in animals and humans.

ataxia Coordination problems, such as clumsy or awkward movements and unsteadiness, often beginning with difficulty in walking.

human leukocyte antigen B27 A genetically determined self-antigen (protein) localized on tissue cell surfaces.

ophthalmoparesis A weakness of the eye muscles.

toxic megacolon An acute distension of the large bowel, which may complicate severe colitis.

Spiral-shaped bacteria were first seen by Theodor Escherich in 1886 in the gut of children suffering from "cholera infantum" but effective isolation techniques were not developed until the later part of the 20th century by Butzler and Skirrow. The genus *Campylobacter*, including 16 species, belongs to the family Camplyobacteriaceae in the rRNA superfamily VI of the epsilon subdivision of the alpha proteobacteria. The most important species are *C. jejuni, C. coli*, and *C. fetus ssp. fetus*. *Campylobacter* are microaerophilic, gram-negative, spirally curved rods, which are rendered highly motile by two polar flagella. The complete genome sequence of the *C. jejuni* reference strain NCTC11168 consists of 1,641,481 bp and is predicted to encode 1654 proteins and 54 RNA species.

EPIDEMIOLOGY

Campylobacter infection is a worldwide anthropozoonosis, with *C. jejuni* probably the most important bacterial causative agent of infectious diarrhea in humans. *C. jejuni* and *C. coli* live predominantly as commensals in a wide range of wild and domestic birds and mammals, including poultry, dairy cows, and domestic pets. Sources of human *Campylobacter* infection therefore are animals and animal products, especially raw milk and poultry meat, as well as untreated water, which is frequently contaminated by birds or farm animals. Water, milk, and poultry have also been involved in community outbreaks. The infectious dose is 500 organisms in milk. *C. jejuni* is an important cause of traveler's diarrhea.

CLINICAL FEATURES AND COMPLICATIONS

Enteritis is the most common illness associated with *Campylobacter* infection. A prodromal period with malaise, headache, back pain, and myalgia is followed by the acute phase, often starting with a short peak of fever up to 40°C, severe abdominal cramps, and

 Copyright 2004, Elsevier (USA). All rights reserved.

Lo, C. W., Paris, P. W., and Clemens, T. L. (1985). Vitamin D absorption in healthy subjects and in patients with intestinal malabsorption syndromes. *Am. J. Clin. Nutr.* **42**, 644–649.

North American Menopause Society (2001). The role of calcium in peri- and postmenopausal women: Consensus opinion of the North American Menopause Society. *Menopause* **8**, 84–95.

Pattanaungkul, S., Riggs, B. L., Yergey, A. L., Vieira, N. E., O'Fallon, W. M., and Khosla, S. (2000). Relationship of intestinal calcium absorption to 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] levels in young versus elderly women: Evidence for age-related intestinal resistance to 1,25$(OH)_2$D action. *J. Clin. Endocrinol. Metab.* **85**, 4023–4027.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food, and Nutrition Board, Institute of Medicine (1997). Calcium. *In* "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride," pp. 71–145. National Academy Press, Washington, DC.

Weaver, C. M. (1994). Age related calcium requirements due to changes in absorption and utilization. *J. Nutr.* **124**, 1418S–1425S.

Webb, A. R., Kline, L., and Holick, M. F. (1988). Influence of season and latitude on the cutaneous synthesis of vitamin D3. *J. Clin. Endocrinol. Metab.* **67**, 373–378.

Campylobacter

MANFRED KIST
University Hospital Freiburg, Germany

anthropozoonosis A transmissible infection that occurs in animals and humans.

ataxia Coordination problems, such as clumsy or awkward movements and unsteadiness, often beginning with difficulty in walking.

human leukocyte antigen B27 A genetically determined self-antigen (protein) localized on tissue cell surfaces.

ophthalmoparesis A weakness of the eye muscles.

toxic megacolon An acute distension of the large bowel, which may complicate severe colitis.

Spiral-shaped bacteria were first seen by Theodor Escherich in 1886 in the gut of children suffering from "cholera infantum" but effective isolation techniques were not developed until the later part of the 20th century by Butzler and Skirrow. The genus *Campylobacter*, including 16 species, belongs to the family Camplyobacteriaceae in the rRNA superfamily VI of the epsilon subdivision of the alpha proteobacteria. The most important species are *C. jejuni, C. coli*, and *C. fetus ssp. fetus*. *Campylobacter* are microaerophilic, gram-negative, spirally curved rods, which are rendered highly motile by two polar flagella. The complete genome sequence of the *C. jejuni* reference strain NCTC11168 consists of 1,641,481 bp and is predicted to encode 1654 proteins and 54 RNA species.

EPIDEMIOLOGY

Campylobacter infection is a worldwide anthropozoonosis, with *C. jejuni* probably the most important bacterial causative agent of infectious diarrhea in humans. *C. jejuni* and *C. coli* live predominantly as commensals in a wide range of wild and domestic birds and mammals, including poultry, dairy cows, and domestic pets. Sources of human *Campylobacter* infection therefore are animals and animal products, especially raw milk and poultry meat, as well as untreated water, which is frequently contaminated by birds or farm animals. Water, milk, and poultry have also been involved in community outbreaks. The infectious dose is 500 organisms in milk. *C. jejuni* is an important cause of traveler's diarrhea.

CLINICAL FEATURES AND COMPLICATIONS

Enteritis is the most common illness associated with *Campylobacter* infection. A prodromal period with malaise, headache, back pain, and myalgia is followed by the acute phase, often starting with a short peak of fever up to 40°C, severe abdominal cramps, and

 Copyright 2004, Elsevier (USA). All rights reserved.

watery stools, sometimes with gross blood. Symptoms usually resolve within a few days, but excretion of bacteria may continue for several weeks. Fatal outcome is rare; case fatality rates range between 0.05 and 0.3 per 1000 infections. Severe colitis is one of the more frequent acute complications, whereas toxic megacolon, appendicitis, pancreatitis, and cholecystitis are rare. Bacteremia may occur in approximately one case per 1000 infections. Septic abortion has been reported occasionally. Neonatal meningitis can be a life-threatening complication in such cases. Reactive arthritis as a late complication can occur 1–2 weeks after onset of bowel symptoms in approximately 1% of patients, especially in human leukocyte antigen B27-positive individuals, and in up to 20% is complicated by urethritis and uveitis (Reiter's syndrome).

Postinfectious Guillain-Barré syndrome (GBS) and Miller–Fisher syndrome (MFS), although rare, are the most important complications following *Campylobacter* enteritis in approximately 1 to 3 per 1000 cases. GBS as the predominant condition is characterized by a bilateral ascending paralysis, which can involve cranial nerves, resulting in respiratory insufficiency with a need for ventilation; MFS occurs with ophthalmoparesis, ataxia, and tendon areflexia. A remarkably high proportion of *Campylobacter*-related GBS patients develop serum antibodies against gangliosides, sialic acid-containing glycosphingolipids that are highly concentrated in the nervous tissues, especially in the myelin sheaths of the nerve axons.

C. fetus ssp. fetus in compromised patients is involved in extraintestinal infections, such as septicemia, endocarditis, thrombophlebitis, meningoencephalitis, and osteomyelitis.

LABORATORY DIAGNOSIS

Primary plating of stool samples on selective medium is the optimal method for recovering *Campylobacter* during the acute stage of enteritis. Identification of *Campylobacter* at the species level can be accomplished by using classical biochemical methods, species-specific polymerase chain reaction, and whole-cell fatty acid gas chromatography.

TREATMENT

Most *Campylobacter* enteritis patients do not require specific treatment other than oral replacement of fluid and electrolytes. Antibiotics are indicated in cases with protracted febrile illness, if infections relapse, and in immunocompromised patients. The drug of choice is erythromycin stearate 500 mg bid. Alternative antibiotics are fluoroqinolones, such as ciprofloxacin 250 mg bid, or amoxicillin 500–1000 mg tid or tetracycline 250 mg tid. All drugs are given for 5–7 days.

See Also the Following Articles

Diarrhea, Infectious • Foodborne Diseases • Food Safety • Traveler's Diarrhea

Further Reading

Kist, M., and Bereswill, S. (2001). *Campylobacter jejuni. Contrib. Microbiol.* **8**, 150–165.

Nachamkin, I., and Blaser, M. J. (eds.) (2000). "*Campylobacter*," 2nd Ed. ASM Press, Washington, DC.

On, S. L. (2001). Taxonomy of *Campylobacter, Arcobacter, Helicobacter* and related bacteria: Current status, future prospects and immediate concerns. *Symp. Ser. Soc. Appl. Microbiol.* **30**, 1S–15S.

Skirrow, M. B., and Blaser, M. J. (1995). *Campylobacter jejuni. In* "Infections of the Gastrointestinal Tract" (M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant, eds.), pp. 825–848. Raven Press, New York.

Van Vliet, A. H., and Ketley, J. M. (2001). Pathogenesis of enteric *Campylobacter* infection. *Symp. Ser. Soc. Appl. Microbiol.* **30**, 45S–56S.

Cancer, Overview

WILLIAM M. GRADY
Vanderbilt University Medical Center

apoptosis Form of cell death that follows a characteristic pattern of molecular and cellular events, resulting in nuclear condensation and cell membrane disruption.
carcinogenesis Process in which normal tissue develops into cancer.
DNA caretaker gene Sequence on a gene encoding a product that normally maintains the fidelity of genomic DNA.
epigenetic Genomic code other than the base pair code of the DNA sequence. The epigenetic code consists of DNA methylation and histone modifications. e.g., CpG DNA methylation.
genomic instability Loss of fidelity of DNA in cells, resulting in changes to the genomic DNA code.
neoangiogenesis Growth of new blood vessels in tissues.
oncogene Genetic sequence encoding a product that promotes cell transformation or cancer formation. Oncogenes usually result from gain-of-function mutations of normal genes, called proto-oncogenes.
tumor suppressor gene Genetic sequence encoding a product that normally suppress tumor formation; inactivation by mutation or transcriptional silencing contributes to tumor formation.

Cancer is a pathologic condition that results from the abnormal growth and death of cells. Cancer causes disease because of mass effects from the growth of the primary tumor or from tumor invasion into adjacent tissues and/or tumor metastases. At the biological level, defining features of cancer include (1) unregulated (or autonomous) cell growth, (2) unregulated programmed cell death (apoptosis), (3) immortalization/loss of senescence, (4) genomic instability, (5) neoangiogenesis, and (6) invasive growth and metastatic behavior. All neoplasms, which include benign and malignant neoplasms, display unregulated growth. However, only malignant neoplasms invade tissues and spread to distant sites in the body (metastasize). Malignant neoplasms that originate in epithelial cells are called carcinomas, and cancers of mesenchymal cell origin are called sarcomas.

INTRODUCTION

The complex process of cancer formation is characterized by alterations in the morphology and behavior of normal cells; the alterations occur as a consequence of alterations in the cells' genomic DNA. In light of the myriad abnormalities in cancers, it has taken decades to begin to arrive at an understanding of the fundamental processes that drive cancer formation. The initial aspect of cancer that was understood and characterized was the histologic and anatomic aberrations found in the affected tissues. This original understanding of cancer as a histologic and biologic process has given way to an understanding that cancer is fundamentally a molecular and a genetic disease. Indeed, it is now appreciated that the histologic changes characteristic of cancer are the consequence of fundamental alterations in the biology of cancer cells that in turn result from the accumulation of mutations in genomic DNA. The genetic and epigenetic alterations of genomic DNA, which can broadly be considered mutations, play a key role in cancer formation by altering the behavior of proto-oncogenes or tumor suppressor genes. For a cancer to form, however, mutations, usually of both oncogenes and tumor suppressor genes, must first occur in a clonal cell population. These alterations allow the cells to bypass a complex regulatory system that normally precisely controls cell proliferation and behavior. It takes a critical number of DNA alterations that, in aggregate, supersede these regulatory mechanisms and lead once-normal cells to behave as cancer cells. This process has been termed multistage or multistep carcinogenesis and is considered to be the most common way that cancers form.

FUNDAMENTAL ASPECTS OF CARCINOGENESIS

The adult human is composed of approximately 10^{15} cells, and many of them replicate on a daily basis. These cells undergo regulated proliferation in order to maintain the homeostasis of the tissues in the various organs in the body. Examples of organs in which cell proliferation is occurring at high rates include the intestines, where the epithelium turns over approximately every 7 days, and the bone marrow, which has stem cells that can divide as often as every 24 hours. Those cells that have the capacity to proliferate and replenish the cellular mass in a tissue are called stem cells. It is estimated

 Copyright 2004, Elsevier (USA). All rights reserved.

that there are 10^{12} stem cell divisions every day, which normally maintains the integrity of the organ by replenishing cells that either die or are shed from the organism. In light of the constant proliferation of these cells, there are multiple checks and balances on the system to prevent inappropriate cell growth and also to regulate programmed cell death, which is also called apoptosis. Cell proliferation and apoptosis are regulated by processes such as cell : cell contact growth inhibition, which halts cell division when epithelial cells come into contact with one another; anchorage dependence, which prevents epithelial cells from growing unless they are in contact with extracellular matrix; and growth factor dependence.

As a consequence of the multiple mechanisms that control cell growth regulation, cancers do not arise as the result of a single mutational event. Rather, cancer cells arise from the accumulation of multiple mutations, which in sum total allow a cell to circumvent the multiple regulatory mechanisms that maintain homeostasis in the organ. These mutations occur through a variety of different mechanisms, including (1) inheritance of mutated alleles from parent to child [e.g., familial adenomatous polyposis (FAP) with germ-line mutations in *APC*], (2) exposure to carcinogenic agents, such as ionizing radiation, ultraviolet radiation, or chemical carcinogens, and (3) endogenous processes such as errors in DNA replication and the intrinsic chemical instability of certain DNA bases. Normally, cells can repair these mutations and, in fact, have evolved a variety of overlapping repair mechanisms to be able to do so. However, on occasion, these repair processes do not function correctly and DNA base changes, or mutations, occur and are retained in the genomic DNA of the cell. These mutations can accumulate over time in clonal populations and result in these cells becoming independent of the regulatory mechanisms that normally control their behavior. In fact, in cultured cells, which are already immortalized, it has been shown to take at least two added genetic changes to transform the cells into cells with tumorigenic competence.

MUTAGENESIS

Mutations can occur because DNA is a complex molecule that is susceptible to chemical damage. The damage to DNA can cause base pair changes in the DNA and alter the genomic code of the cell. There are a variety of different types of damage, or mutations, that can occur and result in changes in the proteins that are ultimately encoded for by the DNA. These alterations to DNA include the following events: (1) point mutations, which cause amino acid substitutions (missense mutations), frameshift mutations, or mutations that convert amino acid codons to stop codons and lead to protein truncation (nonsense mutations), (2) chromosomal imbalance or instability, resulting in gene amplification or overexpression of a gene, (3) gene deletions secondary to chromosomal loss or chromosomal breakage and fusion, and (4) epigenetic alterations, such as methylation of cytosines in CpG islands, leading to transcriptional silencing.

The most common proximal causes of these DNA alterations are believed to be exposure to environmental mutagens, such as benzopyrene in tobacco smoke, and spontaneous damage to DNA from reactive molecules that are generated during normal cell metabolism or during DNA replication. Thus, it has been estimated that approximately 70% of cancer in Western populations can be attributed to diet and lifestyle, with tobacco exposure accounting for 30% of this cancer incidence. In addition, dietary deficiencies of certain micronutrients such as folate and selenium are believed to influence DNA mutation rates. Spontaneous DNA damage is also a frequent event because of the inherent instability of the DNA molecule. In fact, depurination from breakage of the N-glycosidic bond connecting purines to deoxyribose occurs at the rate of 10^4 events/cell/day and deamination of cytidine to uridine occurs about 20 times/cell/day. The most dangerous mutagenic event that occurs as a consequence of reaction of DNA with by-products of oxidative metabolism is the production of 8-hydroxydeoxyguanosine, which occurs at a rate of $2 \times 10^4 - 10^5$ lesions/cell/day. The net effect of these mutations is to lead to alterations in oncogenes and tumor suppressor genes that ultimately disrupt signaling pathways and transform cells.

GENERAL CLASSES OF CANCER GENES

The genes that are mutated in cancer promote the tumorigenesis process through a variety of different mechanisms that alter cell behavior. As originally proposed by Bishop and Varmus, the genes involved in tumorigenesis can be classified as oncogenes or tumor suppressor genes. More recently, Vogelstein has proposed two additional categories, DNA caretaker genes and landscaper genes.

Classic oncogenes are genes that can deregulate cell growth through gain-of-function mutations. They are referred to as proto-oncogenes until they are mutated, at which time they are called oncogenes. The activation of these genes can occur through point mutations, gene amplification, or chromosomal rearrangements. A classic example of an oncogene that is commonly mutated in colon cancer is *KRAS*, which encodes a protein that

regulates signal transduction from growth factor receptors; its mutation commonly results in increased signal transduction pathway activity and subsequent cell proliferation (see Fig. 1). Oncogenic mutations in *KRAS* occur in codons 12, 13, and 61 and result in loss of GTPase activity of the protein, which results in it always being in an active conformation.

Tumor suppressor genes, on the contrary, are genes that normally regulate a variety of processes that maintain normal tissue homeostasis and cell growth. The loss of function of tumor suppressor genes is the key event that contributes to tumor formation. Usually, a tumor suppressor gene will show inactivation of both alleles in cancers. This principle is known as Knudson's two-hit hypothesis and postulates that biallelic inactivation of a tumor suppressor gene is necessary to eliminate function of the gene. Mechanisms through which this can occur include genomic DNA deletion, also known as loss of heterozygosity (LOH), biallelic somatic mutations, and aberrant DNA methylation. Examples of tumor suppressor genes include *TP53, CDH1* (the gene for E-cadherin), and *APC*. Each one of these genes regulates normal cell growth in a specific manner. For instance, E-cadherin is a membrane-associated calcium-dependent homophilic protein that regulates cell:cell adhesion and cell polarity. Many tumors, including diffuse gastric cancer and prostate cancer, show mutational inactivation or transcriptional silencing of *CDH1*, which is believed to cause enhanced tumor cell motility and invasiveness.

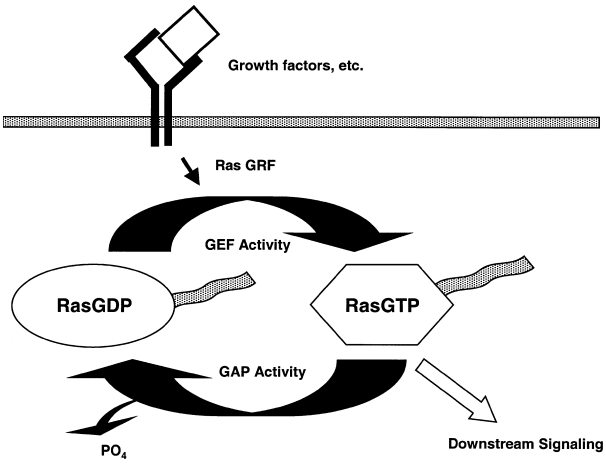


FIGURE 1 Schematic diagram of activation of ras. Ras is activated by phosphorylation of GDP secondary to activation of one of several different pathways, such as the mitogen-activated protein kinase pathway. The phosphorylated ras can then transmit the signal to downstream signaling proteins, which eventually mediate cell proliferation. Ras has endogenous GTPase activity that normally restores it to the inactive state. Mutations in codons 12 and 13, which are the most common type of mutation observed in colon cancer, do not have GTPase activity and result in a constantly activated ras. GRF, GTPase-releasing factor; GEF, guanine nucleotide exchange factor; GAP, GTPase-activating protein.

There is at least one additional category of genes that can be considered to have a discrete role in carcinogenesis. This class of genes includes genes that encode proteins that regulate DNA fidelity, and they have been referred to as DNA caretaker genes. Their inactivation is believed to contribute to the hypermutable state observed in most cancer cells that leads to genomic instability. A classic example of a DNA caretaker gene is the mutation mismatch repair gene, *MLH1*. The product of *MLH1* is one member of a class of proteins called mutation mismatch repair proteins, which correct base–base mismatch and insertion–deletion loop errors that occur during DNA replication. *MLH1* inactivation results in a form of genomic instability called microsatellite instability, increasing the mutation rate up to 1000-fold in expressed sequences. Microsatellite instability results in frameshift mutations in a number of tumor suppressor genes that carry microsatellite sequences in coding regions of the genes, including *TGFBR2, MSH3*, and *IGF2R* (see Fig. 2). Other DNA repair genes that are inactivated in cancers include *MGMT*, which encodes O^6-methylguanine DNA methyltransferase; *ATM*, which causes ataxia–telangiectasia; *FANCA*, which causes Fanconi anemia, *BLM*, which causes Bloom syndrome; and *CKN1, XPA* (causing Cockayne syndrome and xeroderma pigmentosa 1, respectively), and other members of the mutation mismatch repair family, such as *MSH2* and *MSH6*. These genes encode proteins that regulate a variety of DNA repair processes, such as pyrimidine dimer excision (e.g., *XPA*), DNA cross-linking repair (e.g., *FANCA*), and DNA alkylation repair (e.g., *MGMT*).

MULTISTEP MODEL OF CANCER FORMATION

It is well appreciated that cancer cells typically contain hundreds of mutations. In fact, it has been shown that an average cancer may even have up to 11,000 mutations/genome. The carcinogenesis process, though, is initiated by the occurrence of a single mutation in a critical gene in a stem cell. The likelihood of such an event occurring during a cell division at a particular genetic locus is low, approximately 10^{-10}, which means that only 1 cell in 10 billion is mutated per cell division. Thus, it is only as a result of the large number of stem cell divisions that occur every day

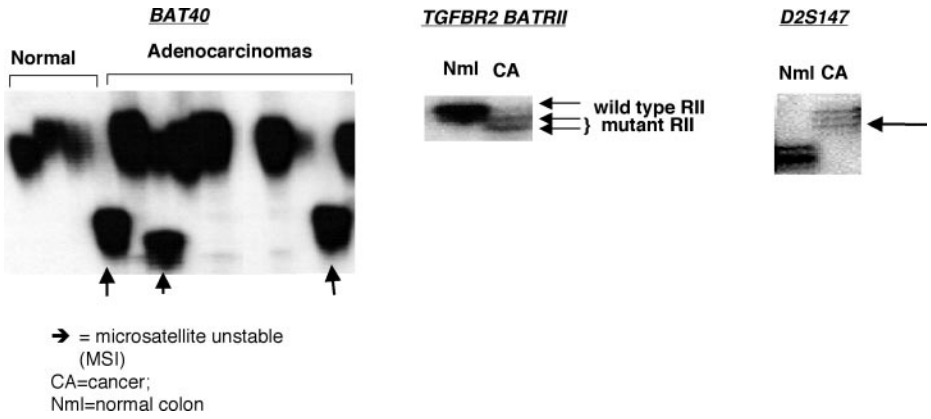


FIGURE 2 Representative example of microsatellite instability analysis at *BAT40*, *TGFBR2 BATRII*, and *D2S147*. Microsatellite instability is most commonly demonstrated by a polymerase chain reaction-based method that generates labeled products amplified from different microsatellite loci, e.g., *BAT40*. The labeled products are subjected to polyacrylamide gel electrophoresis and then visualized. The unstable loci are identified because of their size difference from the normal loci. In this example, the polymerase chain reaction product has been end-labeled with [^{32}P-]ATP and visualized with autoradiography. The microsatellite unstable loci (arrows) are recognized by their size difference from the normal tissue loci.

(approximately 10^{12}/day) that mutations of DNA can have any consequence. Indeed, even though it is estimated that most adult humans have mutated cells inside of them, these people do not have cancers because of the multiple regulatory mechanisms in operation at the cellular level that prevent functional mutations from having a transforming effect on the cells. However, over time, these stem cells can sequentially acquire additional mutations that can eventually overcome these regulatory mechanisms. The exact number of mutations needed to form a cancer cell is not known, but it appears from studies of colon cancer that at least five mutations are needed. Thus, in light of the low rate of DNA mutation and the need for multiple mutations in crucial tumor suppressor genes and oncogenes, it becomes clear that forming a cancer is not a simple process. Furthermore, for a cancer to form, the initial mutations must be coupled to cell proliferation and must promote the clonal expansion of the cells. The clonal expansion is believed to be important in the tumorigenesis process because it creates a cadre of cells that have one mutation in either an oncogene or tumor suppressor gene that can then acquire additional mutations that favor progression to a frank cancer. This process of serial mutations coupled with subsequent waves of clonal growth of the mutant cells is believed to generate a recognizable precancer phase that precedes the appearance of frank carcinomas. For example, in colon cancer, it has been shown that a precancer phase (colon adenoma) usually precedes the cancer phase (colon adenocarcinoma), and that this precancer phase progresses through at least two histological stages before becoming a cancer. Furthermore, this adenoma–carcinoma sequence has easily recognized histologic phases that are accompanied by specific genetic alterations, which presumably drive the cancer formation process (see Figs. 3 and 4). Thus, the carcinogenesis process follows the principles of Darwinian evolution, with cells that acquire mutations that promote their proliferation outcompeting other cells in the precancer. The cancer cells are those cells that show the highest level of reproductive fitness, which at the cellular level is measured by cell proliferation. A dramatic example of this process is the development of cancer cell lines that are used in laboratories around the world. The vast majority of these cell lines, such as HeLa cells, which is a cell line derived from a cervical cancer, or HCT116, a colon cancer cell line, have even outlived the individuals from whom they were derived. Thus, from an evolutionary perspective at the cellular level, these cancer cells have been highly successful.

The concept of multistep carcinogenesis has been most clearly shown in the context of cervical cancer and colon cancer formation, but is likely true of all cancers. Important implications of this model of cancer formation are as follows: (1) there is a precancer phase that may be more amenable to treatment because the cells

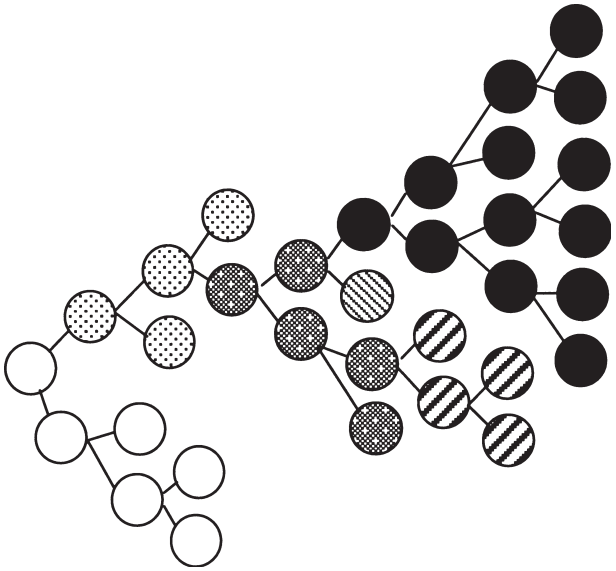


FIGURE 3 Multistep carcinogenesis and Darwinian evolution of cancer. The circles represent cells, and each cell division is shown as two cells to the right of the parent cell, connected to it by lines. The clonal evolution occurring during the cell proliferation is shown by the acquisition of new patterns, representing the acquisition of new genetic makeup or epigenetic alterations that favor tumor cell growth. It can be appreciated that some clones in the cell mass stop proliferating secondary to lethal mutations, whereas others proliferate better than the others do.

have not become as independent of the regulatory mechanisms as they will be when they become frank cancer cells and (2) there may be multiple parallel evolutionary paths for individual cancers as well as for cancer cell clones in a single cancer. Indeed, the phenomenon of intratumoral genetic heterogeneity has been appreciated indirectly by our limited ability to treat most cancers effectively. The polyclonal nature of tumors combined with their Darwinian behavior makes them very effective at acquiring resistance to most chemotherapeutic agents. Furthermore, many cancers, such as Hodgkin lymphoma, have a bimodal age distribution for peak incidences of disease consistent with the idea that there are at least two independent "evolutionary" pathways through which cancers can arise (see Fig. 3). Some of the central biological effects of these mutations of tumor suppressor genes, oncogenes, and DNA caretaker genes on the promotion of tumor formation will be assessed further in the following discussions.

REGULATION OF CELL GROWTH

The mechanisms that control cell division and cell cycle control in dividing cells are highly conserved and are often abnormally regulated in neoplastic cells. Furthermore, a variety of mutations in the genes that govern the cell cycle have been found in many different cancers, providing further evidence that deregulation of cell cycle control contributes to cancer formation. The cell cycle is classically divided into four phases, G_1, S, G_2, and M, and cells not in any one of these phases are in G_0, a resting state. Cells begin dividing by progressing from G_0 to G_1 (gap 1), at which time cell growth occurs, and the cell "determines" if it is ready to initiate DNA synthesis. The cell must transition through a critical point in the cell cycle, called the restriction checkpoint, in order to move from G_1 to the next phase, S. Once the cell passes the restriction checkpoint, DNA synthesis is initiated and the genomic DNA of the cell is replicated. The DNA replication process involves the unfolding of the chromatin and nucleosome complexes and the recruitment of DNA polymerases and primases, DNA helicases, single-strand binding proteins, and topoisomerases to replicate the DNA efficiently at a rate of approximately 50 nucleotides/second. After the DNA is replicated, the cell cycle transitions to G_2 (gap 2), during which the fidelity of the DNA is determined and replication errors are corrected. The cell then must go through a second restriction point, the G_2/M restriction checkpoint, in order to enter M phase and undergo mitosis, which is the process of the replicated chromosomes physically separating into two nuclei and the cytoplasm undergoing cytokinesis. The transitions from G_1 to S and G_2 to M are tightly regulated in order to ensure the orderly proliferation of cells.

The checkpoints are regulated by a family of serine/threonine protein kinases called cyclin-dependent kinases (cdks) and kinase-associated proteins called cyclins. The activation of the cdks is dependent on the cyclins, the levels of which fluctuate throughout the phases of the cell cycle. In addition, proteins that activate the cdks [cdk-activating kinases (CAKs)], phosphatases that remove inhibitory phosphates (cdc25), and cdk inhibitors (p15, p16, p18, p19, p21, p27, and p57) all play essential roles in regulating cell cycle progression and cell division. Finally, the regulation of passage through the G_1/S checkpoint is usually dependent on the phosphorylation status of the nuclear phosphoprotein, retinoblastoma (Rb). Rb binds transcription factors E2F and DP1 in a phosphorylation-specific manner, resulting in their inactivation. The cdk–cyclin complexes can phosphorylate Rb, causing the release of E2F, which can then act as a transcription factor for genes that drive cell cycle progression into S phase. A second pathway regulated by p53 can interact with the Rb pathway to ensure the fidelity of DNA

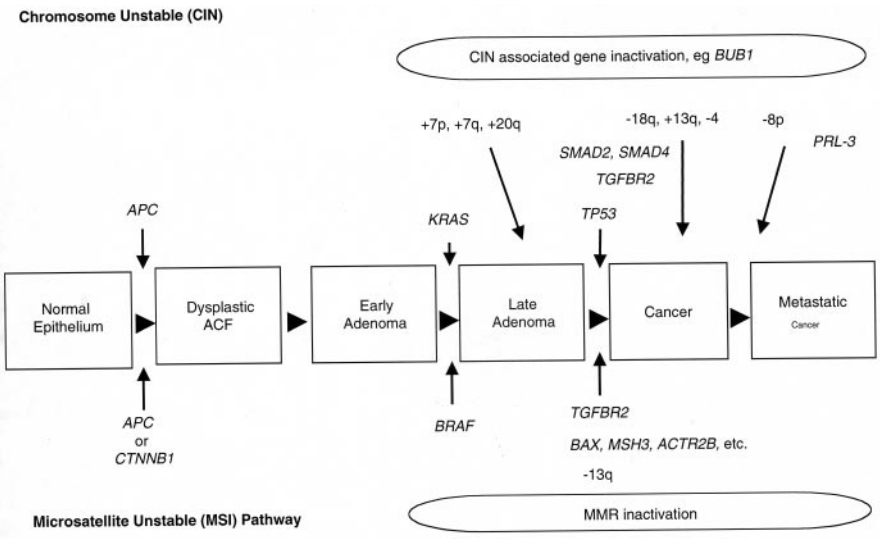


FIGURE 4 An example of multistep carcinogenesis. A schematic representation of the colon adenoma–adenocarcinoma sequence indicates the different stages of colon cancer progression, from the benign neoplasm stage (adenoma) to the malignant neoplasm (cancer). Different patterns of genetic alterations can be identified in colon cancers, depending on whether they display microsatellite instability or chromosome instability (CIN). Microsatellite instability occurs when the mutation mismatch repair (MMR) system is inactivated either through mutations or epigenetic alterations. ACF, Aberrant crypt foci. From Grady (2002). Colorectal cancer: Genetic alterations. *In* "Gastrointestinal Oncology: Principles and Practice" (D. Kelsen, J. M. Daly, S. E. Kern, eds.), pp. 685–701. Copyright Lippincott Williams & Wilkins, reprinted with permission.

replication by preventing transition through the G_1/S restriction point if DNA damage is present.

Cancers have been found to carry activating mutations in the cdks [e.g., *CDK4* mutations in melanomas, cyclin D1 (*CCND1*) overexpression in B cell chronic lymphocytic leukemia secondary to a t(11;19)(q13;p13) chromosomal translocation], inactivating mutations in the cdk inhibitors (*CDKN2A*/p16 mutations in melanomas), and mutations in *RB1*. In addition, many other genetic alterations in tumor suppressor genes and oncogenes have been shown to directly result in altered expression and function of the cyclin-associated proteins.

REGULATION OF CELL DEATH

In addition to mutations in genes that regulate cell proliferation, cancers also commonly carry mutations in genes that regulate cell death. The deregulation of the mechanisms that regulate cell death is important in cancer formation because the maintenance of homeostasis in tissues not only depends on the regulation of cell proliferation, but also on the regulation of cell death. Programmed cell death, also known as apoptosis, is an important mechanism through which organs maintain an appropriate tissue mass and organization. For example, intestinal epithelial cells undergo apoptosis as they reach the luminal surface and are consequently shed. This process, in conjunction with proliferation, maintains the normal villus/crypt architecture in the epithelial layer. Tumor cells are often resistant to apoptosis, and this evasion of apoptosis appears to play a significant role in the clonal expansion of tumor cells and in the ultimate formation of cancers.

Virtually all normal cells are susceptible to apoptosis and can be induced to undergo this process by a variety of different physiologic signals, including hypoxia and cytokine exposure [e.g., transforming growth factor-β (TGF-β), tumor necrosis factor α (TNFα)]. These signals trigger a series of well-orchestrated events that disrupt cellular membranes, degrade the cytoplasmic and nuclear cytoskeletons, and then degrade chromosomes and fragment the DNA. The proteins involved in mediating apoptosis can be roughly divided into sensors and effectors. The sensors detect alterations in the extracellular environment [e.g., interleukin-3 (IL-3) binding to IL-3R] and/or the intracellular environment, and, if appropriately cued, trigger the effectors, which consist of a family of proteases called caspases as well as proteins that ultimately regulate mitochondrial release of cytochrome c and effect mitochondrial membrane potential states. Proteins that play a role in this process

include members of the Bcl-2 family, including Bax, Bid, Bcl-XL, Bcl-W, and Bim, and APAF1. The evidence to date suggests that there are multiple, redundant regulatory and effector components involved in apoptosis.

Cancer cells acquire resistance to apoptosis through an array of different mechanisms. One of the most frequent means through which the apoptotic machinery is deregulated is through *TP53* mutations, which occurs in approximately 50% of all cancers. p53 normally plays a central role in mediating apoptosis in response to genotoxic stress. Another common mechanism through which cancers deregulate apoptosis is via disruption of the phosphatidylinositol 3-kinase (PI 3-kinase)/Akt pathway, which regulates antiapoptotic survival signals and frequently plays a role in preventing apoptosis in a significant portion of cancers. For example, *PTEN*, a gene that encodes for a protein that inactivates Akt, is commonly mutated in a variety of cancers, causing increased PI 3-kinase/Akt signaling pathway activity. It is likely that virtually all cancers have acquired genetic or epigenetic alterations that mitigate against apoptosis.

GENOMIC INSTABILITY

Although it is now widely appreciated that cancer is a disease that results from genetic alterations that deregulate cell proliferation, apoptosis, and senescence, the mechanisms responsible for the acquisition and accumulation of these gene alterations are less well understood. Another apparently fundamental process in cell behavior that is altered in virtually all cancers is the maintenance of DNA fidelity or genomic stability. The role of genomic instability in the process of cancer formation has been an area of investigation since 1914, when Theodor Boveri postulated that cancer cells might be aneuploid. In 1993, Aaltonen *et al.* identified a form of genomic instability called microsatellite instability (MSI) in colon cancers arising in people with the cancer family syndrome, hereditary nonpolyposis colon cancer (HNPCC); these individuals have germ-line mutations in the mutation mismatch repair (MMR) genes, *MLH1* or *MSH2*, giving further support to the concept that instability of genomic DNA plays an important role in carcinogenesis. This discovery of MSI in HNPCC colon cancers provided evidence for the mutator phenotype hypothesis proposed by Loeb in 1974. This hypothesis argues that for potential tumor cells to acquire the number of mutations needed to attain a malignant state, they must have a higher than normal mutation rate, which would presumably occur through inactivation of the mechanisms that normally regulate DNA fidelity. In fact, if one assumes the known non-germline mutation rate of approximately 10^{-8}, it is impossible for a cell to accumulate enough gene mutations during a human lifetime to acquire a malignant phenotype. However, the role of genomic instability as an integral mechanism in cancer formation is not universally accepted. Others have concluded that the genetic events observed in cancer can occur under normal mutation rates, without the need to invoke genomic instability, as long as the normal mutation rates are coupled with rounds of clonal expansion that select for gene alterations that promote tumor formation. Nonetheless, despite the controversy over the exact role of genomic instability in colon tumorigenesis, a wealth of data generated over the past 6–8 years has provided significant indirect support for the concept that genomic instability is a common mechanism that plays a central role in the formation of most, if not all, cancers, including cancers of the gastrointestinal system. This loss of genomic stability appears to occur very early in the process of cancer formation and to play a central role in generating a hypermutable state that leads to the accumulation of enough mutations to transform a cell.

Several forms of genomic instability can be observed in cancer in general. Lengauer *et al.* proposed four major categories of alterations: (1) subtle sequence changes, including base substitutions, deletions, or insertions, and microsatellite instability, (2) alterations in chromosome number (aneuploidy, also termed chromosomal instability), (3) chromosome translocations, and (4) gene amplification. The first two categories occur commonly in colon and stomach cancers. The genome of colon cancers, like many solid tumors, is often marked by a complex pattern of chromosome translocations that appears nearly random. This form of genomic instability, "translocation instability," is likely mediated by the same mechanisms that cause aneuploidy, but its role in colon tumorigenesis is still being defined. Gene amplification occurs rarely in colon cancer and thus does not appear to be an important process in colon cancer formation.

SIGNALING PATHWAYS AND NETWORKS IN CANCER FORMATION

As the signaling pathways of many of the growth factor-coupled receptors, integrin receptors, etc. have been elucidated, it has become appreciated that the ultimate biological effects of mutations in cancers are not only a result of the deregulation of the function of the single gene products, but also the result of disruption of the signaling pathway in which the gene products operate.

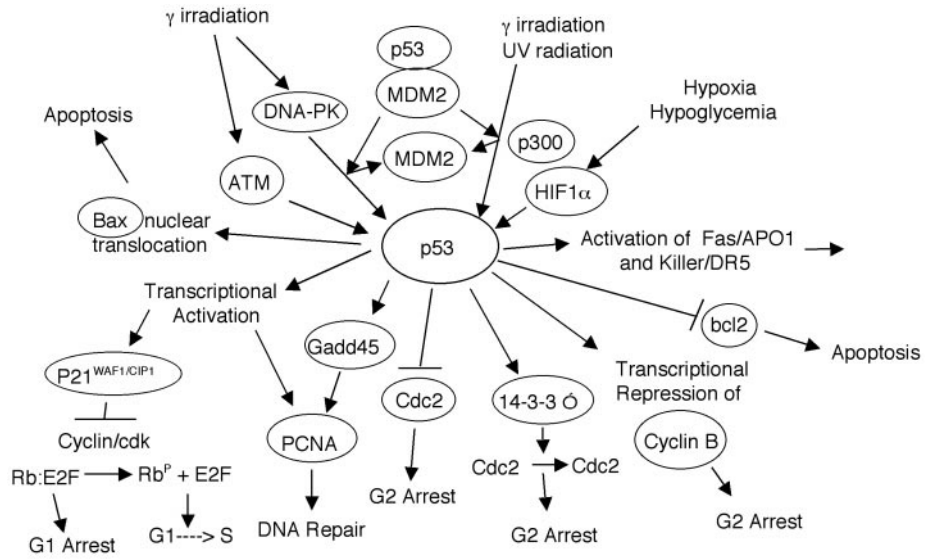


FIGURE 5 Protein network involving p53. p53 interacts with a variety of proteins on exposure of cells to genotoxic events, such as ultraviolet (UV) radiation. These protein–protein interactions result in the phosphorylation and activation of p53, which then either transcriptionally regulates other genes, such as the gene for proliferating cell nuclear agent (PCNA), or directly interacts with the proteins to regulate cell cycle progression or apoptosis. Only a few representative pathways and p53-interacting proteins are shown.

In addition, the multiple mutations in cancers appear to cooperate through effects on these pathways to alter the output of the networks of signaling pathways. Thus, it is now well recognized that cancers result from the mutation of oncogenes and tumor suppressor genes at least in part because these mutations result in the disruption of signaling pathways that regulate cell proliferation and apoptosis. Indeed, it is not uncommon to observe a specific signaling pathway deregulated in a specific cancer type through the inactivation of one of a variety of different members of the signaling path. For instance, in colon cancer, the TGF-β signaling pathway has been shown to be disrupted through mutations of the TGF-β receptor type II (*TGFBR2*), mutations in postreceptor signaling elements (*SMAD2* and *SMAD4*), or up-regulation of pathway inhibitors such as Smad7. Similarly, deregulation of the Wnt signaling pathway can occur through mutations in *APC* or *CTNNB1* (β-catenin), and mutations in one or the other of these genes appear to be nearly universal in colon cancers. Furthermore, the deregulation of these signaling pathways has a multitude of indirect effects, because these pathways also appear to interrelate at multiple levels in the cell, resulting in networks of protein–protein interactions. These networks are often affected in cancers in ways that result in a common effect on target proteins, such as overexpression *CCND1*/cyclin D1. For example, *TP53*, a tumor suppressor gene that is commonly mutated in cancers, interacts with a variety of proteins both in regard to its own regulation and in regard to the regulation of downstream proteins. These effects ultimately regulate cell cycle progression and apoptosis. Consequently, *TP53* mutations have both direct and indirect effects that promote tumorigenesis (see Fig. 5). As our understanding of these signaling pathways and networks improves, our understanding of the global impact of the mutations of different genes on the molecular biology of the cells and cancer formation will improve.

SUMMARY

In general, cancer arises through a complex process of mutagenesis of oncogenes and tumor suppressor genes, followed by clonal expansion of the mutant cells. As a result of the serial acquisition of mutant tumor suppressor genes, oncogenes, and DNA caretaker genes, normal cells transform through a multistep process into malignant cells, resulting in cancer. These gene mutations ultimately lead to cancer formation because they disrupt the regulation of fundamental cellular processes of proliferation, programmed cell death, senescence, and maintenance of DNA fidelity.

Acknowledgments

This work was supported by National Cancer Institute Grant K08 CA77676, Vanderbilt Physician-Scientist Development Award, Glaxo Institute of Digestive Health Basic Research Award, Damon Runyon Lilly Clinical Investigator Award, and a Veterans Administration Advanced Research Career Development Award.

See Also the Following Articles

Anal Cancer • Cholangiocarcinoma • Colorectal Adenocarcinoma • Colorectal Cancer Screening • Diet and Environment, Role in Colon Cancer • Esophageal Cancer • Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD • Familial Risk of Gastrointestinal Cancers • Gallbladder Cancer • Gastric Cancer Surveillance • Genetic Counseling and Testing • Hepatocellular Carcinoma (HCC) • Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) • Pancreatic Cancer • Pancreatic Ductal Adenocarcinoma • Stomach, Adenomas and Carcinomas of the

Further Reading

Aaltonen, L., Peltomaki, P., Leach, F., Sistonen, P., Pylkkanen, L., *et al.* (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816.

Anderson, G. R., Stoler, D. L., and Brenner, B. M. (2001). Cancer: the evolved consequence of a destabilized genome. *BioEssays* **23**, 1037–1046.

Bertram, J. S. (2000). The molecular biology of cancer. *Mol. Aspects Med.* **21**, 167–223.

Boland, C. R., Sato, J., Appelman, H. D., Bresalier, R. S., and Feinberg, A. P. (1995). Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nat. Med.* **1**, 902–909.

Cahill, D., Lengauer, C., Yu, J., Riggins, G., Willson, J., *et al.* (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303.

Eshleman, J., Lang, E., Bowerfind, G., Parsons, R., Vogelstein, B., *et al.* (1995). Increased mutation rate at the *hprt* locus accompanies microsatellite instability in colon cancer. *Oncogene* **10**, 33–37.

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* **100**, 57–70.

Knudson, A. G., Jr., and Strong, L. C. (1972). Mutation and cancer: a model for Wilms' tumor of the kidney. *J. Natl. Cancer Inst.* **48**, 313–324.

Lengauer, C., Kinzler, K., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* **396**, 643–649.

Loeb, L., Springgate, C., and Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. *Cancer Res.* **34**, 2311–2321.

McCormick, F. (1999). Signalling networks that cause cancer. *Trends Cell Biol.* **9**, M53-M56.

Shih, I. M., Zhou, W., Goodman, S. N., Lengauer, C., Kinzler, K. W., and Vogelstein B. (2001). Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.* **61**, 818–822.

Stoler, D. L., Chen, N., Basik, M., Kahlenberg, M. S., Rodriguez-Bigas, M. A., *et al.* (1999). The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15121–15126.

Tomlinson, I., Novelli, M., and Bodmer, W. (1996). The mutation rate and cancer. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14800–14803.

Vogelstein, B., and Kinzler, K. W. (1993). The multistep nature of cancer. *Trends Genet.* **9**, 138–141.

Candidiasis

PAUL F. LEHMANN
Medical College of Ohio

chlamydospore Thick-walled fungal spore, considered a survival structure in most fungi.
fungistatic antifungal Drug that inhibits fungal growth but does not kill fungi.
fungitoxic antifungal Drug that kills fungi.
neutropenic Deficient in neutrophils.

Yeast species of genus *Candida* are the most common etiologic agents causing fungal infections of the gastrointestinal tract. Numerous species are reported and, though *Candida albicans* predominates, it is not rare to find Candida parapsilosis and *Candida tropicalis*. Other species are less commonly isolated and can differ in their susceptibility to typical therapeutic antifungals, including the fungistatic azoles, the fungitoxic polyenes, and caspofungin (Table I). Even in species that have been reported as being susceptible, isolates with resistance to azole antifungals, especially fluconazole, are being reported more commonly. Though there are some differences reported in the virulence of *Candida* species (for example, *C. tropicalis* appears to be more invasive than *C. albicans* in neutropenic mice), the diseases produced by the different species are very similar.

GROWTH PATTERNS OF *CANDIDA ALBICANS*

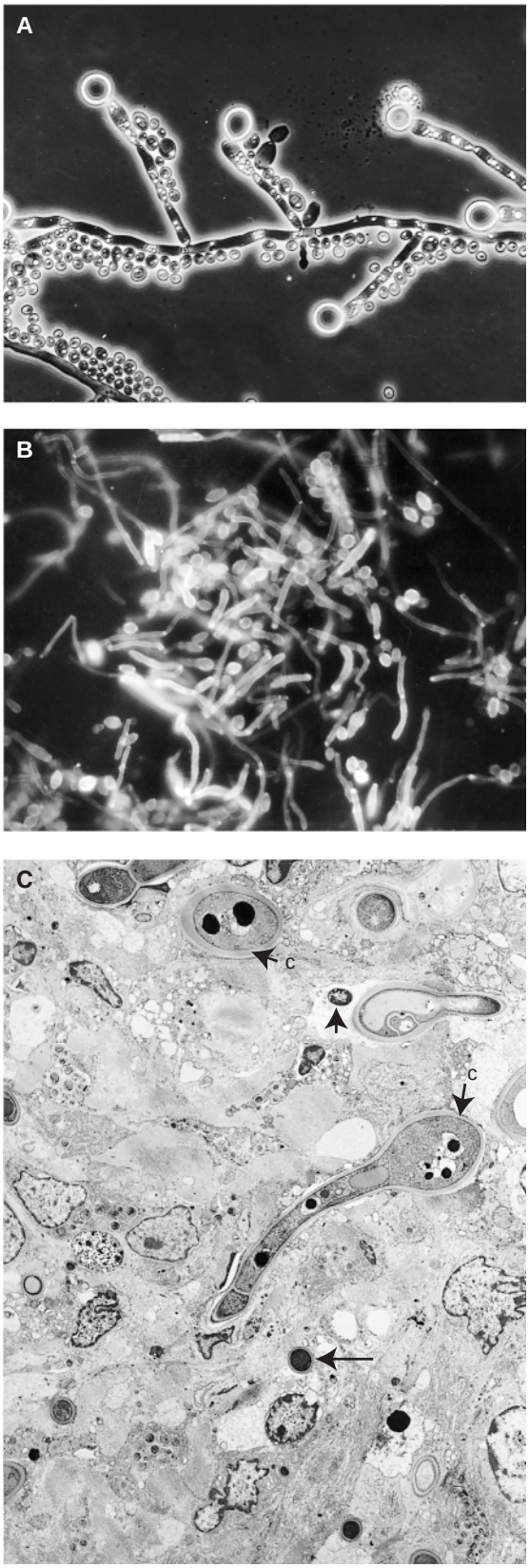
Candida albicans has a variety of morphological forms. It may be found growing as single cells in the form of a budding yeast. It may also grow in a filamentous morphology (Fig. 1). The most invasive form of the fungus has been associated with the filamentous form, but in most situations a mixture of morphological forms is found. Switches between these different morphologies may be necessary to allow both colonization of new epithelial surfaces and release of new cells from colonized epithelium. Indeed, experiments show that epithelial cells tend to bind to the filamentous forms more strongly than to the budding yeasts. The chlamydospore, a morphological form that is produced is culture, is helpful in identifying *C. albicans*. A similar structure has been reported within the intestinal tissues of immunosuppressed mice infected with *C. albicans*. These chlamydospore-like cells show a differential staining similar to that found for chlamydospores produced *in vitro*. Chlamydospores are typically considered to

TABLE I ***Candida*** **Species with Classes and Examples of Antifungal Drugs Active against Them**[a]

	Polyenes	Azoles	Echinocandins
	Amphotericin B	Fluconazole, itraconazole, voriconazole	Caspofungin
Molecular target	Ergosterol	Cytochrome P450–lanosterol demethylase	β-Glucan synthase
C. albicans	+	+[b]	+
C. dubliniensis	+	+[b]	+
C. parapsilosis	+	+	+
C. tropicalis	+	+	+
C. krusei	+	−	+
C. lusitaniae	+[b]	+	+
C. glabrata	+	−	+

[a] +, Normally susceptible; −, usually resistant.
[b] Resistance occasionally found or developing during treatment.

 Copyright 2004, Elsevier (USA). All rights reserved.



be "resting spores," or resistant structures; they might allow the fungus to survive during antifungal therapy and to later recolonize the gastrointestinal tract. Almost never reported in humans, this may be because they have been overlooked or because they are formed only under unusual circumstances. For example, chlamydospores have been detected in bronchoalveolar lavage fluid from one AIDS patient.

PREDISPOSITION AND TISSUE DAMAGE

Candidal diseases vary in their presentation and are significantly modified by predisposing factors in the patient. Frank invasion through the gastrointestinal wall, followed by dissemination of *Candida* via the blood and subsequent invasive infection of other areas of the body, such as the kidneys, is typically only seen when the defenses provided by neutrophils are absent or greatly diminished. This can be the situation when the patient has received cytotoxic therapy for certain cancers or in preparation for stem cell or bone marrow transplantation. High doses of steroids or drugs that cause profound neutrophil dysfunction may also predispose invasive disease. In contrast, when there is adequate neutrophil function, any disease is largely restricted to the mucosal tissues of the gastrointestinal tract. Even so, the overgrowth of the yeasts leads to inflammation and significant discomfort and the disease may spread to adjacent regions, including the oropharynx, the edges of the mouth, and the skin surrounding the anus. All these sites become painful and inflamed. This condition is fairly common in patients with diabetes or with T cell dysfunctions, including AIDS, and it can be responsible for severe diaper (nappy) rash in infants. Candidiasis of the oral mucosa was one of the first signs used to define AIDS; the patients may develop such extensive inflammation of the oral mucosa and esophagus that they find it difficult to eat. In some patients with AIDS, there can be perforations in the esophageal wall. In addition to

FIGURE 1 Morphological forms of *Candida albicans*. (A) Yeasts, filaments, and chlamydospores formed *in vitro*. The chlamydospores, which are approximately 15 μm in diameter, are the spherical terminal spores. (B) Tissue from esophagitis that has been macerated gently, then smeared onto a slide and stained with a fluorescent brightener that binds fungal cell walls. Both yeasts and filaments are present in the lesions. (C) Transmission electron microscopy showing thick-walled chlamydospore-like structures (c) in the gastric submucosa of immunodeficient mice infected with *C. albicans*. The unlabeled arrows indicate a section of a single yeast cell or a filament. Micrographs are provided courtesy of Dr. G. T. Cole, Medical College of Ohio.

morphological switching, production of extracellular phospholipase and proteinase appears important for allowing tissue invasion by *C. albicans*. Complement plays an important role in the induction of inflammation; not only does the yeast activate the alternative pathway directly, but, because antibodies to *C. albicans* cell wall mannoproteins are present in all persons, activation via the classical pathway can also occur.

YEAST MICROFLORA IN HUMANS AND ANIMALS

It can be difficult to prove that some *Candida* species are causes of gastrointestinal infections due to their presence on healthy gastrointestinal mucosa. The microbial flora associated with the gastrointestinal tracts of a large number of different mammals, including aquatic mammals, birds, amphibians, and reptiles, includes *C. albicans* and, most likely, its sibling species *Candida dubliniensis* which until recently was not distinguished as a separate entity. Usually frank invasion of tissues needs to be demonstrated via histology before it is certain that *Candida* is the cause of disease (Fig. 1).

OTHER GASTROINTESTINAL DISEASES CAUSED BY FUNGI

In addition to infections, fungus-induced disease of the gastrointestinal system can be a result of exposure to fungal allergens and toxins, including the mycotoxins, ergot, and mushroom toxins. In persons deficient in trehalase, the trehalose disaccharide found in mushrooms can produce symptoms that are similar to those associated with lactose intolerance. Unusually, drunkenness may follow ingestion of food by persons in which a gastrointestinal stenosis has occurred, because yeasts can produce ethanol via fermentation during digestion.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Fungal Infections • Gastric Infection (non-*H. pylori*) • Microflora, Overview

Further Reading

Calderone, R. A. (2001). "Candida and Candidiasis." American Society for Microbiology Press, Washington, D.C.

Centers for Disease Control and Prevention. (2003). Public Health Image Library. Web site retrieved at http://phil.cdc.gov/phil/default.asp.

Cole, G. T., Halawa, A. A., and Anaissie, E. J. (1996). The role of the gastrointestinal tract in hematogenous candidiasis: From the laboratory to the bedside. *Clin. Infect. Dis.* **22**(Suppl. 2), S73–S88.

Doctor Fungus. (2003). Web site retrieved at http://www.doctorfungus.org/mycoses/human/candida/candida_index.htm.

White, T. C., Holleman, S., Dy, F., Mirels, L. F., and Stevens, D. A. (2002). Resistance mechanisms in clinical isolates of *Candida albicans*. *Antimicrob. Agents Chemother.* **46**, 1704–1713.

World-Wide Web Virtual Library-Biosciences. (2003). *Candida albicans* information. Web site retrieved at http://www.yeastgenome.org/VL-yeast.html

Carbohydrate and Lactose Malabsorption

RICHARD J. GRAND*, ROBERT K. MONTGOMERY*, DENESH K. CHITKARA*, AND HANS A. BÜLLER†
Children's Hospital Boston and Harvard Medical School, and †Sophia Children's Hospital, Rotterdam, The Netherlands

carbohydrate intolerance Clinical symptoms induced by the ingestion of specific forms of carbohydrate (e.g., lactose intolerance, sucrose intolerance, fructose intolerance) or carbohydrates in general.

carbohydrate malabsorption Poor absorption of any given carbohydrate; detected by specific testing (e.g., lactose absorption test, lactose breath test).

lactase deficiency Very low or absent lactase activity; determined by assay of an intestinal biopsy sample (a similar definition holds for deficiency of sucrase or maltase-glucoamylase)

milk intolerance Clinical symptoms induced by the ingestion of milk; can be due to either lactose intolerance or milk protein allergy.

Carbohydrate digestion reduces complex nutrients into smaller components that are suitable for absorption. Starch digestion is accomplished in two phases—an initial intraluminal phase mediated by salivary and pancreatic amylase and a mucosal phase characterized by surface digestion at the intestinal microvillus membrane. Disaccharides undergo surface hydrolysis followed by uptake of monosaccharides by microvillus membrane carriers. Intolerance to carbohydrates occurs when there is any digestive–absorptive imbalance or when bacterial overgrowth occurs in the small intestine.

PHYSIOLOGY OF CARBOHYDRATE ABSORPTION

Carbohydrate Intake

In infants, carbohydrates account for 35–55% of daily calories ingested, and these are mainly as lactose. As weaning foods are introduced, carbohydrate intake varies and approaches the composition commonly found in adults. The average adult ingests 300 g of carbohydrates per day in approximately the following distribution: 52% of daily calories as starch (mainly cereals and potatoes), 37% as sucrose, 5% as lactose (mainly in milk), and 3% as fructose (in fruit and honey). Glycogen, glucose, and maltose are minor constituents of the diet, and cellulose accounts for approximately 4 g of carbohydrates per day.

Digestion

Salivary amylase initiates starch hydrolysis in the mouth, and this process accounts for not more than 30% of total starch hydrolysis. Because salivary amylase is inactivated by an acid pH, no significant hydrolysis of carbohydrates occurs in the stomach. The intraluminal intestinal phase of starch digestion depends on pancreatic amylase to complete hydrolysis, yielding oligosaccharides of varying lengths. This process is extremely rapid; 75% is completed in the proximal 2.5 feet of jejunum within 10 minutes after passage of starch into the small intestine.

The mucosal phase is characterized by surface digestion of oligosaccharides released by amylase. It also includes hydrolysis of disaccharides (maltose, sucrose, and lactose) by specific disaccharidases (maltase–glucoamylase, sucrase–isomaltase, and lactase). The rates of maltose and sucrose hydrolysis are rapid because these disaccharides are readily cleaved, and the released monosaccharides are rapidly absorbed. Lactose digestion is slower, and hydrolysis is the rate-limiting step for the overall process of absorption. The final uptake of monosaccharides is accomplished by the sodium-dependent glucose transporter (SGLT1).

Colonic Salvage

When carbohydrates are not absorbed by the small bowel, they are passed rapidly into the colon as a consequence of the osmolarity of the intraluminal oligosaccharides, the influx of water, and the increase in motility. In the colon, they are converted to short-chain fatty acids and hydrogen gas by the bacterial flora, producing acetate, butyrate, and propionate. The short-chain fatty acids are absorbed by the colonic mucosa, and this route salvages malabsorbed carbohydrates for energy utilization by colonic epithelial cells. This is a mechanism by which the adult colon salvages

 Copyright 2004, Elsevier (USA). All rights reserved.

carbohydrate, especially from wheat starch, and the newborn colon salvages lactose. This fermentative process not only conserves nutritionally important carbohydrate, but it also serves as the basis for the breath hydrogen test (discussed later).

CLINICAL PRESENTATION

Symptoms and Signs of Carbohydrate Malabsorption

Considering clinical symptoms induced by the ingestion of carbohydrates, the term "carbohydrate intolerance" is often applied. Symptoms may be induced by the ingestion of carbohydrate (as in cystic fibrosis) or specific sugars (e.g., lactose, sucrose, or fructose). Either nonspecific carbohydrate or specific sugar intolerances (e.g., lactose intolerance) can produce similar symptoms. These include abdominal pain, cramps, or distension; nausea; flatulence; and diarrhea or vomiting. The abdominal pain may be crampy in nature and may be periumbilical or lower quadrant. Borborygmi may be audible. Carbohydrate intolerance generally produces abnormal stools, which are usually bulky, frothy, and watery. In severe cases, mostly in infants, acidosis and dehydration may be a problem. Vomiting after lactose ingestion is often seen in adolescents.

Clinical symptoms may mimic those of the irritable bowel syndrome. Careful clinical history may point to the correct diagnosis; however, this may be established only after appropriate testing.

PATHOPHYSIOLOGY

Several factors account for the variability of symptoms produced by carbohydrate ingestion in people who are intolerant. This pathophysiology is of particular importance in those with lactose intolerance accompanying low lactase activity, but similar mechanisms apply to people with sucrose intolerance and other sugar intolerances, and to those with intraluminal phase defects in carbohydrate digestion. Important factors include the osmolarity and fat content of the food in which the sugar is ingested, the rate of gastric emptying, individual sensitivity to intestinal distension produced by the osmotic load of unhydrolyzed carbohydrate in the upper small bowel, the rate of intestinal transit, and the response of the colon to the carbohydrate load. In general, the higher the osmolarity of gastric contents and the higher the fat content of the diet containing the specific sugar involved, the slower the gastric emptying and the lesser the symptoms induced by the carbohydrate or sugar.

Different individuals appear to have more or less sensitivity to abdominal distension and patients complain differently when ingested carbohydrates stimulate an influx of water into the lumen of the small intestine or the production of gas that leads to distension of the colon. Those with greater tolerance will report fewer symptoms. These subjective responses are difficult to quantify. Intestinal transit is also influenced by the quality of the diet and individual motility patterns. Accordingly, some lactose- or sucrose-intolerant people experience very rapid movement of sugar to the cecum, whereas others have slower motility. Fecal flora are known to adapt to maldigested carbohydrate. Thus, if an offending carbohydrate or sugar (e.g., lactose) is provided slowly over a long period of time in many "intolerant" people, the flora may adapt to the load, and symptoms produced by gas and acid in the colon may be reduced or eliminated. This mechanism of lactose tolerance commonly occurs in people with low lactase levels; this accounts for the discrepancy between measured "lactase deficiency" or "lactose malabsorption" and lactose tolerance.

The term "carbohydrate malabsorption" is generally reserved for those patients for whom the intestinal malabsorption of sugar or complex carbohydrates has been investigated using an appropriate test of absorption (e.g., lactose or sucrose absorption test) or malabsorption (lactose or sucrose breath hydrogen test, or breath hydrogen production after the ingestion of complex carbohydrate). Carbohydrate malabsorption may be due to pancreatic insufficiency with impairment of the intraluminal phase of starch digestion or to disaccharidase deficiency characterized by an absence or very low levels of specific microvillus membrane enzymes (e.g., sucrase or lactase). In very rare circumstances, glucose malabsorption produces clinical symptoms, as in the disorder known as glucose–galactose malabsorption.

DIAGNOSIS OF CARBOHYDRATE MALABSORPTION

Confirmatory Tests

The diagnosis of carbohydrate malabsorption is based on the combination of clinical findings and results of appropriate tests. The presence of low fecal pH or reducing substances indicates carbohydrate malabsorption, but these tests are valid only when carbohydrate has been recently ingested, intestinal transit time is rapid, stools are collected fresh, and assays are performed immediately, and when bacterial metabolism of colonic carbohydrate is incomplete. In general,

carbohydrate malabsorption is best confirmed using more specific tests, especially because sucrose is a nonreducing sugar, will not produce reducing substances, and requires special techniques for detection in stool.

The capacity for sugar absorption can be measured using a lactose or sucrose absorption test. The patient ingests water containing the sugar to be tested and blood glucose is measured before and at 30-minute intervals for 3 hours after a dose of 2 g/kg body weight (maximal dose in adults, 50 g). In adults, the test has a sensitivity of 75% and a specificity of 96%. However, in children, it is cumbersome, invasive, and time consuming, and has largely been replaced by the breath hydrogen test.

The breath hydrogen test really measures carbohydrate (usually lactose or sucrose) nonabsorption (rather than carbohydrate or sugar hydrolysis and monosaccharide uptake). Its sensitivity and specificity are superior compared to absorption tests, and it is simple and noninvasive. The breath hydrogen test can be performed in people of all ages. The dose is customarily 2 g of carbohydrate (or lactose, sucrose, or glucose) per kilogram of body weight (maximum 25 g). Breath hydrogen is sampled prior to the ingestion of sugar and at 30-minute intervals following the ingestion of sugar for 3 hours. It is customary to use a value of $\sim$10 parts per million as normal, comparing samples obtained after carbohydrate ingestion to the baseline value. Values between 10 and 20 parts per million may be indeterminate unless accompanied by symptoms, but values over 20 parts per million are representative of carbohydrate malabsorption. False positive tests are seen with inadequate pretest fasting and when the patient has smoked recently or has swallowed toothpaste just prior to the test. False negative results are obtained when patients have recently used antibiotics or are nonhydrogen producers (approximately 1% of the population). In children less than 5 years of age, an abnormal lactose breath hydrogen test always signifies abnormal intestinal mucosa or bacterial overgrowth, both of which require further definition by appropriate diagnostic tests. A normal breath hydrogen test does not rule out an intestinal mucosal lesion, and it cannot be used to avoid an intestinal biopsy. The glucose breath hydrogen test can be used in the diagnosis of bacterial overgrowth syndromes.

The assay of disaccharidase activity in small bowel biopsy samples establishes the presence of disaccharidase deficiency, and has been used to define populations with low lactase levels and to establish the diagnosis of sucrase-isomaltase deficiency. However, when low lactase activity accompanies intestinal injury, the lesion may be focal or patchy; consequently, intestinal biopsy samples may not yield an abnormal result on disaccharidase assay. Clinical, biochemical, and breath test data must always be compared to obtain the correct diagnosis.

SPECIFIC DISORDERS AND THEIR TREATMENT

Approach to the Patient

Patients who have symptoms and signs of carbohydrate malabsorption should be evaluated in a systematic fashion (Table I). The clinical findings may not immediately suggest a diagnosis of carbohydrate malabsorption, because many patients with this diagnosis actually have a clinical pattern more like that seen in irritable bowel syndrome. Secondary causes for carbohydrate malabsorption must be pursued and appropriate confirmatory tests obtained. When considering lactose malabsorption, especially in infants and young children, the possibility of milk protein allergy must be ruled out.

It is important to remember that lactose malabsorption may occur in patients with other disorders (for example, irritable bowel syndrome or celiac disease). Thus, a lactose breath hydrogen test may be a valuable

TABLE I Differential Diagnosis of Carbohydrate Malabsorption

Intraluminal phase defects
Primary
Cystic fibrosis
Shwachman–Diamond syndrome
Secondary
Pancreatic insufficiency due to alcohol or chronic pancreatitis
Bacterial overgrowth
Mucosal phase defects
Primary
Lactose
Genetic
Developmental
Congenital
Sucrose
Sucrase–isomaltase deficiency
Glucose–galactose malabsorption
Secondary
Bacterial overgrowth
Mucosal injury
Infectious enteritis
Giardiasis
Celiac disease
Drug-induced enteritis
Inflammatory bowel disease
Radiation enteritis
Acquired glucose malabsorption

part of the evaluation of patients suspected of having irritable bowel syndrome (it should not be used to avoid small bowel biopsy in patients suspected of having celiac disease).

Intraluminal Phase Defects

Pancreatic amylase deficiency is seen with hereditary diseases of the exocrine pancreas (particularly cystic fibrosis and Shwachman–Diamond syndrome), pancreatic insufficiency due to alcohol or chronic pancreatitis, and bacterial overgrowth. The majority of patients with cystic fibrosis have absent or low levels of amylase activity; however, their symptoms are usually more attributable to fat than to carbohydrate maldigestion. Shwachman–Diamond syndrome, consisting of exocrine pancreatic insufficiency and hematologic and skeletal abnormalities, is the second most frequent cause of hereditary pancreatic insufficiency in children. Amylase activity is low or absent in such patients, and they have diarrhea and malabsorption. Low amylase levels in adults with pancreatic insufficiency have been well described, and bacterial overgrowth may produce intraluminal fermentation. Amylase deficiency is generally successfully treated by the administration of pancreatic supplements, achieving virtually complete carbohydrate digestion. Elimination of complex carbohydrates from the diet is not acceptable. Bacterial overgrowth syndrome is treated with appropriate antibiotics.

Mucosal Phase Defects

Primary Lactose Intolerance

Primary lactose malabsorption occurs in three clinical settings: (1) genetic lactase decline, (2) developmental lactase deficiency, and (3) congenital lactase deficiency. Genetically controlled lactase decline with lactose malabsorption is the most common form of genetically determined reductions of lactase activity. This clinical finding has been termed "lactase deficiency," although this term is really a misnomer, because a majority of the world's populations develop low intestinal lactase levels during midchildhood (approximately age 5 years). This finding is most prominent in Asian, African, and indigenous populations (Table II). The exact molecular basis of this pattern is unknown, but it has been shown that lactase mRNA content and lactase activity are both extremely low, strongly suggesting that the genetic regulation is at the level of transcription. In striking contrast, peoples of Scandinavian or Caucasian genetic background have a genetically regulated preservation of intestinal lactase activity and a high degree of lactose tolerance as adults. The molecular basis of this phenotype is also unknown. No matter what the genetic background, normal lactase activity is found in all children until about 5 years of age. Lactase deficiency or lactose intolerance detected in children before this age usually indicates an underlying mucosal lesion or bacterial overgrowth syndrome. Recently, genetic polymorphisms in the introns of a gene lying 5′ to the lactase gene on chromosome 2q22 have been associated with lactase persistence or nonpersistence. Similar associations have been found among polymorphisms in the lactase coding region and proximal 5′ untranslated region. The functional significance of these polymorphisms is at present unknown.

TABLE II Prevalence of Lactose Intolerance in Selected Populations

Country	Prevalence of lactose intolerance in adults (%)
Scandinavia	0–1
Austria	20
United States	
Caucasians	22
France	
North	32
South	44
Italy	
North	50
South	72
United States	
African Americans	65
Native Americans	95
Asians	95–100

There is no evidence that lactase deficiency or lactose malabsorption is a normal part of the aging process. Thus, in the mixed population of Caucasian extraction, the normal aging process does not lead to lactase deficiency. However, alterations in motility, secondary lactase reduction due to medications or to other disorders, or intestinal injury may produce lactose intolerance.

Developmental lactase deficiency is a consequence of gestational age. During fetal development, lactase activity rises late in gestation so that 28- to 32-week premature infants have reduced lactase activity. If they are otherwise healthy, the colon can salvage unabsorbed carbohydrate so that these infants are not nutritionally compromised and do not have diarrhea.

Primary lactase deficiency is characterized by the absence of lactase activity in the small intestine, by normal histology, and by normal levels of other disaccharidases. The disorder has been described mainly in

Finnish populations and is associated with a genetic abnormality in the DNA adjacent to the extended lactase gene, but not in the coding region. This very rare syndrome is associated with diarrhea accompanying lactose ingestion from birth in affected infants. The use of other sugars (e.g., sucrose, fructose) in infant formula leads to elimination of symptoms.

Treatment of Lactose Malabsorption

The treatment of lactose malabsorption includes four general principles: (1) reduction or restriction of dietary lactose, (2) substitution of alternative nutrient sources to avoid reductions in energy and protein intake, (3) regulation of calcium intake, and (4) the use of a commercially available enzyme substitute. When lactose restriction is necessary, the patient must be instructed to read labels of commercially prepared foods, because hidden lactose may be difficult to identify. Table III summarizes the lactose content of selected foods. Complete restriction of lactose-containing foods should be necessary for only a limited period to ascertain the specificity of the diagnosis. Because some patients can tolerate graded increases in lactose intake, small quantities of lactose may subsequently be reintroduced into the diet, careful attention being paid to associated symptoms. Because of its high sugar and fat content, ice cream may be a good way to introduce lactose into the diet. Carefully controlled studies of lactose intolerance have shown that when people complaining of lactose intolerance ingest milk or a control milk containing hydrolyzed lactose, without knowing which sample they ingest, their symptom scores are identical. This indicates an important "anticipatory" component to the symptoms in many people.

Calcium is supplemented in the form of calcium carbonate; tablets such as Tums (Norcliff Thayer, Inc., Tarrytown, NY) or Viactiv (McNeil Nutritionals, Fort Washington, PA) are popular and effective. Standard preparations contain 500 mg of calcium carbonate equivalent to 200 mg of elemental calcium, which is 20% of the United States Recommended Dietary Allowance (RDA) for adults. In infants and young children, liquid calcium gluconate is readily tolerated and available. When complete lactose restriction is recommended, the RDA for calcium should be provided as a supplement. Prolonged lactose restriction predisposes to reduced calcium intake and increases the risk for long-term bone mineral depletion.

TABLE III Lactose Content of Selected Foods

Food	Lactose content
Cow milk	9–13 g per 8 oz
Milk sherbet	4 g per 8 oz
Cottage cheese	5–6 g per 8 oz
Light cream	0.6 g per tablespoon
Natural cheddar cheese	0.4–0.6 per 28 g
Processed Swiss cheese	0.4–0.6 per 28 g
Ice cream	2–6 g per 4 oz

Commercially available "lactase" preparations are actually bacterial or yeast β-galactosidases. When added to lactose-containing food or when ingested with meals containing lactose, these are effective in reducing symptoms and breath hydrogen values in many people who have low lactase levels and complaints of lactose intolerance. However, these products are not capable of completely hydrolyzing all dietary lactose, and the results achieved in individual patients are variable. Some of the commercial "lactase" preparations are listed in Table IV. LactAid liquid (LactAid, Inc., Pleasantville, NJ) may be added to milk (14 drops/quart), which is then refrigerated overnight before use. The resulting hydrolysis of lactose (which is approximately 90% effective) produces a sweeter taste compared to milk containing lactose. Commercial LactAid products are also sold in most markets in the United States. Lactrase (Kremers-Urban Company, Milwaukee, WI) capsules may be taken orally with lactose-containing foods, as can the other products listed in Table IV, but the individual dose required and responses to individual products must be tested in each patient. It should be noted that "acidophilus milk" is not sufficiently lactose depleted. Live-culture yogurt, which contains endogenous β-galactosidase, is a useful alternative source for calcium and calories, and may be well tolerated by a number of lactose-intolerant patients. However, yogurts that contain milk products added back after fermentation may produce symptoms. Although consumption of yogurt alone by individuals with low lactase activity and lactose intolerance reduces symptoms, consumption of yogurt together with additional lactose does not lead to reduced symptoms.

Sucrose Intolerance

Sucrase–isomaltase deficiency is an uncommon autosomal recessive disorder characterized by very low levels of this microvillus membrane enzyme. The genetic basis appears to be mutations in the gene that lead to alterations in the posttranscriptional processing of the enzyme. Sucrose intolerance presents at different ages. In infants, symptoms may appear when sucrose is introduced into the diet; in younger children, chronic diarrhea may occur, alternating with constipation and confusing the diagnosis. Occasionally, the first manifestations appear in adulthood with complaints

TABLE IV Some Commercial "Lactase" Substitutes

Name	Dose form	Supplier
LactAid	Liquid/tablets	LactAid, Inc.
Lactrase	Capsules	Kremers-Urban
LactAce	Capsules	Nature's Way Products, Inc.
DairyEase	Tablets	Glenbrook Laboratories
Lactrol	Caplets	Advanced Nutritional Technology

of carbohydrate malabsorption or may mimic irritable bowel syndrome. The diagnosis is most easily achieved by means of the sucrose breath hydrogen test. Treatment is accomplished by restriction of dietary sucrose. Some studies suggest that provision of exogenous sucrase activity ("sacrosidase") using viable yeast cells (*Saccharomyces cerevisiae*) may assist in sucrose tolerance.

Glucose–Galactose Malabsorption

Glucose–galactose malabsorption is a rare, autosomal recessive disorder detected in neonates who develop diarrhea; their stools are acidic and contain fecal reducing substances usually after the first feeding of glucose water (e.g., Pedialyte or similar preparations). The diagnosis in these patients can be confirmed with a flat glucose and a normal fructose tolerance test. Confirmation of the diagnosis can be achieved *in vitro* by measurement of glucose transport in small bowel biopsy or by intraluminal perfusion. Treatment is based on the elimination of glucose and galactose and the use of fructose-containing formula.

Fructose Malabsorption

Fructose malabsorption is an uncommon cause of serious disease. However, the ingestion of large amounts of fructose (particularly in the form of fruit juices) in infants and young children may lead to diarrhea and symptoms of carbohydrate malabsorption because of an ingested load of fructose in excess of the capacity of the small intestine to absorb the sugar. Fruit juice ingestion in infants and young children may produce the syndrome of "chronic nonspecific diarrhea" or "toddler's diarrhea." This can be rapidly reversed by reducing dietary fructose. Fructose content of foods varies, but considerable quantities can be found in figs (31 g/100 g edible portion), dates (24 g), prunes (15 g), and soft drinks containing high-fructose corn syrup (37.5 g/18 oz of soda).

The sugar alcohol of fructose (sorbitol) is not well tolerated by some people. It is poorly absorbed in the small intestine and its transport to the colon may be associated with symptoms of carbohydrate malabsorption. Sorbitol can be found in "diet" products, such as chewing gums (1.3–2.2 g/piece), and "low-sugar" fruits (pears, 4.6 g/100 g dry weight), fruit juices (apple juice 0.5 g/100 g), and carbonated beverages. In the appropriate clinical setting, a search for sorbitol-containing foods in the diet and elimination of offending substances may lead to resolution of symptoms of carbohydrate malabsorption.

Secondary Carbohydrate Malabsorption

Bacterial overgrowth or stasis syndromes may be associated with increased fermentation of dietary carbohydrates in the small bowel. Clinical symptoms of carbohydrate intolerance are often found. The diagnosis may be suspected when a very early peak of breath hydrogen is detected during a lactose, sucrose, or glucose breath test.

Carbohydrate malabsorption frequently occurs after gastrointestinal tract mucosal injury involving villus flattening or damage to the intestinal epithelium. Disorders that often produce this lesion are listed in Table I. When the mucosa is damaged, lactase is usually the first affected disaccharidase, presumably because of its distal location on the villus. Treatment of the primary disorder is mandatory for the return of lactase activity, which often lags behind the return of normal intestinal morphology. Prolonged lactose intolerance, which may persist for months after healing starts, is unique to this disaccharidase, and its biochemical basis is unexplained.

Secondary sucrase-isomaltase deficiency is usually of less clinical significance compared to secondary lactase deficiency. Enzyme activity is not totally lost and usually does not reach the low levels of primary sucrase-isomaltase deficiency; thus, patients tend to tolerate some sucrose intake. Maltase-glucoamylase deficiency is related to the severity of the mucosal lesion and is usually of minimal clinical significance in most patients.

The principles of treatment in carbohydrate malabsorption in patients with secondary disorders are identical to those for the primary disorders, in addition to the treatment of the underlying problem.

Miscellaneous Observations

Patients demonstrating symptoms compatible with the diagnosis of carbohydrate malabsorption, but for whom specific testing fails to reveal an abnormality, may have symptoms related to fiber ingestion. The custom of eating extremely high-fiber foods may predispose some people to symptomatic carbohydrate malabsorption. Flatulence associated with the ingestion of a variety of beans and other vegetables considered to be

"gas producing" is a common complaint. When dried beans are used in food preparation, they can be soaked for 12–24 hours prior to cooking to eliminate sugars not absorbed in the small intestine. This often reduces "gas"-related symptoms.

Patients whose main complaints are increased flatulence may or may not have carbohydrate malabsorption, but they should have this possibility ruled out by appropriate testing. It should be remembered that swallowed air may lead to flatulence as well. However, air swallowing is more often accompanied by excessive belching. The ingestion of large quantities of air may be due to psychogenic factors, to crying in infants, to gum chewing, or to the consumption of carbonated drinks. Appropriate dietary recommendations may lead to resolution of symptoms.

Acknowledgments

Supported in part by National Institutes of Health Research Grant R37 DK 32658, and NIH Research Training Grant T32 DK 07471. Portions of this chapter are reproduced from *Current Therapy in Gastroenterology and Liver Disease* (T. M. Bayless, ed.), Mosby-Year Book, Inc., St. Louis (1994), with permission of the publisher.

See Also the Following Articles

Amylase • Bacterial Overgrowth • Breath Tests • Carbohydrate Digestion and Absorption • Cow Milk Protein Allergy • Cystic Fibrosis • Diarrhea • Flatulence • Galactosemia • Glycogen Storage Disease • Hereditary Fructose Intolerance • Malabsorption • Tyrosinemia

Further Reading

Büller, H. A., and Grand, R. J. (1990). Lactose intolerance. *In* "Annual Review of Medicine: Selected Topics in the Clinical Sciences," Vol. 41 (W. P. Creger, C. H. Coggins, and E. W. Hancock, eds.), pp. 141–148. Annual Reviews Inc., Palo Alto.

Corazza, G. R., Strocchi, A., Rossi, R., *et al.* (1988). Sorbitol malabsorption in normal volunteers and in patients with coeliac disease. *Gut* **29**, 44–48.

Di Stefano, M., Veneto, G., Malservi, S., *et al.* (2001). Lactose malabsorption and intolerance in the elderly. *Scand. J. Gastroenterol.* **36**, 1274–1278.

Enattah, N. S., Sahi, T., Savilahti, E., *et al.* (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237.

Escher, J. C., de Koning, N. D., van Engen, C. G. J., Arora, S., Büller, H. A., Montgomery, R. K., and Grand, R. J. (1991). Molecular basis of lactase levels in adult humans. *J. Clin. Invest.* **89**, 480–483.

Harms, H. K., Bertele-Harms, R. M., and Bruer-Kleis, D. (1987). Enzyme substitution therapy with the yeast *Saccharomyces cerevisiae* in congenital sucrase-isomaltase deficiency. *N. Engl. J. Med.* **316**, 1306–1309.

Harvey, C. B., Hollox, E. J., Poulter, M., *et al.* (1998). Lactase haplotype frequencies in Caucasians: Association with the lactase persistence/non-persistence polymorphism. *Ann. Hum. Genet.* **62**(Pt. 3), 215–223.

Hoekstra, J. H., van Kempen, A. A. M. W., and Kneepkens, C. M. F. (1993). Apple juice malabsorption: Fructose or sorbitol? *J. Pediatr. Gastroenterol. Nutr.* **16**, 39–42.

Jarvela, I., Sabri Enattah, N., Kokkonen, J., *et al.* (1998). Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene. *Am. J. Hum. Genet.* **63**, 1078–1085.

Jarvis, J. K., and Miller, G. M. (2002). Overcoming the barrier of lactose intolerance to reduce health disparities. J. *Natl. Med. Assoc.* **94**, 55–66.

Martin, M. G., Turk, E., Lostao, M. P., *et al.* (1996). Defects in Na^+/glucose cotransporter (SGLT1) trafficking and function cause glucose–galactose malabsorption. *Nat. Genet.* **12**, 216–220.

Rao, D. R., Bello, H., Warren, A. P., and Brown, G. E. (1994). Prevalence of lactose maldigestion. Influence and interaction of age, race, and sex. *Dig. Dis. Sci.* **39**, 1519–1524.

Rosado, J. L., Solomons, N. W., Lisker, R., *et al.* (1984). Enzyme replacement therapy for primary adult lactase deficiency. Effective reduction of lactose malabsorption and milk intolerance by direct addition of beta-galactosidase to milk at mealtime. *Gastroenterology* **87**, 1072–1082.

Scrimshaw, N. S., and Murray, A. B. (1988). The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. *Am. J. Clin. Nutr.* **48**, 1079–1159.

Suarez, F. L., Adshead, J., Furne, J. K., and Levitt, M. D. (1998). Lactose maldigestion is not an impediment to the intake of 1500 mg calcium daily as dairy products. *Am. J. Clin. Nutr.* **68**, 1118–1122.

Suarez, F. L., Saviano, D. A., and Levitt, M. D. (1995). A comparison of symptoms after the consumption of milk or lactose-hydrolyzed milk by people with self-reported severe lactose intolerance. *N. Engl. J. Med.* **333**, 1–4.

Vesa, T. H., Seppo, L. M., Marteau, P. R., *et al.* (1998). Role of irritable bowel syndrome in subjective lactose intolerance. *Am. J. Clin. Nutr.* **67**, 710–715.

Carbohydrate Digestion and Absorption

ERIC SIBLEY
Stanford University School of Medicine

glycosidic linkage Covalent chemical bond between the monosaccharide units of disaccharides, oligosaccharides, and polysaccharides formed by the removal of a molecule of water. The bond is normally formed between the C-1 on one sugar and the C-4 on the other. An α-glycosidic bond is formed when the OH group on C-1 is below the plane of the glucose ring and a β-glycosidic bond is formed when it is above the plane.

oligosaccharide Carbohydrate compound in which monosaccharide units are joined by glycosidic linkages. The term "oligosaccharide" is commonly used to refer to defined structures having three to nine monosaccharide units.

polysaccharide Macromolecule carbohydrate consisting of more than nine monosaccharide residues joined to each other by glycosidic linkages.

Carbohydrates consumed in the diet must be digested and absorbed in the gastrointestinal tract prior to utilization as nutrient energy sources. The processes of carbohydrate digestion and absorption are accomplished by the action of digestive enzymes and carrier proteins with specialized functions and specificities.

INTRODUCTION

Carbohydrates account for 40–60% of the average caloric energy intake in humans. The primary carbohydrates in the human diet are starches, sucrose, and lactose. Starches and complex polysaccharides comprise 60% of the digestible carbohydrates in the average adult Western diet. Sucrose and lactose comprise 30 and 10%, respectively, of carbohydrates in the adult diet. The various carbohydrates must be digested to their monosaccharide components prior to transport across the surface membrane of the absorptive intestinal epithelial cells, enterocytes. The process of carbohydrate digestion (in the case of starches) is initiated by salivary and pancreatic amylase in the lumen of the gut. Digestion is completed by the action of carbohydrate-specific hydrolases bound to the brush border membrane of the absorptive enterocytes lining the villi of the small intestine. The digestive and absorptive pathways for the three major nutrient carbohydrate sources in the human diet (starches, sucrose, and lactose) are described in the following discussions.

Significant amounts of nondigestible carbohydrates are also consumed in the human diet. These carbohydrates consist of nondigestible starches, largely in the form of plant fiber; nondigestible sugars such as stachyose and raffinose present in legumes; and sugar alcohols such as sorbitol and mannitol used as sweeteners. Humans lack digestive enzymes capable of efficiently hydrolyzing these carbohydrates. Thus, these dietary fibers and nondigestible sugars pass through the gastrointestinal tract undigested or are partially digested by enzymes produced by intestinal bacterial flora to yield short-chain fatty acids, hydrogen, carbon dioxide, and methane.

DIGESTION OF STARCHES

Luminal Hydrolysis by α-Amylases

The two major starch polysaccharides in the human diet are amylose and amylopectin. Amylose is a linear chain of glucose molecules linked by α-1,4 glycosidic bonds and accounts for 20% of dietary starch. Amylopectin is a branching chain of α-1,4-linked glucose molecules with an α-1,6 linkage occurring approximately every 20 glucose molecules. Amylopectin accounts for 80% of dietary starch. Glycogen is a polysaccharide found in animal tissues and has a structure similar to that of amylopectin. In order for these complex carbohydrates to be utilized as a nutrient source they must be digested to smaller oligosaccharides and then ultimately must be hydrolyzed to individual monosaccharides. Food starches are present in association with hydrophobic proteins that limit access to enzymes responsible for starch hydrolysis. Prior to consumption, methods of food preparation, such as cracking grains, milling, and cooking, facilitate starch utilization. The process of enzymatic starch digestion is initiated in the lumen of the gastrointestinal tract by the action of α-amylases secreted by the salivary gland and the pancreas. Human salivary amylase (94% identical to pancreatic amylase) tends to be degraded at acidic pH and thus contributes minimally to starch digestion in

 Copyright 2004, Elsevier (USA). All rights reserved.

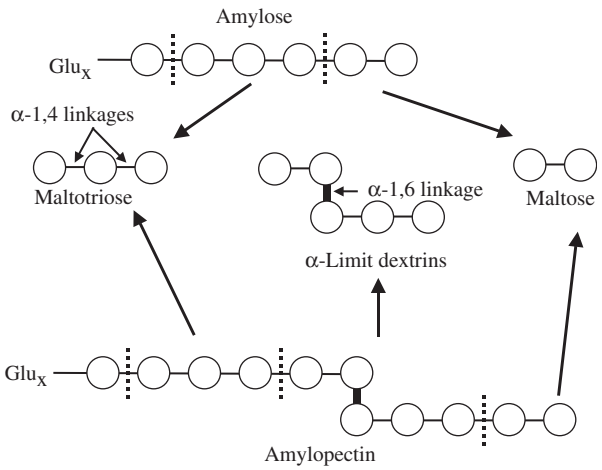


FIGURE 1 Digestion of amylose and amylopectin forms of starch by α-amylase. Glu$_x$ and the open circles represent glucose residues; thin solid lines indicate α(1−4) glycosidic linkages; thick solid lines indicate α(1−6) glycosidic linkages; broken vertical lines indicate α-amylase hydrolysis. Adapted from Gray (1981). "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Copyright Lippincott Williams & Wilkins, with permission.

the small intestine. Pancreatic amylase is secreted into the intestinal lumen in the duodenum and is the major enzyme of intraluminal starch digestion in adults. In neonates with decreased pancreatic secretion, however, salivary amylase supports a significant amount of starch digestion.

Amylase functions to cleave the internal α-1,4 glycosidic bonds of amylose and amylopectin but is inactive against the α-1,6 linkages in amylopectin and glycogen. The products of luminal starch digestion by amylase are maltose (a glucose−glucose disaccharide), maltotriose (a glucose trisaccharide), and α-limit dextrins. The enzymatic action of the α-amylases on starches is shown in Fig. 1.

Brush Border Membrane Hydrolysis by Oligosaccharidases

The oligosaccharide products of luminal starch digestion are hydrolyzed to glucose monomers by the membrane-bound brush-border digestive enzymes, maltase-glucoamylase, sucrase, and isomaltase (also called α-dextrinase). Maltase-glucoamylase removes single glucose residues from the α-1,4 chains of oligosaccharides, and from maltotriose and maltose. Sucrase-isomaltase is initially synthesized in the enterocyte as a single glycoprotein chain. After insertion in the brush-border membrane, it is cleaved into sucrase and isomaltase units that reassociate noncovalently. Sucrase is capable of hydrolyzing the α-1,4 glycosidic bonds of short oligosaccharides, including maltose and maltotriose. Isomaltase is the only enzyme capable of hydrolyzing the α-1,6 glycosidic linkage in the α-limit dextrins. The free glucose products of oligosaccharidase hydrolysis are transported into the enterocyte as described in the following sections. The pathway for starch digestion is summarized in Fig. 2.

DIGESTION OF DISACCHARIDES

Sucrose and lactose are the major disaccharides in the human diet. These sugars, along with maltose and other disaccharides consumed in minimal amounts, cannot be absorbed across the enterocyte surface membrane intact. Unlike starch digestion, there is no luminal digestion of disaccharides. Rather, they are hydrolyzed to their monosaccharide components by specific brush border membrane hydrolases expressed in the small intestine. Maltase catalyzes the enzymatic hydrolysis of maltose as previously described. Trehalase is a brush border enzyme that specifically hydrolyzes trehalose, a disaccharide found predominantly in mushrooms. The brush border membrane hydrolysis of sucrose and lactose is described in the following section and is summarized in Fig. 2.

Sucrose Digestion

Sucrose, the major carbohydrate in most fruits, is a disaccharide composed of a glucose molecule linked via an α-glycosidic linkage to a fructose molecule. As with the other oligosaccharides, sucrose must be digested to its constituent glucose and fructose monosaccharides prior to absorption across the absorptive intestinal cell membrane. The enzymatic hydrolysis of the

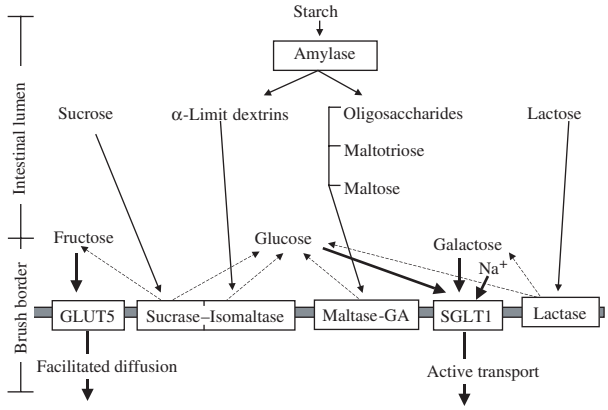


FIGURE 2 Major pathways for digestion and absorption of dietary carbohydrates. GA, glucoamylase. Adapted from Gray (1975).

α-glycosidic bond in sucrose is mediated by the intestinal epithelial cell surface membrane-bound enzyme, sucrase. Sucrase is a membrane-bound glycoprotein associated with the digestive enzyme isomaltase. In most mammalian species, the expression of sucrase-isomaltase is minimal at birth and then is induced dramatically at the time of weaning. In humans, however, sucrase-isomaltase expression is induced prenatally (10–14 weeks of gestation) and is maximal at birth. Such a precocious expression pattern in humans allows for the digestion and absorption of solid sugars prior to weaning in human infants.

Clinical deficiency of sucrase-isomaltase is a rare autosomal recessive condition that results in the inability to digest and absorb dietary sucrose. The condition becomes manifest during infancy on the introduction of sucrose in fruits and juices. The inability to digest the sugars results in signs and symptoms of malabsorption, including diarrhea, increased gas production, and abdominal distension. Treatment for congenital sucrase-isomaltase deficiency consists of avoidance of sucrose in the diet.

Lactose Digestion

Lactose is the predominant carbohydrate in breast milk. The lactose disaccharide consists of a glucose and a galactose molecule linked by a β-1,4 glycosidic bond. The digestive enzyme lactase-phlorizin hydrolase (β-galactosidase) catalyzes the hydrolysis of lactose to yield glucose and galactose, which can then be absorbed across the intestinal mucosa membrane. Lactase-phlorizin hydrolase, lactase, is a brush border membrane protein with enzymatic activity against glycolipids as well as lactose. The lactase-phorizin hydrolase gene is specifically expressed by the small intestinal enterocytes of most mammals. Lactase enzyme activity is maximal in preweaned mammals and declines markedly during maturation. The maturational decline in lactase activity renders most of the world's adult human population intolerant of milk and other dairy products. Lactose present in dairy products cannot be digested in the small intestine, due to the reduced lactase level, and instead is fermented by bacteria in the distal ileum and colon. The fermentative products result in symptoms of diarrhea, gas bloat, flatulence, and abdominal pain. Treatment consists primarily of avoiding lactose-containing foods. Lactase enzyme supplements may be helpful. In a minority of adults, high levels of lactase activity persist in adulthood. This hereditary persistence of lactase is common primarily in people of northern European descent and is attributed to inheritance of an unidentified autosomal-dominant mutation that prevents the normal maturational decline in lactase expression. Linkage disequilibrium and haplotype analyses in humans have recently allowed for identification of two genetic variants located upstream of the human lactase gene that are associated with hereditary lactase persistence.

MONOSACCHARIDE ABSORPTION

Glucose and Galactose Transport

The products of luminal and membrane-bound hydrolysis of starches and sugars are the constituent monosaccharides, glucose, galactose, and fructose. These monosaccharides, present at the enterocyte brush border membrane, must be transported across both the apical and basolateral enterocyte membranes and into the portal circulation. The apical membrane absorption of glucose and galactose is carried out predominantly by a Na^+-dependent active transport mechanism in which each monosaccharide is transported along with two Na^+ molecules (see Fig. 2). The sodium/glucose cotransporter, SGLT1, has been characterized. SGLT1 is a 664-amino-acid transmembrane protein expressed on the apical surface of intestinal epithelial cells. The protein has 12 membrane-spanning domains. Sodium molecules enter the enterocyte via an electrochemical gradient that is maintained by Na^+, K^+-ATPase (the sodium pump) that extrudes sodium molecules out the basolateral surface. By coupling transport to the sodium gradient, the glucose and galactose molecules are able to be transported actively into the cell. Glucose and galactose are transported out of the enterocyte into the portal circulation by the Na^+-independent glucose transporter, GLUT2, located on the basolateral membrane.

Congenital defects of the SGLT1 protein resulting from missense and nonsense mutations have been reported as rare clinical causes of severe glucose–galactose malabsorption in infancy. Such children present in the neonatal period with severe, watery, acidic diarrhea while consuming breast milk or standard infant formulas. The symptoms of diarrhea and dehydration are life threatening. Treatment consists of rehydration and initiation of a glucose- and galactose-free diet. Because fructose is tolerated, most of the carbohydrate initially can be given as fructose.

Fructose Transport

The fructose monosaccharide is transported across the intestinal brush border membrane via a Na^+-independent facilitated diffusion mechanism. GLUT5 is a 501-amino-acid transmembrane protein that transports fructose and glucose molecules. Fructose

is transported from the enterocyte into the portal circulation via the basolateral membrane GLUT2 transporter. Fructose is not as well absorbed as is glucose. Consequently, ingestion of high levels of fructose in the diet can lead to carbohydrate intolerance. In children, drinking excessive amounts of juices high in fructose may result in nonspecific diarrhea and recurrent abdominal pain. In adults, fructose malabsorption has been associated with irritable bowel syndrome.

See Also the Following Articles

Amylase • Carbohydrate and Lactose Malabsorption • Digestion, Overview • Small Intestine, Absorption and Secretion

Further Reading

Alpers, D. H. (1994). Digestion and absorption of carbohydrates and protein. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1723–1749. Raven Press, New York.

Buller, H. A, and Grand, R. J. (1990). Lactose intolerance. *Annu. Rev. Med.* **41**, 141–418.

Burant, C. F., Takeda, J., Brot-Laroche, E., Bell, G. I., and Davidson, N. O. (1992). Fructose transporter in human spermatozoa and small intestine is GLUT5. *J. Biol. Chem.* **267**, 14523–14526.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., and Jarvela, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237.

Gray, G. M. (1975). Carbohydrate digestion and absorption: Role of the small intestine. *N. Engl. J. Med.* **292**, 1225.

Gray, G. M. (1981). Carbohydrate absorption and malabsorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1064. Raven Press, New York.

Gray, G. M. (1992). Starch digestion and absorption in nonruminants. *J. Nutr.* **122**, 172–177.

Levin, R. J. (1994). Digestion and absorption of carbohydrates—From molecules and membranes to humans. *Am. J. Clin. Nutr.* **59**, 690S–698S.

Sitrin, M. D. (1996). Nutritional physiology. *In* "Gastrointestinal, Hepatobiliary, and Nutritional Physiology" (E. B. Chang, M. D. Sitrin, and D. B. Black, eds.), pp. 121–140. Lippincott-Raven, Philadelphia, Pennsylvania.

Southgate, D. (1995). Digestion and metabolism of sugars. *Am. J. Clin. Nutr.* **62**, 203S–211S.

Turk, E., Martin, M. G., and Wright, E. M. (1994). Structure of the human Na+/glucose cotransporter gene SGLT1. *J. Biol. Chem.* **269**, 15204–15209.

Celiac Disease

AHMAD S. ABDULKARIM* AND JOSEPH A. MURRAY†

**Marshfield Clinic, Wisconsin, and †Mayo Clinic and Foundation*

gluten Wheat protein responsible for damage of small intestine in celiac disease.

gluten-free diet Regimen with no gluten or similar proteins in any meal component.

malabsorption Impaired intestinal absorption of nutrients.

Celiac disease, a chronic inflammatory condition associated with small intestinal injury induced by gluten exposure, results in malabsorption of different nutrients. It is associated with multiple other medical conditions. The diagnosis relies on characteristic findings of small intestinal biopsy. Patients with celiac disease usually respond quickly to a gluten-free diet. The disease requires a lifelong commitment to a gluten-free diet to prevent recurrence of symptoms and other potential consequences.

INTRODUCTION

Celiac disease (CD) was described in the first century A.D. However, modern knowledge of the relationship between CD and the damaging effect of gluten on the small intestine is credited to Dicke, who, in the 1940s, first observed improvement of children with CD when there was shortage of wheat during World War II. Dicke noted the recurrence of symptoms when food supplies became available again.

The disease may induce no symptoms at all for a long time, yet in some cases subtle symptoms can be detected and in others a more debilitating form of the disease may be encountered. The disease is being recognized more often nowadays thanks to increasingly available testing

 Copyright 2004, Elsevier (USA). All rights reserved.

is transported from the enterocyte into the portal circulation via the basolateral membrane GLUT2 transporter. Fructose is not as well absorbed as is glucose. Consequently, ingestion of high levels of fructose in the diet can lead to carbohydrate intolerance. In children, drinking excessive amounts of juices high in fructose may result in nonspecific diarrhea and recurrent abdominal pain. In adults, fructose malabsorption has been associated with irritable bowel syndrome.

See Also the Following Articles

Amylase • Carbohydrate and Lactose Malabsorption • Digestion, Overview • Small Intestine, Absorption and Secretion

Further Reading

Alpers, D. H. (1994). Digestion and absorption of carbohydrates and protein. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1723–1749. Raven Press, New York.

Buller, H. A, and Grand, R. J. (1990). Lactose intolerance. *Annu. Rev. Med.* **41**, 141–418.

Burant, C. F., Takeda, J., Brot-Laroche, E., Bell, G. I., and Davidson, N. O. (1992). Fructose transporter in human spermatozoa and small intestine is GLUT5. *J. Biol. Chem.* **267**, 14523–14526.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., and Jarvela, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237.

Gray, G. M. (1975). Carbohydrate digestion and absorption: Role of the small intestine. *N. Engl. J. Med.* **292**, 1225.

Gray, G. M. (1981). Carbohydrate absorption and malabsorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1064. Raven Press, New York.

Gray, G. M. (1992). Starch digestion and absorption in nonruminants. *J. Nutr.* **122**, 172–177.

Levin, R. J. (1994). Digestion and absorption of carbohydrates—From molecules and membranes to humans. *Am. J. Clin. Nutr.* **59**, 690S–698S.

Sitrin, M. D. (1996). Nutritional physiology. *In* "Gastrointestinal, Hepatobiliary, and Nutritional Physiology" (E. B. Chang, M. D. Sitrin, and D. B. Black, eds.), pp. 121–140. Lippincott-Raven, Philadelphia, Pennsylvania.

Southgate, D. (1995). Digestion and metabolism of sugars. *Am. J. Clin. Nutr.* **62**, 203S–211S.

Turk, E., Martin, M. G., and Wright, E. M. (1994). Structure of the human Na+/glucose cotransporter gene SGLT1. *J. Biol. Chem.* **269**, 15204–15209.

Celiac Disease

AHMAD S. ABDULKARIM* AND JOSEPH A. MURRAY†

**Marshfield Clinic, Wisconsin, and †Mayo Clinic and Foundation*

gluten Wheat protein responsible for damage of small intestine in celiac disease.

gluten-free diet Regimen with no gluten or similar proteins in any meal component.

malabsorption Impaired intestinal absorption of nutrients.

Celiac disease, a chronic inflammatory condition associated with small intestinal injury induced by gluten exposure, results in malabsorption of different nutrients. It is associated with multiple other medical conditions. The diagnosis relies on characteristic findings of small intestinal biopsy. Patients with celiac disease usually respond quickly to a gluten-free diet. The disease requires a lifelong commitment to a gluten-free diet to prevent recurrence of symptoms and other potential consequences.

INTRODUCTION

Celiac disease (CD) was described in the first century A.D. However, modern knowledge of the relationship between CD and the damaging effect of gluten on the small intestine is credited to Dicke, who, in the 1940s, first observed improvement of children with CD when there was shortage of wheat during World War II. Dicke noted the recurrence of symptoms when food supplies became available again.

The disease may induce no symptoms at all for a long time, yet in some cases subtle symptoms can be detected and in others a more debilitating form of the disease may be encountered. The disease is being recognized more often nowadays thanks to increasingly available testing

 Copyright 2004, Elsevier (USA). All rights reserved.

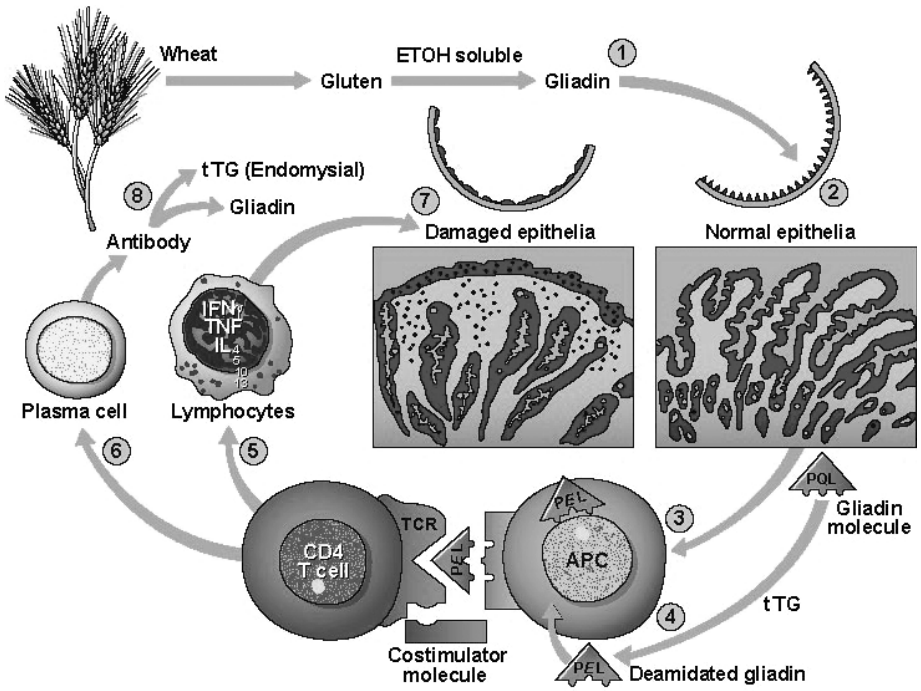


FIGURE 1 The alcohol-soluble gliadin (1) fraction of wheat is highly antigenic. After gliadin peptides are absorbed (2), they are taken up by an antigen-presenting cell (3). The antigen-presenting cell (APC) processes the peptides and presents the gliadin fragments, in conjunction with human leukocyte antigen DQ2 or DQ8, to a T cell expressing T cell receptors (3). Deamidation of the gliadin peptide ligand (PL), which is mediated by tissue transglutaminase (tTG), makes the binding of the ligand to DQ2 or DQ8 stronger by converting the glutamine (Q) in the ligand (PQL) to glutamic acid (E; PEL), inducing a stronger T cell response (4). Activation of T cells leads to activation of lymphocytes (5), which produce a variety of damaging cytokines, and activation of plasma cells (6), which are responsible for the production of antibodies. These actions result in a cascade of effects and ultimately in mucosal damage (7) and antibody production (8). IFN, Interferon ; TNF, tumor necrosis factor; IL, interleukin. Copyright Mayo Foundation, 2002.

methods and better awareness of the disease and its complications.

EPIDEMIOLOGY

Currently, the precise prevalence of CD is hard to predict. CD in the United States is more prevalent than what was initially thought. It was thought that the prevalence of clinically diagnosed CD was 1:300–4500 in parts of Europe and 1:10,000 in the United States. On the other hand, the prevalence in more recent population screening studies ranges between 1:100–500 in Europe and 1:100–200 in the United States. The prevalence of CD is higher in first-degree relatives of patients with CD (close to 10%).

PATHOGENESIS

In genetically susceptible individuals, ingested gluten peptides are presented by the antigen-presenting cells in association with human leukocyte antigen (HLA) DQ2 or DQ8 (Fig. 1). The gluten antigen is presented to T cells expressing α/β T cell receptors, initiating activation and production of lymphocytes, which are responsible for releasing cytokines (e.g., interferon γ and tumor necrosis factor α). Cytokines are thought to be the factors that damage the small intestinal epithelia. The damaged mucosa triggers the release of tissue transglutaminase (tTG), a cytosolic enzyme that deamidates gluten and catalyzes deamidation between gluten–gluten and possibly gluten–tTG; the cross-linking in turn augments the gluten presentation by HLA DQ2 or DQ8. In addition, released cytokines increase the expression of HLA DQ2 on small intestine epithelia, allowing increased antigen presentation to sensitized lymphocytes.

All alcohol-soluble protein components of wheat gluten (gliadins), barley (secalins), and rye (hordeins) are damaging in patients with CD. Rice, corn, and soybean flours are nontoxic. Oats are thought to be safe in most patients.

CLINICAL FEATURES

Children

Symptoms in children usually start after introduction of cereals into the diet, usually after the age of 6 months. Symptoms may include failure to thrive, diarrhea, vomiting, muscle wasting, abdominal distension, abdominal pain, and occasionally constipation. In older children, the disease can present with anemia, rickets, short stature, dental enamel defects, poor performance in school, or behavioral disturbances.

Adults

The disease in adulthood may present as adult onset or may be a clinically silent disease that was present since childhood but produced no symptoms. The most common presenting symptoms in adults with CD are chronic diarrhea and iron-deficiency anemia. Diarrhea may be absent in 50% of patients. Severe steatorrhea is less common and indicates more severe disease. Because of better awareness and test protocols, celiac disease is currently being diagnosed more frequently in asymptomatic patients who are found to have iron deficiency anemia, folate deficiency, or osteoporosis. Other less common features of CD are presented in Table I.

Dermatitis herpetiformis is associated with celiac disease, presenting with intestinal lesions of various degrees of severity in most cases and characterized by blistering skin lesions on the elbows, knees, and buttocks. Intestinal symptoms are not always present when CD presents as dermatitis herpetiformis. The skin disease usually responds well to a gluten-free diet (GFD), although initially dapsone might be needed.

TABLE I Findings in Celiac Disease

Dermatitis herpetiformis	Short stature
Delayed puberty	Abdominal pain
Abdominal distension	Osteopenia
Osteoporosis	Hepatic steatosis
Recurrent abortions	Steatorrhea
Folate-deficiency anemia	Vitamin K deficiency
Thrombocytosis	Infertility (male/female)
Anxiety	Depression
Arthralgia or arthropathy	Dental enamel hypoplasia
Ataxia	Alopecia
Isolated hypertransaminasemia	Aphthous stomatitis
Recurrent pericarditis	Polyneuropathy
Vasculitis	Dilated cardiomyopathy
Hypo/hyperthyroidism	Clubbing
Weight loss	Nausea/vomiting
Ecchymosis	Cheilosis
Epilepsy with or without intracranial calcification	

TABLE II Diseases Associated with Celiac Disease

Type 1 diabetes	IgA deficiency
Autoimmune thyroiditis	Down's syndrome
Primary biliary cirrhosis	Myasthenia gravis
Psoriasis	Autoimmune hepatitis
Addison's disease	Autoimmune atrophic gastritis
Turner syndrome	Congenital heart defects
Rheumatoid arthritis	Sarcoidosis
Cystic fibrosis	Schizophrenia

ASSOCIATED DISEASES

CD has been associated with a number of medical conditions, many of which occur with a higher frequency in patients with CD than in the general population. A list of such associated diseases is shown in Table II.

DIAGNOSIS

Small Intestinal Biopsy

The gold standard diagnostic test for CD is a small intestinal mucosal biopsy obtained during upper endoscopy, with multiple biopsies taken from the second or third part of the duodenum. Although on occasion they might give a hint for the diagnosis, flattened or scalloped duodenal folds seen at endoscopy are not specific for CD. The characteristic findings on the biopsy include (1) partial or complete villous atrophy, (2) crypt hyperplasia, and (3) increased intraepithelial lymphocytes and plasma cells. In the past, a diagnostic small intestinal biopsy on a gluten-containing diet followed by complete healing on a GFD, then deterioration on gluten challenge, was required for the diagnosis of CD. Currently, the criteria for the diagnosis of CD require a single biopsy with characteristic findings while on gluten-containing diet, and an unequivocal clinical response to a GFD.

Villous atrophy alone can be found in other diseases of the gastrointestinal tract; when not related to CD, villous atrophy requires different diagnostic methods and treatment. Some of the diseases that can be associated with villous atrophy include cow's milk allergy, lymphoma of the small intestine, autoimmune enteropathy, tropical sprue, and graft-versus-host, disease among others.

The small intestinal lesion of CD usually decreases distally from the duodenum to the ileum, with the first part of the small intestine usually healing last following institution of a GFD. The patient classically improves within a period of a few weeks of starting a GFD, with

histological improvement that might take up to 1 year before complete resolution of the characteristic lesions and return to a normal or near-normal architecture. In the case of refractory celiac disease, there is persistence of symptoms or recurrence of symptoms after initial improvement following institution of a GFD, and persistence of histological damage on small intestinal biopsy. This may be associated with ulceration and can progress to lymphoma.

Serological Markers

Antigliadin Antibodies

Gliadins, the alcohol-soluble fraction of gluten, elicit a strong humoral response that originates in the submucosa. Immunoglobulins of two subclasses (IgA and IgG), or antigliadin antibodies (AGAs), are produced in the small intestine and are used as an adjunct for diagnosis. Their lack of specificity in CD and the development of newer tests have rendered them not useful clinically for specific diagnosis. However, they can be used to follow the response to a GFD. The IgG subclass may help identify CD cases that are associated with IgA deficiency, whereby IgA AGAs, IgA antiendomysial antibodies, and antitransglutaminase are absent.

Antiendomysial Antibodies

Endomysium is a connective tissue protein found in the collagenous matrix of human and monkey tissue. IgA antibodies directed against the endomysium have been found in association with CD. These antibodies are highly sensitive and specific for CD, using both human umbilical cord or monkey esophagus as a substrate. The test is based on immunofluorescence findings of reticular staining when antiendomysial antibody binds to the endomysium. Although highly specific when positive, this binding is still subjective because of variations in interpretation from one laboratory to another. Like IgA AGAs, antiendomysial antibodies will be absent in IgA-deficient patients with CD.

Tissue Transglutaminase Antibodies

Tissue transglutaminase is a cytosolic protein that is released during wounding and serves as a cross-linker of different extracellular matrix molecules. It has been recently found by Diederich that tTG is the autoantigen of CD. IgA enzyme-linked immunosorbent assay (ELISA)-based kits are now available for tTG antibodies, using either human red cell or recombinant tTG.

Test Sensitivity/Specificity

The test sensitivities/specificities for immunoglobulin antigliadin antibodies, tissue transglutaminase, and antiendomysial antibodies (AEMAs) are as follows:

- IgA AGAs: 70–90%/85–94%
- IgG AGAs: 69–85%/88–92%
- IgA tTG: 90–98%/94–98%
- IgA AEMAs: 85–98%/98–100%

Hematologic and Biochemical Findings

Anemia in CD may be secondary to iron, folate, and, occasionally, vitamin B_{12} deficiencies. Anemia with iron deficiency is one of the most common presentations of CD. Low calcium and vitamin D levels can lead to osteopenia and osteoporosis. Leukopenia and thrombocytopenia can be secondary to severe folate deficiency, whereas thrombocytosis is usually seen with hyposplenism. Low vitamin K levels can result in increased prothrombin. Severe malabsorption and diarrhea can be associated with low sodium, albumin, potassium, magnesium, and zinc levels. IgA deficiency is seen in 2–10% of patients with CD. Persistent hypertransaminasemia in a patient with no other diagnosed chronic liver disease should raise the suspicion of CD.

Radiographic Findings

The role of radiographic evaluation in the initial diagnosis of celiac disease is limited. The nonspecific findings of thickened mucosal folds, dilatation of the small intestine, straightening of the valvulae conniventes, and excessive secretion of fluid into the small intestine on small intestinal X ray suggest malabsorption rather than CD per se. On the other hand, small intestinal X ray can be used to rule out serious complications of CD (lymphoma, ulcerative jejunitis, or carcinoma) in patients who are on a strict GFD and presenting with abdominal pain, weight loss, anemia, hypoalbuminemia, or symptoms of obstruction.

Computer tomography is also used to rule out lymphomas and adenocarcinomas and may reveal hyposplenism, ascites, or cavitating mesenteric lymphadenopathy that can be associated with CD.

Stool Studies

The 72-hour stool fat collection and D-xylose absorption tests are nonspecific and play only a minor role in the specific diagnosis of CD.

TREATMENT

A GFD is the mainstay of the treatment of CD and it must be a lifelong regimen. No drug therapy has been proved to suppress the disease. Involvement of the patient in the treatment plan is crucial. Physicians caring for patients with CD should educate the patient about celiac disease. They should provide dietary consultation regarding a GFD and changes in lifestyle and must also encourage patients to join support groups. Patients should be encouraged to lead as normal a lifestyle as possible. Other responsibilities of the physician include assessing the patient's nutritional status and response to interventions, screening for other known associated diseases, followup to ensure compliance with the GFD, obtaining baseline bone mineral density, and being on watch for potential complications of CD.

Gluten-Free Diet

A GFD as a treatment for CD was established by Dicke. A few points about the GFD require further clarification:

1. Even small amounts of gluten can result in injury to the intestinal lining.
2. Patient education and motivation are crucial for successful treatment of CD.
3. The diet requires extensive education of patients and their families by both doctors and experienced dietitians.
4. A trial of a GFD as the sole means of diagnosis should be abandoned.
5. Patients need to know that a GFD not only improves their symptoms but may also reverse anemia, infertility, depression, and osteoporosis and may help prevent small intestinal malignancy.

Clinical improvement is evident in most patients within 2–3 weeks of initiating a GFD. Gluten contamination of food in the diet is very common and is the leading reason for persistence of symptoms. Flours allowed in a GFD include rice, soybean, corn, soy, potato, tapioca, arrowroot, bean, sorghum, quinoa, millet, buckwheat, tef, nut, and amaranth flours. Products made from these flours are safe for use by celiac disease patients.

Wheat, barley, rye, malt, triticale (wheat–rye hybrid), couscous, kamut, spelt, and semolina flours are not allowed. Recent studies have clearly demonstrated that oats are nontoxic for patients with celiac disease; however, contamination of oat products with other prohibited flours during harvesting, milling, or processing of oat products is a concern.

Some items that might be overlooked as a source of contamination include broth, breading, ketchup, mustard, candy bars, cheese spreads, chip and dip mixes, flavoring in meat products, soy sauce, hydrolyzed plant protein, hot chocolate mixes or cocoa, imitation bacon or seafood, instant coffee and tea, salad dressing, malt, modified food starch, nondairy creamer, pasta, peanut butter, processed meat, sausage products, sauce and soup bases, stuffings, tomato sauce, vegetable gum, vegetable protein (thickener), and yogurt with fruit. Medications, vitamins, and mineral supplements may also contain gluten as an inactive ingredient.

A GFD may be associated with either weight gain or loss. It may also be associated with constipation, and the patient should be counseled about these side effects. Special attention should be paid to deficiency states and supplementation with specific agents as needed. Also, after improvement of intestinal injury and consequent improved absorption, adjustment in the dose of medications may be needed.

Support groups provide assistance to newly diagnosed celiac disease patients, particularly by providing name and contact information of companies that produce GFD items and of local vendors; support groups are also sources of up-to date information on the disease and on the GFD, in addition to a wide variety of materials, books, and recipes that can be used with physician/dietitian guidance.

COMPLICATIONS

Serious complications of CD may arise because of the severity of malabsorption and steatorrhea; in addition to osteomalacia and osteoporosis, complications include mineral deficiency states. Patients with CD are at increased risk of developing lymphomas (enteropathy-associated T cell lymphoma) and small intestinal adenocarcinoma, though the risk is still very small. Refractory sprue is defined as initial failure of a GFD to improve clinical symptoms and histological abnormalities of the small intestine or secondary failure after initial improvement. This complication is hard to treat and harbors the risk of progressing to lymphoma. Another serious complication is ulcerative jejunitis, which can be associated with duodenal strictures or intestinal perforation.

See Also the Following Articles

Calcium, Magnesium, and Vitamin D Absorption, Metabolism, and Deficiency • Endomysial and Related Antibodies • Immunodeficiency • Malabsorption • TH1, TH2 Responses • Tropical Sprue

Further Reading

Abdulkarim, A. S., and Murray, J. A. (2002). Celiac disease. *Curr. Treat. Options Gastroenterol.* **5**, 27–38.

Ciclitira, P. J. (2001). AGA technical review on celiac disease. *Gastroenterology* **120**, 1526–1540.

Dieterich, W., Ehnis, T., Bauer, M., Donner, P., Volta, U., *et al.* (1997). Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* **3**, 797–801.

Molberg, O., McAdam, S. N., Korner, R., Quarsten, H., Kristiansen, C., *et al.* (1998). Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* **4**, 713–717.

Mulder, C. J. J. (2001). When is a coeliac a coeliac? *Eur. J. Gastroenterol. Hepatol.* **13**, 1–6.

Celiac Disease, Pediatric

ALAN M. LEICHTNER

Harvard Medical School and Children's Hospital Boston

gliadin Alcohol-soluble protein fraction of gluten.

gluten Water-insoluble protein-rich residue remaining after wheat starch has been extracted from the dough made from wheat flour.

latent celiac disease Celiac disease in patients with who have a normal small intestinal biopsy on a gluten-containing diet, but at some time before or since have an enteropathy that resolved with gluten restriction.

refractory celiac disease Celiac disease not responsive to 6 months of a strict gluten-free diet.

silent celiac disease Asymptomatic celiac disease in patients with an abnormal intestinal biopsy.

Th1 response Helper T lymphocyte response that results in the production of interferon γ and other proinflammatory cytokines.

Celiac disease is an autoimmune enteropathy triggered by the ingestion of wheat or the related grains, rye and barley, in susceptible individuals. The clinical manifestations of this disorder are extremely variable and individuals with significant small bowel villous atrophy may have no gastrointestinal symptoms or be completely asymptomatic. Recognition of atypical and silent cases of celiac disease and the application of screening serological tests have led to the relatively recent discovery that the prevalence of this disorder has been greatly underestimated previously and may exceed 1 in every 200 individuals in Europe and North America.

INTRODUCTION

Although celiac disease has been recognized for centuries, the first person to link the disorder to ingestion of gluten was a pediatrician named Dicke in the 1940s. In fact, the classic description of celiac disease has derived from the childhood disease and much of our knowledge of the disorder has been gained from the study of children. The processes of growth and development combine to make children special, and the manifestations of celiac disease must be understood in this context. The unique aspects of childhood celiac disease will be emphasized in the following discussion.

EPIDEMIOLOGY

In the era before serological screening, the highest prevalence of symptomatic celiac disease was reported from Western Europe, approaching 1 in 300 individuals in Western Ireland. Celiac disease was also recognized in North America and Australia, both of which were sites of European emigration and where wheat remained a staple food. However, based on studies of patients undergoing small bowel biopsy because of gastrointestinal symptoms, estimates of prevalence in the United States during this era, were significantly lower, approximately

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Abdulkarim, A. S., and Murray, J. A. (2002). Celiac disease. *Curr. Treat. Options Gastroenterol.* **5**, 27–38.

Ciclitira, P. J. (2001). AGA technical review on celiac disease. *Gastroenterology* **120**, 1526–1540.

Dieterich, W., Ehnis, T., Bauer, M., Donner, P., Volta, U., *et al.* (1997). Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* **3**, 797–801.

Molberg, O., McAdam, S. N., Korner, R., Quarsten, H., Kristiansen, C., *et al.* (1998). Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* **4**, 713–717.

Mulder, C. J. J. (2001). When is a coeliac a coeliac? *Eur. J. Gastroenterol. Hepatol.* **13**, 1–6.

Celiac Disease, Pediatric

ALAN M. LEICHTNER

Harvard Medical School and Children's Hospital Boston

gliadin Alcohol-soluble protein fraction of gluten.

gluten Water-insoluble protein-rich residue remaining after wheat starch has been extracted from the dough made from wheat flour.

latent celiac disease Celiac disease in patients with who have a normal small intestinal biopsy on a gluten-containing diet, but at some time before or since have an enteropathy that resolved with gluten restriction.

refractory celiac disease Celiac disease not responsive to 6 months of a strict gluten-free diet.

silent celiac disease Asymptomatic celiac disease in patients with an abnormal intestinal biopsy.

Th1 response Helper T lymphocyte response that results in the production of interferon γ and other proinflammatory cytokines.

Celiac disease is an autoimmune enteropathy triggered by the ingestion of wheat or the related grains, rye and barley, in susceptible individuals. The clinical manifestations of this disorder are extremely variable and individuals with significant small bowel villous atrophy may have no gastrointestinal symptoms or be completely asymptomatic. Recognition of atypical and silent cases of celiac disease and the application of screening serological tests have led to the relatively recent discovery that the prevalence of this disorder has been greatly underestimated previously and may exceed 1 in every 200 individuals in Europe and North America.

INTRODUCTION

Although celiac disease has been recognized for centuries, the first person to link the disorder to ingestion of gluten was a pediatrician named Dicke in the 1940s. In fact, the classic description of celiac disease has derived from the childhood disease and much of our knowledge of the disorder has been gained from the study of children. The processes of growth and development combine to make children special, and the manifestations of celiac disease must be understood in this context. The unique aspects of childhood celiac disease will be emphasized in the following discussion.

EPIDEMIOLOGY

In the era before serological screening, the highest prevalence of symptomatic celiac disease was reported from Western Europe, approaching 1 in 300 individuals in Western Ireland. Celiac disease was also recognized in North America and Australia, both of which were sites of European emigration and where wheat remained a staple food. However, based on studies of patients undergoing small bowel biopsy because of gastrointestinal symptoms, estimates of prevalence in the United States during this era, were significantly lower, approximately

Copyright 2004, Elsevier (USA). All rights reserved.

1 in 2000–3000. The advent of serological testing has permitted screening of larger populations, and recent series of blood donor tests both in Europe and in the United States have demonstrated strikingly higher prevalence rates. Furthermore, these studies have also revealed that a majority of patients are either asymptomatic or have previously unrecognized atypical symptoms of celiac disease. An Italian study of school-age children revealed more than six times as many unexpected cases of celiac disease as established cases. These studies have led to the concept of a celiac disease "iceberg," with classical symptomatic cases being far outnumbered by atypical and silent cases below the "water's" surface (Fig. 1). Currently, the prevalence in Western Europe, estimated to be greater than 1 in 100, and estimates for the United States, at more than 1 in 250, are similar. Furthermore, the prevalence of this disorder has probably been underestimated in certain non-European countries and areas to which Europeans have not emigrated in high numbers, including the northern Sudan and northern India and other parts of Asia.

Early childhood experiences may affect the incidence of celiac disease and its clinical manifestations. The epidemiology of celiac disease has been carefully studied in Sweden, especially in relationship to gluten introduction and breast feeding. The incidence of celiac disease in that country rose rapidly in the period from 1985 to 1987, when infant gluten consumption was increased through the addition of gluten to baby formula, despite the postponement in cereal introduction from 4 to 6 months of age. Starting in 1995, there was a dramatic decline in early childhood celiac disease in Sweden, coincident with national recommendations to promote breast feeding and the introduction of gluten in smaller amounts while babies were still nursing. In London, a similar decline in incidence of celiac disease in the 1980s and 1990s followed recommendations for later introduction of gluten in the infant diet. Experts have speculated that the delayed introduction of large amounts of gluten in the diet may not have actually decreased the overall incidence of celiac disease, but just changed its presentation. Instead of developing classical symptoms of diarrhea and failure to thrive in the first 2 years of life, more at-risk individuals were presenting in later childhood or as adults, often with atypical symptoms. An independent effect of breast feeding on the occurrence of childhood celiac disease is not clear and merits further study.

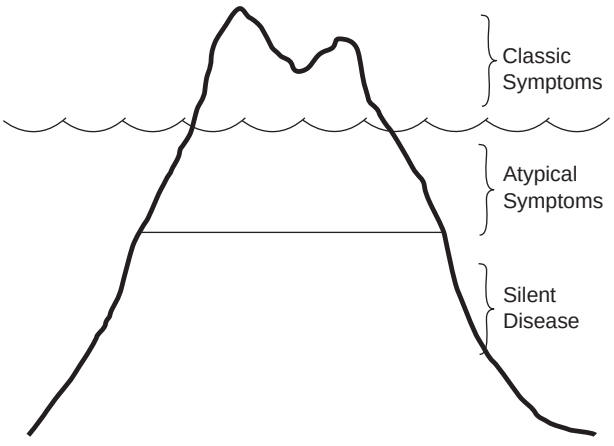


FIGURE 1 The celiac iceberg. Classic symptoms are the tip of the celiac iceberg; most patients have atypical symptoms or silent disease.

Substantial epidemiological evidence of a genetic basis for celiac disease now exists. Approximately 10% of children with celiac disease have a family member who is also affected by the disease and the incidence in monozygotic twins is up to 70% in some series. Histocompatibility genes clearly account for part of this susceptibility; 95% of patients with celiac disease are positive for human leukocyte antigen (HLA) DQ2 or DQ8, as compared to 30% of control patients. The fact that the concordance rate in monozygotic twins is higher than that in HLA identical siblings suggests that other genes may play an important role in the expression of celiac disease.

PATHOPHYSIOLOGY

Major advances have been recently made in understanding the pathogenesis of celiac disease. The toxic components of wheat are proteins, gliadins, which have a rich content of proline and glutamine. These proteins are modified by the enzyme tissue transglutaminase and the resulting peptides are subsequently bound by antigen-presenting cells in the context of the HLA DQ2 molecule. As a consequence, CD4-positive T cells are stimulated to proliferate in a Th1 response. The release of proinflammatory cytokines results in destruction of small intestinal epithelial cells. The unique sequence of the gliadin-derived peptides is required both for the interaction with HLA molecules and for resistance to gastric and intestinal proteolysis.

Abnormal intestinal permeability may also be an important factor in the pathophysiology of celiac disease. One possible mechanism for alteration of permeability is zonulin, a recently discovered protein that can reversibly alter tight junction assembly, thereby increasing permeability via the paracellular pathway. The expression of zonulin has been noted to be increased in intestinal tissues during the acute phase of celiac

disease. Regardless of the exact mechanism, if increased intestinal permeability is an early event in pathogenesis of celiac disease, it might permit exposure of gliadin to the mucosal immune system. Some researchers speculate further that increased permeability of the intestine in celiac disease may permit entry of other luminal factors important in the pathogenesis of associated autoimmune diseases, such as diabetes mellitus.

Despite the elucidation of the key steps in the pathogenesis of celiac disease, many questions remain. The event that triggers the development of celiac disease in a susceptible host is not clear, but presumably could be an infection or other stress resulting in an increase in intestinal permeability to gliadin peptides. Whether the antibody response to tissue transglutaminase observed in active celiac disease plays an important role in the genesis of intestinal injury or is merely an epiphenomenon also remains to be elucidated.

PATHOLOGY

Immune-mediated injury results in the classical histopathologic features of celiac disease, which are subtotal or total villous atrophy, crypt hyperplasia, and an increase in intraepithelial lymphocytes and lamina propria lymphocytes and plasma cells. As cases are diagnosed earlier or in patients with few symptoms, more subtle changes in the intestinal mucosa are being recognized, and some biopsies may demonstrate focal or partial villous atrophy, or only an increase in intraepithelial lymphocytes. In some patients with celiac disease, the immune-mediated injury is not confined to the small intestine and there may be associated lymphocytic gastritis and/or lymphocytic colitis.

CLINICAL MANIFESTATIONS

The typical symptoms of celiac disease follow the introduction of wheat, rye, or barley in to the diet and consist of diarrhea, abdominal pain, and weight loss (or failure to gain weight appropriately). Typically, children between the ages of 9 and 18 months present with frequent, foul-smelling bowel movements, weight loss, poor appetite, a distended abdomen, thin extremities with loss of subcutaneous tissue and sometimes muscle wasting, and irritability (Fig. 2). In severe cases, the child may be pale, listless, and have peripheral edema. In infants younger than 9 months, vomiting may be a prominent symptom. Paradoxically, some children present with constipation, rather than diarrhea, and celiac disease should be considered in the differential diagnosis of intractable constipation.

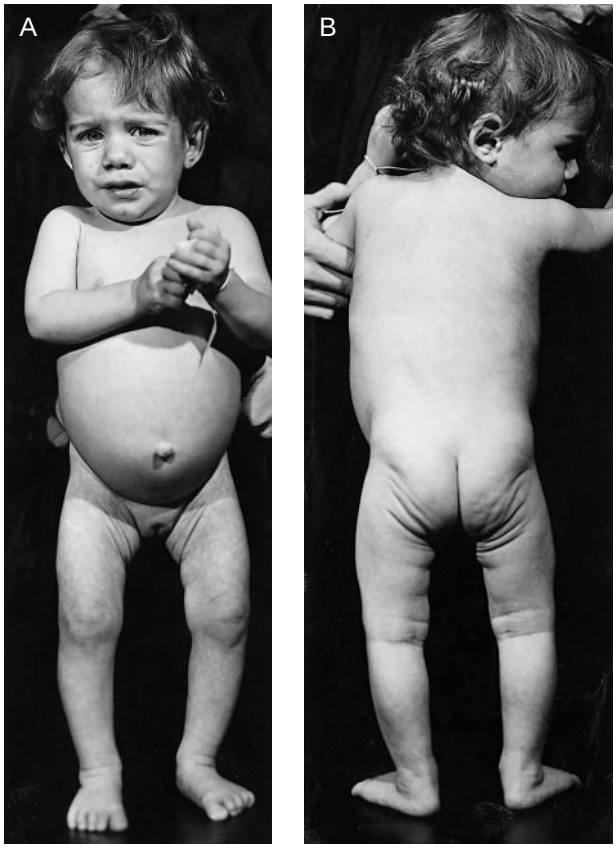


FIGURE 2 An 18-month-old child with celiac disease. Note the anxious look, the distended abdomen, and loss of subcutaneous tissue in the thigh folds and buttocks. Reproduced with permission from H. Shwachman (1954). Mucoviscidosis and the Celiac Syndrome. *Pediatr. Clin. North Am.* **54**.

Although diarrhea and weight loss may occur at any age, older children are more apt to have an atypical presentation. It is extremely important to recognize that isolated short stature, delayed puberty, anemia resistant to iron therapy, osteoporosis, primary amenorrhea, and unexplained elevation of serum transaminase levels may be the presenting complaint of children or adolescents with celiac disease. Other less common signs and symptoms include oral ulcers, behavior problems, dental hypoplasia, and seizures associated with cerebellar calcification.

The intolerance to gluten in celiac disease should be considered permanent, although rare cases of transient gluten intolerance have been reported in childhood. Patients with so-called transient gluten intolerance present with gastrointestinal symptoms and an

abnormal small intestine, and respond to a gluten-free diet. With gluten reintroduction, however, the patient remains free of symptoms and the intestinal biopsy remains normal. Because histopathologic relapse can be delayed after challenge of celiac patients with gluten for months to years, some of these patients have latent celiac disease and will eventually become symptomatic.

A wide range of disorders are associated with celiac disease. Children with Down syndrome have an increased prevalence of celiac disease, estimated at 5%. Other genetic disorders recently associated with celiac disease include Turner syndrome and Williams syndrome. Dermatitis herpetiformis is a pruritic papulovesicular rash that is frequently associated with small intestinal injury reminiscent of celiac disease. This disorder usually responds to a gluten-free diet, although treatment with dapsone is sometimes required. Other autoimmune disorders associated with celiac disease include type 1 diabetes mellitus, autoimmune thyroiditis, rheumatoid arthritis, and possibly others. Up to 8% of patients with diabetes have celiac disease. Since these children often have no gastrointestinal symptoms, the only clue to diagnosis may be difficulty in controlling serum glucose levels. Based on the strong association with celiac disease, however, it would seem advisable to screen all patients newly diagnosed with type 1 diabetes. Immunoglobulin A (IgA) deficiency is also associated with celiac disease and may confound serologic diagnosis.

DIAGNOSIS

The diagnosis of celiac disease should always rest on the finding of histologically abnormal small intestinal mucosa. However, because of the invasive nature of small bowel biopsy, serological tests have been developed and are useful as screening tests. These include assays for antigliadin antibodies (AGAs), antireticulin antibodies (ARAs), antiendomysial antibodies (AEAs), and antitissue transglutaminase antibodies (ATTGs). Patients with celiac disease may have IgA and IgG antibodies in each of these classes, but detection of the IgA class is generally a more specific assay for celiac disease and is the basis for most commercial test kits. The IgG AGA assay has a particularly low specificity and may be elevated in a significant number of control subjects and in patients with other intestinal disorders. Until recently, the best test was the AEA assay, which typically has a sensitivity and specificity of greater than 95%. However, because detection of AEAs is based on an immunofluorescent assay using monkey esophagus or human placental tissue as a substrate, there may be significant variability between different laboratories and the test is quite expensive to perform. The enzyme tissue transglutaminase has recently been identified as a specific target of the AEA assay. This discovery has led to the development of enzyme-linked immunosorbent assays (ELISAs) directed at this enzyme. Initially, the ATTG assays used guinea pig enzyme as the substrate, but the most recent generation assay is based on the human protein. Early studies noted a slightly reduced sensitivity and/or specificity of the ATTG when compared to the AEA assay, but it is likely that the more recently developed assays will not detect significant differences. If the equivalency of these assays is confirmed, it would be most prudent for the clinician to order only the ATTG assay, rather than a panel of different antibodies. Because IgA deficiency can be associated with celiac disease, a total serum IgA should be obtained to ensure that a negative assay is not due to IgA deficiency.

The availability of serological assays has obviated the need for tests of malabsorption and for imaging studies in uncomplicated cases. Standard tests for malabsorption, such as D-xylose absorption assays, breath testing to detect secondary lactose intolerance in young children, and qualitative and quantitative analyses of fecal fat, are too insensitive to be of practical use for the diagnosis of celiac disease in children. Likewise, sugar permeability tests fail to identify most children with active celiac disease. Imaging tests are of limited utility in diagnosing celiac disease and are not indicated unless gastrointestinal complications, such as intussusception, are suspected.

Patients with positive serology and those with negative serology but symptoms suggestive of celiac disease should undergo small bowel biopsy. The first small bowel biopsies were obtained using suction-triggered capsules positioned in the jejunum under fluoroscopic guidance. The advent of small-diameter flexible endoscopes has permitted the performance of endoscopic biopsies, even in infants. Anxious children, particularly those under 5 years of age, may require that the endoscopy be performed under general anesthesia. Patients with celiac disease undergoing endoscopy have been described to have scalloped or otherwise irregular small intestinal folds; however this finding is nonspecific. More recently, a technique of flooding the intestinal lumen to provide a magnified view of the intestinal surface and actual visualization of villi has been developed. Multiple small bowel biopsies should be obtained (at least three), because early intestinal changes may be focal. Biopsies should be obtained as far distally as possible and not in the duodenal bulb, where nonspecific changes of duodenitis may confound biopsy interpretation.

TABLE I Diagnostic Criteria for Celiac Disease

Consistent small bowel biopsy
and at least one of the following responses:
1. Clinical response (resolution of gastrointestinal symptoms) to gluten-free diet
2. Serological response to gluten-free diet
or
Histopathological remission on a gluten-free diet and relapse with gluten challenge (positive gluten challenge)

Traditionally, patients with suspected celiac disease have been subjected to a series of three biopsies to confirm the diagnosis and exclude transient gluten sensitivity or other processes that mimic the clinical and histopathological features. A second biopsy has been performed at least 1 year after the diagnosis to confirm resolution of the mucosal injury. The third biopsy has followed a challenge with dietary gluten to reproduce the histopathological changes of active disease. Current criteria for diagnosis are based on recommendations of the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) and include a small bowel biopsy with compatible histopathological features and a complete clinical response to a gluten-free diet (Table I). Because elevated levels of antiendomysial antibodies in patients with celiac disease decrease to normal levels with adherence to a gluten-free diet, return to normal antibody levels constitutes a complete serological response. It is important to note, however, that it usually takes months for these levels to return to normal.

A gluten challenge should still be considered in certain patients, including those who did not have a biopsy performed before starting on a gluten-free diet or who had atypical biopsies, and those with atypical symptoms and negative serology. In children under 2 years of age, a number of other disorders may result in a flat villous lesion mimicking celiac disease (Table II). Therefore, patients presenting at this age should be considered for gluten challenge, especially if serological studies were not performed or were negative. Also patients, typically teenagers, seeming to tolerate some gluten in their diet and intending to discontinue the gluten-free diet should undergo gluten challenge before the diagnosis of celiac disease is abandoned. The actual challenge can be performed by either reintroducing a standard diet or continuing a gluten-free diet, with the addition of gluten powder added to a food such as applesauce. The latter approach may cause less emotional distress in children who fail the challenge. Repeat serology and biopsy are performed with the recurrence of symptoms or, in the absence of symptoms, after a period of 3 months. Caution should be exercised during the gluten challenge, because patients will occasionally develop severe symptoms (celiac crisis) with reintroduction of dietary gluten.

TABLE II Causes of Villous Atrophy in Childhood

Celiac disease	Autoimmune enteropathy
Cow's milk protein sensitivity	Immunodeficiency
Soy protein sensitivity	Microvillous inclusion disease
Severe viral enteritis	Malnutrition
Giardiasis	Chemotherapy
Tropical sprue	Radiation

Given the high incidence of celiac disease in first-degree relatives of children with celiac disease, parents and siblings should be screened with serological testing. Those with positive serology or negative serology but compatible symptoms should undergo small bowel biopsy.

TREATMENT

The treatment of celiac disease is a gluten-free diet, which requires the dietary exclusion of wheat, barley, rye, and other substances that may be contaminated by these grains. The question of whether small amounts of gluten in the diet are harmful has not been adequately answered. Therefore, the goal of the diet should be to eliminate these grains completely. Plant taxonomy can serve as a guide to what other grains should be acceptable in the diet of patients with celiac disease. Grains derived from plants closely related to wheat, such as barley and rye, are toxic, whereas unrelated grains, such as corn and rice, are completely tolerated. Oats, which are a more distant relative to wheat than are rye and barley, have been demonstrated to be safe in both adults and children with celiac disease when ingested in a pure form. The major practical issue is that oats are frequently contaminated by wheat both in the field and in the factory.

Given the ubiquitous occurrence of wheat in the diet, following a gluten-free diet is extremely challenging and requires the assistance of a specially trained dietitian. Determination of whether prepared foods are truly gluten-free demands continual vigilance, because manufacturers do not always explicitly list specific ingredients or may change the source of their ingredients without warning. Organizations and support groups can be very helpful in providing updated information in this regard. Special gluten-free foods are available in specialty food stores and via mail order or the Internet. Baking, however, remains quite difficult, because several unique nongliadin proteins in wheat are responsible for the stickiness critical to the baking

process. Hidden sources of gluten include food additives, emulsifiers, and stabilizers. Importantly, some medications may contain these substances.

Dietary changes other than gluten restriction are generally not necessary. Secondary lactose intolerance is unusual even in recently diagnosed children with celiac disease, although occasional patients may benefit from temporary restriction of lactose. Vitamin and mineral supplements may be recommended, especially if malabsorption or deficiencies of folate, iron, and calcium are suspected.

COMPLICATIONS

Occasionally, patients with celiac disease do not respond to a gluten-free diet alone and require immunosuppressive therapy for so-called refractory celiac disease. In children, such refractory disease is rare. Furthermore, before this diagnosis is made, other complications such as lymphoma need to be excluded. The incidence of enteropathy-associated T cell lymphoma has been documented to be higher in patients with celiac disease than in the general population. Patients with celiac disease also have a greater risk of developing small intestinal, esophageal, and oropharyngeal carcinoma. However, this increased risk of developing certain gastrointestinal cancers may be reduced by strict adherence to a gluten-free diet.

Patients with celiac disease are at increased risk of developing other autoimmune diseases, such as type 1 diabetes mellitus. This risk has been demonstrated to increase with increasing age at diagnosis. Whether the higher risk is the result of a longer period of exposure to gluten or simply the older age of the patient is not clear. Nevertheless, further study may demonstrate that strict adherence to the gluten-free diet may decrease the risk of developing other autoimmune diseases in patients with celiac disease.

PROGNOSIS

Response to a gluten-free diet usually is prompt, with most children demonstrating improvement in symptoms within 1 to 2 weeks. Mucosal healing may be demonstrated in 6–12 weeks. However, complete catch-up growth may take 1 to 2 years. With strict adherence to a gluten-free diet, life expectancy is similar to that of the general population and parents should expect their children with celiac disease to lead productive lives.

Acknowledgments

This work was supported by the Shoolman Fund for Childhood Celiac Disease at Children's Hospital, Boston.

See Also the Following Articles

Diabetes Mellitus • Endomysial and Related Antibodies • Immunodeficiency • Malabsorption • TH1, TH2 Responses

Further Reading

Branski, D., and Troncone, R. (1998). Celiac disease: A reappraisal. *J. Pediatr.* **133**, 181–187.

Farrell, R., and Kelly, C. (2002). Celiac sprue. *N. Engl. J. Med.* **346**, 180–188.

Fasano, A., and Catassi, C. (2001). Current approaches to diagnosis and treatment of celiac disease: An evolving spectrum. *Gastroenterology* **120**, 636–651.

Shan, L., Molberg, O., Parrot, I., Hausch, F., Filiz, F., Gray, G., Sollid, L., and Khosla, C. (2002). Structural basis for gluten intolerance in celiac sprue. *Science* **297**, 2275–2279.

Walker-Smith, J. (2000). *In* "Celiac Disease in Pediatric Gastrointestinal Disease" (W. A. Walker, P. Durie, J. R. Hamilton, J. Walker-Smith, and J. Watkins, eds.), 3rd Ed., pp. 727–746. B. C. Decker, Hamilton, Ontario.

Walker-Smith, J., Guandalini, S., Schmitz, J., Shmerling, D., and Visakorpi, J. (1990). Revised criteria for the diagnosis of coeliac disease. *Arch. Dis. Childhood* **65**, 909–911.

Cestodes

MELISSA HELD AND MICHAEL CAPPELLO
Yale University

hydatid Cyst with daughter cysts, each containing several scolices.
proglottid The segment of the tapeworm that, when gravid, contains eggs.
scolex The head of the tapeworm.
strobilation The process by which a tapeworm grows in an anterior to posterior direction.

The adult forms of cestodes share several common characteristics. Mammalian tapeworms have no internal cavity but are flat with a segmented body and a head or scolex. The scolex may have two sucking grooves as in *Diphyllobothrium* or four circular suckers as in *Taenia*. The adult tapeworm grows distally in a process called strobilation, so that the most posterior segments are the most mature. The segments of the tapeworm are named proglottids and contain egg-filled uteri when gravid. Tapeworms are hermaphroditic, with proglottids containing both ovaries and testes. There are four major groups of cestodes, but only those of the orders Pseudophylidea (*Diphyllobothrium latum*) and Cyclophylidea (*Taenia saginata, Taenia solium, Hymenolepis nana*) are significant parasites in humans. Clinical manifestations of tapeworm infection are attributable only to the adult or larval stages. This article will describe cestode infections of the human gastrointestinal tract and hepatobiliary system.

TAENIA SOLIUM

Life Cycle and Biology

Taenia solium infection is common in Mexico, Central and South America, parts of Africa, and Southeast Asia. *T. solium* can cause disease in both its adult and its larval forms. Intestinal tapeworm infection (taeniasis) is acquired by eating undercooked pork containing the larval stages, whereas ingestion of *T. solium* eggs leads to tissue dissemination of larvae and cyst formation (cysticercosis). Neurocysticercosis, or infection of the central nervous system with *T. solium*, is the leading cause of seizures in tapeworm endemic areas. Importantly, adult tapeworm carriers are the primary source of transmission of the larval stages in humans.

The *T. solium* tapeworm (Fig. 1) has a scolex outfitted with four suckers and a double crown of hooks, a narrow neck, and a large body (strobila) that measures between 2 and 4 m and consists of several hundred proglottids. The life cycle begins when human feces contaminated with *Taenia* proglottids are ingested by pigs. The invasive oncospheres in the eggs are liberated by the action of porcine gastric acid and intestinal fluids. The oncospheres then actively cross the bowel wall, enter the bloodstream, and are carried to the muscles and other tissues, where they develop into larval vesicles or cysticerci. Ingestion of undercooked pork containing cysticerci leads to cyst wall breakdown and attachment of the larval scolex to the intestinal wall. Maturation into the adult worm takes 6–8 weeks, at which time the adult sheds up to 10 proglottids per day. Proglottids may contain 50,000 to 100,000 eggs.

Clinical Features

Symptoms of intestinal taeniasis often include vague abdominal pain, nausea, changes in appetite, or diarrhea. However, most cases are asymptomatic until passage of proglottids in feces is noted.

Diagnosis, Treatment, and Prevention

Anti-cysticercal antibodies can be detected in serum, cerebrospinal fluid, and saliva using several assays, including complement fixation, hemagglutination, radioimmunoassay, enzyme-linked immunosorbent assay (ELISA), and immunoblot. For those patients with neurologic abnormalities, computed tomography (CT) or magnetic resonance imaging (MRI) of the brain will detect lesions diagnostic of neurocysticercosis. Praziquantel in a single dose (5–10 mg/kg) is the drug of choice for intestinal taeniasis. The role of anthelmintic treatment in the management of uncomplicated neurocysticercosis remains controversial. Infection can be prevented by thorough cooking of pork and sanitary measures taken to avoid spread from tapeworm carriers.

 Copyright 2004, Elsevier (USA). All rights reserved.

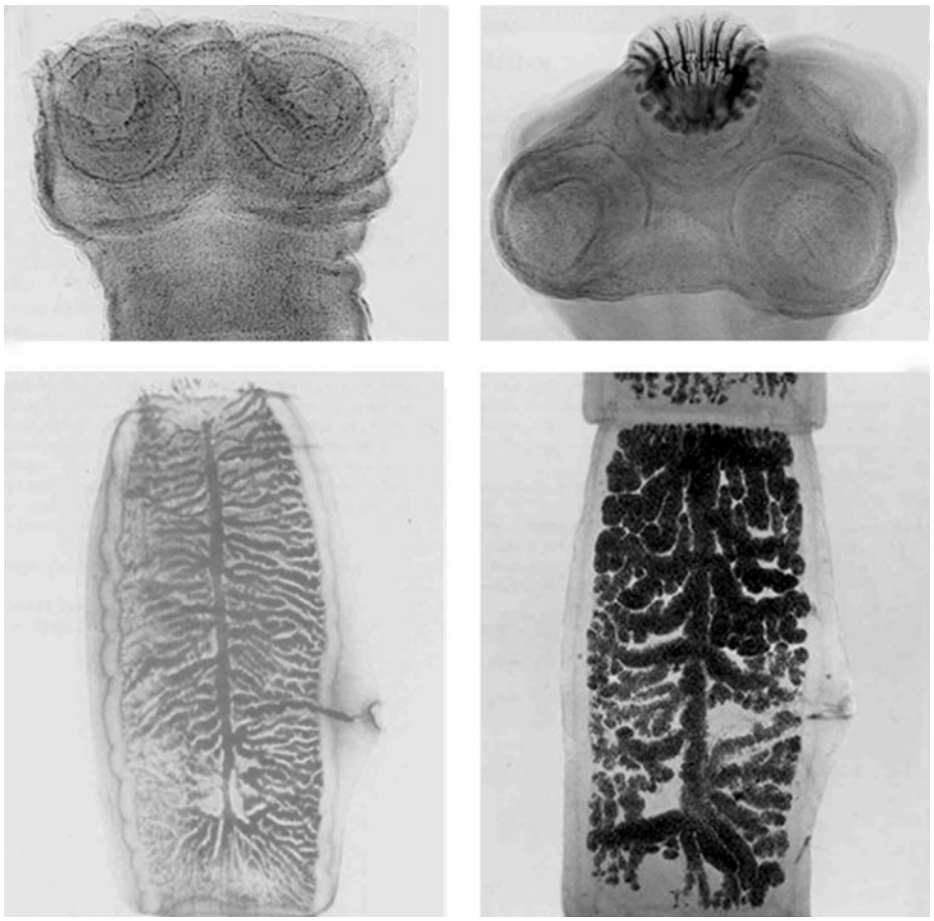


FIGURE 1 Scolices (top) and proglottids (bottom) from the tapeworms *Taenia saginata* (left) and *T. solium* (right). The proglottids are excreted in the feces of infected individuals. Reprinted from Ash, L., and Orihel, T. (1997). "Atlas of Human Parasitology," 4th Ed., ACSP Press, Chicago. Copyright © 1977 by the American Society of Clinical Pathologists. Reprinted with permission.

TAENIA SAGINATA

Life Cycle and Biology

Taenia saginata, the beef tapeworm, is highly endemic in Africa and South America, with a prevalence of greater than 90% in certain areas. The initial step in the life cycle is contamination of pastures or animal feed with human feces that contain *Taenia* eggs. The eggs are ingested by the cattle and hatch in the intestine. They travel via the bloodstream and lymphatics and then develop into cysticerci in the tissues. When undercooked beef containing cysticerci is ingested, the larvae are released from the cyst and attach to the intestine via four suckers (no hooks). The worms then grow distally by strobilation, reaching lengths of 10–25 m. Distal proglottids are shed, with each containing up to 80,000 eggs. The proglottids or eggs usually pass through the large intestine and are excreted. The cycle is completed when the eggs or proglottids are ingested by the intermediate host (cattle).

Clinical Features

The incubation period from the time of infection to the passage of proglottids is approximately 10–12 weeks. Infection with a single tapeworm is typical and many patients are asymptomatic unless they note worm parts being passed from the anus. Symptoms associated with *T. saginata* include nausea, postprandial fullness, abdominal pain, and occasionally vomiting or diarrhea. Eosinophilia of up to 1000 mm^3 occurs in approximately half of patients with *T. saginata* infections. Of note, mild eosinophilia is also common in *H. nana, D. latum*, and intestinal *T. solium* tapeworm infections.

Diagnosis, Treatment, and Prevention

Identification of gravid proglottids is the usual method of diagnosing *T. saginata* infection. They should be collected in saline or fixed in formalin. In order to determine the species of tapeworm, proglottids are pressed gently between two microscope slides, India ink is injected into the lateral pore, and the number of uterine branches is counted. *T. saginata* contains more than 15 primary branches on each side of the central core (Fig. 1). Eggs of *T. saginata* and *T. solium* are morphologically similar and are indistinguishable by light microscopy. Coproantigen detection of *Taenia*-specific antigens by ELISA is a sensitive diagnostic tool, although cross-reactivity with other *Taenia* species can occur. Drug therapy is extremely effective at eradicating tapeworm infections in adults and children. The drug of choice for *T. saginata* infection is praziquantel, which can be given orally in a single dose (5–10 mg/kg). Paromomycin is also effective and is considered safe for use during pregnancy. Proglottid segments may pass for several days after treatment is initiated. Follow-up stool examinations should be performed 3 months after treatment to document cure. Re-infection is quite common in endemic areas. The best ways of preventing infection include thorough cooking of meat and sanitary disposal of human feces. Freezing of beef will kill cysticerci. Unfortunately, these measures of public health control are not always feasible in endemic areas.

DIPHYLLOBOTHRIUM LATUM

Life Cycle and Biology

D. latum can measure up to 15 m in length. Its scolex has two deep grooves rather than suckers. An adult worm can release up to 1 million eggs per day. Once an egg reaches fresh water, a coracidium hatches from the embryonated egg, which is then ingested by a small freshwater crustacean (copepod). A procercoid develops within the copepod. When a fish ingests the infected copepod, the procercoid invades the stomach wall and penetrates the host's muscle tissue. The procercoid develops in the muscle of the fish into a plerocercoid. Humans become infected when they eat fish that contains plerocercoid larvae. The larvae then mature into adult tapeworms in the human intestine. Some mammals that ingest fish may also become definitive hosts for *D. latum*.

Clinical Features

Most patients infected with *D. latum* are asymptomatic. Patients may note worm segments in stool. Significant gastrointestinal symptoms occur rarely, although nausea, diarrhea, abdominal pain or distension, and weight loss have been reported. Chronic infection with *D. latum* is associated with macrocytic anemia secondary to vitamin B12 deficiency.

Diagnosis, Treatment, and Prevention

Diagnosis can be established by examination of the stool for proglottids or eggs. Praziquantel is effective treatment (5–10 mg/kg). The greatest risk for acquiring the infection is the ingestion of raw fish. Not surprisingly, the disease is most common in Asia, in northern Europe, and among Native American populations in Alaska and northern Canada. As sushi has become more popular in the United States, an increased number of cases occurring after patients have eaten raw salmon have been reported.

ECHINOCOCCUS SPECIES

Humans can become infected with the larval stages of *Echinococcus granulosus, Echinococcus multilocularis*, and *Echinococcus vogeli*. The adult worm of *E. granulosus* is found in the intestine of dogs or wolves. Humans become an incidental intermediate host when they accidentally ingest infected eggs from canine feces. Infections occur most frequently in areas where sheep or cattle raising are common.

Life Cycle and Biology

The adult *E. granulosus* tapeworm has a scolex with hooks, a neck region, and one immature, one mature, and one gravid proglottid. Adult tapeworms are found in the intestines of canines such as dogs and wolves. The eggs, which are morphologically identical to those of *Taenia* species, are passed in the feces and ingested by an intermediate host such as sheep. The embryos then hatch from the eggs, penetrate the intestinal mucosa of the host, and enter the bloodstream or lymphatics to be carried to various organs. Larvae then develop into fluid-filled unilocular hydatid cysts with an external membrane and an inner germinal layer. Protoscolices, which precede the development of the adult worm scolex, develop from within the inner capsule of the expanding cyst and can accumulate within the cyst as hydatid sand. Daughter cysts can grow from the inner germinal layer as well as cystic structures termed brood capsules. Cysts may rupture, leading to multifocal dissemination of infectious scolices. Adult worms of *E. multilocularis* are morphologically similar to those of *E. granulosus*, but the larval stage grows by external budding. The disease caused by *E. multilocularis* is

termed alveolar hydatid disease. Foxes are usually definitive hosts and rodents are intermediate hosts for *E. multilocularis*.

Clinical Features

Hydatid cysts are capable of developing in nearly any tissue but almost 90% develop in either the liver or the lung. Most patients will have only a single cyst, but they may develop in multiple sites. Symptoms usually develop once the cyst has become large enough to impinge on surrounding tissues. Cysts may grow slowly and can take up to 20 years to become symptomatic. Lung abscesses may develop if a cyst ruptures into a bronchus or the pleural space. Liver cysts may be asymptomatic calcified cysts or may become large enough to cause right upper quadrant discomfort. Biliary obstruction from a large cyst may present with jaundice, abdominal pain, and fever. In 5–15% of infected adults, hepatic cysts rupture into the biliary tract.

Diagnosis, Treatment, and Prevention

A history of exposure to a canine source in an endemic area is suggestive of echinococcosis. Daughter cysts may appear as highly opaque densities on MRI or CT scan. Certain reference laboratories can perform ELISA, immunoblots, or indirect hemoagglutination tests on serum from infected patients. Surgical removal or drainage of cysts in combination with three or more cycles of albendazole (400 mg twice daily for 1 month) is the recommended treatment regimen. The combination of albendazole with praziquantel may be more effective in some patients.

HYMENOLEPIS NANA

Hymenolepis nana, also called the human dwarf tapeworm, is the cestode that most commonly infects humans. Transmission occurs most commonly via the fecal–oral route and commonly affects young children in endemic areas.

Life Cycle and Biology

The life cycle does not require an intermediate host. Once eggs are ingested, they hatch in the small intestine to release a double-membraned oncosphere. The oncosphere then penetrates the intestinal mucosa and develops into a cysticercoid larva within the lymphatics of the intestinal villi. The cysticercoid then migrates back into the lumen of the intestine, attaches by its scolex to the mucosa, and matures into an adult. The adult tapeworm can measure 20 to 40 mm in length. It has four suckers on its scolex and a retractable rostellum that contains up to 25 hooks. Each proglottid contains three testes and one ovary and may contain up to 200 eggs when gravid.

Clinical Features

Most cases are asymptomatic, but heavy infection with *H. nana* can cause significant intestinal inflammation, especially in children. Diarrhea, abdominal pain, anorexia, and pruritus ani have also been described.

Diagnosis, Treatment, and Prevention

Diagnosis is confirmed by identifying the characteristic double-membraned eggs in a patient's stool. Egg output is irregular and multiple stool examinations may need to be performed. ELISA has been used to detect antibodies to *H. nana* in sera, but the test has poor specificity and is not routinely used. The drug of choice is praziquantel in a single oral dose (25 mg/kg). Prevention of disease is by control of fecal–oral contamination and proper hygienic measures.

UNCOMMON TAPEWORM INFECTIONS

Dipylidium caninum is an unusual cause of intestinal disease in humans. It is a common tapeworm of dogs and cats and is found worldwide. Infection usually occurs when young children ingest an intermediate host such as an infected insect or flea containing cysticercoids. Cases are usually asymptomatic but may be associated with diarrhea, abdominal discomfort, or irritability. Praziquantel is the drug of choice (5–10 mg/kg) in a single dose. Infection with *Hymenolepis diminuta*, a tapeworm of rodents, occurs when humans ingest an infected flea or cockroach. Symptoms are similar to those of *H. nana*. Identification of *H. diminuta* eggs in stool is diagnostic. Praziquantel given in a single dose (10 mg/kg) is the recommended treatment.

See Also the Following Articles

Helminth Infections • Nematodes • Parasitic Diseases, Overview • Trematodes

Further Reading

Allan, J. C., Avila, G., García-Noval, J., Flisser, A., and Craig, P. S. (1990). Immunodiagnosis of taeniosis by coproantigen detection. *Parasitology* **101**, 473–477.

Bruckner, D. A. (1999). Helminthic food-borne infections. *Clin. Lab. Med.* **19**, 639–660.

Kaminsky, R. G. (1991). Albendazole treatment in human taeniasis. *Trans. R. Soc. Trop. Med. Hygiene* **85**, 648–650.

Tsang, V. C. W., Brand, J., and Boyer, E. (1989). Enzyme-linked immunoelectrotransferency blot assay and glycoprotein antigens for diagnosing human cysticercosis. *J. Infect. Dis.* **159**, 50–59.
White, A. C., Jr. (2000). Neurocysticercosis: Updates on epidemiology, pathogenesis, diagnosis, and management. *Annu. Rev. Med.* **51**, 187–206.
Wittner, M., and Tanowitz, H. B. (1999). Cestode infections. In "Tropical Infectious Diseases" (R. L. Guerrant, D. H. Walker, and P. F. Weller, eds.). Churchill Livingstone, New York.
Wolfe, M. S. (1999). Eosinophilia in the returning traveler. *Med. Clin. North Am.* **83**, 1019–1032.

Chagas' Disease

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

megacolon Dilation of the colon.
Trypanosoma cruzi The blood-borne protozoan parasite responsible for Chagas' disease.

Intestinal pseudo-obstruction is one of the pathologic conditions that appears in Chagas' disease. Chagas' disease results from infection with the blood-borne, protozoan parasite *Trypanosoma cruzi*. *T. cruzi* multiplies intracellularly until the loaded cell bursts to release the parasites, which are distributed in the blood to invade the cells of different kinds of tissues. While free in the blood, the parasites are exposed to the immune system where stimulation of antibodies to the parasites occurs. The parasite is transmitted by large blood-sucking bugs of the family Reduviidae and affects an estimated 12–15 million people in Latin American countries, in a range that extends from Mexico to northern Argentina.

INTRODUCTION

Heart muscle and nervous tissue are the favorite targets for the blood-borne, protozoan parasite *Trypanosoma cruzi*. Nearly all fatal cases show inflammatory damage to the heart muscle that results in cardiac failure. Extensive neuronal degeneration is a secondary characteristic in chronic stages of the infection. Debilitating degeneration of neurons occurs in the sympathetic, parasympathetic, and enteric divisions of the autonomic nervous system. Neuropathy in the autonomic nervous system in the later stages of Chagas' disease is associated with megaesophagus and megacolon. The clinical picture in the large intestine is reminiscent of Hirschsprung's disease.

Megaesophagus and megacolon in Chagas' disease reflect impaired transit through an obstructed region and accumulation of the lumenal contents proximal to the obstruction. A marked reduction in the numbers of neurons in the enteric nervous system occurs in the affected regions of gut. Prolongation of gastrointestinal transit time is found also in laboratory animals after infection with *T. cruzi* and the degree of transit prolongation is proportional to the decrease in the numbers of neurons in the animal's enteric nervous system. The obstruction is classified as pseudo-obstruction because the lumen of the obstructed segment is patent (i.e., there is no mechanical obstruction). Pseudo-obstruction may occur in the esophagus or the small and large intestine. Partial surgical resection of the affected segment of intestine provides relief in humans; nevertheless, later recurrences commonly occur.

ETIOLOGY

An early assumption was that the parasite invaded cells in the walls of the viscera and destroyed the intramural neurons by the release of a toxin. Later evidence did not support direct invasion by the parasite as the underlying mechanism for the neuropathy. Current evidence

Copyright 2004, Elsevier (USA). All rights reserved.

Tsang, V. C. W., Brand, J., and Boyer, E. (1989). Enzyme-linked immunoelectrotransferency blot assay and glycoprotein antigens for diagnosing human cysticercosis. *J. Infect. Dis.* **159**, 50–59.
White, A. C., Jr. (2000). Neurocysticercosis: Updates on epidemiology, pathogenesis, diagnosis, and management. *Annu. Rev. Med.* **51**, 187–206.
Wittner, M., and Tanowitz, H. B. (1999). Cestode infections. In "Tropical Infectious Diseases" (R. L. Guerrant, D. H. Walker, and P. F. Weller, eds.). Churchill Livingstone, New York.
Wolfe, M. S. (1999). Eosinophilia in the returning traveler. *Med. Clin. North Am.* **83**, 1019–1032.

Chagas' Disease

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

megacolon Dilation of the colon.
Trypanosoma cruzi The blood-borne protozoan parasite responsible for Chagas' disease.

Intestinal pseudo-obstruction is one of the pathologic conditions that appears in Chagas' disease. Chagas' disease results from infection with the blood-borne, protozoan parasite *Trypanosoma cruzi*. *T. cruzi* multiplies intracellularly until the loaded cell bursts to release the parasites, which are distributed in the blood to invade the cells of different kinds of tissues. While free in the blood, the parasites are exposed to the immune system where stimulation of antibodies to the parasites occurs. The parasite is transmitted by large blood-sucking bugs of the family Reduviidae and affects an estimated 12–15 million people in Latin American countries, in a range that extends from Mexico to northern Argentina.

INTRODUCTION

Heart muscle and nervous tissue are the favorite targets for the blood-borne, protozoan parasite *Trypanosoma cruzi*. Nearly all fatal cases show inflammatory damage to the heart muscle that results in cardiac failure. Extensive neuronal degeneration is a secondary characteristic in chronic stages of the infection. Debilitating degeneration of neurons occurs in the sympathetic, parasympathetic, and enteric divisions of the autonomic nervous system. Neuropathy in the autonomic nervous system in the later stages of Chagas' disease is associated with megaesophagus and megacolon. The clinical picture in the large intestine is reminiscent of Hirschsprung's disease.

Megaesophagus and megacolon in Chagas' disease reflect impaired transit through an obstructed region and accumulation of the lumenal contents proximal to the obstruction. A marked reduction in the numbers of neurons in the enteric nervous system occurs in the affected regions of gut. Prolongation of gastrointestinal transit time is found also in laboratory animals after infection with *T. cruzi* and the degree of transit prolongation is proportional to the decrease in the numbers of neurons in the animal's enteric nervous system. The obstruction is classified as pseudo-obstruction because the lumen of the obstructed segment is patent (i.e., there is no mechanical obstruction). Pseudo-obstruction may occur in the esophagus or the small and large intestine. Partial surgical resection of the affected segment of intestine provides relief in humans; nevertheless, later recurrences commonly occur.

ETIOLOGY

An early assumption was that the parasite invaded cells in the walls of the viscera and destroyed the intramural neurons by the release of a toxin. Later evidence did not support direct invasion by the parasite as the underlying mechanism for the neuropathy. Current evidence

Copyright 2004, Elsevier (USA). All rights reserved.

suggests that an autoimmune mechanism is responsible for the destruction of cardiac and neuronal tissue that occurs in Chagas' disease. The explanation given for the autoreactivity is that the cells of the host and the parasite express common antigenic epitopes. As an immune attack is mounted against the parasite, cross-reactivity develops against tissue cells of the host.

Blood from patients with *T. cruzi* infection contains antibodies that recognize components of cardiac muscle and peripheral and central neurons. An antibody raised against mammalian sensory neurons (i.e., dorsal root ganglion neurons) cross-reacts with antigenic epitopes expressed by the parasite. The same antibody that reacts with the parasite also reacts with and labels neurons in the myenteric and submucosal plexuses of the enteric nervous system. This strongly suggests that the autonomic neuropathy associated with Chagas' disease results from development of antibodies against the parasite that later cross-react with neurons of the host. The autoimmune attack results in the ultimate destruction of the enteric nervous system.

PATHOPHYSIOLOGY

The degenerative inflammatory neuropathy associated with Chagas' disease in the gut leads to pseudo-obstruction. Pseudo-obstruction is a failure of propulsive motility that cannot be explained by mechanical blockage. Because propulsive motility is programmed and organized by the enteric nervous system (i.e., the brain-in-the-gut), progressive autoimmune destruction of the neurons in Chagas' disease removes the gut's brain and its control over the motor functions of the digestive tract. Contractile activity of the intestinal musculature remains in the absence of the enteric nervous system due to the myogenic nature of the smooth muscle. Nevertheless, in the absence of neural control, the contractile patterns are not coordinated and fail to achieve functional propulsion.

See Also the Following Articles

Disinhibitory Motor Disorder • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Pseudo-obstruction • Parasitic Diseases, Overview

Further Reading

Chagas, C. (1909). Über eine neue Trypanosomiasis des Menschen. *Mem. Inst. Oswaldo Cruz* **1**, 158–218.

Köberle, F., and Penha, P. (1959). Chagas' megaesophagus (quantitative studies on the intramural nervous system of the esophagus). *Z. Tropenmed. Parasit.* **10**, 291–295.

Wood, J., Hudson, L., Jessell, T., and Yamamoto, M. (1982). A monoclonal antibody defining antigenic determinants on subpopulations of mammalian neurones and *Trypanosoma cruzi* parasites. *Nature* **296**, 34–38.

Chest Pain, Non-Cardiac

GARY W. FALK
The Cleveland Clinic Foundation, Ohio

achalasia Failure of a muscle (e.g., smooth muscle sphincter) to relax.
Bernstein test A diagnostic test used to determine whether infusion of acid into the esophagus reproduces symptoms of non-cardiac chest pain.
diaphoresis Excessive perspiration.
dysphagia Difficulty in swallowing.
dyspnea Difficult or labored breathing.
edrophonium A drug that when injected stimulates contraction of smooth muscles (e.g., esophageal smooth muscle).
ergonovine A drug that constricts blood vessels and used in provocative tests for sensitivity of smaller blood vessels in the heart to cause chest pain.
gastroesophageal reflux disease A medical disorder caused by acid reflux into the esophagus from the stomach.
myocardial infarction Blockade of blood flow in one or more major vessels of the heart.
myocardial ischemia Reduced blood flow to the heart.
nociceptor Receptor on a sensory nerve that signals pain to the central nervous system.
nutcracker esophagus An esophagus that displays abnormally strong contractions during a swallow.
odynophagia Experience of pain during swallowing.
proton pump inhibitor A drug that blocks acid secretion in the stomach.

Non-cardiac chest pain may be defined as recurrent angina-like or substernal chest pain believed to be unrelated to the heart after a reasonable cardiac evaluation. This condition is also known as unexplained chest pain, as even patients with negative coronary angiograms may experience myocardial ischemia due to microvascular angina.

INTRODUCTION

Recurring substernal chest pain is a common clinical dilemma. Approximately 450,000 new cases are diagnosed annually in the United States. The prognosis for these patients is uniformly excellent with a mortality rate estimated to be less than 1% at 10 years. However, many of these patients continue to have chest pain, which they believe is cardiac in origin. This misunderstanding results in continued health care costs related to emergency room visits, office visits, hospital admissions to exclude myocardial infarction, more diagnostic tests, therapeutic trials, as well as functional impairment in terms of ability to work and quality of life. Thus, the morbidity associated with this condition is quite substantial. The estimated cost of diagnosing and treating these patients approaches two billion dollars annually.

Cardiac and esophageal symptoms may mimic each other. Depending on the criteria used, the esophagus has been implicated as a cause of chest pain in 20 to 60% of these patients.

CLINICAL PRESENTATION

The clinical history does not reliably distinguish between cardiac and non-cardiac causes of chest pain, due in part to the shared innervation of the heart and the esophagus. Both may have the same location (substernal), sensation (squeezing, burning, pressure), and radiation. Furthermore, both may be relieved by nitrates and calcium channel-blocking drugs. Chest pain related to myocardial ischemia typically is related to exertion, whereas esophageal chest pain typically occurs at rest. However, both types of pain may be induced by exercise, because exercise can induce gastroesophageal reflux. Furthermore, both types of pain may also occur at rest. Some features of pain, however, may help to distinguish esophageal from cardiac chest pain. Esophageal causes are more likely to be associated with background esophageal symptoms such as classic heartburn, regurgitation, dysphagia, or odynophagia and respond to antacids or over-the-counter histamine type 2 receptor antagonists. However, these features of esophageal pain are also found in some patients with ischemia-induced chest pain. Similarly, panic attacks, which may include symptoms such as chest pain, dyspnea, and diaphoresis, may be difficult to distinguish from angina-induced chest pain.

 Copyright 2004, Elsevier (USA). All rights reserved.

MECHANISMS OF NON-CARDIAC CHEST PAIN

The mechanisms underlying non-cardiac chest pain remain incompletely understood. The shared innervation of the esophagus and the heart has focused much of the evaluation on the esophagus as the most common cause of non-cardiac chest pain. Depending on the criteria used, the esophagus has been described as the cause of chest pain in 20 to 60% of these patients. This may be due to hypersensitivity to acid, motility abnormalities, intralumenal distension, or visceral hypersensitivity.

Gastroesophageal reflux disease (GERD) has emerged as perhaps the most common esophageal cause of non-cardiac chest pain. GERD may be present in up to 60% of patients with non-cardiac chest pain. Both acid perfusion studies and prolonged pH monitoring have confirmed that intraesophageal acid may trigger chest pain. Furthermore, anti-secretory therapy is effective in decreasing GERD-associated chest pain. However, the mechanism whereby this occurs remains obscure. It is possible that intraepithelial free nerve endings act as acid-sensitive nociceptors. Patients with gastroesophageal reflux-induced non-cardiac chest pain may also develop hypersensitivity to physiologic amounts of intraesophageal acid.

Patients with well-defined motility abnormalities such as achalasia and diffuse esophageal spasm may experience chest pain. However, these disorders are quite uncommon in patients with non-cardiac chest pain. Instead, abnormalities such as high-amplitude, long-duration contractions ("nutcracker esophagus") or nonspecific motility abnormalities are encountered more commonly in these patients. However, the causal relationship between esophageal motility disorders and chest pain is difficult to prove. Patients rarely have chest pain at the time of diagnostic manometric evaluation, even if a motility abnormality is present. Furthermore, provocation of pain with pharmacologic agents is often not accompanied by motility abnormalities. Studies with 24 h ambulatory esophageal manometry reveal that chest pain is associated with motility abnormalities in only 10 to 20% of cases. Administration of medications that decrease esophageal contraction amplitude does not decrease chest pain in patients with nutcracker esophagus. More recent studies suggest an association between nutcracker esophagus and GERD as well as a lower threshold of perception for intraesophageal balloon distension (visceral hypersensitivity), abnormal psychological profiles (depression, somatization, anxiety), and decreased esophageal wall compliance. As such, nutcracker esophagus may simply be a marker of abnormal brain–gut interactions or gastroesophageal reflux disease in patients with non-cardiac chest pain. A more recent observation is that both spontaneous chest pain and chest pain provoked by injection of edrophonium may be preceded by sustained esophageal contractions detectable only by high-frequency intralumenal ultrasonography. These changes are not accompanied by changes in intralumenal pressure.

Abnormal visceral pain perception (visceral hypersensitivity) may contribute to non-cardiac chest pain. Intraesophageal balloon distension reproduces chest pain in these patients at a lower distension volume compared to control subjects who infrequently develop pain. The etiology of visceral hypersensitivity is unknown, but this abnormality of pain perception may be due to abnormal processing of information in the central nervous system. Psychiatric abnormalities are often encountered in patients with unexplained chest pain. The most common diagnosis include depression, anxiety, somatization disorders, and panic attacks. Panic attacks are encountered in approximately one-third of chest pain patients with normal coronary arteries. Panic attacks consist of a generalized anxiety state characterized by chest pain associated with intense fear, palpitations, diaphoresis, paresthesias, dizziness, or breathlessness. Thus, alterations in sensory afferent input, central nervous system processing, and efferent muscle responses may play a role in disturbed brain–gut interactions in some patients with unexplained chest pain.

DIFFERENTIAL DIAGNOSIS

The most common causes of non-cardiac chest pain are shown in Table I. Typically, the cause of non-cardiac chest pain is gastrointestinal, musculoskeletal, or psychological in etiology. The single most important task is to exclude a cardiac source of chest pain first, because of the potential life-threatening nature of cardiac ischemia.

TABLE I Differential Diagnosis of Unexplained Chest Pain

Common causes of unexplained chest pain	Less common causes of unexplained chest pain
Coronary artery disease	Esophageal motility disorders
Gastroesophageal reflux disease	Peptic ulcer disease
Musculoskeletal disorders	Gallstones
Psychological	Mitral valve prolapse
Brain–gut dysfunction	Microvascular angina
Panic attacks	

Once cardiac causes of chest pain are excluded, attention is typically directed to the esophagus. It is now clear that the most common cause of esophageal chest pain is GERD. GERD may present with chest pain as the major and sometimes sole symptom. However, most patients will report heartburn, regurgitation, or both if closely questioned. Acid reflux may coexist and cause chest pain in patients with known coronary artery disease. Furthermore, drugs commonly used to treat coronary artery disease, such as nitrates and calcium channel antagonists, can decrease lower esophageal sphincter pressure and distal esophageal contraction amplitudes, thereby predisposing these patients to gastroesophageal reflux. Some studies report that up to two-thirds of patients with known coronary artery disease with recurrent episodes of atypical chest pain despite maximal medical and surgical therapy have chest pain related to gastroesophageal reflux events.

Chest pain is a common presenting symptom in patients with esophageal motility disorders, especially the spastic variety. However, the majority of patients with unexplained chest pain have normal esophageal motility studies and the remainder have a variety of generally nonspecific abnormalities of questionable significance. In large studies of patients with unexplained chest pain, no more than approximately one-third will have an esophageal motility disorder. Nutcracker esophagus is the most common motility disorder, followed by nonspecific disorders, diffuse esophageal spasm, hypertensive lower esophageal sphincter, and achalasia.

Chest wall muscles, ribs, and cartilage may all cause unexplained chest pain. These patients have pain and tenderness in the anterior chest wall that may radiate across the chest and increase with inspiration. Pressure applied to the anterior chest on physical examination may reproduce the pain. Radiographic studies typically demonstrate no abnormalities. Symptoms may decrease with nonsteroidal anti-inflammatory drug therapy, with steroid injections into costosternal or costoclavicular joints, or with trigger-point injections.

On occasion, patients with peptic ulcer disease will present with unexplained chest pain. Classic dyspepsia may be absent in these patients. Biliary colic is typically described as severe right upper quadrant pain occurring postprandially. Pain may radiate to the chest or back and substernal pain may sometimes be the only manifestation of biliary disease. An ultrasound of the gallbladder is the simplest way to confirm the diagnosis.

Some patients with chest pain and normal coronary arteries may have cardiac ischemia from microvascular angina. These patients have normal epicardial arteries by angiography, but an abnormal coronary vasodilator reserve in response to atrial pacing and ergonovine as well as reduced left ventricular ejection fractions during exercise. Under these conditions, coronary artery resistance increases instead of decreases, causing chest pain. The physiologic significance of microvascular angina is uncertain because similar studies have not been performed in age-matched normal individuals. Chest pain is commonly reported by patients with mitral valve prolapse. Symptoms may be exertional or nonexertional. The duration of pain is highly variable, lasting from minutes to hours. The cause of chest pain in these patients is unknown, although many different mechanisms have been postulated. Many patients with mitral valve prolapse have concurrent panic disorders, but a cause and effect relationship between the two disorders has not been proven. The diagnosis of mitral valve prolapse is made by the characteristic auscultatory findings of a midsystolic click accompanied by a late or holosystolic murmur. The diagnosis is confirmed by echocardiography. Beta-blocker therapy decreases chest pain in some of these patients.

DIAGNOSTIC APPROACH

Diagnostic testing in patients with chest pain should first focus on excluding a cardiac cause of chest pain. The intensity of the evaluation depends on factors such as patient age and risk factors for coronary artery disease. This should be done preferentially under the supervision of a cardiologist who can then choose the best diagnostic study, be it immediate catheterization, echocardiography, or nuclear perfusion imaging.

Once a cardiac cause of chest pain is excluded, the patient's evaluation should shift to possible esophageal causes of chest pain. There are a number of diagnostic studies that can evaluate the structure and function of the esophagus including barium radiography, esophagogastroduodenoscopy, esophageal manometry, 24 h pH testing, 24 h esophageal manometry, and esophageal provocative tests such as edrophonium injection and intraesophageal acid infusion (Bernstein test). However, the yield of each of these tests is low and current guidelines suggest deferring esophageal diagnostic testing until after an adequate trial of proton pump inhibitor therapy.

A trial of proton pump inhibitor therapy is both sensitive and specific for diagnosing GERD in patients with unexplained chest pain. Furthermore, it results in significant cost savings by decreasing the use of diagnostic tests in these patients. The "omeprazole test" consists of omeprazole given as 40 mg in the morning and 20 mg in the evening for 7 days. Should the patient respond, anti-secretory therapy is continued. Others

suggest a trial of any of the proton pump inhibitors given twice daily for at least 1 month prior to considering additional esophageal testing. Should the patient fail to respond, further esophageal diagnostic studies are indicated.

If a proton pump inhibitor trial is negative, 24 h pH monitoring can be considered. Data collection assesses esophageal acid exposure in the supine, upright, and combined positions, and the relationship between reflux events and chest pain can be determined. If the 24 h pH study is noncontributory, esophageal manometry can be performed to determine whether an esophageal motility disorder is present. However, most patients will have normal manometric studies.

Other gastrointestinal causes of chest pain are best evaluated by a careful history and upper endoscopy. The latter study may identify an unsuspected peptic ulcer, esophageal ulcer, or other evidence of gastroesophageal reflux disease that can be appropriately treated. A gallbladder ultrasound is the best way to evaluate the patient for gallstones. Despite this exhaustive evaluation, many patients will still not have a diagnosis. In these patients, formal psychological testing or an empiric trial of antidepressants should be considered.

The physical examination should entail an assessment of tenderness or pain in the vicinity of the sternum and xiphoid. This is recommended because costochondral causes of chest pain are easily treated and often missed.

TREATMENT

Reassuring the patient that cardiac disease is absent is the first step in managing non-cardiac chest pain. GERD is the most common esophageal cause of non-cardiac chest pain and a number of studies demonstrate excellent results in the treatment of non-cardiac chest pain with proton pump inhibitors. In patients with a good response, treatment can be titrated downward to the lowest dose necessary to control symptoms. Should the above measures fail, supportive reassurance based upon the results of additional esophageal testing that demonstrates an esophageal etiology for pain may alleviate some of the patients' symptoms, enable patients to work, and decrease the need for prescription drugs, physician and emergency room visits, and further hospitalizations. None of the smooth muscle relaxants including hydralazine, anticholinergics, nitrates, and calcium channel blockers is reliably effective in patients with non-cardiac chest pain. More aggressive therapy of esophageal motility disorders with pneumatic dilation or surgical myotomy should be avoided. Uncontrolled trials suggest that botulinum toxin injection (100 units) into either the lower esophageal sphincter or along the entire length of the esophagus may be effective in selected patients with diffuse esophageal spasm or other spastic motility disorders accompanied by intractable symptoms.

Tricyclic antidepressants and selective serotonin reuptake inhibitors may be very helpful in patients with lower pain thresholds or psychiatric diagnoses contributing to their chest pain syndrome. In controlled clinical trials, trazadone (100 to 150 mg daily) imipramine (50 mg at bedtime), and sertraline (50 to 200 mg daily) improve symptoms when compared to placebo. Other effective agents include amitryptiline, desipramine, and nortryptiline at doses of 25 to 75 mg daily. Often, these drugs improve symptoms at lower doses than required for treating psychiatric diseases, suggesting that they are working by improving pain thresholds. Patients with panic attacks will improve with tricyclic antidepressants, selective serotonin reuptake inhibitors, or benzodiazepines such as alprazolam. Controlled clinical trials suggest that cognitive behavioral therapy may also be effective in decreasing chest pain frequency and psychological morbidity and disability in these patients. Musculoskeletal causes of chest pain may be treated by injecting a local anesthetic into painful trigger points.

See Also the Following Articles

Achalasia • Barrett's Esophagus • Dysphagia • Gastroesophageal Reflux Disease (GERD) • Manometry • Proton Pump Inhibitors • Swallowing

Further Reading

Achem, S. R., Kolts, B. E., MacMath, T., Richter, J., Mohr, D., Burton, L., and Castell, D. O. (1997). Effects of omeprazole versus placebo in treatment of non-cardiac chest pain and gastroesophageal reflux. *Digest. Dis. Sci.* **42**, 2138–2145.

Balaban, D. H., Yamamoto, Y., Liu, J., Pehlivanov, N., Wisniewski, R., DeSilvey, D., and Mittal, R. K. (1999). Sustained esophageal contraction: A marker of esophageal chest pain identified by intraluminal ultrasonography. *Gastroenterology* **116**, 29–37.

Cannon, R. O., Quyyumi, A. A., Mincemoyer, R., Stine, A. M., Gracely, R. H., Smith, W. B., Geraci, M. F., Black, B. C., Uhde, T. W., Waclawiw, M. A., Maher, K., and Benjamin, S. B. (1994). Imipramine in patients with chest pain despite normal coronary angiograms. *N. Engl. J. Med.* **330**, 1411–1417.

Fang, J., and Bjorkman, D. (2001). A critical approach to non-cardiac chest pain: Pathophysiology, diagnosis, and treatment. *Am. J. Gastroenterol.* **96**, 958–968.

Fass, R., Fennerty, M. B., Ofman, J. J., Gralnek, I. M., Johnson, C., Comargo, E., and Sampliner, R. E. (1998). The clinical and economic value of a short course of omeprazole in patients with non-cardiac chest pain. *Gastroenterology* **115**, 42–49.

Frobert, O., Funch-Jensen, P., and Bagger, J. P. (1996). Diagnostic value of esophageal studies in patients with angina-like chest pain and normal coronary angiograms. *Ann. Intern. Med.* **124**, 959–969.

Katz, P. O., Dalton, C. B., Richter, J. E., Wu, W. C., and Castell, D. O. (1987). Esophageal testing of patients with non-cardiac chest pain or dysphagia: Results of three years' experience with 1161 patients. *Ann. Intern. Med.* **106**, 593–597.

Prakash, C., and Clouse, R. E. (1999). Long-term outcome from tricyclic antidepressant treatment of functional chest pain. *Digest. Dis. Sci.* **44**, 2373–2379.

Rao, S. S., Gregersen, H., Hayek, B., Summers, R. W., and Christensen, J. (1996). Unexplained chest pain: The hypersensitive, hyperreactive, and poorly compliant esophagus. *Ann. Intern. Med.* **124**, 950–958.

Richter, J. E., Barish, C. F., and Castell, D. O. (1986). Abnormal sensory perception in patients with esophageal chest pain. *Gastroenterology* **91**, 845–852.

Varia I., Logue E., O'Connor C., Newby, K., Wagner, R., Davenport, C., Rathey, K., and Krishnan, K. R. (2001). Randomized trial of sertraline in patients with unexplained chest pain of non-cardiac origin. *Am. Heart J.* **140**, 367–372.

Chief Cells

JEAN-PIERRE RAUFMAN
University of Maryland School of Medicine

pepsigogue Agent that stimulates pepsinogen secretion.
pepsinogen Protein precursor of pepsin synthesized and secreted by gastric chief cells.

Gastric chief cells, exocrine cell types localized near the base of gastric glands, synthesize and secrete pepsinogen, the precursor form of the proteolytic enzyme, pepsin.

HISTORICAL OVERVIEW

The Cambridge physiologist John Langley (1852–1925) translated and introduced the term "chief cell" from the "*Hauptzellen*" of the German physiologist Rudolf Heidenhain (1834–1897). Using amphibian peptic cells as a model, Langley elucidated the formation of pepsinogen in chief cell zymogen granules, the secretion of the proenzyme, and the proenzyme conversion to the active acid protease pepsin. Further elucidation of the cellular biology of chief cells has been obtained by using amphibian (bullfrog esophagus) and mammalian (rabbit, rat, and guinea pig) secretory models (isolated gastric glands, dispersed chief cells, and cell culture).

ANATOMY

In mammals, chief cells are located at the base of glands distributed throughout the fundus and corpus of the stomach. It is thought that chief cells derive from mucous neck cells located in the midportion of the glands. The primary function of gastric chief cells is the synthesis and release of the proenzyme pepsinogen, which subsequently, in an acid environment, is converted to the acid protease pepsin. Hence, like other enzyme-secreting cells of the gastrointestinal tract (e.g., pancreatic acinar cells), chief cells contain abundant rough endoplasmic reticulum and apical large dense zymogen-containing granules that occupy more than a third of the cell volume (Fig. 1). The nucleus is located

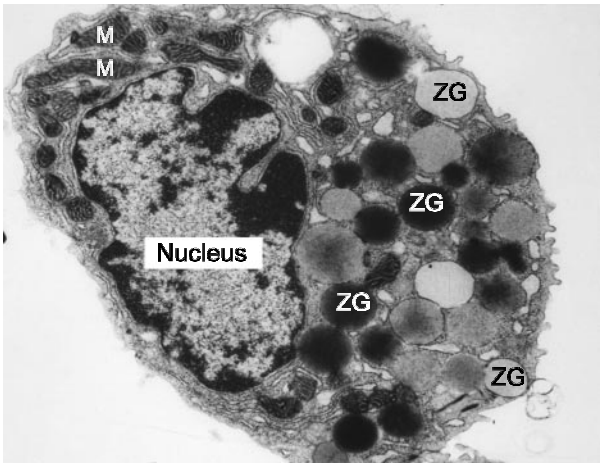


FIGURE 1 Electron micrograph of a dispersed chief cell from guinea pig stomach. Note cell polarity, with the nucleus and zymogen granules (ZG) on opposite sides of cell; M, mitochondria (original magnification ×21,000).

Copyright 2004, Elsevier (USA). All rights reserved.

Katz, P. O., Dalton, C. B., Richter, J. E., Wu, W. C., and Castell, D. O. (1987). Esophageal testing of patients with non-cardiac chest pain or dysphagia: Results of three years' experience with 1161 patients. *Ann. Intern. Med.* **106**, 593–597.

Prakash, C., and Clouse, R. E. (1999). Long-term outcome from tricyclic antidepressant treatment of functional chest pain. *Digest. Dis. Sci.* **44**, 2373–2379.

Rao, S. S., Gregersen, H., Hayek, B., Summers, R. W., and Christensen, J. (1996). Unexplained chest pain: The hypersensitive, hyperreactive, and poorly compliant esophagus. *Ann. Intern. Med.* **124**, 950–958.

Richter, J. E., Barish, C. F., and Castell, D. O. (1986). Abnormal sensory perception in patients with esophageal chest pain. *Gastroenterology* **91**, 845–852.

Varia I., Logue E., O'Connor C., Newby, K., Wagner, R., Davenport, C., Rathey, K., and Krishnan, K. R. (2001). Randomized trial of sertraline in patients with unexplained chest pain of non-cardiac origin. *Am. Heart J.* **140**, 367–372.

Chief Cells

JEAN-PIERRE RAUFMAN
University of Maryland School of Medicine

pepsigogue Agent that stimulates pepsinogen secretion.
pepsinogen Protein precursor of pepsin synthesized and secreted by gastric chief cells.

Gastric chief cells, exocrine cell types localized near the base of gastric glands, synthesize and secrete pepsinogen, the precursor form of the proteolytic enzyme, pepsin.

HISTORICAL OVERVIEW

The Cambridge physiologist John Langley (1852–1925) translated and introduced the term "chief cell" from the "*Hauptzellen*" of the German physiologist Rudolf Heidenhain (1834–1897). Using amphibian peptic cells as a model, Langley elucidated the formation of pepsinogen in chief cell zymogen granules, the secretion of the proenzyme, and the proenzyme conversion to the active acid protease pepsin. Further elucidation of the cellular biology of chief cells has been obtained by using amphibian (bullfrog esophagus) and mammalian (rabbit, rat, and guinea pig) secretory models (isolated gastric glands, dispersed chief cells, and cell culture).

ANATOMY

In mammals, chief cells are located at the base of glands distributed throughout the fundus and corpus of the stomach. It is thought that chief cells derive from mucous neck cells located in the midportion of the glands. The primary function of gastric chief cells is the synthesis and release of the proenzyme pepsinogen, which subsequently, in an acid environment, is converted to the acid protease pepsin. Hence, like other enzyme-secreting cells of the gastrointestinal tract (e.g., pancreatic acinar cells), chief cells contain abundant rough endoplasmic reticulum and apical large dense zymogen-containing granules that occupy more than a third of the cell volume (Fig. 1). The nucleus is located

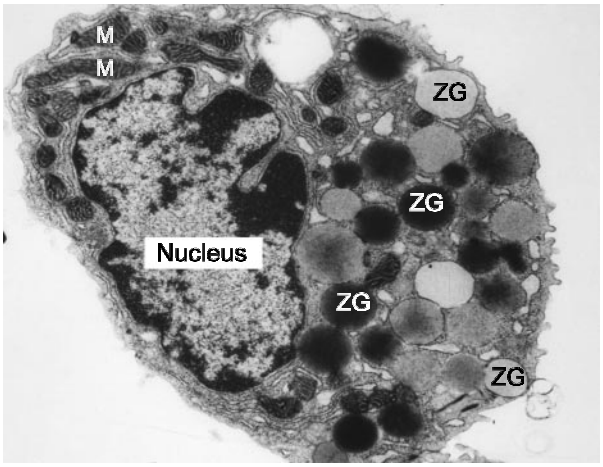


FIGURE 1 Electron micrograph of a dispersed chief cell from guinea pig stomach. Note cell polarity, with the nucleus and zymogen granules (ZG) on opposite sides of cell; M, mitochondria (original magnification ×21,000).

Copyright 2004, Elsevier (USA). All rights reserved.

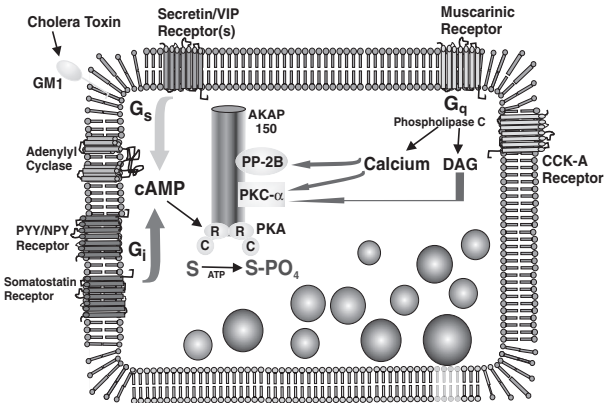


FIGURE 2 The major signal transduction pathways in gastric chief cells and their regulation and interaction. See text for discussion.

at the base of the cell, surrounded by mitochondria (Fig. 1). On stimulation, zymogen granules move toward the apical surface, where they fuse with the plasma membrane and release pepsinogen into the lumen of the gastric gland. The major physiologic control of pepsinogen secretion is by the vagal nerve and intrinsic neural reflexes.

CELL BIOLOGY

The orderly and timely progression of zymogen granules toward the apical membrane of chief cells, so that the active acid protease pepsin can be delivered to the gastric lumen to assist in the digestion of ingested protein, requires an efficient and tightly regulated process. Figure 2 represents a simplified scheme illustrating various steps in this process, from ligand binding to basolateral membrane receptors, mediated by activation of guanosine nucleotide binding proteins (G_q, G_s, and G_i), to phosphorylation of currently unidentified protein substrates. Although the identity of these protein substrates remains obscure, their phosphorylation and dephosphorylation are crucial to exocytosis. These proteins may be related to synaptosomal-associated protein receptors (SNAREs) or may be members of the Rab family of guanosine nucleotide binding proteins that are part of the secretory machinery in other cell types. Rab proteins have been identified in chief cells.

Basolateral membrane receptors for acetylcholine (muscarinic receptors) and for cholecystokinin and gastrin (CCK-A receptor) activate a phospholipase C, which results in an increase in cytosolic calcium concentration and the release of diacylglycerol (DAG). DAG activates the serine–threonine kinase protein kinase C (PKC; the α and ζ isoforms have been identified in guinea pig chief cells). Calcium activates a number of kinases and phosphatases, including PKC, calcium/calmodulin kinase II, and protein phosphatase-2B (PP-2B, also referred to as calcineurin).

Membrane receptors for secretin and vasoactive intestinal peptide (VIP) and for cholera toxin (ganglioside GM1 receptor) are linked to activation of adenylyl cyclase, to increases in cellular cyclic adenosine monophosphate (cAMP), and to activation of protein kinase A (PKA). Activation of receptors for peptide YY and neuropeptide Y (PYY and NPY) and for somatostatin reduces the activity of adenylyl cyclase. Recent studies focusing on the integration of these various signals have elucidated the presence of a 150-kDa A-kinase anchoring protein (AKAP150) that serves as the nidus of a complex of signaling molecules, including PKA, PKC-α, and PP-2B, thereby facilitating concurrent or sequential phosphorylation and dephosphorylation of substrates. Attachment of AKAP150 to the actin cytoskeleton also serves an important role in compartmentalization of these signaling cascades within the cell.

Besides the pepsigogues illustrated in Fig. 2, histamine and cytokines (epidermal growth factor and interleukin-1β) have been reported to stimulate pepsinogen secretion. With the exception of acetylcholine, the importance of these *in vitro* pepsigogues to physiological pepsinogen secretion in response to an ingested meal remains to be determined.

See Also the Following Articles

Exocytosis • Pancreatic Digestive Enzymes • Pancreatic Enzyme Secretion (Physiology) • Pepsin

Further Reading

Raffaniello, R. D., Lin, J., Wang, F., and Raufman, J.-P. (1996). Expression of Rab3D in dispersed chief cells from guinea pig stomach. *Biochim. Biophys. Acta* **1311**, 111–116.

Raufman, J.-P. (1992). Gastric chief cells: Receptors and signal-transduction mechanisms. *Gastroenterology* **102**, 699–710.

Singh, G., Raffaniello, R. D., Eng, J., and Raufman, J.-P. (1995). Actions of rab3-related peptides in dispersed chief cells from guinea pig stomach. *Am. J. Physiol.* **269**, G400–G407.

Xie, G., and Raufman, J.-P. (2001). Association of protein kinase A with AKAP150 facilitates pepsinogen secretion from gastric chief cells. *Am. J. Physiol.* **281**, G1051–G1058.

Cholangiocarcinoma

MARY LEE KRINSKY
University of California, San Diego

bilirubin A bile pigment formed during the catabolism of heme-containing compounds, primarily hemoglobin.
cholecystectomy Surgical removal of the gallbladder.
choledochocyst A congenital segmental cystic dilation of the biliary system located in any portion of the extrahepatic biliary tract.
hepatolithiasis (Oriental cholangiohepatitis) A chronic disease characterized by the formation of primary intrahepatic pigmented stones and sequelae that include recurrent cholangitis, strictures of the intrahepatic bile ducts, hepatic abscesses, portal vein thrombosis, cholangicarcinoma, and sometimes secondary biliary cirrhosis with hepatic failure.
jaundice Excessive accumulation of bilirubin, as a result of enhanced production or impaired elimination, resulting in yellow discoloration of the skin, sclera, and mucous membranes.
sclerosing cholangitis A condition characterized by fibrosis of the biliary tree.
ulcerative colitis Chronic inflammatory condition involving the colon.

Cholangiocarcinoma (CCA) is a rare malignancy that accounts for less than 1% of all cancers. The nomenclature of tumors of the biliary tree has evolved over time. Initially, "cholangiocarcinoma" was intended to define tumors of the intrahepatic biliary tree. In 1965, Gerald Klatskin described tumors of the bifurcation of the extrahepatic ducts, known as "Klatskin" tumors. Currently, CCA includes neoplasms occurring anywhere along the biliary tree and is divided into three groups: (1) those located in the intrahepatic ducts, (2) those located in the perihilar ducts, and (3) those located in the distal extrahepatic ducts. This classification correlates best with the treatment and prognosis. In addition, perihilar tumors have been subclassified further by the Bismuth classification, which identifies the precise location of the tumor.

EPIDEMIOLOGY AND ETIOLOGY

Biliary tract cancers are rare; in the United States, 6800 new cases are estimated to be diagnosed in 2003. Gallbladder cancer accounts for approximately two-thirds of biliary tract cancers and 2000–3000 of these are attributable to cholangiocarcinoma (CCA). The majority of cases occur in patients between the ages of 50 and 70, with a slight male predominance. However, patients with predisposing diseases may present up to two decades earlier. The perihilar and distal extrahepatic bile ducts are the most common sites of involvement. Perihilar tumors account for approximately 70% and distal extrahepatic bile duct tumors for 27% of cases.

The etiology of cholangiocarcinoma in the United States is usually unknown. Genetic predisposition may play a role. It is believed that chronic bile duct inflammation of any etiology may lead to neoplasia of the biliary tree. However, a number of high-risk conditions exist. In the United States, these include patients with chronic inflammation associated with ulcerative colitis with or without primary sclerosing cholangitis (PSC) and chronic intraductal gallstones (hepatolithiasis). Approximately 10–30% of cholangiocarcinomas exist in patients with PSC. Furthermore, congenital and bile duct abnormalities of the biliary tree including choledochocysts, Caroli's disease, bile-duct adenoma, and multiple biliary papillamatosis are associated with cholangiocarcinoma. Caroli's disease consists of numerous intrahepatic cystic dilations of the bile duct with or without stone formation. In Southeast Asia, chronic infections with liver parasites including *Clonorchis sinensis* (in Japan, Korea, and Vietnam) and *Opisthorchis viverrini* (in Laos, Malaysia, and Thailand) are associated with a 25- to 50-fold increased risk of cholangiocarcinoma. Hepatolithiasis (Oriental cholangiohepatitis) is endemic in Southeast Asia and may be associated with congenital ductal abnormalities, chronic infections, and diet. The risk of cholangiocarcinoma in hepatolithiasis is between 1 and 13%. These liver flukes cause chronic inflammation of the intrahepatic bile ducts that leads to stricture and stone formation. In addition, Thorotrast (thorium dioxide) exposure, a radiologic contrast agent used in the 1930s and 1940s, has been associated with the development of cholangiocarcinoma. Environmental factors including multiple chemicals and radionuclides have been recognized as being causal in the development of CCA. Finally, other entertained risk factors may

 Copyright 2004, Elsevier (USA). All rights reserved.

include tobacco use, asbestos exposure, and prior cholecystectomy and biliary sphincterotomy.

PRESENTATION AND DIAGNOSIS

The clinical presentation of a patient with cholangiocarcinoma is dependent on the tumor's location, its invasiveness, and the degree of biliary obstruction. The patient will become symptomatic when the biliary tree is obstructed. Obstructive symptoms are more prevalent with extrahepatic and hilar disease. These include jaundice, pruritus, abdominal pain, nausea, vomiting, anorexia, and weight loss. The majority of patients will present with painless jaundice. This must be differentiated from other common causes, such as pancreatic and ampullary cancer. The physical examination may reveal only jaundice. Less common findings include hepatomegaly, right upper quadrant pain, and a palpable mass (distended gallbladder or the tumor). Biochemical studies will often show elevations in the bilirubin, transaminases, γ-glutamyl transferase (GGT), alkaline phosphatase (AP), 5′-nucleotidase (5′NT), and prothrombin time.

In contrast, patients with intrahepatic involvement may be asymptomatic early on or have abdominal pain. Biochemical testing will reflect elevations of the biliary ductal epithelial enzymes: GGT, AP, and 5′NT. The bilirubin and other parenchymal enzymes (SGOT, SGPT) will usually be normal until the disease is more invasive, involving the majority of the hepatic parenchyma. As the disease progresses, the prothrombin time will be prolonged. The diagnosis of a pancreatic or hepatico-biliary malignancy is often made from the clinical history and biochemical tests. Further imaging, tumor markers, and tissue acquisition will assist in the final diagnosis, tumor staging, and management decisions. No specific tumor marker exists for cholangiocarcinoma. Serum carcinoembryonic antigen (CEA) and cancer antigen 19−9 may be utilized. However, their sensitivity and specificity have been poor and their levels may normalize once the obstruction is relieved. The α-fetoprotein is usually normal.

Diagnostic imaging modalities include transabdominal ultrasound, computed tomography (CT), cholangiography [via endoscopic retrograde cholangiopancreatography (ERCP) or the percutaneous route], magnetic resonance imaging (MRI), positron emission tomography (PET), endoscopic ultrasound (EUS), and intraductal ultrasound. The primary goal of these tools is to define the local extent of the tumor and the presence of distant metastasis. This information will assist in tumor staging and determining prognosis and management strategies.

The initial evaluation of a patient with the above-listed laboratory test abnormalities will commonly include transabdominal ultrasound (US) and/or CT. US is primarily utilized to identify bile duct dilation. Perihilar CCA may appear as intrahepatic dilation. Intrahepatic tumors may appear as segmental intrahepatic dilation and distal extrahepatic tumors may be represented by both intra- and extrahepatic bile duct dilation. CT and MRI commonly identify the location of the obstructing lesion, evidence of metastasis, and bile duct dilation. Intrahepatic CCA may appear as a hepatic mass with intrahepatic segmental bile duct dilation. Extrahepatic tumors, including Klatskin tumors, may appear as a hilar mass with associated biliary dilation. In some cases, the tumor may not be identified and the site of the lesion may be elucidated by locating where the bile duct dilation begins. In one report of 29 patients with histologically proven hilar CCA, 100% of the tumors were detected using intravenous bolus-enhanced (multiphasic) helical CT. However, resectability was accurately determined in only 60% of the patients. Regional lymph nodes may be detected and assist with local regional staging. However, patients with chronic inflammatory conditions of this anatomic region may be associated with benign lymph nodes.

Cholangiography in the past has been a primary tool used to diagnose CCA. The opacification of the biliary tree allows for localization of the obstructing lesion, tissue acquisition, and stenting for biliary drainage. Cholangiography can be performed by ERCP or by percutaneous transhepatic cholangiography (PTC). ERCP can be difficult in proximal and perihilar tumors that completely obstruct the bile duct. The inability to traverse the stricture after contrast injection may lead to an obstructed contaminated duct and the risk of cholangitis. PTC may be limited in patients with PSC due to the marked stricturing of the intrahepatic biliary tree. Other diseases that may be confused with a CCA-associated stricture are gallbladder cancer, Mirizzi's syndrome, or a benign inflammatory stricture caused by prior instrumentation or stone disease. Tissue acquisition is most easily accomplished via brush cytology, which makes the diagnosis in 20−80% of the patients. Other approaches include forceps biopsy, bile aspirate, and fine-needle aspiration (FNA). A cytologic or histologic diagnosis is not essential if the patient is deemed resectable.

Magnetic resonance cholangiopancreatography (MRCP) is a promising noninvasive modality that displays a three-dimensional image of the biliary tree, hepatic parenchyma, and surrounding vessels. This technique may become the primary imaging tool for local regional preoperative staging.

Additional relatively new modalities are being developed to detect, diagnose, and stage CCA. EUS is a relatively noninvasive endoscopic technique that may be utilized to image the distal extrahepatic bile ducts, perihilar tumors, and surrounding structures. Portal vein invasion has been shown to more accurate with EUS than CT, angiography, and US. EUS-guided FNA may be used to obtain a cytologic diagnosis and is associated with much less risk than ERCP. In one study of 10 patients with suspected hilar CCA, EUS-guided FNA diagnosed 7 patients with CCA and 1 with hepatocellular carcinoma. Intraductal ultrasound has been used to image bile duct strictures and to evaluate for portal vein involvement. However, they have a relatively shallow depth of penetration, limiting visualization of large tumors and surrounding lymph nodes and vessels. PET scanning reflects the high glucose metabolism of neoplasia. One study was able to show high uptake in CCA patients versus those who had PSC. Furthermore, one small study suggests a role for improved preoperative staging by identifying distant metastasis. Other modalities include choledochoscopy, radiolabeled antibody tracing, or ligand imaging.

PATHOLOGY

Cholangiocarcinoma arises from the epithelial cells of the bile duct (cholangiocytes). The most common tumor histologic type is adenocarcinoma (>95%) (Table I).

Papillary carcinoma may have the best prognosis because it typically presents early. These tumors have a strong desmoplastic reaction that leads to difficulty in obtaining sufficient cells for a diagnosis. This likely explains the poor yield from endoscopically obtained tissue. Immunohistochemical stains for CEA, mucin, and cytokeratins may be helpful in making the diagnosis.

CCA usually spreads locally to the adjacent structures, abdominal cavity, and lymph nodes. Regional lymph node involvement occurs in more than 75% of patients. Metastases occur late in the disease, most commonly to the liver, peritoneum, lungs, and bones.

TABLE I Histologic Type

Adenocarcinoma[a]
Clear cell adenocarcinoma
Mucinous carcinoma[a]
Papillary carcinoma (invasive, noninvasive)[a]
Cystadenocarcinoma
Signet ring cell carcinoma
Squamous cell carcinoma
Adenosquamous cell carcinoma
Small cell carcinoma
Undifferentiated carcinoma
Leiomyosarcoma
Carcinoid
Granular cell tumors
Rhabdomyosarcoma
Papillomatosis
Kaposi's sarcoma[b]
Lymphoma[b]

[a] Most common histologic types.
[b] AIDS patients.

STAGING AND PROGNOSIS

Extrahepatic biliary tract tumors commonly present early. Early small tumors will partially or completely occlude the small-diameter bile duct and lead to obstructive signs. However, tumors of the hilum have a poor prognosis. They are difficult to resect due to their vascular and hepatic parenchymal invasion. CCA is staged according to the tumor−node−metastasis (TNM) system of classification. Extrahepatic tumors confined to the bile duct or beyond are staged as IA and IB. Those tumors invading the liver, gallbladder, pancreas, or unilateral branches of the portal vein or hepatic artery are Stage IIA. Stage IIB is characterized by any of the above with regional lymph node involvement. Stage III includes tumors that invade the other adjacent organs as well as the main portal vein and common hepatic artery. Stage IV is distant metastasis, regardless of the T and N stages. Alternatively, intrahepatic tumors are currently staged the same as hepatocellular carcinomas because of the limited prognostic information.

TREATMENT STRATEGIES

Surgical resection is the only curative treatment modality for cholangiocarcinoma and is the only modality associated with an improved 5-year survival. Overall, the prognosis is poor, with a 5−10% 5-year survival and a 20-month median survival. However, patients with perihilar disease have a median survival of 12−24 months. Patients with distal extrahepatic CCA have 5-year survival rates of 15−25%. Multiple factors determine a patient's candidacy for resection. Anatomical criteria include the absence of distant metastasis, vascular invasion, lymph node involvement, or extensive disease beyond a feasible surgical margin. Common clinical criteria that prohibit surgical resection include poor performance status, significant cardiopulmonary disease, and cirrhosis.

Patients who undergo surgery and are found to have a tumor approaching or involving the resection margins should be considered for adjuvant therapy. Patients that

are unresectable should be offered palliative biliary drainage and adjuvant therapy. This has been shown to improve quality of life and prolong survival.

The surgical approach depends on the location of the tumor. Intrahepatic cholangiocarcinoma is usually treated by hepatic resection. Perihilar tumors are difficult to resect and less than 40% are resectable. The surgical technique depends on the Bismuth classification (location of the tumor). Distal extrahepatic CCA has the highest cure rate and the most common surgery is pancreaticoduodenectomy (Whipple procedure). Liver transplantation has been disappointing and is discouraged given the poor success rates and shortage of livers.

Adjuvant radiation and chemotherapy have not been well studied. Radiation in patients with involved margins may improve survival and quality of life. However, it may be associated with complications of hepatitis and gastroduodenal obstruction. Chemotherapy preoperatively or postoperatively has not been shown to improve survival or quality of life. Chemoradiotherapy may have a role; however, this remains unclear.

Palliative therapy is pursued in the majority of patients with cholangiocarcinoma. Jaundice is the most common ailment palliated and this can be approached surgically via biliary bypass or with endoscopic/percutaneous biliary stenting. The latter technique is less invasive. Metal stents are preferred over plastic because of their longer patency rates. Additional modalities for biliary palliation include photodynamic therapy, Nd-YAG laser, and microwave tissue coagulation.

SUMMARY

Cholangiocarcinoma is a relatively rare malignancy of the biliary tree. The tumor may be located anywhere along the intrahepatic or extrahepatic ducts. Patients generally present with jaundice, pruritus, weight loss, and pain. Associated conditions include liver fluke infestation in Third World countries, congenital biliary cysts, and primary sclerosing cholangitis. A bilirubin level >12 mg/dl in the setting of obstruction indicates malignancy. The cancers arise from the biliary epithelium and are most commonly adenocarcinomas. A tissue diagnosis may be difficult to make because of a marked desmoplastic reaction. Therefore, vigilance must be advocated in seemingly benign inflammatory strictures. The diagnosis is made from the clinical history, laboratory, and imaging studies. Cholangiography, MRCP, helical CT, EUS, intraductal US, and PET scanning can determine the extent of disease and local regional staging. Tissue acquisition is not necessary preoperatively; however, it can be obtained by EUS-guided FNA, ERCP-directed brushings, or biopsies. Treatment entails either surgical resection for attempted cure or the use of endoscopically or percutaneous placed stents for palliative biliary drainage. The role of adjuvant or neoadjuvant therapy is unclear. The prognosis is poor with a 5–10% 5-year survival and a 20-month median survival.

See Also the Following Articles

Cancer, Overview • Cholangiohepatitis, Oriental • Cholangitis, Schlerosing • Colitis, Ulcerative • Computed Tomography (CT) • Endoscopic Ultrasonography • Gallbladder Cancer • Magnetic Resonance Imaging (MRI) • Percutaneous Transhepatic Choliangiography (PTC)

Further Reading

de Groen, P., Gores, G., LaRusso, N., Gunderson, L., and Naggorney, D. (1999). Biliary tract tumors. *N. Engl. J. Med.* **341**, 1368–1378.

Farrell, J., and Chung, R. (2003). Carcinoma of the gallbladder and extrahepatic biliary tree. *In* "Gastrointestinal Cancers" (J. Crawford and A. Rustgi, eds.), pp. 623–653. Elsevier Science, Philadelphia, PA.

Fritscher-Ravens, A., Bohuslavizki, K., Broering, D., Jenicke, L., and Schafer, H. (2001). FDG PET in the diagnosis of hilar cholangiocarcinoma. *Nuclear Med. Commun.* **22**, 1277–1285.

Fritscher-Ravens, A., Broering, D., Sriram, P., Topalidis, T., *et al.* (2000). EUS-guided fine-needle aspiration cytodiagnosis of hilar cholangiocarcinoma: A case series. *Gastrointest. Endosc.* **52**, 534–540.

Green, F., Page, D., Fleming, I., Fritz, A., *et al.* (2002). Liver (including intrahepatic bile ducts). *In* "AJCC Cancer Staging Handbook," 6th Ed.

Jemel, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E., and Thun, M. (2003). Cancer statistics, 2003. *CA Cancer J. Clin.* **53**, 5–26.

Keiding, S., Hansen, S., Rasmussen, J., Gee, A., Kruse, A., *et al.* (1998). Detection of cholangiocarcinoma is primary sclerosing cholangitis by positron emission tomography. *Hepatology* **28**, 700–706.

Mortele, K., Ji, H., and Ros, P. (2002). CT and magnetic resonance imaging in pancreatic and biliary tract malignancies. *Gastrointest. Endosc.* **56**, S206–S212.

Nakeeb, A., Pitt, H., Sohn, T., Coeman, J., Abrams, R., Piantadosi, S., Hruban, R., *et al.* (1996). Cholangiocarcinoma: A spectrum of intrahepatic, perihilar, and distal tumors. *Ann. Surg.* **224**, 463–475.

Pitt, H., Grochow, L., and Abrams, R. (1997). Cancers of the biliary tree. *In* "Cancer: Principles and Practice of Oncology," 5th Ed., pp. 1114–1127. Lippincott, Philadelphia, PA.

Tilich, M., Mischinger, H., Preisegger, K., Rabl, H., and Szolar, D. (1998). Multiphasic helical CT in diagnosis and staging of hilar cholangiocarcinoma. *Am. J. Roentegenol.* **171**, 651–658.

Cholangiohepatitis, Oriental

SHIH-TING PATRICIA TSAI
Keck School of Medicine, University of Southern California

anaerobe A microorganism that can live and grow in the absence of oxygen.
biliary stent A catheter placed within the bile duct to provide patency of the duct.
cholangicarcinoma An adenocarcinoma, primarily in intrahepatic bile ducts, composed of duct lined by cuboidal or columnar cells.
cholangiography Radiographic exam of the bile duct with contrast medium.
cholangitis Inflammation and infection of the bile duct and biliary system.
choledochoenterostomy Surgical establishment of a communication between the common bile duct and any part of the intestine.
choledochoscopy Endoscopic exam of the common bile duct.
deconjugation The dissociation of the chemical bond between two chemical compounds.
electrohydraulic lithotripsy Electrohydraulic shock wave lithotripsy is a destruction of calculi (stone) by fragmentation using shock wave sent transcutaneously via ultrasound transducers.
endoscopic retrograde cholangiopancreatography An endoscopic procedure that evaluates the biliary system via the sphincter of Oddi.
endoscopy Examination of the interior of intestine by an endoscope.
hemobilia Bleeding into the biliary system as a result of trauma, tumor, or infection.
hepaticoenterostomy Establishment of a communication between the hepatic ducts and the duodenum.
jaundice A yellowish staining of the integument, sclera, deeper tissues, and excretions with bile pigments, resulting from increased levels in the plasma.
lobectomy Excision of a lobe of an organ, e.g., liver.
pancreatitis Inflammation of the pancreas.
portal vein thrombosis Thrombus clot within the portal vein.
Roux-en-Y choledochojejunostomy A surgical anastomosis between the common bile duct and the jejunum.
secondary biliary cirrhosis End-stage liver disease due to obstruction of extrahepatic bile ducts, which leads to cholestasis and proliferation of small bile ducts and fibrosis.
sphincter of Oddi Hepatopancreatic ampullae that open to the common bile duct and pancreatic duct.
sphincterotomy An incision or division of the sphincter of Oddi.
stricture A narrowing or stenosis of a hollow structure, usually consisting of cicatricial contracture or deposition of abnormal tissue.

Oriental cholangiohepatitis, more commonly known as recurrent pyogenic cholangitis, is a chronic disease characterized by the formation of primary intrahepatic pigmented stones and sequelae that include recurrent cholangitis, strictures of the intrahepatic bile ducts, hepatic abscesses, portal vein thrombosis, cholangicarcinoma, and sometimes secondary biliary cirrhosis with hepatic failure. Patients often suffer from recurrent attacks of abdominal pain, fever, jaundice due to intrahepatic ductal stone, and strictures. Two theories are proposed as the initiator of the disease process: parasitic infestation with organisms such as *Clonorchi sinensis* and *Ascaris lumbricoides* and malnutrition theory, which suggests nutritional deficiency of an inhibitor(s) of bacterial bilirubin-deconjugating enzyme, resulting in formation of bile duct stones.

INTRODUCTION

Oriental cholangiohepatitis is a syndrome characterized by recurrent attacks of suppurative cholangitis in which the bile ducts are dilated and contain multiple pigment stones. It was first described among the Chinese in Hong Kong in 1930. This condition is also known as Oriental cholangitis, Oriental biliary obstruction syndrome, intrahepatic pigment stone disease, hepatolithiasis, and primary cholangitis and currently is more commonly known as recurrent pyogenic cholangitis (RPC). As these names suggest, the syndrome is endemic to Southeast Asia. It is also encountered mostly in Asian immigrants in the United States and Canada. Sporadic cases have also been reported among natives of Europe, South Africa, India, and South America.

PATHOLOGY

The bile ducts within the liver (intrahepatic) and outside the liver (extrahepatic) are dilated and contain many bile pigment stones. The common bile duct and left hepatic duct are commonly affected. They are surrounded by extensive fibrosis and inflammation. The

 Copyright 2004, Elsevier (USA). All rights reserved.

papilla of the sphincter of Oddi may be hypertrophied and fibrosed from the repeated passage of stones. As many as 35% of the patients may have biliary strictures. Hilar strictures involving the left hepatic duct are found most frequently, followed by strictures of the common hepatic duct. The stones are composed of bile pigment with color ranging from orange to black and size varies up to 3 cm. However, stones are absent in 20–25% of patients. This may be due to early disease or complete passage of stones.

ETIOLOGY AND PATHOGENESIS

As the name implies, patients often have recurrent pyogenic infection of the biliary system. The common organisms in descending order are *Escherichia coli, Klebsiella, Pseudomonas, Proteus* species, and rarely anaerobes. These enteric organisms are thought to be transported to the biliary system by transient portal bacteremia. Some underlying abnormality, such as ductal epithelial disruption or stasis, promotes seeding of the bacteria and causes infection and secondary pigment stone formation. Stone formation then produces more obstruction and infection and hence the recurrent episode of cholangitis. Two theories are proposed to explain the initiation of the pathogenesis: the parasite theory and the malnutrition theory. The parasite theory proposes that parasitic infestation of the biliary tree results in the epithelial damage. The parasitic debris and ova may act as a nidus. Parasites implicated are liver flukes, such as *Clonorchis sinensis, Opisthorchis viverrini*, and *Opisthorchis felineus*. Infestation with these liver flukes is common in Southeast Asia and up to 45% of patients have evidence of past infestation. *Ascaris lumbricoides* are also associated with RPC found outside of Asia. However, evidence against this theory are the facts that helminthes or ova are recovered from the stools in only 5 to 25% of patients and that RPC is on the decline in some regions that are endemic for *Clonorchis*, such as Japan.

Malnutrition theory suggests that recurrent bacterial gastroenteritis common in lower socioeconomic areas results in portal bacteremia and subsequent seeding of the bacteria in the biliary system. The indigent population often consumes a low-protein diet that may result in deficiency in their bile of glucaro-1:4-lactone, which is an inhibitor of the enzyme β-glucuronidase. Bacterial β-glucuronidase is particularly active at the bile pH and can deconjugate secreted bilirubin, resulting in formation of insoluble calcium bilirubinate and brown pigment stones.

CLINICAL PRESENTATION

RPC is usually seen in patients who are younger than patients with choledocholithiasis in the West. A majority of the patients present in the third to fifth decade. RPC affects men and women equally. Patients most commonly present with Charcot's triad: fever, right upper quadrant pain, and jaundice. The typical clinical course is that of recurrent attacks. Approximately 70–85% of patients have a history of frequent attacks, typically once or twice a year.

COMPLICATIONS

Patients may present with complications such as rupture of the bile duct, causing peritonitis or formation of fistula to nearby organs. Portal vein thrombosis and hemobilia may occur. Systemic sepsis may manifest as brain or lung abscess. Acute pancreatitis occurs in up to 10% of patients. There is an association with an increased risk of cholangiocarcinoma.

DIAGNOSIS

The diagnosis is suggested in patients with the typical history and demographic background. Radiologic studies with ultrasound and computer tomography (CT) are important to make the diagnosis. Ultrasound shows ductal dilation and ductal stone in up to 90% of patients. CT is the best noninvasive diagnostic modality. CT often reveals a characteristic pattern of dilated central intrahepatic ducts with acute tapering of the peripheral ducts. Stones, liver abscesses, and lobar atrophy are best seen on CT scan. Once the patient is stabilized, endoscopic retrograde cholangiopancreatography (ERCP) can be used to demonstrate definitive imaging of the biliary system. The dilation of the intra- and extrahepatic ducts may produce a classic "arrow sign" pattern. However, if the duct was obstructed, cholangiography may be nondiagnostic. ERCP can also assist in sphincterotomy and in stone extraction using balloon and basket and can facilitate bile duct drainage.

MANAGEMENT

The goal of treatment is to treat the acute infection and prevent complications and recurrence. Most patients respond to conservative treatment with broad-spectrum antibiotics. However, up to 20% of patients may require emergent cure surgery. These patients often fail antibiotics with persistent fever, pain, jaundice, or signs

of peritonitis. Emergent surgery often consists of common bile duct exploration, T-tube placement, and cholecystectomy if the gallbladder is also obstructed. ERCP with biliary stent or percutaneous drainage of the intrahepatic duct is also performed in selected patients.

Prevention of future attacks is achieved by removing as many stones and debris as possible and dilating or resecting the stricture. Surgery may be required in patients with gallbladder stones, intrahepatic ductal stones, and disease predominantly localized to the left hepatic ductal system. Surgical techniques include left lobectomy, removal of intrahepatic stone by choledochoscopy and electrohydraulic lithotripsy, and drainage procedures such as choledochoenterostomy or hepaticoenterostomy. In patients with extensive intrahepatic ductal stones requiring frequent endoscopic or radiographic intervention, a Roux-en-Y choledochojejunostomy with the proximal Roux loop attached to the abdominal wall or externalized as a stoma is an increasingly popular approach. Surgical management has a good outcome, with up to 80% of patients remaining symptom-free during a mean follow-up of 7 years. Endoscopic management in one large study reported that 95% of patients were symptom-free during a follow-up period of 2–3 years.

See Also the Following Articles

Cholangiocarcinoma • Cholangitis, Sclerosing • Cholelithiasis, Complications of • Computed Tomography (CT) • Gallstones, Pathophysiology of

Further Reading

Chang, T. (1983). Intrahepatic stones: The Taiwan experience. *Am. J. Surg.* **146**, 241–244.

Fan, S. (1993). Hepatic resection for hepatolithiasis. *Arch. Surg.* **128**, 1070–1074.

Fan, S. (1993). Appraisal of hepaticocutaneous jejunostomy in the management of hepatolithiasis. *Am. J. Surg.* **165**, 332–335.

Horton, J. (1993). Recurrent pyogenic cholangitis. *In* "Gastrointestinal and Liver Disease" (M. Sleisenger and B. Scharschmidt, eds.), 6th Ed., Vol. 1, pp. 1018–1025.

Hutson, D. (1984). Balloon dilatation of biliary strictures through a choledochojejunocutaneous fistula. *Ann. Surg.* **199**, 637–647.

Lim, J. (1990). Oriental cholangiohepatitis: Sonographic findings in 48 cases. *Am. J. Roentgenol.* **155**, 511–514.

Nakayama, F. (1984). Hepatolithiasis: Present status. *World J. Surg.* **8**, 9.

Ong, G. (1962). A study of recurrent pyogenic cholangitis. *Arch. Surg.* **84**, 63–89.

Pitt, H. (1994). Intrahepatic stones: The transhepatic team approach. *Ann. Surg.* **219**, 527–537.

Sauerbuch, T. (1992). Non-surgical management of bile duct stones refractory to routine endoscopic measures. *Clin. Gastroenterol.* **6**, 799–817.

Sperling, R. (1997). Recurrent pyogenic cholangitis in Asian immigrants to the United States: Natural history and role of therapeutic ERCP. *Digest. Dis. Sci.* **42**, 865–871.

Stain, S. (1995). Surgical treatment of recurrent pyogenic cholangitis. *Arch. Surg.* **130**, 527–533.

Cholangitis, Sclerosing

KHALDOUN A. DEBIAN
University of Southern California

cholangiocarcinoma A malignant transformation of cholangiocytes that carries a poor prognosis.
cholestasis A pathologic process that prohibits the clearance of bile salts from the hepatocyte to the lumen of the gastrointestinal tract.
fibrosis End result of an inflammatory response of an organ or tissue to an insult, resulting in scar formation.
hepatic cirrhosis End-stage condition of the liver recognized histologically by extensive fibrosis and clinically by hepatic cellular dysfunction and portal hypertension.
ischemia Loss of blood supply to the cell, resulting in hypoxia and subsequent cellular injury.
sclerosing cholangitis A condition characterized by fibrosis of the biliary tree.
ulcerative colitis Chronic inflammatory condition involving the colon.

Sclerosing cholangitis (SC) refers to a host of chronic hepatobiliary disorders characterized by an irreversible diffuse fibrosis of the biliary system, leading to the formation of strictures of both the intra- and the extrahepatic biliary trees. The first description of SC is credited to Delbet in 1924. SC used to be discovered at surgery in the course of investigation of a patient with liver disease. It was not until the advent of endoscopic retrograde cholangiopancreatography in the early 1970s that SC became more recognized, which led to improved research and understanding of the disease. The etiology of SC is varied. When the etiology of the biliary fibrosis is known, the SC is referred to as secondary. When the etiology is not known, it is referred to as primary sclerosing cholangitis (PSC). PSC, however, seems to stand as a separate entity, a chronic and progressive fibrosing liver disease with clinical and serologic features suggestive of an underlying autoimmune dysfunction. Most of the published data pertain to PSC.

PRIMARY SCLEROSING CHOLANGITIS

Primary sclerosing cholangitis (PSC) is part of a spectrum of diseases referred to as sclerosing cholangitis. PSC is characterized by chronic, active, and progressive inflammation affecting the intra- and extra-hepatic bile ducts, leading to an irreversible fibrosis of the biliary tree, stricture formation, and bile duct obliteration and loss; it manifests with cholestasis and eventually leads to liver cirrhosis.

Epidemiology

PSC affects both males and females with a male : female ratio of 2 : 1; most patients present between 25 and 45 years of age. PSC has been reported in patients as early as the neonatal period and as late as the eighth decade. The prevalence of PSC in the United States is estimated to be 1 to 6 cases per 100,000. However, because of the strong association with inflammatory bowel disease (IBD), the prevalence of PSC is higher in populations where IBD is more prevalent, as in the Scandinavian countries.

Etiology/Pathophysiology

The etiology of PSC has not been defined. Many possible mechanisms and associations have been proposed but the evidence is inconclusive. The association between PSC and IBD and other immune system diseases, such as thyroiditis and type I diabetes mellitus, increased levels of serum immunoglobulins, the presence of autoantibodies, and an association with human leukocyte antigen (HLA) haplotypes are strong evidence that immune-mediated malfunction plays a major role in the pathogenesis of PSC. HLA-A1, -B8, and -DR3 as well as -DR52a, -DQ2, and -DQ6 haplotypes have been found to have a strong association with PSC. HLA-A1, -B8, and -DR3 haplotypes seem to confer an risk of developing PSC that is independent of ulcerative colitis (UC) but the risk is increased in its presence. The genetic inheritance pattern favors pleomorphism in the gene of the major histocompatibility complex. However, PSC does not have many of the characteristics of a classic autoimmune disease, such as having a female preponderance and response to immunosuppressive therapy.

Thus far, five different human HLA haplotypes have been associated with PSC: three with an increased risk and two with a reduced risk of the disease. Those that are

 Copyright 2004, Elsevier (USA). All rights reserved.

associated with an increased risk are MIC-A5.1 and MIC-B24.

Nonimmunogenic mechanisms have been proposed but supportive and reproducible evidence has been lacking. These mechanisms include chronic portal bacteremia, toxic bile acid metabolites produced by enteric flora, toxins produced by enteric bacteria, chronic viral infection, and ischemic vascular damage.

Portal Bacteremia

The association between PSC and UC led investigators to question whether chronic portal bacteremia could constitute a cause for PSC based on findings in animals. Further research failed to find evidence of portal vein phlebitis, a typical feature of portal bacteremia.

Toxic Bile Acid Metabolites

Lithocolic acid is hepatotoxic in animals. It is formed from chenodeoxycholic acid by bacterial 7-α-dehydroxylation in the colon. However, an abnormal concentration of bile acids was not found in the bile and portal blood of patients with PSC. Furthermore, lithocolic acid is rapidly sulfated and rendered nontoxic in humans.

Bacterial Toxins

When *N*-formyl L-methionineL-leucineL-tyrosine, a peptide produced by enteric flora, was introduced in rat colons, it produced a form of colitis and the biliary tree of the rats showed pathological changes similar to those occurring in PSC. Similar changes were also seen when killed *Escherichia coli* and other bacteria cell wall polymers were injected into the portal vein of rabbits. However, the natural history of PSC does not support this hypothesis. Antibiotic treatment was not effective; the course of PSC is distinct from that of UC and not altered even after proctocolectomy.

Viral Infection

Infection with human immunodeficiency virus (HIV) and cytomegalovirus (CMV) can produce changes in the biliary tree similar to those seen in PSC. However, there has been no isolation of CMV or Reovirus3 (a virus that has trophism to the biliary epithelium) from patients with PSC.

Ischemia

The pathology in patients with PSC does not show any involvement of the vasculature of the liver or any ischemic changes in the bile duct as seen in patients with vasculitis or intrahepatic arterial infusion of chemotherapeutic drugs.

Complications and Natural History

What is striking about the natural history of PSC is the variability in its clinical course and progression as well as the dissociation between the cholangiographic appearance of the disease and the clinical symptoms. Some patients might present with cholangitis that resolves spontaneously and does not recur for a long period of time; others, with a marked sclerotic appearance of the bile duct, may be asymptomatic and have only an elevated alkaline phosphatase or have advanced liver disease with cirrhosis. This discrepancy between symptoms and signs makes the formulation of a reliable outcome scoring system hard to achieve. Nevertheless, multiple scoring systems have been developed based on univariate and multivariate predictors to aid in monitoring the clinical course of PSC and the timing of liver transplantation.

The most serious complications that face patients with PSC are the development of liver cirrhosis and the development of cholangiocarcinoma (CC). Also, patients with PSC and concomitant UC seem to have a higher risk of developing colon cancer than patients with UC without PSC. Other complications include the formation of choledocholithiasis, cholangitis, malabsorption, pruritus, fatigue, and chronic pancreatitis.

Hepatic Cirrhosis

In PSC, the development of liver cirrhosis is a consequence of chronic biliary cholestasis. The constant inflammatory process that affects the bile ducts leads to progressive fibrosis of the liver parenchyma and bridging fibrosis. The exact mechanism that leads to this stage is poorly understood. Whether bile retention and copper retention play a role has yet to be elucidated. Once cirrhosis sets in, patients are subject to the usual complications of liver cirrhosis: portal hypertension, ascites, variceal bleeding (esophageal, gastric, or at the anastomotic site in patients with IBD who underwent surgery), peritonitis, encephalopathy, and liver failure; the last complication is the most common cause of death.

Cholangiocarcinoma

The prevalence of CC in patients with PSC ranges between 4 and 20%. The prevalence at autopsy is higher and approximates 40%. In patients undergoing liver transplantation, the prevalence of incidental CC is between 23 and 33%. The tumor is usually multicentric, arises around the proximal main bile duct and its bifurcation, and is less common distally along the duct. Factors such as age at presentation, sex, PSC involvement of the intra- or extrahepatic ducts, or the presence of IBD

do not affect the development of CC. Alcohol consumption is an additional risk factor for developing CC in PSC.

CT scan of the abdomen is valuable in a deteriorating patient to rule out the presence of CC. Ultrasound has low sensitivity in this regard. Positron emission tomography with [^{18}F]fluoro-2-deoxy-D-glucose might be of value in the diagnosis of CC with a sensitivity of 92% and a specificity of 93%.

Patients with tumors that display p53 overexpression and k-Ras mutation have a worse prognosis and shorter survival. However, these markers are not helpful in the diagnosis of CC. For example, k-Ras is found in 30% of patients with PSC without CC. DNA cytometry, a method that detects DNA aneuploidy, was found to be promising in the detection of CC in patients with PSC. Cells of CC showed DNA aneuploidy in 80% of the tissues compared to 12% in tissues of bile ducts of patients with PSC but without CC and 1% in tissues of gallbladders with chronic inflammation.

CA 19-9 and carcinoembryonic antigen (CEA) are not sensitive or specific for the diagnosis of CC. A model using both of these tests (CA 19-9 $-$ CEA $\times$ 40) has yielded 86% accuracy, 66% sensitivity, and 100% specificity in the diagnosis of CC.

Obtaining biopsies and cytology from brushing at the time of endoscopic retrograde cholangiopancreatography (ERCP) has resulted at best in a sensitivity of 80% but the average is approximately 45%. Currently, performing ERCP or any of the tests mentioned above as part of surveillance for CC has not been shown to be beneficial and, in the case of ERCP, is not recommended.

Choledocholithiasis

Cholelithiasis and choledocholithiasis are well described in patients with PSC. Gallstones are usually of the calcium bilirubinate variety. Pigment stones may form in the extra- and intrahepatic ducts, particularly in patients with recurrent attacks of cholangitis. In a retrospective review of 85 patients with sclerosing cholangitis, 20% of patients had biliary stones, 8% had cholelithiasis, 11% had choledocholithiasis, and 7% had both cholelithiasis and choledocholithiasis. The presence of cholelithiasis has no additional clinical significance compared to patients who have cholelithiasis without PSC. The development of cholangitis or a worsening clinical status in a patient with PSC should direct the clinic

Cholangitis

Acute ascending cholangitis can occur in patients with PSC. It occurs mainly in those patients who had endoscopic or surgical manipulation of the biliary system and is less likely in those without endoscopic stenting, dilation, or surgical drainage procedures. In the latter subgroup of patients, a less severe type of cholangitis occurs more frequently (10 to 20%). Usually it is self-limiting but can be recurrent.

Pruritus

Pruritus can be one of the most frustrating complications of cholestasis. Symptoms can become so severe as to interfere with sleep and cause skin abrasions. The pathogenesis of the pruritus is uncertain. The evidence that retention of endogenous bile acids causes cholestatic pruritus is deduced and indirect. Furthermore, recent findings suggest that central events in the brain that lead to an increased opioidergic tone may be implicated.

The view that bile acid retention causes pruritus is old. Varco1 in 1947 noted that biliary drainage reduced pruritus in patients with extrahepatic biliary obstruction and when bile was fed to patients, their pruritus returned. It is known that biliary drainage improved cholestatic pruritus in patients with intrahepatic cholestasis. Processes and medications that bind bile salts, such as the administration of cholestyramine, passage of plasma over charcoal or anion-exchange resins, and extracorporeal albumin dialysis, have been shown to diminish cholestatic pruritus. In all of these procedures, retained substances in addition to bile acids could have been removed at the same time so that cause and effect relationships are uncertain.

Whether peripheral events, such as the accumulation of bile acids in interstitial fluid of the skin, initiate the neural events that mediate this form of pruritus is uncertain. Three lines of evidence support this hypothesis: (1) Opioid receptor ligands with agonist properties (e.g., morphine) mediate pruritus. (2) Endogenous opioid-mediated neurotransmission/neuromodulation in the central nervous system is increased in cholestasis. (3) Opiate antagonists induce amelioration of the behavioral consequence of the pruritus of cholestasis (scratching activity). Serotonergic neurotransmission may also contribute to the pruritus.

Other Complications

Chronic pancreatitis is also associated with PSC. The frequency of this association differs widely in the literature, ranging from 8 to 77%. In general, the pancreatic exocrine function is well preserved and these patients rarely develop steatorrhea secondary to pancreatic insufficiency.

Colonic neoplasia seems to be more prevalent in patients with IBD and concomitant PSC than in patients who have IBD but not PSC.

Clinical Presentation

Symptoms

Patients with PSC usually present in an insidious manner. Most patients suffer from vague symptoms for approximately 2 years before diagnosis. These symptoms can include fatigue, right upper quadrant (RUQ) pain, and eventually jaundice and pruritus. Presentation of a patient with PSC differs between those who have an associated IBD and those who do not. In a Swedish study, 92% of asymptomatic patients had an associated IBD at the time of diagnosis versus 74% of symptomatic patients. This reflects the awareness of the association and therefore the early diagnosis. In all patients, 56% were symptomatic, 37% had RUQ pain, 30% had jaundice, 30% had pruritus, 4% had ascites, and 4% presented with esophageal variceal bleeding. Forty-six percent had histologic stages 1 and 2, 28% were stage 2, and 26% were stage 4. In another study, 10 to 15% had symptoms suggestive of recurrent cholangitis, such as fever, chills, and RUQ pain. Along the course of PSC, symptoms of weight loss or worsening liver test should raise the possibility of CC and, if fever and RUQ pain develop, of cholangitis.

Physical Exam

Early in the course of the disease and when PSC is diagnosed incidentally, physical signs specific for PSC are absent. Patients with advanced liver disease will manifest signs of portal hypertension and cirrhosis, such as spider angiomas, large spleen, muscle wasting, and icterus. Skin scratches secondary to pruritus can be seen. If the patient has cholangitis, abdominal exam reveals tenderness in the RUQ; otherwise, the abdominal exam is normal.

Laboratory Tests

The most common abnormal laboratory value is alkaline phosphatase (AKP). It is seen in more than 90% of patients with PSC. Usually it is elevated 3 to 5 times above the normal range but values as high as 20 times elevated have been reported. Normal values still could be seen in 6 to 8.5% of patients with PSC. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are rarely elevated more than three times their normal levels. Bilirubin is usually normal but is elevated in bouts of cholangitis or intermittent obstruction and in advanced cirrhotic patients. Elevated prothrombin time usually denotes vitamin K deficiency and, less likely, cirrhosis. Albumin is usually normal unless the patient has advanced liver disease.

Blood counts are normal, but eosinophilia can be observed on rare occasions. Serum levels of immunoglobulins are increased. The immunoglobulin M (IgM) level is elevated in 45% of patients followed by an elevation in IgG and IgA. Antinuclear cytoplasmic antibody in a perinuclear pattern is seen in 77 to 88% of patients with PSC. Smooth muscle antibodies are also seen. Anti-mitochondrial antibody is typically absent. Other miscellaneous laboratory findings include elevated levels of cholesterol, lipoprotein X, ceruloplasmin, and copper.

The differential diagnosis of cholestasis is long and is best approached by putting the patient's presentation in the context of his or her past medical history and life events in order to expose predisposing factors. A list of disorders that can present with cholestasis is outlined in Table I.

Diagnosis

The diagnosis of PSC requires a high index of suspicion since patients can be asymptomatic. All patients with IBD should be screened with liver tests. If alkaline phosphatase is elevated, the possibility of PSC should be ruled out. The diagnosis of PSC is established by demonstrating the characteristic changes affecting the biliary tree. The different modalities available to make the diagnosis are discussed below.

TABLE I Differential Diagnosis of Cholestasis

Acute
Biliary obstruction (benign and malignant)
Severe sepsis
Total parenteral nutrition
Medications[a] (chlorpromazine, estrogen, flucloxacillin, chlorpropramide, etc.)
Ischemic cholangitis[a] (hepatic artery injury, hepatic intrarterial chemotherapy)
Cholestatic hepatitis A
Chronic
Primary sclerosing cholangitis
Primary biliary cirrhosis
Autoimmune cholangitis
Idiopathic adulthood ductopenia
Ischemic cholangitis
Liver transplant-related (post transplant rejection, preservation injury, hepatic artery thrombosis, biliary anastomotic stricture)
Infectious (hepatitis C, AIDS)
Benign recurrent intrahepatic cholestasis
Recurrent pyogenic cholangitis
Infiltrative cholestasis (lymphoma, sarcoidosis, amyloidosis)
Cystic fibrosis
Malignancy (cholangiocarcinoma, hepatocellular carcinoma)

[a] Can also be chronic.

Imaging

Endoscopic retrograde cholangiopancreatography Because of the lack of specificity of many of the changes seen on liver biopsies, cholangiography is considered the gold standard for the diagnosis of PSC. Although percutaneous transhepatic cholangiography (PTC) is available, because of the nature of the small and attenuated bile ducts and the formation of biliary strictures in PSC, ERCP is the preferred method for diagnosing PSC. In addition to establishing the diagnosis, ERCP can provide a means for therapeutic intervention, such as stricture dilation and obtaining cytological brushing from dominant and suspicious strictures. Also, it can provide information about the pancreatic duct and rule out associated chronic pancreatitis. The classic appearance of affected bile ducts in PSC consists of diminished and attenuated intrahepatic ducts (pruning) with multifocal strictures alternating with areas of normal or mildly dilated segments (cholangiectasia), giving the typical beaded appearance. The finding of webs and diverticula in the biliary tree is not pathognomonic for PSC since it can be seen equally in other conditions. The majority of patients will demonstrate strictures of both intra- and extrahepatic ducts (seen in 67% of patients). Twenty-seven percent have intrahepatic involvement alone and 6% have extrahepatic duct involvement only. The gallbladder and the cystic duct are usually spared.

Adequate visualization of the intrahepatic ducts is needed to delineate the anatomy and exclude other etiologies. Using a balloon and injecting contrast under occlusion are typically used to achieve opacification of the ducts. For example, an underfilled intrahepatic tree in Caroli's disease can give an appearance of PSC. Bile ducts of patients with cirrhosis are attenuated but do not show the beading appearance seen in PSC.

Computed tomography The findings seen on CT scan in patients with PSC are not specific for this disease although they do suggest the diagnosis. However, CT is useful in the follow-up of patients with PSC once the diagnosis has been established. In 16 of 19 patients with involvement of the extrahepatic duct seen on cholangiography, CT demonstrated stenosis of the common bile duct. It also depicted intrahepatic ductal dilation/stenosis in all patients and detected 3 cases of CC. CT is superior to cholangiography in detecting CC with a sensitivity and specificity of 82 and 80%, respectively, compared to 54 and 53% for cholangiography.

Magnetic resonance cholangiography Magnetic resonance cholangiography (MRC) is a rapidly developing imaging modality that is gaining popularity in the diagnostic work-up of patients with hepatobiliary and pancreatic disease. The main advantages it offers over ERCP and PTC are its noninvasive nature and its ability to provide information about the extrabiliary structures. As regards PSC, the current technology of MRC is not yet able to detect the small changes of the intrahepatic ducts seen on cholangiography. Although MRC is not recommended for the diagnosis of PSC, it can be helpful in the work-up of a patient with PSC to rule out CC or as a road map for ERCP.

Ultrasonography For practical purposes, ultrasonography proves to be of little value in the diagnosis and follow-up of patients with PSC. It is of great value in the initial work-up of patients presenting with hepatobiliary symptoms and helps to rule out other etiologies of cholestasis, such as major biliary strictures, postoperative complications, choledocholithiasis, liver cirrhosis, and ischemic insults to the hepatic arteries and veins.

Pathology

Gross specimens of liver from patients with PSC show thickened and hardened bile ducts. Histology reveals evidence of fibrosis involving all layers of the bile ducts from the mucosa to the adventitia. Perfusion studies of livers removed at the time of liver transplantation demonstrated intrahepatic tubular and saccular dilation (cholangiectasia) with transformation of bile ducts into fibrous cords with complete or partial obliteration of the duct lumen. These changes give the typical beading and pruning appearance of the bile ducts seen on cholangiography.

Microscopically, four stages of biliary and hepatic involvement have been established and carry different prognostic values. Early in the course of the disease, the interlobular ducts are infiltrated with small and large lymphocytes, neutrophils, plasma cells, occasional macrophages, and eosinophils. Degeneration of the ductular epithelium is seen. The portal tracts are inflamed and edematous. The ductules show a periductular inflammation and concentric fibrosis representing the described "onion skin" appearance. This finding occurs in less than 50% of liver biopsy specimens. The portal tracts in PSC are typically less inflamed than in primary biliary cirrhosis (PBC). In stage 2, the above-mentioned inflammatory process extends outside the portal tract to involve the periportal hepatocytes. The earlier involved bile ducts start to disappear and bile ductopenia is prominent. There is interference with bile flow and features of cholestasis begin to develop in the hepatocytes adjacent to the portal tract. Accumulation of bile, copper, and copper-associated proteins and formation of Mallory bodies become apparent. Periductular fibrosis is less prominent. In stage 3, bridging fibrosis is seen,

bile ducts disappear, and cholestasis may be prominent, mainly in periportal and periseptal hepatocytes. Stage 4 is frank liver cirrhosis and the histologic feature is similar to what is seen resulting from other etiologies of liver cirrhosis. The parenchyma of the liver displays an increase in copper staining. It is important to note that the pathologic changes seen in PSC, especially in the early stages, vary in severity from one part of the liver to another, which makes staging on the basis of liver biopsy alone difficult.

In less than 5% of patients with IBD and histological changes typical of PSC, the cholangiogram is normal. These patients are thought to have small-duct PSC.

Management

Management of patients with PSC is aimed at addressing two main issues. The first is an attempt to slow the progression of the disease process and subsequent hepatic dysfunction. The second is to treat the complications of PSC.

Treatment of the Disease Process

No treatment has been shown to reverse the inflammatory process. At best there has been improvement in the liver tests and liver biopsy histology but without any proven reproducible evidence of delaying the progression of hepatic dysfunction.

Despite the prevailing concept of the autoimmune basis of PSC, immunosuppressive therapy has been disappointing. Results with oral corticosteroids are inconsistent. However, in a subgroup of patients with PSC with pronounced features of autoimmune disease, corticosteroids may be beneficial. Studies using budesonide had mixed results. AST and alkaline phosphatase (AKP) improved, whereas bilirubin increased. The combination of prednisone and ursodeoxycholic acid (UDCA) showed minimal improvement. Whether adding azathioprine to the latter regimen is beneficial needs further investigation. Cyclosporine A has been used with disappointing results. The use of tacrolimus, methotrexate, pentoxifylline, and cladibrine (a nucleoside analogue with anti-lymphocytic activity) still needs the verification of controlled and larger trials.

Medications that affect bile acid physiology have been used. Cholestyramine (a bile acid-binding resin) has been used with equivocal results. Similar to results obtained in treating PBC, UDCA has resulted in improvement in the laboratory tests (AST, ALT, AKP, and bilirubin) but the effect on histology is less apparent, especially in patients with cirrhosis. The cholangiographic abnormalities do not change and there is no proof that chronic therapy with UDCA alters survival. Cessation of treatment results in an increase in all laboratory values. The beneficial dose was between 15 and 25 mg/kg per day. Higher doses seem to exacerbate pruritus with no added benefit; a lower dosage, e.g., 10 mg/kg, did not have the same beneficial effect.

UDCA is the 7β-epimer of chenodeoxycholic acid and it makes a small fraction of the normal human bile acid pool. The mechanism whereby UDCA induces improvement in liver tests in PSC and other cholestatic diseases is unknown. Potential effects could be related to the following: (1) reduced concentration of the hydrophobic (toxic) bile acid; (2) a hepatoprotective effect; (3) stimulation of choleresis; (4) influence of bile acid transport; or (5) an immunomodulatory effect.

The low side-effects profile of UDCA resulted in its liberal use in PSC. However, long-term benefit from this therapy has not been shown yet. For this reason, and because of the slow progression of PSC, treatment of asymptomatic patients or those with early stage PSC should be carried out as part of a controlled clinical study when possible.

Treatment of Complications of PSC

By the time cirrhosis has developed in patients with PSC, liver transplantation is the only therapy that is associated with improved survival. PSC is the third most common cause of liver transplantation in adults. One-year survival after liver transplantation has been reported to range from 71 to 88%. Studies from the Mayo Clinic and the University of Pittsburgh have shown that the 5-year survival of patients with PSC post-liver transplant was 73% versus 28% for patients without liver transplantation. Unfortunately, the finding of an unsuspected CC in the explant was associated with decreased survival.

Short of liver transplantation, the complications of advanced liver disease and portal hypertension are treated as in any other cirrhosis. The complication of peristomal variceal bleeding occurs to a greater extent in patients with PSC with ileostomy secondary to colectomy from UC. This bleeding can be severe and hard to manage, eventually requiring portosystemic shunts or liver transplantation.

Osteoporosis develops in patients with cholestatic liver disease. Pathogenesis is not fully understood and therapy is not fully satisfying. Exercise, calcium, and vitamin D supplements should be given to all patients with PSC. Therapy with calcitonin and alendronate should be given to those with documented osteopenia or osteoporosis although improvement in bone density has not been consistent.

Steatorrhea usually develops late in the course of PSC. Deficiencies of fat-soluble vitamins, such as

vitamins A and K and, to a lesser extent, vitamins E and D, can occur. A low-grade pancreatic insufficiency can develop and therefore supplementation with pancreatic enzymes, in addition to maintaining a low-fat diet, can be helpful.

Endoscopic Treatment

As yet, there is no controlled study in patients with PSC to show a survival benefit from endoscopic intervention and dilation of the accessible biliary strictures. In addition, once the biliary milieu is violated with instrumentation, the incidence of cholangitis increases and these patients are bound for recurrent intervention. Therefore, ERCP should be reserved for patients who develop complications from PSC, such as cholangitis or dominant biliary strictures. Preliminary data have shown that patients with dominant ductal strictures who underwent scheduled endoscopic dilations had a longer interval before needing liver transplantation than controls.

Outcome Predictor Models

Survival analysis from a large number of asymptomatic patients with PSC showed that most patients have progressive disease. Moreover, their survival is decreased when compared to matched controls in the same geographic area. Data from the Mayo Clinic suggest that 33 to 40% of patients with PSC die within 5 to 7 years after their diagnosis, with an average survival of 11.9 years. Because of the latter finding, risk score models that reflect disease severity and predict survival have been developed. These models, despite their incomplete nature, do provide an objective means of assessing the severity of the disease in patients with PSC and therefore help in their management and in the timing of liver transplantation. The Mayo Clinic PSC model was based on data from 426 patients and was validated. Predictors of survival were as follows: age, serum bilirubin, histologic stage on liver biopsy, and splenomegaly. According to this model, patients who are in the low-risk category had a probability of 92% for 5-year survival. Moderate-risk patients have a 5-year survival of 55% and those in the high-risk category had a 1-year survival of 70% and a 5-year survival of 18%. Still, this model does not correct for the fact that the development of CC does not directly correlate with the duration or the stage of PSC.

Relation to Inflammatory Bowel Disease

It is estimated that 75% of patients with PSC have IBD; 87% of these patients have UC and 13% have Crohn's disease. It is known that 2.5 to 7.5% of patients with UC have or will develop PSC during their illness. It is more common in patients who have diffuse colonic involvement than in those with distal involvement. Patients with IBD tend to be older at the time of diagnosis (44 versus 38 years of age). Sex, pattern of distribution of the biliary strictures, and time to develop CC did not differ between patients with and those without an associated IBD. Several studies have shown that the presence of PSC in patients with UC increases the incidence of colorectal neoplasia. Currently it is suggested that these patients undergo yearly surveillance colonoscopy. Whether the risk of colon cancer increases in these patients after liver transplantation is still controversial.

SECONDARY SCLEROSING CHOLANGITIS

Biliary tract pathology similar to PSC on laboratory and cholangiographic studies can develop secondary to a variety of insults. Injury affecting the vascular supply of the biliary tree (the hepatic artery) can cause ischemic cholangiopathy, and infectious conditions, such as acquired immunodeficiency syndrome (AIDS), can affect the biliary tree. Direct exposure of the biliary tree to different forms of scolecide, such as concentrated sodium chloride solution (30%) or alcohol, during therapy of liver *Echinococcus* cysts (hepatic hydatic disease) can also cause sclerosing cholangitis.

A different form of inflammatory change that affects the biliary tree in a segmental fashion is seen in recurrent pyogenic hepatitis. In this disorder, the intrahepatic ducts show "beading" and dilation with formation of intrahepatic stones of different sizes and number, affecting the left hepatic ducts more frequently than the right.

Ischemic Cholangitis

Ischemic cholangitis is a term used to describe a sclerosing cholangitis that is due to an injury of the hepatic artery, its major branches, and all the way to the level of the microscopic peribiliary plexus. The biliary tree is totally dependent on arterial circulation for its blood supply and any injury to this system can lead to ischemia followed by stricturing and beading of the affected area. Table II lists the causes of ischemic cholangitis.

Clinically, the diagnosis is suspected when patients have the predisposing conditions and present with increased bilirubin and AKP. Cholangiogram shows a picture that is indistinguishable from PSC. On histology, the biliary epithelium of the major ducts displays atrophy and erosion and, when severe, cholangiectasia.

TABLE II Causes of Ischemic Cholangitis

Hepatic artery thrombosis after liver transplantation
Cold and warm preservation injury post-OLT[a]
Postcholecystectomy arterial injury
Intrahepatic artery infusion of floxuridine, epirubicin
Vasculitis (temporal arteritis, polyarteritis nodosa)
Paroxysmal nocturnal hemoglobinuria

[a] Orthotopic liver transplantation.

The smaller ducts and the intrahepatic ducts show similar findings and sometimes ductopenia. Depending on the etiology of ischemic injury, the surrounding arteries can show evidence of vasculitis, thrombosis, and signs of chronic rejection or they can be normal.

In the setting of liver transplantation, the biliary tree is at risk for serious complications. Anastomotic strictures used to be more common in the early days of transplantation but decreased in frequency with improvements in surgical techniques. However, it is still one of the most common complications in living-related donor liver transplantation. Nonanastomotic strictures can be due to hepatic artery thrombosis or prolonged cold or warm ischemia time during the processes of organ procurement and vascularization, respectively. Complications from hepatic artery thrombosis (HAT) carry high rates of morbidity and mortality (30–40%). HAT usually occurs in 3% of liver transplantation patients but has been reported to occur in as many as 19% of patients. Diagnosis is suspected when patients present with fever, elevated liver tests, or frank cholangitis. Occasionally, patients can be asymptomatic or can present with fulminant hepatic failure. This complication can occur early (less than 4 weeks) or late (more than 4 weeks) posttransplantation. Diagnosis could be established with a Doppler ultrasound study (sensitivity of 53%) but occasionally an angiogram must be performed (sensitivity of 82%). Management is usually by retransplantation (71%). Occasionally, intra-arterial urokinase treatment is effective. If the graft was salvaged, treatment of the biliary stricture proves to be difficult, requiring multiple sessions of endoscopic dilation and stenting, most frequently with hepatico-jejunostomy being the final treatment.

Infusion of different chemotherapeutic materials directly into the hepatic artery has been associated with biliary strictures in 50% of reported series. The most recognized agent is floxuridine. The speculated mechanism for inducing the biliary lesion is direct toxicity and secondary thrombosis to the hilar hepatic arteries used.

Infectious Cholangiopathy

Secondary cholangitis has been seen in association with certain infectious diseases. The best described is AIDS cholangiopathy, seen in patients with advanced AIDS. The prevalence of AIDS cholangiopathy is not well documented. In one study examining patients with AIDS-related diarrhea, cholangiopathy was seen in 30% of these patients. The etiologic factor is believed to be infectious. The most commonly isolated organism is *Cryptosporidium parvum*. Other organisms found include Microsporidia, *Mycobacterium avium* complex, CMV, *Isospora belli*, *Salmonella*, *Enterobacter*, and *Candida albicans* species. Also, the HIV virus itself is known to involve the biliary tree but clinical significance is not proven. In addition to infectious agents, Kaposi's sarcoma and Burkitt's lymphoma can cause similar results in this population.

Patients with AIDS cholangiopathy can be asymptomatic, with the only abnormality being an elevated AKP. Other patients can present with frank cholangitis with RUQ pain, fever, and leukocytosis. Abdominal pain is the most common symptom encountered in AIDS cholangiopathy. Bilirubin is usually less than 3 mg/dl, with higher levels seen in patients with biliary strictures secondary to Kaposi's sarcoma, lymphoma, or extrinsic bile duct compression. AST and ALT can be mildly elevated. AIDS cholangiopathy has been classified into three subgroups: (1) diffuse biliary involvement (37%); (2) combined biliary involvement and papillary stenosis (50–60%); or (3) papillary stenosis alone (6–8%).

Patients with AIDS and elevated AKP should be suspected of having AIDS cholangiopathy. Diagnosis is by demonstrating a cholangiogram with beading and narrowing of the biliary tree or with papillary stenosis. ERCP proves to be the diagnostic modality of choice. It offers diagnosis, a means to collect samples for cultures, and a therapeutic modality in case of the presence of papillary stenosis. The cholangiogram shows the main bile duct to have a rugged and irregular contour. The intrahepatic ducts show a pattern of narrowing and dilation similar to that seen in PSC. Isolated strictures of the main bile duct are usually not present, but if they are present, they are usually related to a malignancy (lymphoma, Kaposi's sarcoma) or due to external compression. However, ERCP should not be performed in asymptomatic patients and should be reserved for those in whom papillary stenosis is suspected. Isolating the culprit organism is not straightforward. Combining the cultures of bile aspirate with biopsies from the biliary tree and from the peripapillary region can increase the yield.

The prognosis of patients with AIDS cholangiopathy is poor. However, the survival of these patients is not different from patients with AIDS with the same opportunistic infections but without cholangiopathy.

Medical therapy for various organisms involved (paromomycin for *Cryptosporidium*, ganciclovir for CMV) has failed to clear the biliary infection, change the course of the cholangiopathy, or affect survival. Patients usually die from AIDS-related complications other than the cholangiopathy. ERCP and endoscopic sphincterotomy do not alter the natural history of AIDS cholangiopathy; however, they do relieve abdominal pain in 75% of patients with associated papillary stenosis.

Secondary sclerosing cholangitis and subsequent cirrhosis were reported in three patients following severe systemic (extrahepatic/extrabiliary) infection. Also, a case of sclerosing cholangitis in association with *Clonorchis sinensis* (a parasite) has been reported.

Other Etiologies

Hypereosinophilic sclerosing cholangitis is a rare disease caused by eosinophilic infiltration of the gallbladder and biliary tract seen in idiopathic hypereosinophilic syndrome. Therapy with steroids has been reported to improve symptoms, biochemical parameters, and cholangiographic parameters. Pathophysiology is not fully known.

Langerhans cell histiocytosis is a rare clonal disorder that consists of single or multiple mass lesions composed of cells with an abnormal Langerhans cell phenotype seen mainly in the pediatric population. Adult patients with this disorder can have involvement of the liver and biliary tree in a picture similar to sclerosing cholangitis.

Sarcoidosis is a multisystem granulomatous disorder of unknown cause that presents most commonly with hilar adenopathy, pulmonary infiltrates, and skin or eye lesions. Liver involvement is seen in more than 90% of these patients, manifesting with cholestasis. Biliary involvement, however, is not common but a case resembling sclerosing cholangitis has been reported. Improvement of the biliary stricture was seen after therapy with prednisone. Finally, tumors affecting the biliary tree can present with a cholangiographic appearance of sclerosing cholangitis and should be ruled out.

In summary, sclerosing cholangitis is a spectrum of disease that affects the intrahepatic and extrahepatic biliary trees. More often than not, it is an irreversible, slow, and progressive fibrosing process that leads to hepatic cirrhosis. Unfortunately, there is no effective therapy short of liver transplantation. Further understanding of the physiology of inflammation and repair of the biliary and hepatic system might lead to improvements in therapy.

See Also the Following Articles

AIDS, Biliary Manifestations of • Cholangiocarcinoma • Choletithiasis, Complications of • Cholestatic Diseases, Chronic • Cirhosis • Colitis, Ulcerative • Crohn's Disease • Gallstones, Pathophysiology of • Liver Transplantation • Pruritus of Cholestatis

Further Reading

Bass, N. M. (1998). Sclerosing cholangitis. *In* "Sleisenger and Fordtrand's Gastroenterology and Liver Disease," 6th Ed., Chap. 59, p. 1006.

Batts, K. P. (1998). Ischemic cholangiopathy. *Mayo Clin. Proc.* **73**, 380–385.

Broome, U., Olson, R., *et al.* (1996). Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. *Gut* **38**, 610–615.

Campell, W. L., Peterson, M. S., Federal, M. P., *et al.* (2001). Using CT and cholangiography to diagnose biliary tract carcinoma complicating primary sclerosing cholangitis.

Holtmeier, J., and Leuschner, U. (2001). Medical treatment of primary biliary cirrhosis and primary sclerosing cholangitis. *Digestion* **64**, 137–150.

Jones, E. A., and Bergasa, N. V. (1999). The pathogenesis and treatment of pruritus and fatigue in patients with PBC. *Eur. J. Gastroenterol. Hepatol.* **11**, 623–631.

Kim, W. R., Ludwig, J., and Lindor, K. D. (2000). Variant forms of cholestatic diseases involving small bile ducts in adults. *Am. J. Gastroenterol.* **95**, 1130–1137.

Loinaz, P. D., Garcia, S. C., Jimenez, G. I., *et al.* (2000). Diagnosis and treatment of hepatic artery thrombosis after liver transplantation. *Chirurgia Ital.* **52**, 505–525.

Olson, R., Danielsson, A., Jarnerot, G., *et al.* (1991). Prevalence of primary sclerosing cholangitis in patients with ulcerative colitis. *Gastroenterology* **100**, 1319–1323.

Scheur, P. J. (1998). Pathologic features and evolution of primary biliary cirrhosis and primary sclerosing cholangitis. *Mayo Clin. Proc.* **73**, 179–183.

Wiesner, R. H. (1998). Liver transplantation for primary biliary cirrhosis and primary sclerosing cholangitis: Predicting outcomes with natural history models. *Mayo Clinic Proc.* **73**, 575–588.

Wilcox, C. M., and Monkemuller, K. E. (1998). Hepatobiliary diseases in patients with AIDS: Focus on AIDS cholangiopathy and gallbladder disease. *Digest. Dis.* **16**, 205–213.

Cholecystectomy

PATRICK J. JAVID * AND DAVID C. BROOKS *,†
*Brigham and Women's Hospital and †Harvard Medical School

cholangiogram Radiograph of the gallbladder and bile ducts; may be performed intraoperatively by either an open or a laparoscopic technique.
cholecystitis Inflammation of the gallbladder.
cholecystostomy Drainage of the gallbladder, usually by percutaneous tube placement, in patients with cholecystitis who are too unstable to tolerate a traditional cholecystectomy.
choledocholithiasis Stones in the common bile duct.
gallstones (cholelithiasis) Stones or pebbles within the gallbladder.
pneumoperitoneum Insufflation of carbon dioxide gas into the abdominal cavity; performed in all laparoscopic procedures to increase the "work space" available inside the abdomen.

Cholecystectomy, defined as surgical removal of the gallbladder, can be performed for a variety of indications. The traditional approach has been through an open incision in the abdomen, but a revolutionary change in the field of general surgery took place with the introduction of the laparoscopic cholecystectomy in the 1980s. Today, laparoscopic cholecystectomy is one of the most common surgical procedures performed worldwide. After intensive retrospective and prospective review, both surgical approaches to cholecystectomy achieve excellent results and a high degree of patient satisfaction.

HISTORICAL BACKGROUND

Although the existence of gallstones was first recognized in the fifth century, the first documented cholecystectomy took place on July 15, 1882, by Carl Langenbuch in Berlin. After extensive anatomical dissection on cadavers and live animals, Langenbuch's first human patient was a 43-year-old male who suffered from longstanding biliary colic so debilitating that he had become a morphine addict. The operation was deemed a success; not only did the patient survive, but his biliary colic was fully relieved and he gained over 10 pounds during his postoperative stay alone. The first successful cholecystectomy in the United States was performed by Justus Ohage in St. Paul, Minnesota, in 1886.

Open cholecystectomy would remain the gold standard for treatment of gallbladder pathology for the next 100 years. In 1987, Mouret performed the first laparoscopic cholecystectomy in France, thereby introducing a revolutionary change in biliary surgery. Over the past decade and a half, the laparoscopic cholecystectomy has become the new gold standard in the treatment of a wide range of gallbladder disease.

INDICATIONS FOR CHOLECYSTECTOMY

The indications for cholecystectomy include cholelithiasis, cholecystitis, biliary dyskinesia, and gallbladder cancer. Most patients with cholelithiasis are asymptomatic, and there is sufficient evidence that prophylactic cholecystectomy for asymptomatic disease is not warranted. Although the annual conversion rate of asymptomatic to symptomatic gallstones is 1–4% per year, a majority of patients develop symptoms for some time before they develop the complications of cholelithiasis, such as cholecystitis or gallstone pancreatitis. Moreover, studies that have followed patients with asymptomatic gallstones over several years have found that less than 20% of these patients will develop biliary symptoms, and the patients with symptoms can be safely treated at the time of presentation. Therefore, the risk of maintaining observation of asymptomatic patients is less than or equal to the risk of prophylactic cholecystectomy, and surgery is not recommended in these patients. One exception to this rule is the pediatric patient with gallstones; a young patient with gallstones will most likely develop symptoms over time. Another important exception is the patient who is undergoing an intraabdominal operation for morbid obesity; a patient with rapid postoperative weight loss is prone to symptomatic cholelithiasis, and data show that open cholecystectomy at the same time as the bariatric surgery is both clinically prudent and cost-effective.

Symptomatic cholelithiasis is the primary indication for cholecystectomy. The operation is performed to avoid recurrent biliary colic and to prevent complications such as choledocholithiasis, cholecystitis, and gallstone pancreatitis. The majority of patients with one episode of biliary colic will suffer from recurrent

 Copyright 2004, Elsevier (USA). All rights reserved.

pain. In addition, symptomatic patients have a higher risk of complications from their gallstone disease. In one large population-based study, 6.5% of symptomatic patients developed serious complications of gallstone disease that required surgical intervention over a 10-year period.

Acute cholecystitis mandates a surgical procedure on the gallbladder. The natural history of untreated cholecystitis involves gangrenous necrosis of the gallbladder and ultimately life-threatening perforation. In most cases, a laparoscopic or open cholecystectomy is indicated to remove the gallbladder as the primary source of infection. Although in the past many surgeons have treated patients suffering from acute cholecystitis with 4- to 6-week courses of antibiotics before operating, current data show that early operation within 48–72 hours of the onset of symptoms yields the most favorable outcomes. In unstable patients for whom the formal operating room is contraindicated, percutaneous cholecystostomy tube placement and future interval cholecystectomy are indicated.

Acute acalculous cholecystitis also warrants surgical management of the gallbladder. In the stable patient, cholecystectomy is the treatment of choice. However, the patient subset most prone to the acalculous variety of cholecystitis also has a high incidence of hemodynamic instability and severe cardiovascular disease and may be more appropriately treated with cholecystostomy.

Biliary dyskinesis is defined as symptoms of biliary colic with the absence of gallstones and is often associated with abnormal functional emptying of the gallbladder. It is believed that more than 90% of patients with documented biliary dyskinesia (as determined by abnormal gallbladder emptying in response to a fatty meal) have improvement or resolution of their symptoms following cholecystectomy.

Gallbladder cancer is an uncommon indication for cholecystectomy. Most often, cancer of the gallbladder is unsuspected and is found on routine pathologic examination of the gallbladder specimen following cholecystectomy performed for a separate indication. If gallbladder cancer is suspected preoperatively, the operation of choice is open cholecystectomy, given the risks of port site metastasis in laparoscopic cholecystectomy and the need for complete resection of the gallbladder and the adjacent hepatic tissue. For gallbladder cancer confined to the muscular layer of the gallbladder wall (Stage I), simple cholecystectomy is sufficient. For more extensive malignancies, an extended cholecystectomy with resection of portions of the liver around the gallbladder bed is indicated.

OPERATIVE STRATEGY

In preparation for both laparoscopic and open cholecystectomy, all patients should first undergo a thorough history and physical examination. Routine laboratory studies are ordered, including liver function studies; normal liver function confirms that the common bile or hepatic ducts are not obstructed by the patient's gallstone disease. The presence of nonobstructing stones, however, is not ruled out. The risks of the procedure to be discussed in detail with the patient should include infection, bleeding, possibility of bile duct injury, and conversion to open procedure if laparoscopic cholecystectomy is planned. A prophylactic dose of an intravenous first- or second-generation cephalosporin is given preoperatively. All patients are positioned in the supine position on the operating room table, and the entire abdomen is prepped with sterile technique.

Laparoscopic Cholecystectomy

The laparoscopic cholecystectomy involves extensive instrumentation and monitors. In general, a laparoscope, laparoscopic electrocautery and suction devices, and gas flow line are connected at the start of the procedure. Two monitors are placed on either side of the patient, usually toward the head of the bed so that both the operating surgeon and the first assistant have available lines of vision. The operating surgeon stands on the patient's left side. Finally, a nasogastric tube is placed to decompress the stomach.

The first maneuver in a laparoscopic cholecystectomy is the establishment of pneumoperitoneum. This can be done percutaneously with a Veress needle or by an open incision directly into the abdomen. This is performed through a 1-cm incision that holds a 10-mm trocar containing the laparoscope. The incision is often made just below the umbilicus in a transverse fashion but may also be performed vertically above and below the umbilicus. Through this incision, dissection is performed first to the fascia, which is sharply incised, and then to the abdominal cavity. The surgeon can place a finger through the incision and probe into the open abdomen. A 10-mm trocar is placed through this incision, and the gas flow is connected to the trocar to achieve pneumoperitoneum. The laparoscope is then inserted through the trocar when pneumoperitoneum is adequate.

At this point, the surgeon inspects the abdominal cavity. The camera is rotated around the abdomen, looking for adhesions to the anterior abdominal wall as well as additional intraabdominal pathology. Next, two 5-mm incisions are made on the patient's right

anterior abdominal wall, one in the right upper quadrant in the midclavicular line and one laterally in the anterior axillary line at the level of the umbilicus. The 5-mm trocars are placed through these incisions under direct visualization from the laparoscope to avoid injury to the liver or bowel. Grasping forceps are inserted via these trocars. A final incision is made through the midline 3 cm below the xiphoid, and a blunt dissector is inserted.

Using the grasping forceps, the fundus of the gallbladder is retracted toward the diaphragm and tipped posteriorly. This lifts the posterior liver to expose the infundibulum of the gallbladder. Detailed inspection of the gallbladder is carried out; occasionally, due to extensive adhesions around the gallbladder, it is clear at this point that the operation would carry a prohibitively high risk of complications if continued laparoscopically, and an open procedure is then initiated. If the gallbladder appears suitable for laparoscopic removal, the lower grasping forcep is attached to the apex of the gallbladder fundus and is retracted cephalad and held firmly in this position. The remaining grasping forcep is placed on the lateral infundibulum of the gallbladder and is retracted anterolaterally. In this way, the gallbladder is stretched out in both a vertical and transverse axis so that the triangle of Calot is splayed out, isolating and exposing the bile ducts and the vascular structures.

Gently using the blunt dissector, any omental adhesions to the gallbladder infundibulum are lightly grasped and retracted downward. After all adhesions are removed, the peritoneum overlying the gallbladder infundibulum is then dissected downward. These actions serve to expose the region of the cystic duct and artery. Once exposed, the next goal is to dissect around the cystic duct and artery separately, so that both may be individually clipped and ligated. During this portion of the operation, the operating surgeon should carefully trace out the cystic duct as well as the common bile duct. The surgeon should plan on ligating the cystic duct close to the gallbladder so as to avoid any injury to the common bile duct. If this anatomy cannot be readily discerned, a cholangiogram and/or conversion to an open procedure should be strongly considered.

When the cystic duct and/or artery have been fully exposed and dissecting instruments are easily passed around the structure, a clip applicator is inserted through the subxiphoid trocar and placed around the structure. It is recommended to place two clips on the inferior side of the structure and one clip on the superior aspect. The applicator is then removed, and endoscissors are used to divide the duct or artery between the upper and lower clips, maintaining a small stump on the downward side. After one structure has been clipped and ligated, attention turns toward the remaining structure for dissection and ligation.

With the cystic structures ligated, the gallbladder remains attached only to the liver bed. The peritoneum on the lateral and medial sides of the gallbladder is scored with electrocautery and the gallbladder is separated from the liver bed using a combination of anterior retraction and dissection via electrocautery. Any bleeding areas from the liver bed are coagulated with electrocautery as soon as they are visualized, because they often bleed profusely and retract when the gallbladder is fully removed. Once the gallbladder is completely detached from the liver, the camera is placed through the subxiphoid trocar, and a large grasping forceps or specimen bag is inserted through the umbilical port. The assistant then places the gallbladder in the forceps or bag, which is withdrawn via the umbilical port. At this point, the area is inspected for bleeding, and irrigation is performed. The umbilical incision is closed first at the fascia and then all incisions are closed with running subcuticular absorbable sutures.

Open Cholecystectomy

Elective open cholecystectomy is performed when it is decided preoperatively that laparoscopic cholecystectomy would be too difficult or contraindicated. This most often occurs in the patient with multiple previous abdominal surgeries. The procedure is also performed when a laparoscopic cholecystectomy is found to be too difficult and risky to the patient intraoperatively, and the procedure is therefore converted to an open cholecystectomy. Finally, open cholecystectomy may be performed in conjunction with other procedures, such as bariatric surgery or pancreatic and duodenal procedures.

Open cholecystectomy is performed using either the "fundus-down" or the "retrograde" approach to the gallbladder. Both approaches utilize the same incision and exposure of the right upper quadrant of the abdomen. A right subcostal incision is made approximately two fingerbreadths below the costal margin. Alternatively, an upper midline incision may be used if additional abdominal pathology exists or is in question prior to the operation. Once the abdominal cavity has been entered, the gallbladder is palpated to examine for stones in the common bile duct. A clamp is placed on the fundus of the gallbladder and lifted cephalad to expose the gallbladder in full.

In the "fundus-down" approach, the fundus is grabbed securely with the clamp and an incision is

made superficially in the gallbladder serosa around the fundus. This allows for exposure of a plane between the visceral layer of gallbladder serosa and the liver bed. This plane is dissected using electrocautery starting at the tip of the fundus and working down across the back side of the gallbladder toward the body and neck. In this way, a small amount of serosa is left on the liver side of the gallbladder bed and the liver is not dissected. As dissection is continued, the neck of the gallbladder leads to the cystic duct, which is then visualized, palpated for stones, and ultimately clamped with a right angle. If stones are felt in the cystic duct, they are pushed into the gallbladder neck rather than distally into the common bile duct prior to clamping. The cystic duct is then suture ligated using a nonabsorbable suture. The cystic artery is also exposed in this way, clamped, and suture ligated. The exposure usually allows for visualization of the common bile duct; if difficulty exists in defining the cystic and common bile ducts, a cholangiogram is strongly recommended.

In the retrograde approach, following skin incision and exposure of the gallbladder, attention is turned to the region of the porta hepatis, which contains the cystic and common bile ducts as well as the cystic artery. Omental fat or adhesions to these structures are taken down carefully to facilitate exposure. Once the structures are adequately isolated, the cystic duct and artery are dissected and secured with ties but not divided until all other structures have been dissected and identified. This prevents inadvertent division or injury to the common bile duct. At this point, a second clamp can be placed on the body of the gallbladder to help retract the gallbladder away from its liver bed. This exposes the aforementioned avascular plane between the gallbladder and the liver and allows for dissection using electrocautery. This method of dissection is similar to that in the laparoscopic cholecystectomy in that it proceeds from the neck to the fundus of the gallbladder. Although the retrograde approach commonly results in less bleeding because the cystic artery is ligated at the beginning of the dissection, primary exposure of the duct and artery structures can be more difficult. Both approaches have shown equivalent success rates.

Drainage of the gallbladder bed is controversial. Numerous studies have shown that drainage for an uncomplicated procedure does not significantly improve wound infection rate or overall morbidity. Most surgeons agree that drainage is necessary only if the case is complicated by significant bleeding or concern over the integrity of the biliary tree. Placing a drain in complicated procedures allows for a controlled fistula if a bile leak ensues, as well as early recognition of excessive postoperative bleeding.

OUTCOMES OF CHOLECYSTECTOMY

Cholecystectomy has been performed for over 100 years and is a safe operation with a low rate of morbidity and minimal mortality. At the advent of the current laparoscopic era, several studies reviewing tens of thousands of patients confirmed that the open cholecystectomy was indeed the gold standard for the treatment of symptomatic gallstones. The overall mortality rate in these studies averaged less than 0.5%, with one population-based study of 42,474 patients documenting a mortality rate of 0.17%. Patient age was the most consistent single factor associated with operative mortality; other influential factors include comorbid conditions and nature of the biliary disease. In fact, the mortality rate from open cholecystectomy for acute cholecystitis is three times the rate for simple biliary colic but still approximates 0.6%. Cardiovascular disease, including perioperative myocardial infarction, stroke, congestive heart failure, and pulmonary embolism, was the leading cause of postoperative death in this patient group. Technical complications of the procedure account for the minority of all postoperative mortality.

The morbidity rate of open cholecystectomy remains 5–15%, although most complications are nonbiliary in nature and relatively minor. Approximately half of these complications result in no disability and do not prolong the patient's hospital stay, and serious disabling morbidity occurs in only 0.16% of patients. Pulmonary difficulties are common and often secondary to the right subcostal muscle-splitting incision, which makes deep inspiration painful and contributes to risk of perioperative pneumonia. Bile leakage or fistula has been found in 0.3–0.6% of patients. Injury to the common bile duct, which often requires extensive operative reconstruction and increases overall rate of mortality, has been found to occur in only 0.1–0.2% of cases.

Laparoscopic cholecystectomy has proved equal to open cholecystectomy in mortality and morbidity, with the additional advantages of increased patient comfort, shorter hospital stays, and quicker recovery periods. The procedure is frequently done as day surgery—significantly shorter than the typical 3- to 5-day stays following an open cholecystectomy. Studies have shown an average time of 8 days from day of operation to return to work. Overall conversion from a laparoscopic to an open procedure occurs in 2–5% of all laparoscopic cholecystectomies.

Mortality from laparoscopic cholecystectomy is exceedingly low, ranging from 0 to 0.15%. In general, the mortality rate for laparoscopic cholecystectomy is lower than open cholecystectomy. However, of the few deaths

from laparoscopic cholecystectomy, a sizable number are related to technical complications. Complications that increase the risk of mortality include bowel perforations and vascular injuries from percutaneous trocar placement as well as trauma to extrahepatic bile ducts. In some cases, these complications are not observed during the initial procedure, and the delay in the diagnosis results in an unstable and critically ill patient.

The overall risk of complication from laparoscopic cholecystectomy is roughly equal to that of open cholecystectomy. Studies have documented morbidity rates of 2–5%, again with a predominance of minor complications. Wound infection is the most common complication, at 0.3–1%, and is usually found at the umbilical trocar site where the gallbladder is removed from the abdomen. Postoperative hernias, again most common at the umbilical incision, have been found in 0.3–0.5% of patients. Major complications are infrequent but again are often technical in nature. Excessive bleeding (from the gallbladder bed, incision site, or vascular injury) is found in 0.5% of patients, and bowel perforations have been recorded in 0.1–0.3% of patients. Major bile duct injuries have been seen in 0.3–0.7% of patients, slightly higher than in the open cholecystectomy. A majority of patients with major bile duct injuries require laparotomy either at the time of the initial surgery or at a later date for reconstruction. Complications requiring laparotomy are seen in approximately 1% of all patients undergoing a laparoscopic cholecystectomy.

Although laparoscopic cholecystectomy was once thought too risky for patients with acute cholecystitis, the procedure is now used routinely for these patients early in the course of their infection. Initial retrospective reviews revealed that laparoscopic cholecystectomy can be performed in this patient group with acceptable morbidity and no mortality. More structured studies in patients with acute cholecystitis have found that the overall complication rate for laparoscopic cholecystectomy is equivalent to or less than that for open cholecystectomy. Conversion rate, however, is higher in these patients, with a rate of 7–28%, but appears dependent on the timing of the operation. Recent data have shown conclusively that early laparoscopic cholecystectomy (within 48–72 hours of presentation of acute cholecystitis) results in acceptable outcomes, with a lower risk of conversion to an open procedure.

Within 5 years of its inception in the operating room, the laparoscopic cholecystectomy has become the new gold standard for cholecystectomy and continues in that role today. Open cholecystectomy is reserved for patients with difficult anatomy or advanced cholecystitis.

See Also the Following Articles

Cholelithiasis Complications of • Gallbladder Cancer • Gallstones, Pathophysiology of • Laparoscopy

Further Reading

Deziel, D. J. (1994). Complications of cholecystectomy. *Surg. Clin. N. Am.* **74**, 809–823.

Gracie, W. A., and Ransohoff, D. F. (1982). The natural history of silent gallstones. *N. Engl. J. Med.* **307**, 798–800.

Johnson, D. E., and Kaplan, M. M. (1993). Medical progress: Pathogenesis and treatment of gallstones. *N. Engl. J. Med.* **328**, 412–421.

Lujan, J. A., Parrilla, P., Robes, R., Marin, P., Torralba, J. A., and Garcia-Ayllon, J. (1998). Laparoscopic cholecystectomy vs open cholecystectomy in the treatment of acute cholecystitis. *Arch. Surg.* **133**, 173–175.

Roslyn, J. J., Binns, G. S., Hughes, E. F. X., Saunders-Kirkwood, K., Zinner, M. J., and Cates, J. A. (1993). Open cholecystectomy: A contemporary analysis of 42,474 patients. *Ann. Surg.* **218**, 129–137.

Soper, N. J., Stockmann, P. T., Dunnegan, D. L., and Ashley, S. W. (1992). Laparoscopic cholecystectomy: The new gold standard? *Arch. Surg.* **127**, 917–921.

Strasberg, S. M., and Clavien, P. A. (1993). Overview of therapeutic modalities for treatment of gallstone disease. *Am. J. Surg.* **165**, 420–426.

The Southern Surgeons Club. (1991). A prospective analysis of 1518 laparoscopic cholecystectomies. *N. Engl. J. Med.* **324**, 1073–1078.

Cholecystokinin (CCK)

RODGER A. LIDDLE
Duke University

gastrin Gastrointestinal polypeptide hormone produced by G cells of the gastric antrum; stimulates gastric acid secretion.

G-protein-coupled receptor Family of cell surface receptor proteins consisting of seven transmembrane regions. The intracellular region interacts with guanosine triphosphate-binding proteins that, on ligand binding, transduce signals within the cell, leading to a cellular response (e.g., secretion, motility, and growth).

microvilli Fingerlike extensions along the apical surface of intestinal mucosal cells. On enterocytes, microvilli increase the absorptive surface of the cell; on endocrine cells, microvilli allow potential stimuli greater exposure to their targets (e.g., receptors).

myenteric (Auerbach's) plexus System of nerves and ganglia lying within the longitudinal and circular muscle layers of the intestine. The nerves of the system innervate numerous targets, including the myenteric externa, mucosa, and sympathetic prevertebral ganglia.

sphincter of Oddi Muscular region surrounding the distal ends of the common bile duct and pancreatic duct as they enter the duodenum. When constricted, this sphincter prevents flow of bile and pancreatic juice into the duodenum and restricts reflux of duodenal contents back into the bile and pancreatic ducts.

submucosal plexus Network of nerves and small ganglia found in the submucosa of the intestine. It is composed of outer and inner layers and transmits secretomotor and vasodilator stimuli to the mucosa. Primary sensory nerves are contained in this plexus, which also communicates with the myenteric plexus.

Cholecystokinin (CCK), a peptide hormone, is produced and secreted by endocrine cells of the upper small intestine following food ingestion. CCK is the major hormone responsible for stimulating pancreatic enzyme secretion and gallbladder contraction. CCK promotes satiety, delays gastric emptying, potentiates insulin secretion, and may regulate bowel motility; CCK may also play a role in learning and memory, anxiety, analgesia, and thermoregulation.

INTRODUCTION

Cholecystokinin (CCK) was discovered by Ivy and Oldberg in 1928 when they recognized that intestinal extracts could stimulate gallbladder contraction in dogs; they named the substance cholecystokinin ["cholecyst" (gallbladder); "kinin" (to move)], the term now commonly used. In 1943, Harper and Raper noted that a similar extract, which they named "pancreozymin," stimulated pancreatic enzyme secretion. It was not until CCK was purified and its amino acid sequence determined by Mutt in 1968 that it was proved that CCK and pancreozymin were the same hormone, both possessing the ability to stimulate the gallbladder and pancreas.

Over the past three decades, CCK has been found to have many other biological effects. In experimental animals and human volunteers, CCK has been shown to delay gastric emptying, potentiate insulin secretion, and regulate bowel motility. One of the most noteworthy actions of CCK is its ability to induce satiety and reduce food intake. Until the development of reliable assays for measuring blood levels of CCK, the physiological effects of CCK remained controversial. However, it has now been shown in humans that physiological levels of CCK stimulate gallbladder contraction and pancreatic enzyme secretion, inhibit gastric emptying, potentiate insulin secretion, and reduce food intake. Although yet to be proved physiologically, CCK may regulate bowel motility and, in certain species, promote pancreatic growth. Less well-described but fascinating actions of CCK include effects on learning and memory, anxiety, analgesia, and thermoregulation. In the small intestine, CCK is produced by discrete endocrine cells within the mucosa. However, CCK is even more abundant in the brain and is found in peripheral nerves innervating the intestine, where it functions as a neurotransmitter.

MOLECULAR FORMS

The original CCK peptide isolated from the intestine was a tritriacontapeptide (CCK-33). Several larger and smaller molecular forms of CCK have since been found in the intestine, brain, and blood of animals and humans. The biologically active region of CCK resides in its carboxyl terminus and all forms of CCK possesses an identical carboxyl five-amino-acid sequence (-Gly-Trp-Asp-Met-Phe-NH_2) (Fig. 1). This region is

 Copyright 2003, Elsevier (USA). All rights reserved.

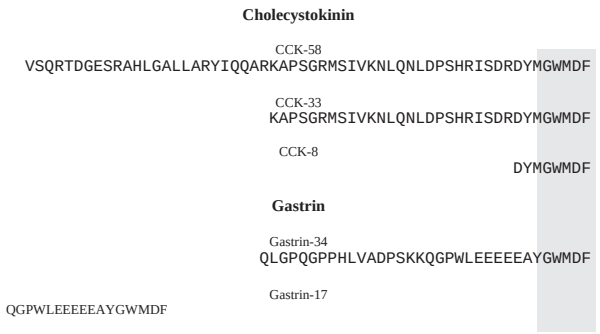


FIGURE 1 The amino acid sequences of the common molecular forms of CCK and gastrin. All biologically active forms of CCK and gastrin share an identical carboxyl-terminal pentapeptide sequence (shaded area).

common to gastrin, thus gastrin has some CCK-like activity (albeit weak) and CCK shares some weak gastrin-like activity. The amino acid sequence shared by the two hormones has made developing assays for CCK difficult because antibodies directed against the biologically active region of CCK may cross-react with gastrin. This problem is accentuated by the finding that circulating levels of gastrin are 10- to 100-fold greater than those of CCK.

CCK is produced from a single gene that encodes a 115-amino-acid preprohormone. By posttranslational processing, molecular forms of CCK ranging in size from 4 to 83 amino acids have been identified in tissues and blood. However, the major molecular forms of CCK are CCK-8, CCK-33, and CCK-58. In humans, the CCK gene is located in the 3q12–3pter region of chromosome 3. CCK expression is both tissue specific and developmentally regulated. In the intestine, the CCK gene is expressed prenatally and is regulated after birth primarily by ingestion of foods that stimulate CCK secretion. In the central nervous system, stimuli regulating neuronal CCK gene transcription include growth factors, second messengers such as cyclic AMP, the neurotransmitter dopamine, and hormones such as estrogen.

DISTRIBUTION

Cholecystokinin cells are individual flask-shaped cells that are scattered throughout the mucosa of the small intestine. The concentration of CCK cells is greatest in the proximal small intestine and diminishes in a gradient fashion toward the distal jejunum and ileum. CCK cells arise from progenitor cells in the intestinal crypts and, along with enterocytes, migrate up the villus. Residing in the mucosa, the apical surface of CCK cells is open to the lumen of the intestine. Here cells can actually "sample" luminal contents such as food and releasing factors. These enteroendocrine cells also possess microvilli that increase the exposed surface area, thus allowing greater exposure to potential stimuli.

Like other gastrointestinal hormones, CCK is a "brain–gut" peptide, meaning that the same transmitter is found in both the central nervous system and intestine. In the brain, CCK is highly concentrated in the striatum, hippocampus, and cerebral cortex. CCK-containing neurons may also synthesize dopamine. Such nerves have been shown to extend to the limbic forebrain and ventromedial hypothalamus, where they may participate in controlling food intake. CCK has been demonstrated to modulate dopamine release, dopamine-mediated reward, and receptor binding and function. These actions have implications for a role of CCK in drug abuse and neurologic and psychiatric disease.

CCK is prevalent in peripheral nerves of the gastrointestinal tract. It is most abundant in nerves innervating the colon, with fewer CCK nerves in the ileum. In these locations, CCK is present in myenteric and submucosal plexi, where it innervates ganglionic bodies. CCK is also abundant in the vagus nerve and celiac plexus. In the intestine, CCK stimulates acetylcholine release and causes smooth muscle contraction. Postganglionic CCK-containing neurons also terminate around the islets of Langerhans, where CCK may stimulate islet hormone secretion (e.g., insulin and glucagon release).

CCK RECEPTORS

CCK exerts its biological actions by binding to specific receptors on its target tissues. Gastrointestinal CCK receptors reside on tissues of the pancreas, gallbladder, stomach, lower esophageal sphincter, ileum, and colon. In the nervous system, CCK receptors are abundant in brain and on some peripheral nerves. Two types of CCK receptors have been identified: CCK alimentary (CCK-A) receptors are the primary CCK receptor and mediate most of the effects of CCK in the gastrointestinal tract. The CCK brain (CCK-B) receptor is identical to the gastrin receptor and is the major CCK receptor subtype in the nervous system. It is also abundant in the stomach. Both receptors are G-protein-coupled, seven-membrane-spanning proteins, but arise from different genes. CCK receptor antagonists have been extremely useful in pharmacological and physiological studies to define the physiological role of CCK. The first CCK receptor antagonists useful for *in vitro* studies were cyclic nucleotide analogues (e.g., dibutyryl cyclic guanosine monophosphate). Subsequently, amino acid derivatives (e.g., CR-1409), carboxyl-terminal

CCK analogues, and substituted benzodiazepines (e.g., devazepide) were developed and would be used *in vivo*. Clinical studies have shown CCK receptor antagonists to inhibit CCK- and meal-stimulated gallbladder contraction, accelerate gastric emptying, and induce hunger (i.e., reverse satiety), indicating that CCK has important physiological roles in each of these processes.

CHOLECYSTOKININ RELEASE

CCK is secreted from specialized endocrine cells of the mucosa (known as I cells) into the extracellular space, where it is taken up into the bloodstream. It is by this mechanism that circulating CCK reaches distant target tissues such as the pancreas and gallbladder. Enteric endocrine cells package CCK in secretory granules that are stored along the basolateral surface of the cell, thus allowing CCK to be secreted into the interstitium when the cell is stimulated. *In vivo*, ingested proteins and fat are the major dietary stimulants of CCK release. However, in some species, including humans, and in cell preparations *in vitro*, partially digested nutrients such as amino acids and peptides and fatty acids are potent releasers of CCK, indicating that these components directly interact with the CCK cell.

Circulating blood levels of CCK average approximately 1 p*M* in the fasting state and increase to between 5 and 8 p*M* after eating. Postprandial levels remain elevated for 3–5 hours as food empties from the stomach into the upper small intestine. Gastric distension does not influence CCK release. Although fat and protein are the primary stimulants of CCK, carbohydrate has a modest effect on secretion.

It is well recognized that inactivation of protease activity in the lumen of the small intestine of rodents stimulates CCK release and pancreatic exocrine secretion. This phenomenon is now known as negative feedback control of CCK release. Not only has this principle now been demonstrated in rodents, but it also applies to other species, including humans. Thus, CCK release is controlled in part by the presence or absence of pancreatic enzymes in the intestine. (Fig. 2). This concept indicates that there exist intestinal releasing factors that are secreted into the intestine and stimulate CCK secretion. When pancreatic enzymes are present in the intestine, these CCK releasing factors are inactivated; however, when pancreatic enzymes are inhibited or pancreatic secretion is reduced, these CCK releasing factors are active and stimulate CCK secretion. Similarly, with ingestion of a meal that temporarily binds trypsin and other digestive enzymes, CCK releasing factors are also available to stimulate CCK secretion. To date, a human counterpart of a CCK-specific releasing factor has not been identified.

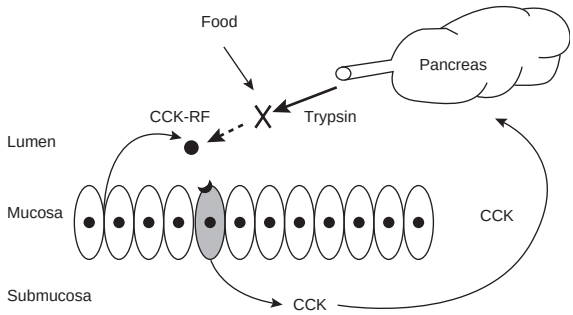


FIGURE 2 Model for regulation of CCK release by an intraluminal releasing factor. One or more CCK releasing factors (CCK-RF) stimulate intestinal CCK secretion. CCK-RF is normally secreted in the intestinal lumen, where it is exposed to pancreatic enzymes. Under basal conditions, CCK-RF is inactivated by even small amounts of enzyme; following a meal, however, food competes for enzyme binding, allowing CCK-RF to interact with CCK cells to stimulate hormone secretion. CCK in the circulation stimulates pancreatic secretion, which, following digestion of food, restores CCK-RF and CCK secretion to basal levels. Modified from Liddle (1995).

BIOLOGICAL ACTIONS OF CCK

CCK is the major hormonal regulator of gallbladder contraction. Coincident with this effect, CCK also relaxes the sphincter of Oddi, which also promotes bile secretion into the intestine. In humans, the predominant CCK receptor type in the pancreas is CCK-B. In most species, however, CCK-A receptors predominate in pancreas, and CCK is a potent stimulant of pancreatic secretion. Therefore, in humans, although CCK stimulates pancreatic secretion, its role may be limited.

At physiological blood concentrations that occur after a meal, CCK delays gastric emptying; this may be important for its ability to reduce food intake and induce satiety. Due to its effects on gallbladder contraction, pancreatic secretion, and gastric emptying, CCK coordinates many digestive processes. Thus, CCK plays an important role in the ingestion and digestion of a meal. Although it has been shown that CCK causes relaxation of the lower esophageal sphincter and promotes intestinal motility, it appears that these effects are neural rather than hormonal. CCK receptors have been found on some gastrointestinal and lung cancers, but it remains unknown whether CCK plays a role in human cancer growth.

CLINICAL USES OF CCK

CCK is used along with secretin as a test of pancreatic function. In patients with pancreatic insufficiency, low levels of pancreatic juice are recovered following intravenous injection of these hormones. CCK can also be used clinically to stimulate gallbladder contraction and is helpful in radiographic testing of gallbladder function. For diagnostic purposes, CCK has facilitated the collection of bile and pancreatic juices for cytological examination.

CCK injections have been administered therapeutically to patients who are unable to eat (e.g., parenteral alimentation) in order to stimulate gallbladder contraction. This therapy has been effective in reducing gallbladder sludge and in preventing gallstone formation.

Low blood levels of CCK have been reported in patients with celiac disease, bulimia nervosa, and in conditions that delay gastric emptying. The defect in celiac disease is likely due to reduced CCK secretion from diseased small intestinal mucosa. The cause of abnormal CCK responses in bulimia is unknown, but may be related to alterations in gastric emptying because normal postprandial CCK release is dependent on delivery of food from the stomach to the small intestine. It is not known whether CCK deficiency contributes to any specific pathological consequences. There are no known diseases of cholecystokinin excess.

See Also the Following Articles

Gastric Acid Secretion • Gastric Emptying • Gastrin • Pancreatic Enzyme Secretion (Physiology) • Pancreatic Function Tests

Further Reading

Calam, J., Ellis, A., and Dockray, G. J. (1982). Identification and measurement of molecular variants of cholecystokinin in duodenal mucosa and plasma. *J. Clin. Invest.* **69**, 218–225.

Doty, J. E., Pitt, H. A., Porter-Fink, V., and Denbesten, L. (1985). Cholecystokinin prophylaxis of parenteral nutrition-induced gallbladder disease. *Ann. Surg.* **201**, 76–80.

Gibbs, J., Young, R. C., and Smith, G. P. (1973). Cholecystokinin decreases food intake in rats. *J. Comp. Physiol. Psychol.* **84**, 488–495.

Hansen, T. V. O. (2001). Cholecystokinin gene transcription: Promoter elements, transcription factors and signaling pathways. *Peptides* **22**, 1201–1211.

Harper, A. A., and Raper, H. S. (1943). Pancreozymin, a stimulant of secretion of pancreatic enzymes in extracts of the small intestine. *J. Physiol.* **102**, 115–125.

Ivy, A. C., and Oldberg, E. (1928). A hormone mechanism for gallbladder contraction and evacuation. *Am. J. Physiol.* **65**, 599–613.

Jorpes, E., and Mutt, V. (1966). Cholecystokinin and pancreozymin, one single hormone? *Acta Physiol. Scand.* **66**, 196–202.

Konturek, S. J., Kwiecien, N., Obtulowicz, W., Kopp, B., Oleksy, J., and Rovati, L. (1990). Cholecystokinin in the inhibition of gastric secretion and gastric emptying in humans. *Digestion* **45**, 1–8.

Liddle, R. A. (1994). Cholecystokinin. *In* "Gut Peptides: Biochemistry and Physiology" (J. H. Walsh and G. J. Dockray, eds.), pp. 175–216. Raven Press, New York.

Liddle, R. A. (1995). Regulation of cholecystokinin secretion by intraluminal releasing factors. *Am. J. Physiol.* **269**, G319–G327.

Liddle, R. A., Goldfine, I. D., Rosen, M. S., Taplitz, R. A., and Williams, J. A. (1985). Cholecystokinin bioactivity in human plasma: Molecular forms, responses to feeding, and relationship to gallbladder contraction. *J. Clin. Invest.* **75**, 1144–1152.

Lund, T., GeurtsvanKessel, A. H. M., Haun, S., and Dixon, J. E. (1986). The genes for human gastrin and cholecystokinin are located on different chromosomes. *Hum. Genet.* **73**, 77–80.

Rozengurt, E., and Walsh, J. H. (2001). Gastrin, CCK, signaling, and cancer *Annu. Rev. Physiol.* **63**, 49–76.

Spannagel, A. W., Green, G. M., Guan, D., Liddle, R. A., Faull, K., and Reeve, J. R. Jr. (1996). Purification and characterization of a luminal cholecystokinin-releasing factor from rat intestinal secretion. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4415–4420.

Szelenyi, A. (2001). Cholecystokinin and thermoregulation—A minireview. *Peptides* **22**, 1245–1250.

Cholelithiasis, Complications of

ARSHAD H. MALIK* AND PETER F. MALET†

University of Arkansas for Medical Sciences, Little Rock, and †University of Texas Southwestern Medical Center, Dallas

acute cholecystitis Acute inflammation of the gallbladder.
biliary pain Symptom due to transient obstruction of the biliary tract, including the gallbladder.
cholangitis Inflammation and infection of the biliary ductal system due to obstruction.
choledocholithiasis Stones in the common bile duct.
gallstone pancreatitis Inflammation due to a gallstone obstructing pancreatic outflow.
gallstones (cholelithiasis) Precipitates (stones or pebbles) within the gallbladder.
ileus (gallstone ileus) Bowel obstruction due to a large stone.
laparoscopic cholecystectomy Removal of the gallbladder via a laparoscopic approach.
microlithiasis (sludge) Microscopic precipitates in bile that can be visualized with ultrasonography or bile microscopy.
Mirizzi syndrome Obstruction of the common hepatic duct due to extrinsic pressure by a stone in the cystic duct or the gallbladder neck.

Gallstones, which occur commonly and usually remain asymptomatic, can result in a variety of symptoms and presentations ranging from mild to extremely severe. The most common symptom from gallstones is abdominal pain, termed biliary pain, which generally occurs in the right upper quadrant of the abdomen. More serious complications include acute cholecystitis, biliary duct obstruction with or without acute cholangitis, and acute pancreatitis. More unusual problems caused by gallstones include gallstone ileus, enteric fistulae, and Mirizzi syndrome.

DIAGNOSIS AND COMPLICATIONS

Ultrasonography is the best method for diagnosing gallbladder stones; its sensitivity in detecting gallbladder stones is about 95%. Although biliary duct stones may sometimes be demonstrated by ultrasonography (about 30–50% sensitivity), these stones are more often diagnosed by the duct obstruction and subsequent duct dilation that they cause. Cholangiography, usually with endoscopic retrograde cholangiopancreatography (ERCP), can directly image ductal obstruction due to stones. ERCP is >90% sensitive and offers the potential for therapeutic intervention. Magnetic resonance cholangiopancreatography (MRCP), however, is being used more often to diagnose bile duct stones.

In selected situations only, percutaneous transhepatic cholangiography (PTC) or intraoperative cholangiography performed during cholecystectomy can be performed to look for ductal stones. Endoscopic ultrasonography (EUS) can visualize ductal stones and sludge or stones in the gallbladder, but is not often used for this purpose. Computer tomography (CT) scanning can often provide more extensive information than ultrasonography about masses and abscesses, but its sensitivity in detecting gallstones is significantly lower.

NATURAL HISTORY

It is now known that a majority of gallstones cause no symptoms after formation. Surveys of large populations using ultrasonography have shown that from 60 to 80% of persons with gallstones are asymptomatic. Only a small proportion of patients with gallstones develop symptoms on a yearly basis on long-term followup. The incidence rate at which asymptomatic patients develop biliary pain is approximately 1–2% annually. The risk of first presenting with a complication rather than just biliary pain is very low. Because of this low rate of development of symptoms, the consensus is that asymptomatic patients with gallstones should not undergo cholecystectomy because many or even most of them will never develop symptoms.

Asymptomatic patients with gallstones may be followed expectantly with surgery or other interventions reserved for those who eventually develop biliary pain. There is still some variation to this approach based on individual physician and patient preference. There is one condition in which there is still uniform agreement that cholecystectomy should be performed as soon after diagnosis as possible, and that is porcelain gallbladder (calcification in the gallbladder wall) because of its risk of developing malignancy, that is, gallbladder cancer.

 Copyright 2004, Elsevier (USA). All rights reserved.

BILIARY PAIN

The most common complication of gallstones is episodic upper abdominal pain. The term "biliary colic" is still commonly used to describe pain due to gallstones. However, the term "biliary pain" is preferred because the pain due to gallstones is not a true colic, that is, of a waxing and waning nature like intestinal colic. Biliary pain can have somewhat varied features among patients, but the hallmark of true biliary pain is its episodic nature.

The term "chronic cholecystitis" is still sometimes used to describe the condition in which a patient has experienced repeated episodes of biliary pain. Strictly speaking, chronic cholecystitis describes the histologic changes in the gallbladder. There is frequently an inexact correlation between the pathologic changes in the gallbladder and the symptoms the patient has experienced.

Etiology and Pathogenesis

Biliary pain is thought to arise from transient obstruction of the cystic duct by stones or sludge. It may also develop due to microlithiasis or biliary sludge, presumably due to similar mechanisms. Sometimes, patients experience what appears to be biliary pain with no demonstrable gallstones or sludge. The explanation for this is not completely known; in some cases the pain may result from disordered biliary motility at the gallbladder or sphincter of Oddi level, but in other cases, no explanation is found.

Clinical Features of Biliary Pain

Biliary pain is typically characterized by its location in the right upper quadrant or epigastrium and may radiate around to the right subscapular area or midback or occasionally to the right shoulder. Sometimes, the pain may be felt in the middle lower chest or even in the left abdomen. The pain may range from fairly mild to very severe, and may be described, for example, as feeling like a pressure, stabbing, heavy weight, cramping, and so forth. Sometimes, the pain is compared to a toothache or childbirth. Patients often state that they are restless or cannot get comfortable during an attack of biliary pain and often walk around the room or shift position around on the couch or bed waiting for the pain to end. Patients with milder discomfort may only describe it as a pressure or a heavy feeling or a localized bloating and not a "real" pain. Some patients may experience nausea during an episode, but vomiting is uncommon. There are no systemic signs of toxicity such as fever or chills.

Biliary pain usually has a fairly clear onset with no antecedent warning. The pain then reaches a plateau of intensity ranging from quite mild to extremely severe. It may last from 15 to 30 minutes up to typically 3 to 4 hours. Sometimes, the duration may be 6 to 8 hours or more, but over 12 hours is unusual unless acute cholecystitis is developing. After the pain abates, the patient may feel a residual abdominal soreness for a while. In some patients, what appears to be otherwise typical biliary pain may last for 15 minutes or less. Some patients even describe episodes lasting for a few minutes; it is difficult to know whether these brief episodes represent true biliary pain.

Biliary pain is typically episodic in occurrence; the interval between episodes varies from daily to once every few weeks or months. Some patients have only one episode every year or so. The clearly episodic nature of most biliary pain is why it is often described in terms of "attacks" or "bouts." Once a patient experiences the first episode of biliary pain, it is highly likely that subsequent episodes will occur. Approximately 75% of such patients will have at least one further attack within a 2-year period.

Differential Diagnosis of Biliary Pain

In some cases, biliary pain caused by gallbladder stones may not be typical. The pain may vary in one or more features, such as its location, frequency, duration, or character. Because there is no "gold standard" for confirming biliary pain, it is not uncommon to have difficulty in deciding whether the discomfort patients are experiencing is due to gallstones. There are many conditions that need to be considered in the differential diagnosis of right upper quadrant abdominal pain, including: peptic ulcer disease, pancreatitis, gastroesophageal reflux, angina, bowel obstruction, liver disease, rib pain, and irritable bowel syndrome. In such situations, clinical judgment, further diagnostic testing, followup over time, and possible treatment for gastroesophageal reflux or irritable bowel syndrome usually help elucidate the clinical picture. When the relationship between stones and symptoms is not entirely clear, decisions regarding cholecystectomy need to be carefully considered based on the patient's wishes and on surgical consultation.

It has been fairly convincingly shown that symptoms generally categorized as dyspepsia (intolerance to fatty foods, flatulence, bloating, belching, and heartburn) cannot be attributed to gallstones. Cholecystectomy does not have a reliably beneficial effect on such symptoms. The term "postcholecystectomy syndrome" has been used to describe symptoms that remain after

cholecystectomy; most often, such residual symptoms were not due to the stones in the first place.

Management

During an actual episode of biliary pain, analgesics may be used for pain control. Cholecystectomy is the definitive treatment of choice for symptomatic gallstones. Once a patient experiences the first episode of biliary pain, it is very likely that future episodes will occur and it is for symptomatic relief that cholecystectomy is recommended for such patients. Laparoscopic cholecystectomy is usually performed; occasionally, the traditional open approach is used either as a conversion from the laparoscopic approach or as a planned procedure. Cholecystectomy removes the stones and the gallbladder and prevents future episodes of pain.

Oral bile salt therapy with ursodiol is an option that is rarely used now. Extracorporeal shock-wave lithotripsy combined with ursodiol is utilized occasionally outside the United States.

Biliary Pain and the Risk of Developing More Serious Complications

Patients with symptomatic gallstones (that is, patients who have experienced biliary pain at least once) are at risk for developing a major complication, such as acute cholecystitis, cholangitis, or pancreatitis, at any time. The yearly rates at which these complications develop have not been well studied because most symptomatic patients with gallstones undergo definitive treatment and do not enter long-term studies in which no therapy is given. But from the available data, it appears that the yearly risk of developing a serious complication is in the range of 1–3%.

A patient with previously asymptomatic gallstones may present for the first time not just with simple biliary pain, but with a serious complication. However, the annual incidence rate of this type of presentation, although not known with certainty, is much lower than that for with patients with symptomatic gallstones.

ACUTE CHOLECYSTITIS

Acute cholecystitis is the most common acute complication of gallbladder stones. Usually, the diagnosis is straightforward based on a typical clinical scenario and confirmatory radiologic studies. But in some cases, the diagnosis is more difficult because of an atypical clinical presentation. A high index of suspicion for acute cholecystitis is warranted in any patient with upper abdominal pain.

Etiology and Pathogenesis

Cystic duct obstruction is the precipitating event that results in inflammation of the gallbladder by cytokines and other mediators (lysolecithin, prostaglandins) of inflammation. The obstruction is usually caused by a stone, but mucus, sludge, and viscous bile may also play a role. In addition, the normally protective mucus layer over the luminal surface is disrupted, allowing relatively toxic bile salts to come in contact with the mucosa. Secondary bacterial infection may later ensue.

Clinical Features

The majority of patients present with at least moderate pain in right upper quadrant; the pain may radiate to the right shoulder or scapula. The pain usually has been present for at least several hours before the patient presents. Fever may be present, although temperature is usually not above 102°F; higher temperatures suggest bacteremia or an abscess. There is right upper quadrant tenderness and Murphy's sign (accentuated tenderness to palpation during inspiration) is often present. Clinical presentation may vary, however, and can range from deceptively mild to quite severe with sepsis. There may be minimal pain and minimal tenderness, particularly in the elderly or an obtunded patient. Other patients may present in a very toxic manner with high fever, severe abdominal pain and tenderness, bacteremia, and marked leukocytosis, sometimes in shock.

Laboratory studies show a white blood cell (WBC) count of usually 10,000–15,000/ml. Aspartate aminotransferase (AST), alanine aminotransferase (ALT), and alkaline phosphatase as well as bilirubin may be normal or just slightly elevated. If the alkaline phosphatase is disproportionately elevated relative to the transaminases, choledocholithiasis should be considered.

With more extensive complications, such as empyema, gangrene, localized or free perforation, peritonitis, or emphysematous cholecystitis (air in the gallbladder wall), the clinical scenario is more severe and represents more of a surgical emergency. Although unusual, the combination of acute cholecystitis and either cholangitis or pancreatitis may occur, thus making the clinical picture more complex.

Diagnosis

The diagnosis of acute cholecystitis is made primarily based on typical clinical findings and secondarily based on radiologic studies. Either ultrasonography or radionuclide hepatobiliary scanning with iminodiacetic acid (HIDA) derivatives may be used. Ultrasonography

is preferred because it is easier to perform and provides additional information regarding the bile ducts and other organs. The typical findings seen are gallstones, dilated gallbladder, thickened gallbladder wall, edema within the gallbladder wall, or pericholecystic fluid; sludge may also be seen. The sensitivity of ultrasonography for acute cholecystitis is approximately 90–95% and specificity is about 80%.

Management

It is common practice to administer antibiotics to patients with acute cholecystitis. Coverage for enterococci and gram-negative aerobic organisms is usually enough. For patients who are extremely toxic, coverage for anaerobic organisms should be added. Patients should be kept at a "nothing per oral" (NPO) status; intravenous fluids should be started and mild analgesia may be necessary for pain.

Cholecystectomy is the treatment for acute cholecystitis and may be performed as soon as the diagnosis is made and the patient is stable enough. Generally, this is within the first 2 to 3 days after admission. In the great majority of cases, laparoscopic cholecystectomy is possible. In those cases in which a complication, such as an abscess, gangrenous wall, or perforation, is suspected or is encountered, an open approach is needed. If the patient has a seriously decompensated medical condition, consideration may be given to either surgical or radiologic-guided cholecystostomy as a temporizing measure.

Acute Acalculous Cholecystitis

Acute cholecystitis may occur without the presence of gallstones in the gallbladder, in which case it is termed acute acalculous cholecystitis (AAC). AAC accounts for $<5\%$ of all cases of acute cholecystitis and is most frequently encountered in patients who are hospitalized and frequently already seriously ill. Cystic duct occlusion is thought to be the principal pathophysiologic event in the development of AAC. The occlusion may be due to sludge or microlithiasis, viscous bile, or mucus. Gallbladder stasis is a major factor in many cases and gallbladder ischemia is also considered to play a significant role. Patients with AIDS are a special subgroup because infectious causes for cholecystitis, especially viral or fungal, may be seen.

The diagnostic approach is similar to that for calculous cholecystitis but is complicated by the fact that many patients with AAC are already seriously ill and may have atypical signs and symptoms. Treatment is with open cholecystectomy, which should be performed expeditiously. In some cases, especially if the patient is septic, an ultrasound-guided percutaneous puncture of the gallbladder can be performed and a catheter can be left in place for drainage. AAC has a high mortality, especially if treatment is delayed, because gangrene, perforation, and abscess formation are more frequent.

CHOLEDOCHOLITHIASIS AND OBSTRUCTIVE JAUNDICE

Choledocholithiasis refers to gallstones in the common bile duct and is present in approximately 10% of patients with cholelithiasis. These stones can cause partial or complete obstruction of the biliary system, causing pain and ductal dilation, obstructive jaundice, favorable conditions for acute bacterial cholangitis, and acute gallstone pancreatitis.

Clinical Features

Clinical manifestations suggestive of choledocholithiasis include biliary pain, nonspecific abdominal pain, obstructive jaundice, cholangitis, or acute biliary pancreatitis. Individuals may also harbor asymptomatic stones that are diagnosed serendipitously. Physical examination may reveal icteric discoloration of the skin and visible mucous membranes and abdominal tenderness in the right upper quadrant on palpation.

Diagnosis

Laboratory Tests

Choledocholithiasis cannot reliably be diagnosed on the basis of the patient's history and physical examination alone. Almost all patients will have an elevated alkaline phosphatase and γ-glutamyl transpeptidase (GGT) level. Bilirubin and aminotransferase levels are elevated in 70–90% of patients at the onset of symptoms. However, normal liver chemistries do not completely exclude the possibility of choledocholithiasis. In patients with biliary obstruction, alkaline phosphatase rises faster and often precedes the elevation of serum level of bilirubin.

Radiologic Imaging

Transcutaneous ultrasound has been the traditional noninvasive test of choice for diagnosing choledocholithiasis. It is very specific but has low sensitivity, thus a negative ultrasound result should be interpreted with caution. Dilation of extra- or intrahepatic ducts can be a useful indirect sign of intraductal stones. In cases with high likelihood of ductal obstruction with stones, ERCP should be the next step after ultrasonography and offers the potential for therapeutic intervention.

However, there is a 3–6% complication rate associated with therapeutic ERCP for stone removal. EUS is very sensitive and specific and, in cases of unsuccessful cannulation of the major duodenal papilla, can be a valuable diagnostic alternative to ERCP. MRCP can also be an alternative to ERCP to screen for possible choledocholithiasis in high-risk patients with complicated medical problems. The reported sensitivity, specificity, and overall accuracy of MRCP is 91.6, 100, and 96.8%, respectively.

Therapy

Patients with common bile duct (CBD) stones complicated by obstructive jaundice or cholangitis need urgent ERCP. Uncomplicated choledocholithiasis requires elective ERCP. For large stones, mechanical or laser lithotripsy may be required. The overall therapeutic success rate of endoscopic treatment is 80–95%. If endoscopic therapy fails, laparoscopic or open surgical CBD exploration should be considered.

ACUTE CHOLANGITIS

Cholangitis is a clinical syndrome consisting of stasis in and infection of the biliary ductal system; gallstones are the cause in 80% of cases. Acute cholangitis can be a serious and life-threatening illness and requires prompt recognition and treatment.

Etiology and Pathogenesis

The most common cause of acute cholangitis in the United States is choledocholithiasis. Bile is usually sterile but biliary obstruction disrupts the host defense mechanisms, allowing infecting microbes to reach the biliary tree, typically by direct spread from the duodenum into the bile duct or hematogenous spread by translocation of organisms through the bowel wall into the portal venous system. The presence of a foreign body, such as a stone or stent, can then act as a nidus for bacterial colonization. Once bacteria colonize bile, biliary stasis favors multiplication. High pressure promotes the migration of bacteria from bile into the systemic circulation, resulting in septicemia. Bile cultures are positive in over 90% of cases, yielding a mixed growth of gram-negative and gram-positive bacteria. *Escherichia coli* is the major gram-negative bacterium isolated (25–50%), followed by *Klebsiella* (15–20%) and *Enterobacter* species (5–10%). The most common gram-positive bacteria are *Enterococcus* species (10–20%). *Pseudomonas* may be found after interventional endoscopy or a surgical procedure. Anaerobes, such as *Bacteroides* and *Clostridia*, are usually present in a mixed infection, but are rarely the sole infecting organisms.

Clinical Features

The hallmark of cholangitis is the classic "Charcot's triad" of fever, right upper quadrant pain, and jaundice, but this is fully present in only 50–75% of cases. Later manifestations include mental obtundation and hypotension, producing "Reynold's pentad," which is associated with significant morbidity and mortality. Hypotension may be the only presenting symptom in elderly patients, and septic shock in severe cases can lead to multiorgan failure.

Diagnosis

Diagnosis can usually be made on clinical grounds alone, with laboratory tests and imaging studies playing a confirmatory role.

Laboratory Tests

Routine laboratory tests typically reveal an elevated WBC count with neutrophil predominance, and a cholestatic pattern of liver function test abnormalities with elevations in serum alkaline phosphatase, GGT, and bilirubin. AST and ALT levels are usually mildly elevated, although a pattern of acute hepatocyte necrosis can occasionally be seen. Mildly increased serum amylase occurs in up to one-third of patients, with 10% also having clinical acute pancreatitis.

Radiologic Imaging

Ultrasonography is usually the first imaging study in patients suspected of having cholangitis to look for CBD dilatation and presence of stones. Ultrasonography should usually be followed by ERCP both to confirm the diagnosis and to intervene therapeutically with sphincterotomy, stone extraction, or stent insertion as needed. If ERCP is not possible, PTC can serve to drain the obstructed biliary system.

Management

General Support

The initial treatment usually includes stopping all oral intake and intravenous fluids, frequent monitoring of vital signs for evidence of sepsis and multiorgan failure, administering parenteral vitamin K to counteract any coagulopathy due to hepatic dysfunction, and treating with antibiotics. Pressor agents may be indicated in severely ill patients. The mainstays of therapy are antibiotics and establishment of biliary drainage.

Antibiotics

Antibiotics for the treatment of acute cholangitis should be given for 7–10 days. A broad-spectrum antibiotic regimen is required (e.g., ampicillin and gentamicin, third-generation cephalosporins, imipenem, or ciprofloxacin). In sick patients, metronidazole is often added to cover anaerobes.

Biliary Drainage

Eighty percent of patients with acute cholangitis will respond to conservative management and antibiotic therapy. Biliary drainage can then be performed on an elective basis. In 15–20% of cases, cholangitis fails to improve over the first 24 hours with conservative therapy alone, requiring urgent biliary decompression. Endoscopic sphincterotomy with stone extraction and/or stent insertion is now the treatment of choice for establishing biliary drainage in acute cholangitis.

Role of Surgery

Emergency surgery for acute cholangitis has largely been replaced by nonoperative biliary drainage. Elective surgery carries a very low morbidity and mortality risk as opposed to emergency surgery. If emergent surgery is needed due to failure of a nonsurgical drainage procedure, choledochotomy with placement of a large-bore T tube carries a lower mortality risk compared to cholecystectomy with CBD exploration.

The rendezvous procedure combines the endoscopic technique with percutaneous cholangiography. It is used when endoscopic cannulation of the papilla is unsuccessful and surgery is indicated, but the risk associated with operation is high. The combined technique increases the success rate of biliary tract cannulation and thus facilitates the diagnosis and treatment of biliary tract disorders.

Prognosis

Reported mortality rates for acute cholangitis vary from 13 to 88%. With effective antibiotics and biliary drainage, the prognosis for mild to moderate cholangitis is much improved. However, the mortality rate remains very high (approximately 50%) for patients with severe cholangitis (Reynold's pentad).

GALLSTONE PANCREATITIS

Gallstone pancreatitis is an episode of acute pancreatitis caused by a gallstone. Gallstones are a common cause of acute pancreatitis, being responsible for 30–75% of all cases. The presence of gallstones increases the relative risk for pancreatitis 12- to 35-fold. Gallstone pancreatitis may sometimes be associated with cholangitis.

Etiology and Pathogenesis

It is now accepted that gallstone-associated pancreatitis results from the passage of stones through the sphincter of Oddi into the duodenum. Many cases of presumed idiopathic pancreatitis are caused by gallstones too small to be visualized. Recent advances in ultrasonography and the wider use of ERCP for diagnosis have demonstrated that in cases that previously would have been classified as "idiopathic," small microcalculi are indeed present in up to 75% of such cases. The pathway by which ductal obstruction by the passage of a stone leads to pancreatitis is still unclear. It is believed that changes of acute pancreatitis are most likely related to an injury from the digestive enzymes secreted from the pancreas; a series of chemical events leads to premature activation of inactive digestive enzymes that are usually activated in the duodenum.

Clinical Features

Patients with acute gallstone pancreatitis present with complaints of severe diffuse epigastric or right upper quadrant abdominal pain that typically radiates to the back. This is usually constant and severe and may last many hours. Severe nausea and vomiting may occur. Patients with severe biliary pancreatitis appear ill with signs of systemic toxicity (tachycardia, hypotension, fever, and tachypnea). Skin and visible mucous membranes may be icteric. Palpation of the abdomen usually shows severe tenderness in the epigastric area and the right upper quadrant, often with guarding and rebound tenderness.

Diagnosis

Laboratory Tests

Blood tests for amylase and lipase are sensitive but not specific for acute pancreatitis. Serum amylase is less specific than lipase. The use of a cutoff level three times the upper limit of normal serum value increases the specificity of amylase and lipase in the diagnosis of acute pancreatitis. The WBC count is usually elevated in patients with acute gallstone pancreatitis. Threefold elevations of serum ALT and AST are highly specific but not very sensitive (only about 50%) for gallstone pancreatitis. Total bilirubin and alkaline phosphatase are elevated in patients with gallstone pancreatitis and obstruction of the extrahepatic bile ducts.

Imaging

Transabdominal ultrasound is recommended for all patients with suspected biliary pancreatitis. Dilation of the biliary and pancreatic duct can be an indirect ultrasonic sign of a stone impacted in the major duodenal papilla. The CT scan is currently the best test available to evaluate the size and structure of the pancreas (edema, necrosis, phlegmon, and development of pseudocyst), the condition of the gallbladder and biliary ducts, and potential intraabdominal complications (intraabdominal or intrapancreatic abscess, retroperitoneal fluid collection, and pancreatic ascites).

ERCP is the best technique to visualize the pancreatic and biliary ducts and can be complemented by therapeutic manipulations to relieve the obstruction and to extract the offending stone(s). ERCP does carry an associated morbidity and should be performed in patients with suspected impacted stones in the distal CBD or duodenal papilla only if therapeutic intervention is envisaged.

Management

General clinical management is the same as for pancreatitis of any other cause. In severe cases and those in which there is any evidence of biliary obstruction, the use of prophylactic antibiotics and the early application of ERCP are warranted.

General Measures

General supportive measures include intravenous fluids, bowel rest, histamine-2 (H2) receptor antagonists or proton pump inhibitors, parenteral analgesia, antiemetic medications, and parenteral nutrition. Systemic antibiotics reduce septic complications and mortality in severe cases. Development of sepsis or other systemic complications requires admission to the intensive care unit. A majority of patients have a relatively mild illness that clinically resolves within 1 week with this conservative management strategy.

The modified Glasgow or acute physiology and chronic health evaluation (APACHE II) scoring systems can be used to predict which patients are likely to follow a severe disease course, to develop systemic complications, and to benefit from more aggressive interventional therapies.

Biliary Decompression

Bile duct decompression is beneficial in patients with predicted severe pancreatitis. The currently recommended treatment for patients with gallstone-associated pancreatitis and either jaundice or cholangitis includes ERCP with endoscopic sphincterotomy (ES) within the first 2–3 days. Three of four published randomized trials have shown a significant reduction in local and systemic complications of severe gallstone-associated pancreatitis by the use of urgent ERCP plus ES. Emergent ERCP is not indicated in patients with predicted mild or moderate gallstone pancreatitis.

On the basis of available information, ERCP is recommended within the first 2 to 3 days of hospitalization for patients with gallstone pancreatitis who have evidence of biliary sepsis or organ failure. Stones that are found in the common bile duct should be removed endoscopically.

Role of Surgery

Elective cholecystectomy is recommended for patients with cholelithiasis after gallstone pancreatitis is resolved to prevent future relapses of pancreatitis. It is advisable for cholecystectomy to be performed during the same hospital admission. Cholangiography should be performed during the cholecystectomy to exclude residual common bile duct stones. The one exception involves patients with severe gallstone pancreatitis who have a prolonged clinical course. Several weeks are allowed to elapse between the hospitalization for acute pancreatitis and readmission for laparoscopic cholecystectomy.

MISCELLANEOUS COMPLICATIONS

Calcification in the gallbladder wall (termed porcelain gallbladder) may develop over many years in gallbladders that contain stones. Gallbladder carcinoma is found in a significant proportion of gallbladders with wall calcification. Prophylactic cholecystectomy is recommended for patients with porcelain gallbladder.

Gallbladder cancer may also be found in stone-containing gallbladders without wall calcification. It does not occur at a high enough frequency to warrant prophylactic cholecystectomy. Mirizzi syndrome is the obstruction of the common hepatic duct as a result of extrinsic pressure by a gallstone in the cystic duct or the gallbladder neck. The clinical scenario is usually that of acute cholangitis. Fistulae between the gallbladder and other organs may develop as a result of gallstone erosion through the gallbladder wall. The most common sites of fistulization are into the duodenum or colon, but other sites may be involved. Occasionally, a fistula may develop between the common bile duct and another organ, usually the duodenum. Cholecysto- or choledochoenteric fistulae may or may not require specific intervention, depending on the patient's symptoms.

Gallstone ileus is a mechanical bowel obstruction resulting from a stone, typically fairly large, eroding into

usually the duodenum and becoming impacted at the ileocecal valve or some other strictured area in the gastrointestinal tract. Gallstone ileus is not rare and most often affects older patients. The diagnosis may be suspected based on the finding of air in the biliary tract in the setting of a bowel obstruction.

SUMMARY

Most gallbladder stones are asymptomatic. When they do become symptomatic, biliary pain is the most common manifestation. The hallmarks of biliary pain are its episodicity and location in the upper abdomen, usually in the right upper quadrant. Because gallstones are fairly common, other conditions may coexist with gallstones and may actually be responsible for the symptoms/signs initially attributed to the stones. The treatment of choice for symptomatic gallbladder stones is laparoscopic cholecystectomy; when this approach is not feasible, open cholecystectomy is the alternative.

Acute cholecystitis is the most common acute syndrome that gallstones may cause. Although cholecystectomy is the treatment of choice, there are often difficult diagnostic and management decisions to be made and consultation between internist/gastroenterologist, surgeon, and radiologist is frequently warranted in order to arrive at the most efficient plan for care. Acute acalculous cholecystitis requires a high index of suspicion for diagnosis; patients are usually quite ill and rapid therapy is usually necessary. When managing acute gallstone complications, whether the patient is improving, stabilizing, or worsening will dictate the pace of the diagnostic evaluation and the number and types of diagnostic studies that may be performed in order to confirm the presumptive diagnosis. The patient's course will also influence the threshold for intervening therapeutically as well as the choice of intervention.

See Also the Following Articles

Bilirubin and Jaundice • Cholecystectomy • Computed Tomography (CT) • Gallbladder Cancer • Gallstones, Pathophysiology of • Pancreatitis, Acute • Ultrasonography

Further Reading

Barkun, A. N. (2001). Early endoscopic management of acute gallstone pancreatitis—An evidence-based review. *J. Gastrointest. Surg.* **5**(3) 243–250.

Bilhartz, L. E., and Horton, J. D. (1998). Gallstone disease and its complications. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., Vol. 1, pp. 948–972. W. B. Saunders, Philadelphia.

Hanau, L. H., and Steigbigel, N. H. (2000). Acute (ascending) cholangitis. *Infect. Dis. Clin. North Am.* **14**, 521–546.

Kadakia, S. C. (1993). Biliary tract emergencies. Acute cholecystitis, acute cholangitis and acute pancreatitis. *Med. Clin. North Am.* **77**(5), 1015–1036.

Lillemoe, K. D. (2000). Surgical treatment of biliary tract infections. *Am. Surg.* **66**(2), 138–144.

Malet, P. F. (1996). Complications of cholelithiasis. *In* "Liver and Biliary Diseases" (N. Kaplowitz, ed.), 2nd Ed., pp. 673–691. Williams and Wilkins, Baltimore.

Raraty, M. G., Finch, M., and Neoptolemos, J. P. (1998). Acute cholangitis and pancreatitis secondary to common duct stones: management update. *World J. Surg.* **22**(11), 1155–1161.

Cholera

LARS ECKMANN
University of California, San Diego

cholera Diarrheal disease caused by toxigenic strains of *Vibrio cholerae*.
phage Bacterial virus that can transmit genetic material between different bacteria.
pili Bacterial surface appendages involved in attachment to host cells.
serogroups Categorization of bacteria into groups with similar carbohydrate-based surface antigens. Tests are done by agglutination with defined antisera.

Cholera is among the leading causes of diarrheal disease in developing countries of Africa, Asia, and South America. Despite the recognition of its infectious etiology more than a 100 years ago and the existence of effective control measures, cholera remains a major public health problem due to the lack of adequate water supply and infrastructure, financial resources, and education in many areas of the world.

MICROBIOLOGY

Cholera is caused by infection with toxigenic strains of *Vibrio cholerae*, gram-negative, motile, curved, rod-shaped bacteria of the family Vibrionaceae. Of the 200 recognized O serogroups, only two serogroups, O1 and O139, have been associated with severe disease and pandemics. *Vibrio cholerae* are natural inhabitants of the aquatic ecosystems of rivers and coastal waters, which form a large bacterial reservoir not accessible to exposure control measures. Toxigenic strains differ from nonpathogenic strains by the presence of specific virulence genes, some of which can be transmitted horizontally between different strains. The major virulence genes, which code for toxins and colonization factors, are categorized into two groups. The toxin-coding genes are located in a single-stranded filamentous phage integrated into the bacterial chromosome as prophage DNA, which can be excised and transmitted as a virion to other bacteria. Many of the genes required for bacterial colonization are located in a gene cluster, termed the *Vibrio* pathogenicity island (VPI); the expression of this cluster is tightly controlled by environmental conditions. The entire *V. cholerae* genome comprises two circular chromosomes of 3 and 1.1 megabases, the sequences of which have been determined completely. This provides an important basis for understanding the biology and virulence of these bacteria.

EPIDEMIOLOGY

Transmission is by the fecal–oral route. Infection is initiated by ingestion of contaminated water or, less commonly, food. Contact with contaminated utensils and houseflies and direct person-to-person contact also have been implicated, but are less important routes of transmission. The most common causes of water contamination are insufficient or absent disinfection and breaks in the integrity of water supply systems, leading to mixing of fecally contaminated and fresh water. These modes of transmission point to the primary importance of hygienic measures in preventing infection, both at the personal and the community levels. In addition to environmental exposure, personal risk factors play a role in determining susceptibility. Persons with low gastric acidity due to chronic gastritis, malnutrition, or treatment with gastric acid blockers are at increased risk for cholera infection. Bottle feeding of infants is also associated with increased risk of acquiring infection, whereas breast feeding is protective.

PATHOGENESIS

Vibrio cholerae colonizes the lumen and mucosal surface of the small intestine, but has little capacity to invade the intestinal mucosa. Bacteria attach to the mucous layers and the brush border of the intestinal epithelium via specialized appendages, particularly the toxin coregulated pili (TCP) and bacteria-associated hemagglutinins. Attachment is important for effective bacterial colonization of the small intestine. Intestinal disease is caused mainly by the release of cholera toxin, an AB holotoxin composed of a single A subunit and five B subunits. The B subunits bind to the cell surface of intestinal epithelial cells, whereas the A subunit enters

 Copyright 2004, Elsevier (USA). All rights reserved.

the cells and exerts enzymatic activity. Cholera toxin is taken up by apical endocytosis and subsequently follows a retrograde trafficking pathway through the Golgi cisternae to the endoplasmic reticulum, after which it enters the cytoplasm and binds to a G protein, $G_{\alpha s}$, at the cell membrane. The toxin ADP-ribosylates and thereby activates $G_{\alpha s}$, which in turn stimulates epithelial adenylyl cyclases, leading to increased production of cyclic AMP from ATP. Cyclic AMP activates protein kinase A, which phosphorylates and controls the functions of multiple target proteins in the cells, particularly those involved in regulated ion transport. The most important effects of elevated cAMP levels are stimulation of active chloride secretion and inhibition of sodium absorption, leading to the loss of electrolytes and secondarily water. In addition to cholera toxin, *V. cholerae* elaborates other toxins, such as zonula occludens toxin (ZOT), that contribute to disease pathogenesis by disrupting the epithelial barrier.

DISEASE AND DIAGNOSIS

The majority of infections (50–95%) with *V. cholerae* are asymptomatic or only mildly symptomatic, with a few episodes of watery stools without mucus or blood, and no significant fluid loss. Severe disease (cholera gravis) is characterized by massive watery diarrhea ("rice water" stools), vomiting, and dehydration, in the absence of abdominal pain and fever. Clinical symptoms first appear 1–3 days after infection, and often progress rapidly, leading to hypovolemic shock and anuria within 24 hours after onset. With sustained fluid loss, severe electrolyte imbalances occur, particularly hypokalemia and acidosis due to loss of potassium and bicarbonate in the stool, respectively. Mortality in untreated patients can be as high as 50%, with a higher risk in children compared to adults. If treated appropriately, disease symptoms subside after 4–6 days.

The symptoms of mild to moderate cholera cannot be clearly distinguished clinically from diarrhea caused by other infectious agents. Cholera should be considered in all cases with severe watery diarrhea and vomiting, particularly those with rapid progression and severe dehydration, and the potential for exposure to *V. cholerae*. Confirmation requires laboratory identification of toxigenic *V. cholerae* strains in stool specimens or rectal swabs of patients with diarrhea. The bacteria are grown on selective semisolid media [e.g., thiosulfate citrate bile sucrose (TCBS) agar], which is sufficient for identification in most cases because bacterial numbers in the stool are often high ($>10^6$/ml). Based on their characteristic tolerance to elevated pH, bacteria can be enriched by initial growth in alkaline peptone water before plating on selective agar media. Colonies are further identified by biochemical assays and agglutination tests with specific antisera. For epidemiological studies, different *V. cholerae* strains can be further characterized by determining the presence and/or sequence of specific genes using polymerase chain reaction (PCR)-based methods. This information can be used to follow the spread of infections with specific strains and to reconstruct the natural history of novel strains.

TREATMENT

The treatment of acute cholera is primarily symptomatic and aimed at rapid and appropriate fluid replacement. The time between the onset of symptoms and initiation of therapy critically determines the outcome of infection. Rehydration therapy can be oral or intravenous, depending on the available medical facilities and supplies. Oral rehydration is based on the fact that glucose-dependent sodium and water absorption remains intact in the presence of cholera toxin. Patients are given an oral rehydration solution (3.5 g sodium chloride, 2.5 g sodium bicarbonate, 1.5 g potassium chloride, and 20 g glucose per liter of water, as recommended by the World Health Organization). The volume is adjusted according to the severity of dehydration (as determined by several clinical signs, such as degree of thirst, mental state, pulse, and skin turgor). In severe cases and when ready access to appropriate clinical facilities is available, intravenous rehydration therapy with a balanced electrolyte solution (e.g., Ringer's lactate) is indicated.

Although rehydration is necessary and sufficient for patient management, antibiotics can reduce the severity of symptoms and the duration of bacterial excretion. Tetracycline and doxycycline are the first choice as antimicrobials, with erythromycin, trimethoprim/sulfamethoxazole, and furazolidone as alternatives for infections with tetracycline-resistant strains. The appropriate antibiotics are given orally when vomiting stops, which usually coincides with the completion of initial rehydration. Injectable antibiotics are not necessary. Antisecretory drugs have little or no benefit in cholera. Antidiarrheal, antiemetic, and antispasmodic drugs and corticosteroid medications are contraindicated in the management of cholera, because they have no beneficial effects and can prolong the disease.

PREVENTION

Prevention is mostly focused on improving hygienic and sanitary conditions. Of foremost importance are measures to provide adequate sewage collection and

disposal, the protection of clean water resources, and effective water disinfection. Although the effectiveness of these measures is well established, many regions throughout the world do not have the resources and infrastructure to institute them. On a personal level, hygienic measures can be taken to reduce the risk of infection. These include hand washing after defecation, proper food preparation and storage, and cleaning and drying of kitchen utensils.

Chemoprophylaxis with antibiotics can be effective under individual circumstances but has not been successful in limiting the spread of cholera at the community level. Several vaccines have been developed and tested, but their efficacy is limited. The development of vaccines that confer effective and long-lasting protection against different *V. cholerae* strains without causing excessive disease on inoculation remains a promising but challenging approach to controlling cholera.

See Also the Following Articles

Bacterial Toxins • Diarrhea • Diarrhea, Infectious • Foodborne Diseases • Food Safety

Further Reading

Reidl, J., and Klose, K. E. (2002). *Vibrio cholerae* and cholera: Out of the water and into the host. *FEMS Microbiol. Rev.* **26**, 125–139.
Shears, P. (2001). Recent developments in cholera. *Curr. Opin. Infect. Dis.* **14**, 553–558.

Cholestatic Diseases, Chronic

NATALIE J. TÖRÖK AND KEITH D. LINDOR
Mayo Clinic

autoantigen A "self-antigen" evoking an immunologic response by the host.
alloantigen An antigen that occurs in some but not all members of the same species.
cholangiography Radiographic examination of the bile ducts with contrast medium.
ductopenia Paucity of the bile ducts.
osteopenia Decreased calcification or density of bone.
periductal fibrosis Fibrosis occurring around the bile ducts.
steatorrhea Passage of large amounts of fat in the feces.
xanthoma A yellow nodule of the skin, composed of lipid-laden histiocytes.
xanthelasma Yellow plaques on the eyelids or around the eyes.

Primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) are the most common chronic cholestatic liver diseases in adults. Differential diagnosis of cholestatic diseases in adults also includes drug-induced cholestasis, idiopathic adulthood ductopenia, idiopathic biliary ductopenia, cholestasis of pregnancy, and cystic fibrosis. PBC and PSC represent clinically distinct entities with a presumed autoimmune basis. Both conditions are characterized by progressive destruction of bile ducts leading to chronic cholestasis and consequent complications such as portal hypertension and liver failure. Despite improvements in the management of complications from end-stage liver disease and excellent long-term results from liver transplantation, a continued understanding of the clinicopathologic processes responsible for cholestatic liver disease remains important.

PRIMARY BILIARY CIRRHOSIS

Epidemiology

Primary biliary cirrhosis (PBC) occurs worldwide and predominantly in middle-aged women, with a female to male ratio of 9 : 1. The median onset of disease is 50 years of age, but with a wide range of 21 to 91 years. The true prevalence and incidence of PBC are unclear, because few epidemiological studies have been conducted. The estimates of annual incidence range from 2 to 22 per million individuals. Whether this variation represents a true difference in prevalence or

 Copyright 2004, Elsevier (USA). All rights reserved.

disposal, the protection of clean water resources, and effective water disinfection. Although the effectiveness of these measures is well established, many regions throughout the world do not have the resources and infrastructure to institute them. On a personal level, hygienic measures can be taken to reduce the risk of infection. These include hand washing after defecation, proper food preparation and storage, and cleaning and drying of kitchen utensils.

Chemoprophylaxis with antibiotics can be effective under individual circumstances but has not been successful in limiting the spread of cholera at the community level. Several vaccines have been developed and tested, but their efficacy is limited. The development of vaccines that confer effective and long-lasting protection against different *V. cholerae* strains without causing excessive disease on inoculation remains a promising but challenging approach to controlling cholera.

See Also the Following Articles

Bacterial Toxins • Diarrhea • Diarrhea, Infectious • Foodborne Diseases • Food Safety

Further Reading

Reidl, J., and Klose, K. E. (2002). *Vibrio cholerae* and cholera: Out of the water and into the host. *FEMS Microbiol. Rev.* **26**, 125–139.
Shears, P. (2001). Recent developments in cholera. *Curr. Opin. Infect. Dis.* **14**, 553–558.

Cholestatic Diseases, Chronic

NATALIE J. TÖRÖK AND KEITH D. LINDOR
Mayo Clinic

autoantigen A "self-antigen" evoking an immunologic response by the host.
alloantigen An antigen that occurs in some but not all members of the same species.
cholangiography Radiographic examination of the bile ducts with contrast medium.
ductopenia Paucity of the bile ducts.
osteopenia Decreased calcification or density of bone.
periductal fibrosis Fibrosis occurring around the bile ducts.
steatorrhea Passage of large amounts of fat in the feces.
xanthoma A yellow nodule of the skin, composed of lipid-laden histiocytes.
xanthelasma Yellow plaques on the eyelids or around the eyes.

Primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) are the most common chronic cholestatic liver diseases in adults. Differential diagnosis of cholestatic diseases in adults also includes drug-induced cholestasis, idiopathic adulthood ductopenia, idiopathic biliary ductopenia, cholestasis of pregnancy, and cystic fibrosis. PBC and PSC represent clinically distinct entities with a presumed autoimmune basis. Both conditions are characterized by progressive destruction of bile ducts leading to chronic cholestasis and consequent complications such as portal hypertension and liver failure. Despite improvements in the management of complications from end-stage liver disease and excellent long-term results from liver transplantation, a continued understanding of the clinicopathologic processes responsible for cholestatic liver disease remains important.

PRIMARY BILIARY CIRRHOSIS

Epidemiology

Primary biliary cirrhosis (PBC) occurs worldwide and predominantly in middle-aged women, with a female to male ratio of 9 : 1. The median onset of disease is 50 years of age, but with a wide range of 21 to 91 years. The true prevalence and incidence of PBC are unclear, because few epidemiological studies have been conducted. The estimates of annual incidence range from 2 to 22 per million individuals. Whether this variation represents a true difference in prevalence or

 Copyright 2004, Elsevier (USA). All rights reserved.

results from different methodologies being used is unclear. The prevalence of PBC has increased over time, with 18 cases per million individuals in 1976 rising to 240 cases per million individuals in 1997. This rise may reflect an increase in survival time for patients with PBC, as well as increased awareness of the disease, leading to more frequent diagnosis. In families with one affected member, a prevalence rate of up to 4% has been described among first-degree relatives. An increased prevalence of other autoimmune-related disorders among first-degree relatives of PBC patients is also observed despite the absence of a documented strong genetic association.

Pathogenesis

Although the cause of PBC remains unknown, several lines of evidence suggest an autoimmune process. These include the presence of anti-mitochondrial antibodies (AMA), a frequent association with other autoimmune diseases, involvement of T cells in the destruction of the bile ducts, and numerous defects in immunologic regulation. The mechanisms and agent(s) responsible for the inflammation and bile duct destruction are unknown. The disease seems to be triggered by an immune-mediated response to an alloantigen or autoantigen that leads to progressive destruction of bile ducts and eventual development of biliary cirrhosis. Disease progression is characterized by ductular proliferation and necroinflammation leading to fibrous septa formation. Bridging between portal triads results in nodule formation and cirrhosis. The severity of ductopenia in PBC appears to correlate highly with the degree of fibrosis and is unrelated to periportal or lobular inflammation.

Clinical Features

Asymptomatic Disease

PBC is diagnosed at an asymptomatic stage in as many as 30% of patients with this condition. Such patients are found incidentally to have an elevated serum alkaline phosphatase level and AMA during routine health evaluations. Recently, it has been found that most asymptomatic individuals with AMA and normal liver biochemistries have features in the liver biopsies diagnostic of or consistent with PBC; these patients eventually develop symptoms and laboratory abnormalities of chronic cholestasis.

Symptomatic Disease

The patient who is diagnosed with PBC is typically a woman in the fifth or sixth decade of life with complaints of fatigue or pruritus. Other symptoms may include right upper quadrant abdominal pain, jaundice, and anorexia. Fatigue, although relatively nonspecific in PBC, is the most common symptom, found in about two-thirds of patients, and generally becomes worse as PBC progresses. Pruritus can occur at any time during the course of the illness, developing early, or later as the PBC evolves, or intermittently throughout the course of the disease. Pruritus is generally intermittent during the day and most troublesome at night. The development of xanthomas and xanthalesmas from abnormal lipid metabolism is also associated with PBC. Though uncommon, steatorrhea in PBC is usually from bile salt malabsorption, pancreatic exocrine insufficiency, or concomitant celiac disease. Osteopenia resulting in spontaneous bone fractures is often a manifestation of the metabolic bone disease associated with PBC. Jaundice is a symptom that often heralds the onset of advanced histologic disease and is seen in fewer than 20% of cases at initial diagnosis. Typical complications of cirrhosis such as ascites, variceal bleeding, and hepatic encephalopathy occur. Patients with advanced disease may warrant consideration for surveillance endoscopy to exclude esophageal varices at high-risk for bleeding.

Associated Diseases

Eighty percent of individuals with PBC also have coexistent extrahepatic autoimmune diseases. These include a number of rheumatologic diseases such as sicca syndrome, scleroderma, rheumatoid arthritis, dermatomyositis, mixed connective tissue disease, and systemic lupus erythematosus. Autoimmune thyroiditis and renal tubular acidosis have also been described with PBC. There is an increase in malignancy in patients with PBC; hepatocellular carcinoma is uncommon, but PBC patients have a substantially increased risk of developing this disease compared with the general population.

Diagnosis

Biochemical Changes

Liver biochemical tests show a cholestatic picture. Almost all patients have elevations in serum alkaline phosphatase (three to four times normal) and γ-glutamyltranspeptidase. Increased values for alanine aminotransferase and aspartate aminotransferase (AST) of two to three times normal are common; however, marked elevations are unusual and may suggest PBC—autoimmune hepatitis overlap syndrome or coexisting viral hepatitis. Serum total bilirubin levels often rise during the course of disease progression

with levels occasionally reaching 20 mg/dl. With advanced disease, the combination of elevated total bilirubin, decreased albumin, and increased prothrombin time represent poor prognostic markers for outcome. Serum immunoglobulins, especially IgM levels, are increased, as are serum bile acid levels, in particular cholic and chenodeoxycholic acids, as well as serum cholesterol.

Serologic Diagnosis

Up to 95% of patients with PBC test positive for AMA. The M2 antibody, which is directed against the pyruvate dehydrogenase complex of the inner mitochondrial membrane, has a sensitivity of 98% and a specificity of 96% in the diagnosis of PBC. Anti-centromere antibodies seen in the CREST (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, teleangiectasias) syndrome despite absent clinical findings of scleroderma have also been noted. Seventy percent of PBC patients may also exhibit anti-nuclear antibodies (ANA) with or without anti-smooth muscle antibody (ASMA). The terms "AMA-negative PBC" or "autoimmune cholangitis" have been created to describe patients with clinical features and natural history profiles consistent with PBC but without AMA. ANA or ASMA positivity is often observed in the absence of AMA in this patient group.

Liver Biopsy

Although serum AMA will confirm a suspected diagnosis of PBC, a liver biopsy is needed for determining the stage of disease at diagnosis. Prognostic information such as determining the need for surveillance of complications related to cirrhosis or consideration of liver transplant evaluation may also be provided by liver biopsy. Wide variations in histology can exist even with slight elevations in total bilirubin and alkaline phosphatase. The most important and only diagnostic clue in many cases is ductopenia as defined by the absence of interlobular bile ducts in more than 50% of portal tracts. The florid duct lesion, in which the epithelium of the interlobular and segmental bile ducts degenerate segmentally with formation of noncaseating epitheloid granulomas, is nearly diagnostic of PBC, but is found in a relatively small number of cases (<10%) and mainly in early stages. The two most popular classification systems are those presented by Ludwig and Scheuer, which classify the disease in four stages. Both systems describe progressive pathologic changes, initially beginning in the periportal areas surrounding the bile ducts, with the eventual development of cirrhosis.

Natural History

The natural history of PBC has been described in both asymptomatic and symptomatic PBC.

Mitchison *et al*. and Metcalf *et al*. have reported that 83% of patients who were AMA-positive and who had normal liver biochemistries and no symptoms of liver disease had liver histology compatible with PBC. These patients were followed for a median of 17.8 years, and eventually 83% of patients developed abnormal liver biochemistries and 76% developed symptoms consistent with PBC. These studies clearly demonstrated that asymptomatic patients who are AMA-positive have very early PBC.

Several reports have provided evidence regarding the natural history of asymptomatic patients who are AMA-positive, have abnormal liver biochemistries, and have liver histology compatible with PBC. Patients who remain asymptomatic for years may have significantly longer survival than symptomatic patients, but their life expectancy is still less than that of an age- and gender-matched control population.

In contrast, patients with symptomatic PBC show a more rapid progression to end-stage liver disease and its inherent complications and they have a worse prognosis. Several independent predictors of poor prognosis have been identified in this group of patients with PBC, including advanced age, high serum bilirubin levels, poor synthetic function, hepatomegaly, ascites, variceal bleeding, and advanced histological stage. When untreated, PBC may follow a course that extends over a 15- to 20-year period. However, in patients with serum bilirubin levels greater than 10 mg/dl, the average life expectancy is reduced to 2 years. To predict survival in patients with PBC, several prognostic models have been reported. Of these models, the Mayo risk score based on patient age, serum bilirubin, albumin, prothrombin time, and edema has been cross-validated and widely used in predicting survival and guiding physicians in patient referral for liver transplantation.

Treatment

A number of potential treatments for PBC have been evaluated to date with the primary goals of stabilizing or halting disease progression. A variety of treatment options have been tested in patients with primary biliary cirrhosis. Therapies that have been tested and have not been found to be beneficial or had excessive toxicity included penicillamine, cyclosporin, azathioprine, thalidomide, mycophenolic acid, and chlorambucil. Corticosteroids and colchicine as well as methotrexate have had some advocates, although data supporting the use of any of these are

either quite limited or unconvincing. The drug with which there is currently the most experience and the most optimism about is ursodeoxycholic acid. This drug has been approved by the FDA for use in patients with PBC. Ursodeoxycholic acid is a safe, well-tolerated drug that improves liver biochemistries, improves survival free of transplantation, reduces the risk of developing varices, and improves pruritus. Four randomized, placebo-controlled trials using ursodeoxycholic acid (UDCA) have demonstrated histologic benefit and/or a delay in the time to transplantation or death. At a dose of 13 to 15 mg/kg/day, UDCA has also been shown to reduce the risk of developing esophageal varices and cirrhosis in the long term. This is a cost-effective therapy and is expected to have a real impact on the natural history in patients with primary biliary cirrhosis.

Pruritus as a frequent complication of cholestatic liver disease creates difficult management options as significant impairments in quality of life are observed. Cholestyramine (a bile-acid-binding resin) may decrease the intensity of pruritus but rarely leads to complete symptom resolution. Rifampin (300 to 600 mg daily), antihistamines, and UDCA with or without cholestyramine have been tried, usually with good success. Patients should be instructed to take UDCA several hours after cholestyramine administration to prevent the binding of UDCA and its elimination without absorption.

Osteopenia or metabolic bone disease has been associated with an accelerated rate of bone loss among individuals with PBC. Decreased bone formation rather than increased resorption is considered to be an important mechanism of action.

PRIMARY SCLEROSING CHOLANGITIS

Epidemiology

Seventy percent of affected individuals are men with an average age of 40 years at diagnosis. There have been no epidemiologic studies on the prevalence of PSC in any general population group. Most data were derived from epidemiologic studies on inflammatory bowel disease (IBD) patients. The close association between PSC and IBD has been well established, as 70–80% of PSC cases are associated with IBD, usually chronic ulcerative colitis (CUC). Conversely, approximately 5% of patients with CUC develop PSC. From an estimated prevalence rate for CUC of 40 to 225 cases per 100,000 individuals in the United States, the prevalence of PSC is estimated to be between 2 and 7 cases per 100,000 individuals. These figures, however, underestimate the true prevalence of PSC, since the prevalence of PSC in CUC is likely to be higher than estimated.

Pathogenesis

The cause of PSC is unknown. However, a number of factors that might cause recurring damage to the bile ducts and lead to the development of the disease have been proposed. Periductal fibrosis with inflammation, bile duct obliteration, and ductopenia constitute the main histologic findings in PSC. Similar findings may occur in other conditions such as PBC, autoimmune hepatitis (AIH), or chronic extrahepatic bile duct obstruction, which constitute the primary differential diagnoses. In the pathogenesis of PSC, cellular immune factors may play a role, as suggested by the composition of the portal infiltrate, which is made up mainly of $CD4^+$ T lymphocytes. An immune-mediated mechanism also has been suggested by the findings in an experimental model of immune-mediated cholestasis and shared antibodies against colonic and biliary epithelia.

Anti-neutrophil cytoplasmic antibodies (ANCA) originally detected in serum from patients with ulcerative colitis were subsequently identified in a high proportion of patients with PSC. However, ANCA have also been detected in the serum of patients with other autoimmune liver diseases such as autoimmune hepatitis and PBC. The occurrence of PSC in relatives suggests human leukocyte antigen association and genetic predisposition for developing this condition. In fact, the DR4 allele may be a marker of more rapid disease progression. Although major advances in the understanding of PSC have been made in recent years, the cause of the disease still remains to be elucidated.

Clinical Features

Originally, many patients who were recognized as having this disease presented with features of advanced liver disease. Now, increasingly, it appears that more patients are being discovered incidentally, often during evaluation for abnormal liver tests seen in association with inflammatory bowel disease. Asymptomatic elevations in serum liver biochemistries or the development of symptoms such as fatigue (75%) and pruritus (70%) are often observed with the diagnosis of PSC. Jaundice (60%) and weight loss (40%) are less common and should raise the suspicion of advanced disease or a complication related to PSC. Xanthomatous lesions and hyperpigmentation, however, are less common than in patients with PBC. Fever and abdominal pain in a patient with PSC are highly suggestive of bacterial cholangitis, which can manifest at initial presentation. The presence

of jaundice is not required for the diagnosis of bacterial cholangitis. Over the course of time as the disease progresses, most patients will eventually develop one or more of these symptoms. Over the average period of 10–15 years, patients often progress to end-stage liver disease and suffer the consequences of a failing liver or require liver transplantation, although this course can be highly variable, with some patients surviving 25 or more years with nearly asymptomatic PSC.

Radiographic Features

Endoscopic retrograde cholangiopancreatography (ERCP) is the test of choice for visualizing the biliary tract and is considered the gold standard for making the diagnosis of PSC. Segmental fibrosis of both intrahepatic and extrahepatic ducts with subsequent saccular dilation of normal intervening areas results in the characteristic "beads on a string" appearance seen in PSC (Fig. 1). Approximately 20% of patients will have only intrahepatic and hilar involvement with sparing of the remaining extrahepatic duct. The presence of a significant or dominant stricture on ERCP should raise the question of cholangiocarcinoma as a supervening complication of PSC. Cross-sectional imaging studies are of limited value in patients with suspected primary sclerosing cholangitis. Magnetic resonance cholangiography (MRCP) is a noninvasive test that does not require administration of contrast or use of radiation and appears to be promising. MRCP does not allow histological sampling or therapeutic intervention but may be a cost-effective initial step.

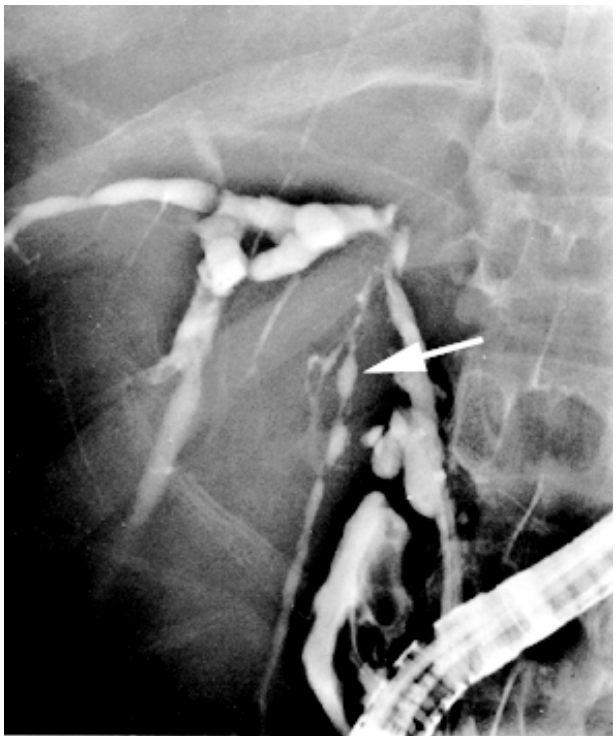


FIGURE 1 ERCP in primary sclerosing cholangitis shows beading irregularities (arrow) in the bile ducts.

Diagnosis

Nearly all patients with PSC have significant elevations in serum alkaline phosphatase (three to five times normal) and a milder increase in serum hepatic transaminases. High levels of total bilirubin, which often fluctuate in PSC, are worrisome for advanced disease when persistent elevations occur. No significant differences in serum liver biochemistry profiles have been seen among asymptomatic and symptomatic individuals at diagnosis. In combination with the appropriate clinical and radiographic findings, an initial diagnosis of PSC can be made without the explicit use of hepatic histology. Liver biopsy is useful in histological confirmation and staging of the disease, but, because of the sampling variability, can be troublesome. A recent analysis by Burak *et al*. suggests that a liver biopsy adds little to the diagnosis of PSC. Unlike PBC, no specific autoantibody has been identified to confirm the diagnosis of PSC (see Table I). A total of 70 to 80% of patients will test positive for ANCA though its clinical significance remains unclear because this autoantibody is common in several other liver diseases. An increased prevalence of extrahepatic autoimmune disorders has been observed for PSC but these are much less commonly seen than in PBC, except for inflammatory bowel disease.

Natural History

The natural history of PSC is not as well defined as with other liver diseases, primarily because the entity has been widely recognized only in the past quarter century. In 1970, a review described all cases of PSC reported to that time, which numbered less than 100. With the advent of ERCP in the mid to late 1970s, the diagnosis became much more commonly established so that some major medical centers are seeing over 100 patients a year with PSC. Because the broad recognition of the disease is relatively recent, some may have the sense that the natural history is changing. The median reported survival in patients with PSC is 10 to 12 years from the time of diagnosis. Several variables such as older age, elevated serum levels of bilirubin, AST, low albumin, hepatosplenomegaly, variceal bleeding, presence of IBD, and histologic stage have been identified as independent risk factors indicating poor prognosis. Child–Pugh classification has been evaluated as a

TABLE I Primary Sclerosing Cholangitis (PSC) and Primary Biliary Cirrhosis (PBC)

	PSC	PBC
Sex	65% male	90% female
Presentation	Fatigue, pruritus, jaundice, cholangitis	Fatigue, pruritus, jaundice
Serum anti-mitochondrial antibodies	Negative or very low titer	Positive
Cholangiogram	Beaded, irregular	Non specific, may be pruned appearance
Associated diseases	Inflammatory bowel disease, immunodeficiency syndromes, cholangiocarcinoma	Autoimmune disorders, thyroiditis, renal tubular acidosis

prognostic indicator for survival in PSC. The study showed that the age-adjusted Child–Pugh model predicts survival before transplantation and it has been used in formulating minimal listing criteria. The recently described entity of small duct PSC also seems to follow a course comparable to classic PSC. These patients have similar clinical, biochemical, and histologic features and may make up fewer than 5% of patients with a histologic diagnosis of PSC. These patients are distinguished by normal cholangiograms in the setting of inflammatory bowel disease. Some of these patients may develop cholangiographic features of classic PSC and eventually require liver transplantation. More work is obviously needed to better understand the natural history of this variant.

Treatment

Due to the variable disease progression seen in PSC, the development of randomized clinical trials for the assessment of therapies has been difficult. As a potential consequence, there is no acceptable medical therapy available today for the treatment of PSC. As seen with PBC, the use of pharmacologic agents such as *d*-penicillamine, colchicine, corticosteroids, and immunosuppressive agents has not been shown to impart a significant clinical benefit. UDCA in standard doses (13 to 15 mg/kg/day) appears to cause biochemical improvement but has no significant impact on histology or survival. Recently, it was shown that UDCA at a higher dose (20 mg/kg) looks promising by showing improvement in liver biochemistries and histology after 2 years of treatment compared to placebo.

PSC is one of the more common indications for liver transplantation in the United States. The cumulative 1-year survival rate after liver transplantation is now greater than 85%. Liver transplantation should be considered before the disease is too advanced because PSC patients with low (<4.4) Mayo risk score have a significantly better survival.

Cholangitis in PSC is most commonly associated with a previous history of biliary tree manipulation by surgical or radiologic/endoscopic interventions in select individuals. Once suspected, empiric intravenous antibiotic therapy to cover gram-negative and anaerobic bacteria should be initiated. ERCP should then be considered to exclude biliary obstruction from benign (choledocholithiasis, dominant stricture) or malignant (cholangiocarcinoma) etiologies.

The presence of a dominant stricture (defined as a high-grade, localized narrowing of the biliary tree) is often manifested by a sudden increase in serum alkaline phosphatase and/or bilirubin in the asymptomatic patient. Progressive jaundice and the development of cholangitis from obstruction are also consequences of dominant stricture formation, requiring cholangiography. Balloon dilation of identified strictures by endoscopic or radiologic approaches often provides satisfactory results. Use of endoscopic stents for 3 to 6 months to prevent recurrent stricture formation may be fraught with an increased rate of infectious complications. Pinch biopsies with brushings for histological analysis of all dominant strictures are required to help exclude the presence of cholangiocarcinoma.

Cholangiocarcinoma is the most feared complication of PSC, occurring in 10–15% of cases. Clinical progression of disease in a rapid fashion with jaundice and marked liver biochemistry elevations often alerts the clinician to the possibility of bile duct cancer. Repeated attempts at diagnosis with biopsy and brushings may be required for definitive exclusion. Occasional surgical intervention is required for full-thickness biopsy in special instances. The overall median survival is approximately 5 months from the diagnosis of cholangiocarcinoma, although liver transplantation in highly selected cases was shown to improve survival.

VARIANT CHOLANGIOPATHIES

Recent observations have documented a group of patients with clinical characteristics suggestive of PBC without AMA but with positive ANA or ASMA. The

terms "autoimmune cholangitis" or "AMA-negative PBC" have been used in referring to these individuals. The clinical course and response to UDCA therapy are the same as in patients with AMA-positive PBC.

The presence of features consistent with both AIH and PBC has resulted in a diagnostic category termed the overlap syndrome. Both ANA and AMA are often positive through serologic testing. Piecemeal necrosis and coexistent periductal/portal inflammation with bile duct destruction is commonly seen. This group appears to benefit from immunosuppressive treatment as well as UDCA. Similar overlap also occurs in PSC and AIH in both adult and pediatric populations. The presence of typical biliary stricturing and dilation seen in PSC is often accompanied by histologic lesions seen in AIH and also by high-titer ANA. The overlap of AIH and PSC may warrant the use of immunosuppressive therapy.

See Also the Following Articles

Cholangiocarcinoma • Cholangitis, Sclerosing • Cirrhosis • Colitis, Ulcerative • Crohn's Disease • Liver Biopsy • Liver Transplantation

Further Reading

Angulo, P., Lindor, K. D., *et al.* (1999). Utilization of the Mayo risk score in patients with primary biliary cirrhosis receiving ursodeoxycholic acid. *Liver* **19**, 115–121.

Angulo, P., Pearce, D. H., *et al.* (2000). Magnetic resonance cholangiography in patients with biliary disease: Its role in primary sclerosing cholangitis. *J. Hepatol.* **33**, 520–527.

Balasubramaniam, K., Grambsch, P. M., *et al.* (1990). Diminished survival in asymptomatic primary biliary cirrhosis: A prospective study. *Gastroenterology* **98**, 1567–1571.

Burak, K. A. P., and Lindor, K. D. (2001). Is there a role for liver biopsy in primary sclerosing cholangitis? *Gastroenterology* **120**(Suppl. 1), A1–A765, AA1–AA88.

De Vreede, I., Steers, J. L., *et al.* (2000). Prolonged disease-free survival after orthotopic liver transplantation plus adjuvant chemoirradiation for cholangiocarcinoma. *Liver Transplant.* **6**, 309–316.

Dickson, E. R., Grambsch, P. M., *et al.* (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology* **10**, 1–7.

Gores, G. J. (2000). Early detection and treatment of cholangiocarcinoma. *Liver Transplant.* **6**(Suppl. 2), S30S34.

Kim, W. R., Ludwig, J., *et al.* (2000). Variant forms of cholestatic diseases involving small bile ducts in adults. *Am. J. Gastroenterol.* **95**, 1130–1138.

Ludwig, J., Dickson, E. R., *et al.* (1978). Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). *Virchows Arch. A Pathol. Anat. Histol.* **379**, 103–112.

Metcalf, J. V., Mitchison, H. C., Palmer, J. M., Jones, D. E., Bassendine, M. F., and James, O. F. (1996). Natural history of early primary biliary cirrhosis. *Lancet* **348**, 1399–1402.

Mitchison, H. C., Bassendine, M. F., *et al.* (1986). Positive antimitochondrial antibody but normal alkaline phosphatase: Is this primary biliary cirrhosis? *Hepatology* **6**, 1279–1284.

Nashan, B., Schlitt, H. J., *et al.* (1996). Biliary malignancies in primary sclerosing cholangitis: Timing for liver transplantation. *Hepatology* **23**, 1105–1111.

Pasha, T., Heathcote, J., *et al.* (1999). Cost-effectiveness of ursodeoxycholic acid therapy in primary biliary cirrhosis. *Hepatology* **29**, 21–26.

Shetty, K., Rybicki, L., *et al.* (1997). The Child-Pugh classification as a prognostic indicator for survival in primary sclerosing cholangitis. *Hepatology* **25**, 1049–1053.

Cholesterol Absorption

DAVID Y. HUI
University of Cincinnati College of Medicine

bile Fluid secreted by the liver and passed into the intestinal lumen, where it aids in lipid solubilization.
brush border Microvilli on the plasma membrane of the intestinal epithelium, which is specialized for absorption.
chylomicrons Large lipid–protein complexes secreted by the intestine into the lymphatics and the blood circulation.
emulsions Large lipid droplets.
enterocytes Intestinal cells.
lipase Enzyme that hydrolyzes lipids.
lipoproteins Lipid–protein complexes in the blood circulation.
lymphatics Vessels that contain or convey lymph; lymphatics of the small intestine are where absorbed fat is drained.
micelle Small structure, usually between 3 and 10 μm in size, made up of polymeric bile salts and fatty acids.
vesicle Structure of lipids dispersed in an aqueous solution, usually surrounded by a phospholipid bilayer.

Dietary cholesterol absorption is defined as the process by which dietary cholesterol is absorbed through the intestinal tract and transported into the blood circulation. In view of the fact that dietary fat and cholesterol intake are increasing at an alarming pace in industrialized countries, there is an increasing interest among clinicians to understand the mechanism of dietary cholesterol absorption. The goal is to design intervention strategies, either by dietary manipulations or therapeutics, to lower the efficiency of dietary fat and cholesterol absorption, thereby reducing the risk of coronary heart disease, cancer, diabetes, and obesity.

INTRODUCTION

Epidemiological studies have clearly established a direct relationship between cholesterol level in blood plasma and coronary heart disease. Elevated plasma cholesterol level also promotes other debilitating diseases, including certain forms of cancer, diabetes, and obesity. Cholesterol levels in the circulation are dependent on several parameters, including endogenous synthesis, secretion, and catabolism of the various plasma lipoproteins. Another major contributor to the amount of cholesterol in plasma is the exogenous cholesterol ingested in the diet. Depending on different individuals, 10–60% of the cholesterol entering the body each day is derived from the diet. In addition to its effect on total plasma cholesterol level, the amount of dietary cholesterol entering the bloodstream also has a direct impact on the level of low-density lipoproteins (LDLs), the "bad" cholesterol that directly contributes to plaque formation in the arterial wall. It has been estimated that a 90% reduction of cholesterol absorption in subjects with a moderately elevated plasma cholesterol level can reduce plasma cholesterol level by 35%, and that a 62% reduction in plasma cholesterol level can be achieved by 100% inhibition of cholesterol absorption.

Cholesterol exists in the diet mainly in the unesterified form, with only 5–10% of the cholesterol being esterified by fatty acids as cholesteryl esters. The intake of cholesterol is usually associated with fat consumption. The process of absorption begins in the stomach with partial fat digestion by preduodenal lipases and emulsification by peristalsis. The crude emulsions are then delivered to the duodenum, where they are mixed with bile and pancreatic juice. The dietary cholesterol appearing in the lumen of the intestine is usually associated with triglycerides and phospholipids in lipid emulsions, and coexists with cholesterol secreted from the liver into the bile, which is mixed with bile salt in the form of micelles. The lipid emulsions are further digested in the intestinal lumen by enzymes secreted by the pancreas, converting the phospholipids and triglycerides to fatty acids, lysophospholipids, and monoglycerides. Cholesteryl esters are also hydrolyzed to unesterified cholesterol and free fatty acids. The unesterified cholesterol is then distributed to phospholipid vesicles and bile salt mixed micelles, and is transported to the brush border of the intestine, where it is taken up by the mucosal cells. A diagrammatic representation of luminal events involved in cholesterol absorption is shown in Fig. 1.

Once the cholesterol is transported to the intestinal cell surface by phospholipid vesicles and bile salt mixed micelles, it can diffuse through the cell membrane into intracellular compartments. Cholesterol embedded in vesicles and micelles can also be taken up into the cell

 Copyright 2004, Elsevier (USA). All rights reserved.

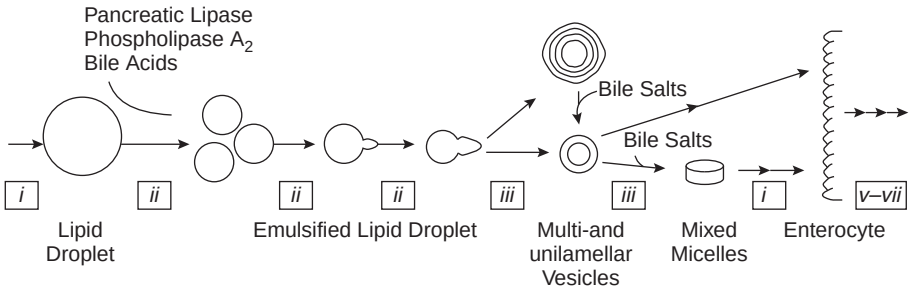


FIGURE 1 Process of dietary lipid absorption in the lumen of the small intestine. Dietary lipid absorption involves a multitude of processes in several different tissues. The processes include (*i*) lipid emulsification in the stomach, (*ii*) lipid digestion within the intestinal lumen by enzymes secreted from the pancreas, (*iii*) micellar solubilization of the lipid components, (*iv*) intestinal uptake, (*v*) intracellular lipid esterification and lipoprotein assembly, (*vi*) enterocyte secretion, and (*vii*) homeostasis of the dietary components in plasma.

interior by an active transport process involving cell surface transporter proteins. The amount of cholesterol entering the enterocytes can also be regulated by specific cell surface transporter proteins, the function of which is to export cholesterol back to the luminal space for excretion out of the body. The cholesterol entering the mucosal cells is rapidly transported to an intracellular compartment, where it is reesterified with fatty acid, packaged into large intestinal lipoproteins called chylomicrons, and then secreted into the lymphatics.

The process of dietary cholesterol absorption, from the point of cholesterol ingestion to its appearance in circulation, thus requires the participation of a number of different proteins. It is not surprising, therefore, that different expression level and/or activity of these proteins can influence the efficiency of cholesterol absorption. Indeed, genetic variations in cholesterol absorption efficiency have been reported in a number of species, including humans. A number of key proteins that participate in the cholesterol absorption process, including those that are required for lipid digestion, cellular uptake and efflux, and intracellular processing of the cholesterol prior to its secretion into the circulation, are the focus of the following discussion. The latest advances indicating genetic regulation and intervention strategies aimed at reducing cholesterol absorption efficiency are also discussed.

LIPID DIGESTION

Cholesterol is ingested in the diet in association with the fatty triglycerides. These lipids are emulsified in the stomach by the action of peristalsis. Some of the dietary triglycerides are partially hydrolyzed by preduodenal lipases into diglycerides and free fatty acids. Depending on the individual species, the preduodenal lipase can be found in the tongue (lingual lipase) or in the stomach (gastric lipase). In humans, the preduodenal lipase is gastric lipase and no lingual lipase is found. Regardless of the tissue source of the preduodenal lipases, these enzymes prefer triglycerides with short- and medium-chain fatty acids and are incapable of hydrolyzing significant amounts of triglycerides with long-chain fatty acids. These lipases are also ineffective in hydrolyzing phospholipids and cholesteryl esters. Thus, these preduodenal lipolytic enzymes do not hydrolyze much of the phospholipid-coated emulsions, and a majority of the dietary fat and cholesterol are transported to the intestinal lumen as large, emulsified lipid droplets.

The lipid emulsions entering the luminal area of the intestine are digested by lipolytic enzymes secreted by the pancreas. The major lipases secreted by the pancreas in response to a meal include the pancreatic triglyceride lipase, a pancreas-specific form of phospholipase A_2, and carboxyl ester lipase. The latter enzyme is also known as cholesterol esterase or as bile salt-stimulated lipase. Two additional proteins, pancreatic lipase-related protein-1 and -2, which are similar in sequence and structure to the pancreatic triglyceride lipase, are also secreted by the pancreas. The important role of pancreatic enzymes in fat and cholesterol absorption is best illustrated by fat and cholesterol malabsorption in cystic fibrosis patients and in patients with other pancreatic diseases, including those with pancreatitis due to chronic alcoholism. These patients suffer steatorrhea and reduced cholesterol absorption and have deficiencies of fat-soluble vitamins and essential fatty acids.

In vitro studies show that pancreatic triglyceride lipase is the primary digestive tract enzyme that hydrolyzes triglycerides in emulsified lipid particles. Its activity is dependent on the presence of its coenzyme,

colipase. The contribution of pancreatic triglyceride lipase/colipase in fat and cholesterol absorption is highlighted by observations that patients with congenital pancreatic triglyceride lipase deficiency or deficiency of colipase suffer from fat and cholesterol malabsorption similar to that seen in patients with pancreatic insufficiency. The mechanism by which pancreatic triglyceride lipase facilitates dietary cholesterol absorption is indirect. *In vitro* cell culture experiments suggest that cholesterol carried by bulky lipid emulsions is not accessible for uptake by intestinal cells and that pancreatic triglyceride lipase hydrolysis of the core triglycerides is required for intestinal uptake of the cholesterol. It is likely that the dietary cholesterol in lipid emulsions needs to be redistributed to smaller phospholipid vesicles and bile salt mixed micelles prior to its transport to the brush border surface, where it can be taken up by the intestinal mucosa. In the absence of pancreatic triglyceride lipase-mediated lipid hydrolysis, the cholesterol is retained in the lipid emulsion, which may be too large for penetration through the unstirred water layer of the brush border membrane.

The remodeling of the cholesterol carrier in the intestinal lumen also requires phospholipase hydrolysis of phospholipids on the surface of emulsions and on lipid vesicles. Thus, phospholipase-mediated hydrolysis of the phospholipid in mixed micelles may also be an important step in the cholesterol absorption process. The role of phospholipase in facilitating cholesterol absorption can be inferred from early *in vitro* organ culture experiments, which showed that increasing phospholipid content of bile salt mixed micelles retarded cholesterol transport to the intestine. Subsequent studies also showed cholesterol output from the lumen to the lymphatics was higher when duodenal infusions were performed with lipid emulsions containing low levels of phosphatidylcholine. In addition, human subjects receiving an intraduodenal infusion of phosphatidylcholine were also found to absorb cholesterol to a lower extent compared with subjects infused with the same amount of cholesterol and safflower oil with a similar level of fatty acids. A direct role of phospholipase A_2 in facilitating cholesterol transport to intestinal cells was suggested by cell culture experiments showing that the addition of the pancreatic Group 1B phospholipase A_2 to the cell incubation medium increased cholesterol uptake by Caco-2 cells. In intact animals, phospholipase A_2 inhibitors effectively reduced cholesterol transport from the intestinal lumen to the lymphatics. Although these experiments clearly demonstrate a role for phospholipase A_2 in mediating cholesterol absorption, the particular type of phospholipase that is involved in this process is less clear. The Group 1B phospholipase A_2 secreted by the pancreas is the predominant phospholipase in the intestinal lumen, but animals lacking this enzyme are not defective in cholesterol absorption under normal dietary conditions. However, under high dietary fat conditions, the phospholipase A_2-deficient mice display reduced fat and cholesterol absorption efficiency. These latter observations suggest that phospholipase A_2 synthesized by the intestine may compensate for the lack of the pancreatic-type phospholipase A_2 in phospholipid digestion in the intestinal lumen and in mediating cholesterol absorption. Thus, the pancreatic phospholipase A_2 is uniquely required among various phospholipases in controlling cholesterol absorption and transport during high dietary fat input, whereas other phospholipases expressed in the intestine may take its place when dietary lipid content is low.

The role of the third lipolytic enzyme secreted by the pancreas, carboxyl ester lipase, in cholesterol absorption has drawn considerable interest over the past three decades because of its ability to influence cholesterol esterification and deesterification directly. This enzyme also catalyzes the hydrolysis of triglycerides and fat-soluble vitamin esters. *In vitro* data reveal that complete hydrolysis of triglycerides to fatty acids and glycerol requires the concerted action of both pancreatic triglyceride lipase and the carboxyl ester lipase. Thus, carboxyl ester lipase works in concert with pancreatic triglyceride lipase in mediating fat digestion. Its direct involvement in the dietary cholesterol absorption process is suggested by early studies that showed that the reduction in cholesterol absorption due to pancreatic diversion can be restored by infusion with pancreatic juice containing the carboxyl ester lipase, but not with juice devoid of this enzyme. The incubation of rat intestinal sacs with cholesterol containing micelles also shows that cholesterol uptake by the intestinal sacs is three- to fivefold higher if carboxyl ester lipase is included in the medium. Specific inhibitors of carboxyl ester lipase, such as the phenoxylphenyl carbamates WAY-121,751 and WAY-121,898, are effective inhibitors of cholesterol absorption in normal and cholesterol-fed animals. Unfortunately, these results have been challenged by more recent studies, which yield contradictory results. It is likely that the carboxyl ester lipase preparations used in the earlier studies contain other lipolytic enzymes that act in concert with carboxyl ester lipase to facilitate cholesterol absorption. The phenoxylphenyl carbamates may also be less specific and may also affect the activity of other pancreatic lipases. The generation of carboxyl ester lipase-deficient mice by the homologous recombination technique conclusively shows that this enzyme plays only a minor role in controlling

the amount of dietary cholesterol absorbed through the intestine. However, carboxyl ester lipase may be more important in dictating the type of lipoproteins being produced by the intestine in response to a fatty meal. Animals lacking a functional carboxyl ester lipase are incapable of secreting the larger sized chylomicrons. Lipoproteins produced in the intestine of these animals are similar in size to the smaller liver-derived very-low-density lipoproteins. It is postulated that carboxyl ester lipase alters intracellular lipid trafficking and/or signaling pathways that are important in transporting exogenous lipids to intracellular compartments where lipoprotein assembly occurs. This hypothesis remains to be tested by more vigorous experiments.

The pancreatic lipase-related protein-1 and -2 (PLRP-1 and PLRP-2) appear to be less important in lipid digestion and absorption under normal conditions in adults. Both of these enzymes have the highest expression level during the developmental period before the onset of pancreatic triglyceride lipase gene expression. The expression profile of these genes suggests that the related proteins may compensate for the low pancreatic triglyceride lipase activity during the suckling period, to ensure efficient fat digestion in newborns. The possible role of PLRP-2 in newborn fat digestion is supported by the observation that PLRP-2-deficient mice display higher fecal fat content and decreased rate of weight gain than do their wild-type counterparts when fed similar diets. In contrast, the role of PLRP-1 in fat digestion and absorption is less certain; *in vitro* analyses fail to detect significant lipolytic activity of this protein. No PLRP-1-deficient animal model is currently available to test for the function of this lipase *in vivo*.

CHOLESTEROL UPTAKE BY INTESTINAL MUCOSAL CELLS

The digestion of the surface phospholipids and the core triglycerides in bulky lipid emulsions results in the liberation of the dietary cholesterol and its transfer to phospholipid vesicles and bile salt micelles. These lipid carriers are responsible for cholesterol transport to the surface of the intestinal mucosa, where cholesterol can be absorbed into the enterocytes. Bile salt present in the lumen helps transport cholesterol across the unstirred water layer, which is a series of water lamellae at the interface between the bulk water phase of the lumen and the enterocyte cell membrane. Thus, abnormal bile acid and biliary lipid composition may affect cholesterol absorption through inhibition of this step of the absorption process. This hypothesis is substantiated by observations that mice defective in bile acid synthesis due to a genetic defect in the cholesterol 7α-hydroxylase gene *Cyp7α1* absorb dietary cholesterol minimally in comparison with wild-type mice. Likewise, defects in bile salt or phospholipid export from the liver to the bile, due to deletion of P-glycoprotein genes, also result in abnormal cholesterol absorption efficiency.

Once cholesterol is transported to the mucosal surface of the intestine by phospholipid vesicles and bile salt micelles, the cholesterol can be taken up by the enterocytes in a rapid process through passive diffusion across the cell membrane. Although recent data suggest the presence of an active transport mechanism for cholesterol uptake, interpretation of the data remains equivocal. The amount of cholesterol absorbed into the enterocytes can also be regulated by membrane surface proteins that are responsible for mediating cholesterol efflux.

Regardless of the precise mechanism by which cholesterol enters the intestinal mucosa, intestinal cholesterol uptake is a two-step process. The first step is a reversible process by which cholesterol in vesicles and micelles is embedded into the plasma membrane of the enterocytes. The amount of cholesterol that can be accommodated in the intestinal cell membrane is correlated with the sphingomyelin content in the membrane. Intestinal cell culture studies *in vitro* reveal that the rate of exogenous cholesterol uptake also decreases when the membrane is depleted of sphingomyelin. Thus, the amount of cholesterol that can be absorbed into the enterocytes is in part regulated by the sphingomyelin concentration in the plasma membrane. By extrapolation, cholesterol absorption and transport efficiency are also dictated by the level and activity of the sphingomyelin hydrolytic enzyme, sphingomyelinase, present in the intestinal lumen.

The second step in the cholesterol uptake process is irreversible; which membrane-bound cholesterol is transported to the endoplasmic reticulum, where the cholesterol can be utilized for lipoprotein assembly and secretion to the plasma compartment. The influx of membrane cholesterol into the cell interior requires membrane sphingomyelin hydrolysis after the exogenous cholesterol is embedded into the cell membrane. A by-product generated from sphingomyelin hydrolysis is ceramide, which regulates intracellular lipid trafficking and thus plays an important role in controlling the rate of lipoprotein being produced by the intestine. Accordingly, enzymes that regulate intracellular ceramide concentration, including the carboxyl ester lipase secreted by the pancreas (which can be taken up by enterocytes into the cell interior), are also

important regulators in determining the rate of cholesterol absorption and the type of lipoproteins produced in response to a fat meal.

Cholesterol uptake by intestinal cells may also be an active process mediated by transporter proteins on the brush border membranes of the enterocytes. The strongest evidence supporting the existence of cholesterol-transporting proteins comes from early observations that cholesterol absorption follows second-order reaction kinetics, with a high- and a low-affinity component. Hydrolysis of proteins off the brush border membranes reduces cholesterol absorption to a first-order reaction with only a low-affinity component. The hypothesis of a protein-mediated cholesterol absorption process is also supported by experiments showing that cholesterol transport activity in liposomes can be reconstituted with proteins extracted from intestinal brush border membranes.

The identification of the putative cholesterol transporter on brush border membranes remains elusive. Although numerous laboratories have attempted to identify the cholesterol transporter, its identity has remained controversial. One promising lead comes from results reported by Helmut Hauser and colleagues in Zurich. Their studies suggest that the scavenger receptor class B type I protein (SR-BI) may function as a cholesterol transporter in the intestine. According to their results, cholesterol uptake by the intestinal-cell-like Caco-2 cells is suppressed by SR-BI antibodies. The binding of apolipoprotein A-I to SR-BI on the brush border membranes also inhibits cholesterol uptake from lipid donor particles. In support of these *in vitro* observations, SR-BI is found to be expressed along the gastrocolic axis of the intestine, with the highest level of expression in the proximal intestine and decreasing to minimal expression level in the ileum. Furthermore, SR-BI is localized to both the apical and the basolateral surfaces of enterocytes. Thus, the localization of SR-BI in the intestine is consistent with the hypothesis of its possible role in dietary cholesterol absorption. Unfortunately, this hypothesis cannot be substantiated, because mice lacking a functional SR-BI gene absorb dietary cholesterol with an efficiency similar to that observed in wild-type mice. Additionally, SR-BI has been shown *in vitro* to modulate cellular uptake of cholesteryl esters and triglycerides in addition to facilitating unesterified cholesterol transport. Neither cholesteryl esters nor triglycerides can be absorbed prior to their digestion by lipases. Thus, the precise role of SR-BI in the intestine remains uncertain.

The controversy surrounding the possible existence of cholesterol transport proteins responsible for dietary cholesterol absorption may stem from differences in interpreting results obtained from *in vivo* and *in vitro* studies. In *in vitro* experiments with enterocyte cell culture or brush border membranes, the cholesterol substrate is usually supplied as vesicles and micelles. Importantly, the substrate concentration is usually kept at a low concentration to maintain cell and membrane integrity. At low concentrations, the higher affinity but lower capacity process predominates in a second-order reaction, thus allowing for the determination of a protein-mediated high-affinity process. In contrast, under *in vivo* conditions, in which there is constant bile flow with its accompanying cholesterol-containing micelles, substrate concentration is relatively high. The higher concentration of cholesterol substrate in the intestinal lumen will favor uptake through the higher capacity, though lower affinity, passive transport process. The failure to observe protein-mediated cholesterol absorption *in vivo*, and the lack of effect on cholesterol absorption by specific transporter knockout mice, may be related to the concentration of the substrate used. In this regard, an intestinal cholesterol transporter may have only an auxiliary function in cholesterol absorption when cholesterol concentration in the intestinal lumen is low. Alternatively, the transporter may target the cholesterol to a specific intracellular compartment in which it can be utilized efficiently for lipoprotein assembly and export. These two possibilities remain to be resolved.

Another line of evidence used previously to support the concept of a protein-mediated cholesterol absorption process is based on observations of the selectivity of cholesterol absorption. Plant sterols, including stigmatosterol, sitosterol, and sitostanol, have chemical structures very similar to that of cholesterol and yet are not absorbed by the intestine. These plant sterols and cholesterol have similar hydrophobicities and are expected to diffuse through the plasma membrane with similar efficiencies. Accordingly, the passive transport process alone cannot account for the selectivity in cholesterol absorption without the concomitant absorption of the plant sterols. However, recent studies have shed light on the selectivity of the cholesterol absorption process. These studies were based on investigations of sitosterolemic patients, who can absorb plant sterols with high efficiency. Results of these studies reveal that plant sterols can indeed diffuse into intestinal cells. The selectivity of cholesterol absorption is due to adenosine triphosphate (ATP)-binding cassette (ABC) proteins, which preferentially pump plant sterols through the apical membranes back to the intestinal lumen in an ATP-dependent mechanism. Sitosterolemic patients have inherited an autosomal recessive mutation in the gene encoding either the ABCG5 or ABCG8

protein. The mutated ATP transporter cannot excrete the plant sterols from intestinal mucosal cells, thereby resulting in plant sterol absorption and secretion with lipoproteins into the circulation. Thus, the normal function of the ABCG5 and ABCG8 proteins is the ATP-dependent excretion of plant sterols back into the lumen, prevention of further plant sterol processing, and sterol transport to the circulation. Interestingly, mutations in either of the genes for ABCG5 and ABCG8 also result in increased cholesterol absorption. Whether the increased cholesterol absorption is due to a direct effect of the normal function of the ABCG5 and ABCG8 proteins on cholesterol efflux or to an indirect effect due to modification of luminal lipid composition remains to be determined.

Another ABC transporter protein that has been implicated to play a role in regulation of cholesterol absorption is ABCA1. This protein was initially identified as a liver protein that regulates high-density lipoprotein (HDL) levels. Tangier disease, which results in the absence of HDL and familial HDL deficiency, is a direct result of ABCA1 gene mutations. Factors that activate ABCA1 gene expression, such as the liver X receptor (LXR) agonist rexinoids, are effective inhibitors of dietary cholesterol absorption in rodents. The latter observation led to the suggestion that ABCA1 may regulate cholesterol absorption by limiting the amount of cholesterol taken up by intestinal cells. However, it is important to note that ABCA1-deficient mice absorb cholesterol with normal efficiency. Moreover, ABCA1 has not been demonstrated to be present in enterocytes. Thus, the role of ABCA1 in cholesterol absorption may be indirect, and may be related to alterations in biliary lipid composition and/or other differences unrelated to lipid uptake by intestinal cells.

INTRACELLULAR CHOLESTEROL PROCESSING AND LIPOPROTEIN ASSEMBLY

Cholesterol entering the intestinal mucosal cells is unesterified and is packaged into lipoproteins as esterified cholesterol before secretion into the lymphatics. Approximately 70–80% of the cholesterol found in the lymph is esterified with fatty acids. Thus, cholesterol reesterification is an important component of the cholesterol absorption process. The enzyme responsible for cholesterol esterification in the intestine has been debated for a number of years. Some evidence suggests that intestinal cholesterol esterification is mediated by the pancreatic carboxyl ester lipase, whereas other evidence suggests that this process is mediated by another enzyme, acyl coenzyme A : cholesterol acyltransferase (ACAT). The discrepancy has been resolved based on studies with the carboxyl ester lipase-deficient mice. As discussed previously, these animals absorb dietary cholesterol with normal efficiency. Cholesterol secreted as chylomicrons in the carboxyl ester lipase knockout mice is mainly esterified with fatty acids, similar that observed in wild-type mice. Thus, this enzyme appears to play at best a supportive role for cholesterol esterification in the intestine.

Attention has also been focused on the role of ACAT in intestinal cholesterol esterification. Progress in definitively proving the involvement of ACAT in intestinal cholesterol absorption has been hampered for many years due to the difficulty in purification of this enzyme. The effect of ACAT inhibitors in blocking intestinal cholesterol esterification and dietary cholesterol absorption varies in different experiments. The discrepancy in these studies has been partially resolved by molecular biology tools, which show the presence of at least two distinct ACAT genes in the mammalian genome. The first ACAT gene, called *ACAT1*, is expressed primarily in adrenals and in macrophages, with minimal expression in liver and intestine. Thus, deficiency in ACAT1 does not have any effect on cholesterol esterification in the latter tissues. Accordingly, ACAT1-deficient mice also absorb dietary cholesterol with normal efficiency, as expected. Similarly, inhibitors that are designed based on suppression of ACAT1 activity may also have less effect on intestinal cholesterol absorption. The second ACAT gene, *ACAT2*, encodes the ACAT enzyme that is primarily responsible for cholesterol esterification in liver and in intestine. Knockout of the *ACAT2* gene in mice results in animals that are resistant to diet-induced hypercholesterolemia and gallstone formation. Interestingly, although the resistance to diet-induced hypercholesterolemia may be related to reduction of cholesterol absorption when the animals are fed a cholesterol-enriched diet, no difference in cholesterol absorption between wild-type and ACAT2-deficient mice can be observed when the animals are fed normal mouse chow. These latter results suggest that although ACAT2 is essential for intestinal cholesterol esterification and efficient cholesterol absorption when fed a high-cholesterol diet, other enzymes, possibly the carboxyl ester lipase or ACAT1, may substitute for ACAT2 in mediating intestinal cholesterol esterification, supporting cholesterol absorption under normal conditions.

GENETICS OF CHOLESTEROL ABSORPTION

The complexity of the dietary cholesterol absorption process, which involves the participation of a multitude

of different proteins, suggests possible genetic influence on regulating the amount of cholesterol absorbed by different individuals. A number of studies have provided evidence to support this hypothesis. First, variations in cholesterol absorption efficiency account for some of the differences between hypo- and hyperresponding rabbits after eating a cholesterol-rich diet. Second, cholesterol uptake by human intestinal biopsy samples can also be divided into three groups, with low, medium, and high rates of cholesterol absorption. Third, genetic differences in cholesterol absorption efficiency have also been demonstrated in nonhuman primates. Although these studies have failed to distinguish genetic and environmental factors as the determining factor in the observed difference in cholesterol absorption, more recent studies with inbred mice have added more strength to the concept of genetic regulation of cholesterol absorption. The various inbred strains of mice are established by repeated inbreeding such that the genetic composition within each strain is identical. Analysis of inbred mice has been used in the past to identify successfully novel genes important for determining atherosclerosis susceptibility. Using similar approaches, by screening for cholesterol absorption efficiency among the various inbred strains of mice housed and fed under identical conditions, the genetic influence of cholesterol absorption efficiency is now well established. The data clearly suggest the involvement of multiple genes in dictating cholesterol absorption efficiency.

Identification of the gene loci that dictate cholesterol absorption efficiency is now underway. The most extensive work thus far is being carried out in the laboratory of David Russell in Dallas and in Martin Carey's laboratory in Boston. Both laboratories use the quantitative trait loci approach to tentatively identify several gene loci that may have influence on cholesterol absorption efficiency. One of these loci is called *Chab1* (cholesterol absorption gene 1). This locus is localized to mouse chromosome 2 and includes genes for the ABC transporter protein ABCB5 and for the liver X receptor-α. The latter gene encodes a transcription factor that regulates expression of lipid metabolism genes. Other putative cholesterol absorption gene loci have been identified on mouse chromosomes 1, 5, 6, 10, and 15. These gene loci may be linked to the *ACAT2* gene and to genes that participate in bile acid metabolism. It remains to be determined whether any of these previously identified genes are directly responsible for the genetic influence on cholesterol absorption, or whether they are only situated in close proximity to authentic cholesterol absorption genes.

The list of candidate cholesterol absorption genes identified by the quantitative trait loci approach to date does not include several genes thought to be involved in the overall cholesterol absorption process. For example, genes for the putative cholesterol transporters SR-BI, ABCA1, ABCG5, and ABCG8 do not map to the absorption gene intervals identified to date. Likewise, the genes in the bile acid synthetic pathway, *Cyp7α1*, *Cyp7β1*, and *Cyp39a1*, do not map to the currently identified putative cholesterol absorption gene loci. These results do not imply that these genes are not involved with cholesterol absorption, but rather that there may be little or no interstrain variation of these genes in terms of their role in cholesterol absorption. The limited screening of inbred strains of mice to date also reduces the power of the quantitative trait loci approach. Nevertheless, these studies point to possibility of additional novel genes that may influence cholesterol absorption efficiency. The identification in these loci of the precise genes that participate in regulation of cholesterol absorption will provide valuable information toward understanding individual differences in cholesterol absorption efficiency and responses to dietary cholesterol.

INHIBITORS OF CHOLESTEROL ABSORPTION

The impact of dietary cholesterol absorption on plasma cholesterol levels, and hence risk of coronary heart disease, has led to increasing interest in lowering cholesterol absorption as a treatment strategy to reduce plasma cholesterol level. Because of the complexity of the cholesterol absorption process, numerous pharmaceutical targets have been designed in attempts to suppress cholesterol absorption. The most effective treatment to date is the use of bile acid sequestrants such as cholestyramine, which binds to bile salts in the intestinal lumen. As a consequence of binding, cholesterol solubility is reduced and the dietary cholesterol is excreted from the body instead of being absorbed by intestinal mucosal cells. Unfortunately, bile salt sequestrants are not well tolerated, thus limiting their use for treatment of hypercholesterolemia.

The ACAT enzymes, responsible for cholesterol esterification in the intestine, comprise another pharmaceutical target for suppression of cholesterol absorption. Several ACAT inhibitors with high potency for inhibiting cholesterol esterification in cultured cells have been synthesized by the pharmaceutical industry. Some of these compounds have been found also to be effective in suppressing cholesterol absorption in

animal models when the animals were fed a cholesterol-enriched diet. However, most of the ACAT inhibitors are not effective in inhibiting cholesterol absorption under more normal dietary conditions. Thus, results of ACAT inhibitor studies mimic those observed with ACAT2 knockout mice showing that ACAT-mediated cholesterol absorption may only be vital for cholesterol absorption during high-cholesterol feeding. Importantly, clinical trials of ACAT inhibitors in humans have been disappointing, and none of the compounds have shown significant reduction of cholesterol absorption and suppression of diet-induced hypercholesterolemia in human volunteers.

More recent studies have revealed the potential of two new classes of compounds in suppression of cholesterol absorption. One class of compounds selectively inhibits cholesterol absorption through interaction with specific sites on the brush border membranes. These inhibitors include the sterol glycosides synthesized by Merck Research Laboratories and ezetimibe, produced by Schering-Plough. The latter compound has also been shown to be effective in suppressing diet-induced atherosclerosis in a rodent model. Both of these compounds bind to brush border membranes in a saturable manner, suggesting the possibility that they may be interacting with cholesterol transporters in suppressing cholesterol absorption. It is also possible, however, that these compounds interact with specific lipid components in the brush border membrane, thereby preventing cholesterol trafficking into and out of the plasma membrane and/or its transport to an intracellular compartment where it can be used for lipoprotein synthesis and secretion. Thus, the exact mechanism by which these compounds suppress cholesterol absorption in animal models remains unknown.

The second class of compounds that show promise in suppressing cholesterol absorption includes reagents that activate ABC transporter expression in the intestine. The goal of the design of this class of compounds is to increase cholesterol export to the intestinal lumen, where it can be excreted out of the body. The design takes advantage of observations that agonists for the oxysterol LXRs and the bile acid farnesoid X receptor (FXR) are potent activators of ABC transporter gene expression. Because LXRs and FXR form obligate heterodimers with retinoid X receptors (RXRs) for their activities, recent studies have used an RXR-specific agonist to demonstrate feasibility of increasing ABC transporter expression in intestine as a mechanism to suppress cholesterol absorption in rodents. However, RXR interacts with a number of different transcriptional factors and can regulate genes that span several metabolic pathways. Thus, specific LXR or FXR agonists may be preferred as more specific regulators of cholesterol transport. The continued development of this class of compounds and testing for the efficacy in reducing cholesterol absorption in human subjects will be valuable.

Acknowledgments

Research on cholesterol absorption performed in the author's laboratory is supported by a Program Project Grant (PO1 DK54504) from the National Institutes of Health.

See Also the Following Articles

Apoproteins • Barrier Function in Lipid Absorption • Fat Digestion and Absorption • Gallstones, Pathophysiology of • Lipoproteins • Small Intestine, Absorption and Secretion

Further Reading

Beynen, A. C., Katan, M. B., and van Zutphen, L. F. M. (1987). Hypo- and hyperresponders: Individual differences in the response of serum cholesterol concentration to changes in diet. *Adv. Lipid Res.* **22**, 115–171.

Carey, M. C., Small, D. M., and Bliss, C. M. (1983). Lipid digestion and absorption. *Annu. Rev. Physiol.* **45**, 651–677.

Chawla, A., Repa, J. J., Evans, R. M., and Mangelsdorf, D. J. (2001). Nuclear receptors and lipid physiology: Opening the X-files. *Science* **294**, 1866–1870.

Field, F. J., Kam, N. T. P., and Mathur, S. N. (1990). Regulation of cholesterol metabolism in the intestine. *Gastroenterology* **99**, 539–551.

Mansbach, C. M., Tso, P., and Kuksis, A. (eds.) (2001). "Intestinal Lipid Metabolism." Kluwer Academic/Plenum Publ., New York.

Schwarz, M., Davis, D. L., Vick, B. R., and Russell, D. W. (2001). Genetic analysis of intestinal cholesterol absorption in inbred mice. *J. Lipid Res.* **42**, 1801–1811.

Wang, D. Q. H., Paigen, B., and Carey, M. C. (2001). Genetic factors at the enterocyte level account for variations in intestinal cholesterol absorption efficiency among inbred strains of mice. *J. Lipid Res.* **42**, 1820–1830.

Wilson, M. D., and Rudel, L. L. (1994). Review of cholesterol absorption with emphasis on dietary and biliary cholesterol. *J. Lipid Res.* **35**, 943–935.

Circulation, Overview

D. Neil Granger* and Peter R. Kvietys†

*Louisiana State University Health Sciences Center, Shreveport, and †London Health Sciences Center and University of Western Ontario, Canada

acinus Smallest functional unit of the liver, which receives blood flow from the portal vein and hepatic artery.
angiogenesis Growth of new blood vessels.
ascites Collection of serous fluid (derived from liver and/or intestine) in the peritoneal cavity.
ischemia A reduction in blood flow of such magnitude as to cause tissue hypoxia.
reperfusion The restoration of blood flow following a period of ischemia.

The gastrointestinal (GI) circulation accounts for over 25% of cardiac output under resting conditions. Blood flow to the GI tract is dynamic, however, with tissue-specific increases caused by the ingestion of a meal and profound reductions in flow elicited by stresses that threaten to deprive the vital organs (e.g., brain, heart) of their normal blood supply. Although the different organs perfused by the GI circulation share common stimuli and mechanisms of blood flow regulation, significant differences related to tissue function have been described. As seen in most other vascular beds, impaired blood perfusion of GI organs can lead to profound organ dysfunction and tissue injury. The stomach, liver, intestine, and pancreas provide excellent examples of the unique and diverse anatomical features and regulatory mechanisms that distinguish the GI circulation from other regional vascular beds.

GASTRIC CIRCULATION

Introduction

Gastric blood flow plays an important role in sustaining the normal physiologic functions of the stomach and it helps to protect the gastric mucosa against ulcer formation. Intrinsic regulatory mechanisms ensure that blood flow is adjusted to meet the energy-demanding processes of gastric secretion and motility. Gastric blood flow also helps to maintain a barrier against back-diffusion of lumenal acid, thereby preventing mucosal damage and ulceration. Impairment of gastric blood flow renders the mucosa vulnerable to the damaging actions of gastric juice as well as ingested agents, such as ethanol, aspirin, and bacteria (e.g., *Helicobacter pylori*).

Anatomy

Oxygenated blood is provided to the stomach via the celiac artery, with deoxygenated blood drained by the portal vein. Small branches of the celiac artery give rise to arterioles in the external muscle layers and the submucosa. Some of the submucosal arterioles pierce the muscularis mucosae to produce the capillary network that supplies the mucosa. Hence, the tone of the submucosal arterioles determines the magnitude of mucosal blood flow. The mucosal capillaries drain into a central vein that begins just beneath the surface epithelial cells. These venules coalesce to form a dense venous plexus in the submucosa. The submucosal venous drainage penetrates the external muscle layers where additional venous blood is provided by the muscle microvasculature.

Hemodynamics

The parallel-coupled capillary networks of the gastric muscular layer and the mucosa are under separate control, responding independently to tissue metabolism, other local factors, and extrinsic neural input. Between meals, blood flow in the mucosal layer is approximately six times higher than that of the muscle layer, and approximately 75% of total gastric blood flow is distributed to the mucosa, with 25% directed to the muscle layer. This intramural distribution of blood flow is altered when either of the two layers becomes functionally active; i.e., when the mucosa is stimulated to produce acid, mucosal blood flow (and its percentage of total flow) preferentially increases.

Blood Flow Regulation

Gastric blood flow is controlled by neural, humoral, and metabolic factors. Sympathetic activation elicits reductions in total gastric blood flow and mucosal flow through arteriolar constriction. Parasympathetic nerves exert a tonic vasodilatory influence on gastric arterioles,

 Copyright 2004, Elsevier (USA). All rights reserved.

with vagotomy resulting in a reduction in blood flow. Gastrin and histamine, both powerful stimulants of gastric acid secretion, increase mucosal blood flow. Oxygen consumption by the stomach increases in proportion to acid production. Changes in both blood flow and oxygen extraction assist in meeting the demand for additional oxygen in the acid-secreting stomach. When gastric blood flow is reduced, however, acid secretion and blood flow may fall in parallel, owing to the fact that the rate of oxygen delivery to the parietal cells is limited by blood flow.

Pathophysiology

The gastric microcirculation contributes to gastric ulcer formation in several ways. Capillary transport of parietal cell-derived bicarbonate normally plays an important role in protecting the surface epithelium against acid-induced injury and ulceration. The mucosal capillaries originating near the gastric pits transport bicarbonate toward the mucosal surface, where it can diffuse into the interstitial compartment beneath the surface epithelial cells. The latter cells transport bicarbonate into the gastric lumen where it can buffer lumenal acid. As a consequence, there is an inverse relationship between gastric mucosal injury and the rate of vascular bicarbonate delivery to the mucosal surface.

Arteriolar constriction and capillary plugging with thrombi or activated leukocytes often accompany the ulceration process and are thought to promote mucosal injury by rendering the tissue ischemic and vulnerable to necrosis. Repair of gastric mucosal ulceration relies on the restoration of blood flow and on the growth of new blood vessels (angiogenesis), which usually sprout from venules.

HEPATIC CIRCULATION

Introduction

The hepatic circulation has many unique features compared to other vascular beds in the splanchnic circulation. There is the unusual presence of both an arterial input and a venous input that combine to deliver a large fraction of cardiac output to the liver. The acinar arrangement of the microvasculature creates a series of microenvironments within the organ that imparts a significant spatial resolution for specific biosynthetic, biotransformation, and detoxification functions of the liver. These and other unique features of the hepatic circulation serve to ensure that the varied and complex functions of the liver are maintained at levels that are optimal for the whole animal.

Anatomy

The liver receives blood from two sources: fully oxygenated blood is derived from the high-pressure, high-resistance hepatic artery and partially oxygenated blood comes from the low-pressure, low-resistance portal vein. Within the liver, these vessels give rise to numerous smaller vessels, called hepatic arterioles and terminal portal venules, which supply a small mass of parenchyma called the liver acinus. The mixture of arterial and venous blood from these terminal vessels drains into the sinusoids, which constitutes the capillary network of the liver. The sinusoidal capillaries are highly permeable to water and plasma proteins. The sinusoids radiate toward the periphery of the acinus, where they connect with the terminal hepatic venules and ultimately into progressively larger branches of hepatic veins and the inferior vena cava.

Hemodynamics

Total hepatic blood flow accounts for approximately 25% of resting cardiac output (1250–1500 ml/min) in the resting adult human. The portal vein supplies the liver with 70–75% of its blood and the hepatic artery provides the remaining 25–30%. Because of the higher oxygen content of arterial blood, the hepatic artery and portal vein contribute roughly equal amounts of oxygen to the liver in the fasting state. The mean blood pressure in the hepatic artery is approximately 90 mm Hg and that in the portal vein is approximately 10 mm Hg, resulting in a sinusoidal capillary pressure of 2–3 mm Hg. The low pressure within these highly permeable capillaries serves to minimize excessive loss of fluid and protein into the liver interstitium and the subsequent leakage of this plasma filtrate into the peritoneal cavity (ascites fluid).

Blood Flow Regulation

Increased blood flow in the portal vein leads to an increased hepatic arterial resistance and a reduction in portal vein flow produces hepatic arterial dilation. Although this reciprocal relationship between hepatic arterial and portal vein blood flows tends to maintain a constant total blood flow through the liver, flow in one system usually cannot fully compensate for the reduction in blood flow in the other system. However, the liver is more effective at maintaining a constant oxygen consumption because the extraction of oxygen from hepatic blood is very efficient. The small diffusion distances for oxygen transport between blood and hepatocytes accounts for this highly efficient oxygen extraction. Hepatic blood flow increases following

ingestion of a meal, largely due to an increased flow in the portal vein.

Capacitance Function

The liver represents the most important blood reservoir in human. It contains approximately 15% of the total blood volume of the body. Hepatic blood volume can nearly double when right atrial pressure is elevated. The capacitance function of the liver plays an important role during hemorrhage. After moderate blood loss, activated sympathetic nerves constrict the hepatic venules and expel enough blood to compensate for as much as 25% of the hemorrhage.

INTESTINAL CIRCULATION

Introduction

The blood circulation plays an important role in the support of intestinal functions, such as propulsion of chyme and assimilation of ingested nutrients. Intrinsic regulatory mechanisms allow the intestine to adjust the distribution of blood flow between the muscular and mucosal layers in accordance with local metabolic needs. An extensive network of collateral channels within and external to the gut wall helps to ensure adequate intestinal blood flow.

Anatomy

The small intestine receives blood primarily from the numerous vascular arcades arising from the superior mesenteric artery. The arterial arcades are connected by extensive collateral channels both within the mesentery (extramural) and within the bowel wall proper (intramural). The major arterial vessels supplying the mucosa, submucosa, and muscularis emerge from an arterial plexus within the submucosa. The submucosa proper has a very sparse vascular supply. The arterioles entering the muscle layer branch into capillaries running parallel to the muscle fibers. The mucosal villi are supplied by a single arteriole running centrally to the tip where the vessel branches into a fountain-like pattern of capillaries that drain into the centrally located vein. The venous drainage of the mucosa and muscularis empties into large veins within the submucosa. These veins enter the mesentery in parallel to the arterial arcades of the superior mesenteric artery and eventually drain into the portal vein.

Hemodynamics

Intestinal blood flow accounts for 10–15% of the resting cardiac output (500–750 ml/min) in the adult human. There appears to be an oral-to-anal gradient in blood flow (milliliters per gram of tissue) along the small intestine. In the resting state, approximately 65% of the total intestinal blood flow is directed to the mucosa, 25% to the muscularis, and the remainder to the submucosa. This distribution of flow within the bowel wall is usually attributed to the greater metabolic demand of the mucosa. Stimulation of mucosal epithelial transport processes favors improved mucosal perfusion, whereas enhanced motor activity redistributes blood flow to the muscle layers.

Blood Flow Regulation

Extrinsic control of intestinal blood flow is exerted by neural and humoral factors. Activation of parasympathetic nerves usually results in vasodilation (increased blood flow) mediated by acetylcholine. Sympathetic nerve stimulation elicits vasoconstriction (decreased blood flow) that is mediated by norepinephrine. This α-adrenergic vasoconstriction is short-lived because intestinal arterioles escape from the constrictor influence of norepinephrine, resulting in partial restoration of normal blood flow (autoregulatory escape). Local release of adenosine appears to mediate this autoregulatory escape. Hormones such as vasoactive intestinal peptide, cholecystokinin, and secretin can induce vasodilation and increase blood flow, whereas angiotensin II and vasopressin are potent constrictors of intestinal arterioles. Indeed, a large proportion of basal vascular tone in the intestine can be attributed to circulating angiotensin II and vasopressin.

Intrinsic control of intestinal blood flow is mediated by both metabolic and nonmetabolic factors. Ingestion of a meal results in an increase in both intestinal oxygen consumption and blood flow. The postprandial hyperemia is directly coupled to the increase in intestinal oxygen consumption. For any given increase in oxygen consumption, the greater the initial oxygen extraction, the greater the postprandial hyperemia. If the initial oxygen extraction is low, then the postprandial hyperemia is minimal and the increased oxygen demand is met primarily by an increase in oxygen extraction. The opposite holds if the initial oxygen extraction is high.

The postprandial hyperemia is confined to that segment of intestine directly exposed to chyme; segments distal to the bolus of chyme have normal resting blood flow. Of the hydrolytic products of food digestion, lumenal glucose and oleic acid are capable of eliciting an intestinal hyperemia. Intralumenal glucose presumably elicits a hyperemia due to stimulation of absorptive processes, since 2-deoxyglucose (which is not absorbed) does not elicit a hyperemia. The glucose-induced

hyperemia is mediated by metabolic factors, such as low tissue pO_2, and adenosine release. The same metabolic factors contribute to the oleic acid-induced functional hyperemia; however, a portion of the hyperemia can be attributed to oleic acid-induced irritation of the mucosa, which is linked to local release of vasoactive intestinal peptide. The importance of active transport of nutrients to the postprandial hyperemia is best exemplified by the differential responses of the jejunum and ileum to lumenal bile or bile salts. In the jejunum, bile does not elicit a hyperemia, whereas in the ileum (where bile salts are actively transported), lumenal bile produces a profound hyperemic response.

Pathophysiology

Occlusion of a major intestinal artery does not result in the expected reduction in intestinal blood flow. In adult animals, occlusion of a branch of the superior mesenteric artery results in only a 30–50% reduction in intestinal blood flow, which is attributed to the extensive network of intramural and extramural collateral channels. In neonatal animals, which have less developed collateral channels, a similar arterial occlusion reduces intestinal blood flow by 70%. This may explain why the neonatal intestine is more vulnerable to ischemic necrosis than adult intestine.

Reperfusion of the ischemic intestine exacerbates the microvascular dysfunction and tissue injury incurred during ischemia. The reperfusion-induced intestinal pathology is comparable to that observed during an intense inflammatory response. The reintroduction of oxygen upon reperfusion results in the local formation of oxygen radicals, which in turn initiate a series of events that promote the recruitment and activation of neutrophils. Extravasated neutrophils release a variety of proteases and reactive oxygen metabolites that mediate the mucosal dysfunction and epithelial necrosis induced by ischemia and reperfusion.

PANCREATIC CIRCULATION

Introduction

The pancreas has both exocrine and endocrine functions. The endocrine functions of the pancreas are localized to the islet cells and the secretory functions are localized to the ducts and acini. The human pancreas can secrete approximately 1 liter of juice every 24 h. The water and electrolytes delivered to pancreatic juice are ultimately derived from the blood. Thus, the pancreas controls its blood flow to cope with the demands of basal secretion and makes adjustments (arteriolar dilation) to ensure an adequate supply of blood flow to maintain maximal secretion. The hormones secreted by the islet cells modulate exocrine function.

Anatomy

The pancreas is supplied by arterial blood from two sources: the superior mesenteric artery and the celiac artery. The arteries entering the gland proper are the inferior pancreaticoduodenal artery (from the superior mesenteric artery) and the superior pancreaticoduodenal artery and branches of the splenic artery (from the celiac artery). Within the pancreas, arterioles supply the three functional components of the gland: the acini, ducts, and islets. The blood draining all three plexi drains into either the splenic vein or the superior mesenteric vein, which ultimately empty into the portal vein.

Portal Circulation

There is also an extensive portal system within the pancreas. Blood vessels draining the capillary plexus within the islets of Langerhans empty into the acinar and ductal plexi. Blood vessels draining the acinar capillary plexus perfuse the ductular plexus.

The functional significance of this intrapancreatic portal circulation is that the hormones secreted by the islets reach the exocrine portion of the gland (the acini and ducts) at very high concentrations. Virtually all of the peptides and hormones secreted by the islets exert an influence on pancreatic secretion. This can be attributed to the low molecular weight of these hormones and peptides, which allows them to readily filter across the fenestrated capillaries surrounding the acini and ducts.

Blood Flow Regulation

Pancreatic blood flow doubles after a meal. The kallikrein–kinin system, cholinergic and purinergic neurotransmitters, and gastrointestinal hormones (cholecystokinin and secretin) have all been implicated in postprandial hyperemia. The hyperemia-associated increase in pancreatic oxygen delivery appears to be linked to the increased metabolic demand imposed on the pancreas by the stimulated secretory processes. The postprandial hyperemia also provides the additional fluid that accompanies the enhanced electrolyte transport in pancreatic acini and ducts. The arteriolar dilation elevates pancreatic capillary hydrostatic pressure, which drives fluid across the capillaries into the interstitial spaces surrounding the ductal epithelium. The importance of these vascular adjustments

postprandially is underscored by the fact that during periods of maximal stimulation, the pancreas can secrete its weight in juice in less than 30 min.

See Also the Following Articles

Hepatic Circulation • Liver, Anatomy • Pancreas, Anatomy • Small Intestine, Anatomy • Stomach, Anatomy

Further Reading

Crissinger, K. D., and Granger, D. N. (1999). Gastrointestinal blood flow. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), Chap. 23, pp. 519–546. Lippincott, Williams and Wilkins, Philadelphia.

Granger, D. N., Grisham, M. B., and Kvietys, P. R. (1994). Mechanisms of microvascular injury. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Chap. 49, pp. 1693–1722. Raven Press, New York.

Granger, D. N., Kvietys, P. R., Korthuis, J., and Premen, A. J. (1989). Microcirculation of the intestinal mucosa. *In* "Handbook of Physiology, Section 6, The Gastrointestinal System, Vol. 1, Motility and Circulation" (J. D. Wood, ed.), Part 2, Chap. 39, pp. 1405–1474. Oxford University Press, New York.

Greenway, C. V., and Lautt, W. W. (1989). Hepatic circulation. *In* "Handbook of Physiology, Section 6, The Gastrointestinal System, Vol. 1, Motility and Circulation" (J. D. Wood, ed.), Part 2, Chap. 41, pp. 1519–1564. Oxford University Press, New York.

Holzer, P., Livingston, E. H., and Guth, P. H. (1994). Neural, metabolic, physical, and endothelial factors in the regulation of gastric circulation. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Chap. 35, pp. 1311–1330. Raven Press, New York.

Kaplowitz, N. (1992). "Liver and Biliary Diseases." Williams and Wilkins, Baltimore, MD.

Kvietys, P. R., Granger, D. N., and Harper, S. L. (1989). Circulation of the pancreas and salivary glands. *In* "Handbook of Physiology, Section 6, The Gastrointestinal System, Vol. 1, Motility and Circulation" (J. D. Wood, ed.), Part 2, Chap. 42, pp. 1565–1595. Oxford University Press, New York.

Lautt, W. W. (1985). Mechanism and role of intrinsic regulation of hepatic arterial blood flow: Hepatic arterial buffer response. *Am. J. Physiol.* **249**, G549–G556.

Lautt, W. W., and Macedo, M. P. (1997). Hepatic circulation and toxicology. *Drug Metab. Rev.* **29**, 369–395.

Tepperman, B. L., and Jacobson, E. D. (1994). Circulatory factors in gastric mucosal defense and repair. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Chap. 36, pp. 1331–1352. Raven Press, New York.

Cirrhosis

IAN R. WANLESS
University of Toronto

adhesion Close approximation of portal tracts and hepatic veins. Adhesions indicate that loss of parenchyma has occurred, usually through the process of extinction.
parenchymal extinction Process of focal loss of hepatocytes, usually by a mechanism of ischemia. An extinction lesion is the histologic product of this process.
regression Process of return of liver parenchyma toward a normal histologic appearance, as in "regression of cirrhosis."
remnant Residual visible structure of a portal tract or hepatic vein after injury and repair.
septum Linear fibrous array that subdivides the parenchyma. Septa are actually planes when seen in three dimensions.
stellate cell Facultative fibroblast residing in the subendothelial space of sinusoids.

Cirrhosis is defined anatomically as a condition in which fibrous septa subdivide the hepatic parenchyma into residual masses with a nodular configuration. The septa are widespread throughout the liver. Cirrhosis develops incrementally with the accumulation of local scars. Thus, in early or focal disease, the diagnosis of cirrhosis depends on arbitrary quantitative limits.

PATHOGENESIS OF CIRRHOSIS

There are two basic mechanisms for the deposition of collagen, stellate cell activation and parenchymal extinction, and each is associated with a distinct histologic pattern of fibrosis. Both are generally initiated by hepatocellular injury.

Stellate Cell Activation

Stellate cells are quiescent fibroblasts that normally reside in sinusoidal walls within the subendothelial space of Disse. Stellate cells are activated by inflammatory mediators to commence collagen synthesis. Simultaneously, there occurs activation of tissue metalloproteinases that degrade collagen. When stellate cells are activated in low-grade disease, the collagen is deposited as delicate fibers in the sinusoidal walls, causing what is known as sinusoidal, or pericellular, fibrosis. This is most easily appreciated in the perivenular regions (Rappaport zone 3). It has been demonstrated in experimental animals that sinusoidal fibrosis is rapidly removed within weeks of the cessation of injury.

Parenchymal Extinction

Parenchymal extinction is defined as the loss of a contiguous group of hepatocytes (Fig. 1). Lesions may be small, involving a few hepatocytes, or larger, involving many acini or even a whole lobe. Extinction occurs

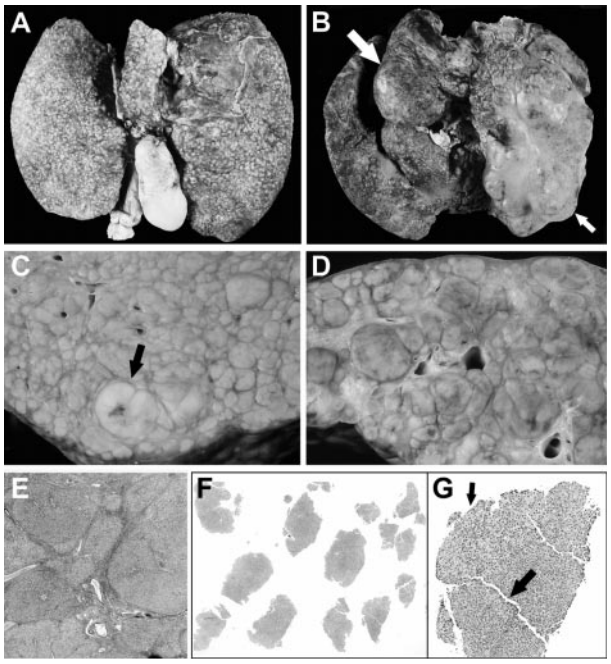


FIGURE 1 Pathology of cirrhosis. (A) Micronodular cirrhosis caused by hepatitis C virus. (B) Alcoholic cirrhosis with regional parenchymal extinction seen as small irregular right lobe with fibrous thickening of the capsule (small arrow). The caudate lobe is hypertrophied (large arrow). (C) Mixed cirrhosis from hepatitis C with a small hepatocellular carcinoma (arrow). (D) Alcoholic cirrhosis with an irregular macronodular pattern. There are delicate septa on the right and broad septa on the left. Macronodular cirrhosis caused by alcohol is found after a long period of inactive disease and remodeling. (E) Cirrhosis with uniform septa of moderate thickness (Masson trichrome). (F) Needle biopsy of a cirrhotic liver showing almost no fibrosis (Masson trichrome). (G) Close-up of F, showing that fracture planes along septa are smooth and curved (small arrow) whereas fractures in nonfibrotic areas (large arrow) are irregular.

 Copyright 2004, Elsevier (USA). All rights reserved.

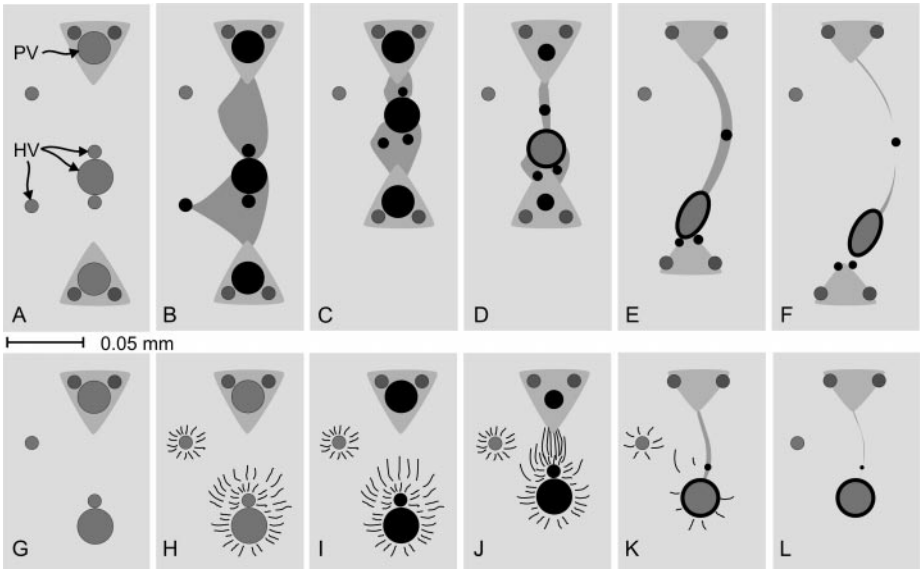


FIGURE 2 Diagrammatic depiction of tissue remodeling in chronic hepatitis (A–F) and in alcoholic liver disease (G–L) during the development and regression of cirrhosis. (A and G) Normal acini; the sequence of events leading to small regions of parenchymal extinction is shown in panels B–F and H–L. Obstructed veins are shown as black circles. (B) Obliteration of small portal and hepatic veins occurs early in the development of cirrhosis in response to local inflammatory damage. The supplied parenchyma becomes ischemic. (C and D) Ischemic parenchyma shrinks and is replaced by fibrosis (process of extinction). The shrinkage is accompanied by close approximation of adjacent vascular structures. (E) Septa are deformed and stretched by unsymmetrical expansion of regenerating hepatocytes. (F) Fibrous septa are resorbed. They become progressively thinner and then perforate before disappearing. Small residual tags may extend from portal tracts. Trapped portal structures and hepatic veins are released from the septa and are recognizable as deformed remnants that are irregularly distributed. Residual hepatic veins are often described as ectopic. Note the absence of portal veins. In alcoholic disease (G–L), the sequence of events may differ from that of other forms of chronic liver disease. (H) Sinusoidal fibrosis is often prominent, with a pericellular pattern of fibrosis prior to the development of parenchymal collapse. (I and J) Inflammation and fibrosis lead to hepatic and portal vein obliteration with secondary condensation of preformed sinusoidal collagen fibers into a septum. (K and L) After prolonged periods of inactivity, sinusoidal fibrosis and septa are resorbed. Modified with permission from Wanless *et al.* (2000) and Wanless (2003).

by a mechanism of ischemia secondary to obstruction of veins or sinusoids. After parenchyma collapses, adjacent regeneration compresses the collapsed regions into linear membranes recognized as septa. Large lesions are recognized by regions with closely approximated portal tracts separated only by collapsed stroma and fibrosis. Small lesions are identified by the close approximation of a small hepatic vein and its adjacent portal tracts (an adhesion).

A number of important consequences arise from the concept of parenchymal extinction: (1) although usually initiated by hepatocellular injury, extinction results when the circulation in a region fails, (2) each lesion of extinction has a natural evolution during the healing process, (3) cirrhosis is the accumulated result of a large number of independent and discrete lesions, and (4) the morphology of the cirrhosis is determined by the size and location of the vascular occlusion events as well as the degree of healing (regression) of each extinction lesion. The obstruction of small veins is a bystander event in response to local parenchymal injury. Thrombosis is most important in larger vessels, where local parenchymal inflammation is unlikely to damage the veins.

Natural History of Cirrhosis

The morphology of cirrhosis is determined by the amount of time that has elapsed from the onset of parenchymal extinction (Figs. 2 and 3). If active injury continues, new lesions of extinction will coexist with old lesions. If the primary disease is inactive and remote,

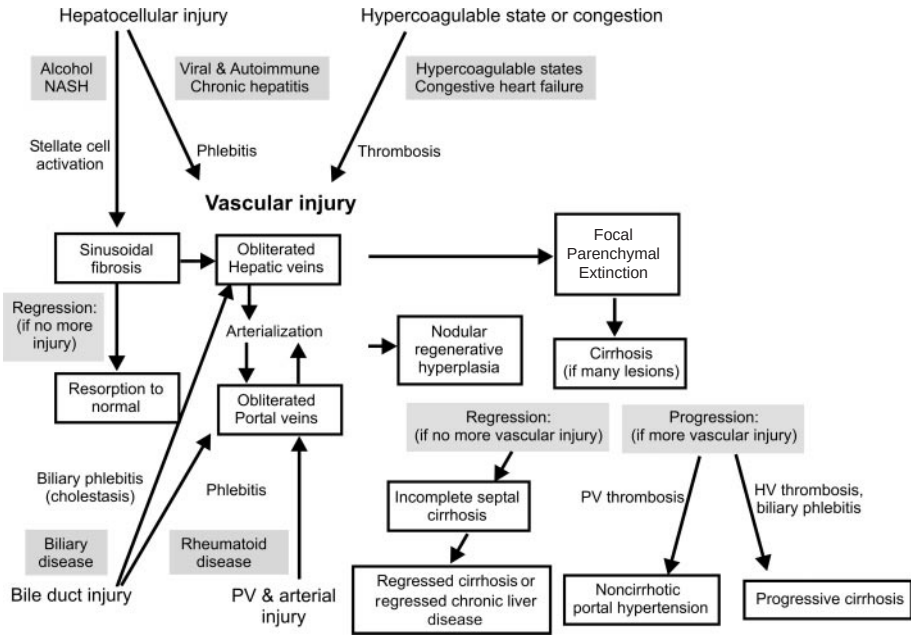


FIGURE 3 Summary of the factors that determine that natural history of chronic liver disease. There are five principal classes of disease leading to chronic liver disease (light gray boxes). Most patients with chronic liver disease have hepatocellular injury. This may lead to local activation of stellate cells and sinusoidal fibrosis that is largely reversible. Those patients developing obliteration of vessels, especially small hepatic veins (HV), develop lesions of parenchymal collapse (extinction) that heal as fibrous septa. When septa are numerous, the histologic features of cirrhosis are present. Biliary disease and rheumatoid disease are associated with predominant portal tract disease. Rheumatoid disease usually affects the portal vessels (PV), leading to multifocal atrophy of parenchyma with minimal or no fibrous septation, a condition recognized as nodular regenerative hyperplasia. In biliary disease, portal inflammation is an early event, leading to portal tract fibrous expansion and sometimes presinusoidal portal hypertension; zone 3 cholestasis occurs later, leading to bile salt injury of small hepatic veins. Thus, the order of vascular events is different but accumulation of extinction lesions eventually leads to cirrhosis. The outcome depends on the time course of disease activity (dark gray boxes). If injury ceases, regression occurs, because all extant lesions heal and new lesions do not appear. If injury continues, new lesions develop and there is progressive collapse and fibrosis, either from the continuing primary injury, secondary thrombotic events, or secondary bile salt injury to hepatic veins. NASH, Nonalcoholic steatohepatitis.

the liver will contain only lesions in late stages of repair. Fresh extinction lesions are represented by bridging necrosis in highly active hepatitis or milder lesions of focal congestion and apoptosis in mild hepatitis.

Once cirrhosis has developed, the hepatic blood flow is chaotic and sluggish. Portal vein flow may be biphasic or continuously retrograde. In addition, loss of anticoagulant function and prothrombotic effects of sepsis and cholestasis may contribute to the increased risk of thrombosis of the hepatic and portal vein. Portal vein thrombosis is found in up to 40% of cirrhotic livers examined at transplantation. Thrombosis of medium or large hepatic veins causes large regions of extinction to occur, explaining the markedly irregular capsular shapes found in a third of livers with late-stage cirrhosis (Fig. 1).

After primary liver disease goes into remission or is successfully treated, old extinction lesions (septa and adhesions) predominate. Fibrosis is progressively removed so that broad septa become delicate and delicate septa become incomplete and may even disappear. Frequent residual lesions are recognized as delicate fibrous spurs on portal tracts. Given sufficient time, micronodular cirrhosis may remodel into macronodular cirrhosis, incomplete septal cirrhosis, and eventually nearly normal livers. After portal tracts are released from septa, they often lack portal veins so that the post-regression liver is largely supplied by arteries. These

findings explain residual portal hypertension in patients with such livers. The major impediments to the reversibility of cirrhosis are continuing primary activity, secondary thrombosis, cholestatic decompensation, and the arterialized state of the liver.

ANATOMIC SUBTYPES

Several recognized patterns of cirrhosis can be understood, from this introduction, to depend on the severity of the underlying liver disease and the duration of inactivity prior to examination (Table I). The main types of cirrhosis are micronodular and macronodular; the former type has most nodules less than 3 mm in diameter and the latter type has most nodules greater that 3 mm in diameter. The World Health Organization (WHO) classification also defines a mixed category in which the nodules are both larger and smaller than 3 mm. Mixed cirrhosis is often found in primary biliary cirrhosis and primary sclerosing cholangitis. Incomplete septal cirrhosis is a highly regressed form of cirrhosis often associated with portal hypertension but normal hepatocellular function. Incomplete septal cirrhosis and noncirrhotic portal hypertension often represent late stages of regression. Portal vein thrombosis is often the complication that leads to portal hypertension in these cases, many of which would otherwise escape clinical attention.

Cirrhosis with regional parenchymal extinction, usually with severe micronodular cirrhosis, is a form of cirrhosis in which large contiguous regions of collapse and fibrosis have occurred (Fig. 1). This may be confused with postnecrotic cirrhosis, which is the fibrotic stage of severe acute hepatitis occurring with large contiguous regions of extinction. Because many of these patients have hepatic failure and jaundice, these livers are usually severely cholestatic.

TABLE I Anatomic Classification of Cirrhosis and Related Forms of Chronic Liver Disease

Micronodular cirrhosis
Macronodular cirrhosis
Mixed cirrhosis
Biliary cirrhosis
Pigmentary cirrhosis (hemochromatosis)
Cirrhosis with regional parenchymal extinction
Postnecrotic cirrhosis (subacute massive necrosis)
Incomplete septal cirrhosis[a]
Highly regressed cirrhosis[a]

[a] Noncirrhotic portal hypertension occurs with these forms.

EVALUATION OF LIVER SPECIMENS

Accuracy in staging liver disease depends on observer and sampling errors. Observer error can be minimized if the pathologist has complete knowledge of the clinical situation and a clear understanding of the natural history of the diseases within the differential diagnosis. Sampling error is impossible to avoid entirely, because the definition of cirrhosis requires involvement of the entire liver and the biopsy represents only a small portion of the total. Also, fibrosis in chronic liver disease is concentrated into fibrous septa that are not distributed uniformly. Sampling error depends on the character of the liver disease and is greatest in those diseases having a low density of diagnostic lesions in the tissue. In macronodular cirrhosis, the septa may be more than 1 cm apart, so that a small-needle biopsy may not have any septa. In contrast, a biopsy 1–2 mm in length may be sufficient for the diagnosis of micronodular cirrhosis. In one study, agreement of stage and corrected sinusoidal pressure was poor for biopsies less than 1.5 cm in length. Transjugular biopsies are generally smaller than percutaneous biopsies in both length and width. However, the advantage of a transjugular biopsy is the opportunity to obtain simultaneously a measurement of portal pressure.

Resection specimens should be examined for weight, color, thickened capsule, relative size of the lobes, regional parenchymal collapse, and patency of ducts, vessels, and any prosthetic stents. Uncomplicated cirrhosis is usually very uniform in appearance. Any heterogeneity should prompt extra sampling of vessels and ducts to discover the cause. Nodules with variance of size, color, or texture may be an indication of neoplasia. Dilated paraumbilical veins or severe congestion of nodules may correlate with Doppler evidence of reversed portal vein flow.

Staging Systems

Many staging systems are in wide use, including the METAVIR and Ishak systems, applicable for most forms of chronic liver disease, and the Scheuer and Ludwig systems for chronic biliary disease. These systems vary in the number of categories (0 to 4, 0 to 6, or 1 to 4) and the category definitions. These systems do not provide categories for the different severities of cirrhosis. The Laennec system attempts to standardize staging for all chronic liver disease on a scale of 0 to 4, with cirrhosis being grade 4 (Table II). An extension divides cirrhosis into subgrades 4A, 4B, and 4C, in recognition of the variable severity among cirrhotic livers. The definitions of each grade are based on the number and width of

TABLE II Laennec Scoring System for Grading Fibrosis in Liver Biopsies

Grade	Name	Criteria (septa thickness and number)	Descriptive examples
0	No definite fibrosis	—	—
1	Minimal fibrosis	+/−	No septa or rare thin septum; may have portal expansion or mild sinusoidal fibrosis
2	Mild fibrosis	+	Occasional thin septa; may have portal expansion or mild sinusoidal fibrosis
3	Moderate fibrosis	++	Moderate thin septa, up to incomplete cirrhosis
4A	Cirrhosis, mild, definite, or probable	+++	Marked septation with rounded contours or visible nodules; most septa are thin (one broad septum allowed)
4B	Moderate cirrhosis	++++	At least two broad septa, but no very broad septa and less than half of biopsy length composed of minute nodules
4C	Severe cirrhosis	+++++	At least one very broad septum or more than half of biopsy length composed of minute nodules (micronodular cirrhosis)

[a]Modified from Wanless and Crawford (2003).

fibrous septa. This simplification decreases the opportunity for interobserver variation. The expanded scale allows quantification of changes with time in patients with cirrhosis. As with other systems, the Laennec system also reports grade of activity on a scale of 0 to 4, as well as noting etiology-specific features.

Stage should be estimated with a connective tissue stain. When the presence of a septum is uncertain, curved contours in the tissue may assist identification. Because all biopsies are susceptible to sampling errors, some measure of diagnostic certainty should be recorded, including comments on the size, fragmentation, and technical quality of a biopsy. When faced with a biopsy of regressed cirrhosis in which there are no visible septa, detection of obliterated veins may distinguish the biopsy from normal. Immunostain for CD34 may detect arterialized sinusoids that would otherwise appear normal.

Evaluation of Etiology

Although etiology is increasingly being determined by serological and other laboratory tests, the biopsy often provides important information. This is especially true when laboratory results are mildly abnormal or conflicting, or when more than one disease process is present.

Chronic Hepatitis

Chronic hepatitis is characterized by an infiltrate of lymphocytes with variable numbers of plasma cells. The infiltrate is both portal and parenchymal. In portal tracts, the infiltrate may be diffuse or concentrated in lymphoid aggregates. The parenchymal infiltrate is diffuse, involving all zones. In low-grade disease (activity grade 1), the infiltrate is not accompanied by visible necrosis. Mild to moderate disease (activity grades 2 and 3) is accompanied by acidophilic bodies and focal dropout of hepatocytes. In severe disease (activity grade 4), there is contiguous dropout that may span portal and venular regions (bridging necrosis). Activity at the portal–parenchymal interface is often called "piecemeal" necrosis or interface hepatitis; because this activity correlates with activity elsewhere in the parenchyma, its presence has no special significance.

Chronic hepatitis is most often caused by hepatitis B virus, hepatitis C virus, autoimmune hepatitis, and, rarely, by Wilson's disease or drug reactions. These forms of hepatitis usually cannot be distinguished histologically so that the diagnosis is made by serological or clinical features. Cytoplasmic inclusions of hepatitis B surface antigen, as detected by the presence of ground glass cells or immunohistochemical staining, are diagnostic of hepatitis B. Untreated autoimmune hepatitis is usually more active, compared to the other types. Plasma cells may be found in chronic hepatitis of any etiology but are more consistently prominent in autoimmune hepatitis. The pattern of fibrosis varies with the activity of the disease. Hepatitis C, as a low-grade disease, usually involves only the smallest hepatic veins and thus the individual extinction lesions are small but involve a large percentage of hepatic veins by the time cirrhosis is present. The resulting cirrhosis is micronodular. Hepatitis B and autoimmune hepatitis

involve larger hepatic veins so that larger extinction lesions occur, leading to cirrhosis with fewer such lesions (and more spared hepatic veins) and thus a macronodular pattern.

Fatty Liver Disease

Fatty liver disease is commonly caused by alcohol abuse or metabolic states with hyperinsulinemia, as seen in obesity or type 2 diabetes mellitus. Fatty liver disease is characterized by macrovesicular steatosis. Progressive disease activity involves steatohepatitis, defined by steatosis plus evidence of active or past necrosis. The minimum requirement for a diagnosis of steatohepatitis is debated. A reasonable definition would be steatosis with ballooning and one other feature of activity, such as Mallory bodies, pigmented macrophages in zone 3, or neutrophilic infiltrate. Fibrosis may be an indicator of previous steatohepatitis. In this setting, the fibrosis characteristically deposits in the walls of zone 3 sinusoids. Small hepatic veins are obliterated with focal collapse and approximation of small portal tracts and adjacent veins. When larger regions of parenchyma collapse, broad fibrous septa develop. The same sequence of events occurs with chronic viral disease, thus sinusoidal fibrosis is not pathognomonic of fatty liver disease.

Alcoholic and nonalcoholic fatty liver disease are often identical in histologic appearance, though generally alcoholic hepatitis is more active compared to nonalcoholic steatohepatitis. Low-grade steatohepatitis often occurs in patients with chronic hepatitis C and may be related to obesity in these patients. After onset of cirrhosis, the fat and activity tend to decline, making histologic diagnosis of fatty liver disease difficult in the late stages.

Biliary Disease

Chronic biliary disease is caused by long-standing obstruction of bile ducts, by cholestatic drug reactions, and, in children, by genetic abnormalities involving transport proteins. Large duct obstruction lasting many months can result in cirrhosis, though this seldom happens because of surgical intervention. In adults, the frequent causes leading to cirrhosis are primary biliary cirrhosis and primary sclerosing cholangitis. In primary biliary cirrhosis, there is inflammatory destruction of ducts less than 40 μm in diameter and these lesions are easily seen in needle biopsies. In primary sclerosing cholangitis, there is inflammation and fibrosis in the ducts larger than 0.4 mm; these lesions are seldom seen in small biopsies. In this disease, the ducts have concentric fibrosis called "onion-skinning." In cystic fibrosis, there are eosinophilic inclusions in the lumina of small ducts and occasionally fibrous strictures in large ducts. When duct lesions are not available for examination, chronic biliary disease can be suspected by prominent fibrous expansion of portal tracts accompanied by swelling of periportal hepatocytes (feathery degeneration), often with periportal Mallory bodies and a ductular reaction characterized by ductular proliferation, portal edema, and neutrophilic infiltration. In addition, there may be bile-stained hepatocytes, clusters of foamy or bile-stained macrophages, and liver cell rosettes. Ductular proliferation alone is not a reliable indicator of biliary disease because it may be a manifestation of regeneration in nonbiliary diseases.

Fibrosis in biliary disease is topographically variable and correlates with regions of high-grade duct obstruction. This correlation is most easily seen in primary sclerosing cholangitis and cystic fibrosis, in which whole segments may undergo extinction distal to sites of obstruction.

Metabolic Diseases

Many uncommon or rare metabolic diseases may result in cirrhosis. In the adult, metal accumulation is a frequent finding. Iron and copper are deposited in the liver in hemochromatosis and Wilson's disease, respectively. The overload is usually genetically determined but may be acquired through ingestion or, in the case of iron, by transfusion. Both metals are thought to injure hepatocytes by catalyzing generation of free radicals. These metals may be detected by histochemistry and the genetic predisposition can be determined by molecular analysis of the genome. Even large amounts of iron cause very slow progression of fibrosis and this progression is often enhanced by coincidental risk factors, such as steatohepatitis or viral infection. Iron and copper may accumulate in severe cirrhosis of any etiology and therefore their presence is not diagnostic of hemochromatosis or Wilson's disease. Copper stain is often negative in Wilson's disease, so quantitative copper content in the biopsy should be considered.

Congestive Cirrhosis

Congestive heart failure causes sinusoidal dilatation and mild hepatocellular atrophy. Fibrosis develops when local thrombosis causes obliteration of small hepatic veins. The fibrosis is therefore focal and is called cardiac sclerosis; complete cirrhosis does not occur from congestive failure alone.

Thrombosis of all three main hepatic veins causes hepatic enlargement and ascites (Budd–Chiari syndrome) with a markedly congested liver. The congestion

and fibrosis are topographically variable, with a dominant venocentric (reversed nodularity) pattern of necrosis and fibrosis. Secondary portal vein thrombosis is found in 20% of patients with Budd–Chiari syndrome. This event may cause the cirrhosis to remodel into a venoportal pattern of fibrosis, with portal tracts incorporated into the fibrous septa in a fashion similar to the more frequent types of cirrhosis. Early Budd–Chiari syndrome has obvious congestion on biopsy, with marked sinusoidal dilatation, hemorrhage into the liver cell plates, and sometimes infarcts. Late congestive cirrhosis may be similar to other forms of cirrhosis, and examination of medium to large hepatic veins may be necessary to confirm the diagnosis.

Venoocclusive disease begins as necrosis of sinusoidal endothelial cells with secondary fibrous obliteration of small hepatic veins. This disease occurs in response to radiomimetic drugs in preparation for bone marrow transplantation and rarely in response to ingestion of pyrrolizidine alkaloids from bush tea or contaminated bread.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Alcohol Metabolism • Autoimmune Liver Disease • Budd–Chiari Syndrome • Cholestatic Diseases, Chronic • Fibrogenesis • Hepatitis B • Hepatitis C • Hepatocellular Carcinoma (HCC) • Liver Biopsy • Liver Transplantation • Portal Hypertension and Esophageal Varices • Portal Vein Thrombosis • Wilson's Disease

Further Reading

Anthony, P. P., Ishak, K. G., Nayak, N. C., Poulsen, H. E., Scheuer, P. J., and Sobin, L. H. (1978). The morphology of cirrhosis. Recommendations on definition, nomenclature, and classification by a working group sponsored by the World Health Organization. *J. Clin. Pathol.* **31**, 395–414.

Arthur, M. J., Iredale, J. P., and Mann, D. A. (1999). Tissue inhibitors of metalloproteinases: Role in liver fibrosis and alcoholic liver disease. *Alcohol Clin. Exp. Res.* **23**, 940–943.

Friedman, S. L., Maher, J. J., and Bissell, D. M. (2000). Mechanisms and therapy of hepatic fibrosis: Report of the AASLD Single Topic Basic Research Conference. *Hepatology* **32**, 1403–1408.

Iredale, J. P. (2001). Hepatic stellate cell behavior during resolution of liver injury. *Semin. Liver Dis.* **21**, 427–436.

Ishak, K., Baptista, A., Bianchi, L., Callea, F., De Groote, J., Gudat, F., *et al.* (1995). Histological grading and staging of chronic hepatitis. *J. Hepatol.* **22**, 696–699.

Ludwig, J. (1988). Etiology of biliary cirrhosis: Diagnostic features and a new classification. *Zentralbl. Allg. Pathol.* **134**, 132–141.

Nakashima, E., Kage, M., and Wanless, I. R. (1999). Idiopathic portal hypertension: Histologic evidence that some cases may be regressed cirrhosis with portal vein thrombosis. *Hepatology* **30**, 218A.

Popper, H. (1977). Pathologic aspects of cirrhosis. *Am. J. Pathol.* **87**, 228–264.

Poynard, T., Bedossa, P., and Opolon, P. (1997). Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* **349**, 825–832.

Scheuer, P. (1967). Primary biliary cirrhosis. *Proc. R. Soc. Med.* **60**, 1257–1260.

Sciot, R., Staessen, D., Van Damme, B., Van Steenbergen, W., Fevery, J., De Groote, J., *et al.* (1988). Incomplete septal cirrhosis: Histopathological aspects. *Histopathology* **13**, 593–603.

Tanaka, M., and Wanless, I. R. (1998). Pathology of the liver in Budd–Chiari syndrome: Portal vein thrombosis and the histogenesis of veno-centric cirrhosis, veno-portal cirrhosis, and large regenerative nodules. *Hepatology* **27**, 488–496.

Wanless, I. R. (2003). Physioanatomic considerations. *In* "Schiff"s Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 9th Ed., p. 43. Lippincott-Raven, Philadelphia.

Wanless, I. R., and Crawford, J. M. (2003). Cirrhosis. *In* "Surgical Pathology of the Alimentary Tract, Pancreas, Biliary Tree, and Liver" (R. D. Odze, J. R. Goldblum, and J. M. Crawford, eds.), in press.

Wanless, I. R., Nakashima, E., and Sherman, M. (2000). Regression of human cirrhosis: Morphologic features and the genesis of incomplete septal cirrhosis. *Arch. Pathol. Lab. Med.* **124**, 1599–1607.

Wanless, I. R., Sweeney, G., Dhillon, A. P., Guido, M., Piga, A., Galanello, R., *et al.* (2002). Lack of progressive hepatic fibrosis during long-term therapy with deferiprone in subjects with transfusion-dependent beta-thalassemia. *Blood* **100**, 1566–1569.

Wanless, I. R., Wong, F., Blendis, L. M., Greig, P., Heathcote, E. J., and Levy, G. (1995). Hepatic and portal vein thrombosis in cirrhosis: Possible role in development of parenchymal extinction and portal hypertension. *Hepatology* **21**, 1238–1247.

Cobalamin Deficiency

MANISHA CHANDALIA AND NICOLA ABATE
University of Texas Southwestern Medical Center, Dallas

homocysteine An amino acid that is converted to methionine with cobalamin as cofactor. Cobalamin deficiency leads to increased levels of homocysteine, which is used to diagnose cobalamin deficiency.

megaloblastic anemia Anemia characterized by peripheral macrocytosis (mean cell volume > 100 fL) and bone marrow containing megaloblasts due to impaired DNA synthesis. Cobalamin deficiency along with folate deficiency leads to megaloblastic anemia.

methylmalonic acid (MMA) A metabolite of methylmalonyl-coenzyme A1 that is excreted in the urine. Cobalamin deficiency leads to the accumulation of MMA; measurement of MMA in plasma is utilized for the diagnosis of cobalamin deficiency.

pernicious anemia Cobalamin deficiency due to the absence of intrinsic factor required for absorption of cobalamin. This deficiency leads to megaloblastic anemia.

vitamin B12 Cyanocobalamin used as a vitamin preparation therapeutically.

Cobalamin is a complex organometallic compound, which is composed of a tetrapyrrolic ring structure in the center of which is a cobalt ion. Cobalamin can be synthesized only by bacteria. It is incorporated into the food chain by certain animals, mainly herbivores, which then act as a source of nutrients for other animals. Liver, kidney, beef, fish, shellfish, eggs, and milk and other dairy products provide most of the cobalamin in a normal diet. The minimum daily requirement of cobalamin is approximately 2.5 μg. Cobalamin plays a critical role in cellular DNA synthesis, and its deficiency leads to anemia, neurologic disease, and other manifestations.

INBORN ERRORS OF COBALAMIN METABOLISM

Several inborn errors of Cbl metabolism have been described, each of which produces a distinct form of cobalamin, which in turn produces a clinical syndrome. These intracellular errors of cobalamin metabolism can be classified as follows:

1. Defective synthesis of adenosylcobalamin causing methylmalonic aciduria: intramitochondrial cobalamin reductase deficiency and intramitochondrial cobalamin transferase deficiency.
2. Defective synthesis of methylcobalamin causing hyperhomocysteinemia: cytosolic methionine synthase reductase defect and defective methionine synthase.
3. Defect of both adenosyl- and methylcobalamin synthesis, causing both methylmalonic aciduria and hyperhomocysteinemia: cytosolic cobalamin reductase defect—early onset; cytosolic cobalamin reductase defect—late onset; and defective cobalamin lysosomal efflux.

ETIOLOGY OF COBALAMIN DEFICIENCY

Dietary Intake

The dietary intake of cobalamin is usually adequate in people consuming at least dairy products. However, in true vegans and their breast-fed infants, dietary intake of cobalamin is insufficient and these individuals are at risk of developing cobalamin deficiency unless they receive supplementation. In addition, women who are only moderate vegetarians may become cobalamin deficient during pregnancy and lactation.

Gastrointestinal Diseases

In most cases, deficiency of cobalamin is due to malabsorption. Malabsorption can result from abnormalities at several levels.

Stomach

Defective release of cobalamin from food For the release of cobalamin bound tightly to proteins in food, acid and pepsin in the stomach are essential. People who are achlorhydric, often the elderly, are commonly unable to release cobalamin from food sources, but they can absorb crystalline B12, which is found in multivitamin preparations. Only a minority of people with this defect develop frank cobalamin deficiency, but many have biochemical changes including low levels of plasma cobalamin and high levels of MMA and homocysteine. Patients on drugs that suppress gastric acid production, such as proton pump inhibitors, may

 Copyright 2004, Elsevier (USA). All rights reserved.

also have a defect in cobalamin release from the food sources, but Cbl deficiency in such patients is very rare.

Pernicious anemia Pernicious anemia is considered to be the most common cause of cobalamin deficiency. The cobalamin deficiency is caused by the absence of intrinsic factor because of either atrophy of the gastric mucosa or autoimmune, cytotoxic T-cell-mediated destruction of parietal cells. It is most common in individuals of northern European descent and African Americans and is least common in Asians and southern Europeans. Patients with pernicious anemia have circulating antibodies to parietal cell antigens (90%) and to IF (60%).

Postgastrectomy Cobalamin deficiency develops following total gastrectomy or extensive damage to the gastric mucosa. Due to lack of IF, oral intake of B12 is ineffective and parenteral administration of B12 is essential to prevent cobalamin deficiency.

Helicobacter pylori Some studies have suggested a link between infection with *Helicobacter pylori* and cobalamin deficiency.

Intestine

Small intestinal colonization with bacteria Any condition leading to small intestinal colonization with bacteria (bacterial overgrowth) can cause cobalamin deficiency. This occurs due to utilization of dietary cobalamin by the bacteria. Intestinal blind loops, strictures, diverticula, and pseudo-obstruction are associated with bacterial overgrowth of the small intestine.

Ileal abnormalities Any condition leading to a defect in the absorptive capacity of the distal ileum can lead to cobalamin deficiency. Cobalamin deficiency is seen in regional enteritis (Crohn's disease), tropical sprue, gluten-sensitive enteropathy, Whipple's disease, tuberculosis, and resection of distal ileum.

CLINICAL MANIFESTATION OF COBALAMIN DEFICIENCY

The clinical features of cobalamin deficiency are megaloblastic anemia, neurological changes, and hyperhomocysteinemia. The major hematologic finding is a megaloblastic anemia with elevated serum bilirubin and lactate dehydrogenase levels that reflect the increased red blood cell breakdown due to ineffective erythropoiesis. The peripheral blood smear shows macrocytes, occasionally megaloblasts, and hypersegmented neutrophils (greater than 5% of neutrophils with five or more lobes or 1% with six or more lobes). Bone marrow aspiration and biopsy reveal a very hypercellular marrow with megaloblastic erythroid hyperplasia and giant metamyelocytes. The inadequate conversion of deoxyuridate to thymidylate, which leads to slowing of DNA synthesis and delayed nuclear maturation, and methionine deficiency play a central role in these abnormalities.

The neurologic features of cobalamin deficiency consist of the classic picture of subacute combined degeneration of the dorsal and lateral spinal columns, which is due to a defect in myelin formation by an unknown mechanism. The neuropathy is symmetrical and affects the legs more than the arms. It begins with paresthesias and ataxia associated with loss of vibration and position sense and can progress to severe weakness, spasticity, clonus, paraplegia, and even fecal and urinary incontinence.

Cobalamin and folate are essential for the metabolism of homocysteine and deficiency of either leads to hyperhomocysteinemia, which is recognized as an independent risk for cardiovascular disease, atherosclerosis, and venous thromboembolism. It appears that mild deficiency of cobalamin, which may not be recognized as classic cobalamin deficiency due to a lack of hematologic features, may cause elevation of plasma homocysteine and risk for atherosclerosis and thromboembolism. The optimum plasma cobalamin levels to achieve an optimum plasma homocysteine level are not defined. Furthermore, there are ethnic differences in the association between plasma levels of cobalamin and homocysteine. Several clinical trials are in progress to assess the role of vitamin B12 supplementation in lowering plasma homocysteine and its impact on cardiovascular disease event and mortality.

DIAGNOSIS OF COBALAMIN DEFICIENCY

Cobalamin deficiency can be evaluated by measuring serum cobalamin levels. The normal level of cobalamin in serum is considered to be >200 pg/ml; values <100 pg/ml indicate a clinically significant deficiency. Elevated serum MMA and homocysteine levels are considered useful in recognizing subtle cobalamin deficiency.

Schilling Test

To determine the pathogenesis of cobalamin deficiency, a Schilling test can be performed, which consists of three parts: In part (1), a patient is given radioactive cobalamin by mouth, followed soon by an intramuscular injection of unlabeled cobalamin. A urine sample is collected 24 h later and the amount of radioactive cobalamin excreted gives an indication of absorption.

Since cobalamin deficiency in most of the cases is due to malabsorption, only a small amount of radioactive cobalamin is typically absorbed and excreted in the urine. In part (2), the patient is given radioactive cobalamin bound to IF and excretion of cobalamin in urine over 24 h is measured. If urinary cobalamin excretion is normalized in the patient, it suggests IF deficiency, due to one of the causes listed above. If excretion of cobalamin does not normalize, it suggests small intestinal bacterial overgrowth or an ileal disease. In part (3), the patient is treated with antibiotics and the Schilling's test is repeated. Normalization of urinary B12 excretion suggests bacterial overgrowth as the cause of Cbl deficiency.

TREATMENT OF COBALAMIN DEFICIENCY

In patients with poor intake (vegans), daily oral administration of cobalamin (2 mg crystalline B12) is very effective. However, in most other cases, since the defect is malabsorption, parenteral therapy is preferred. Parenteral therapy is given as an intramuscular injection of cyanocobalamin. The treatment begins with 1000 μg cobalamin per week for 8 weeks followed by 1000 μg cyanocobalamin per month for the rest of the patient's life. High oral doses of B12 have also been reported to be effective, presumably because small (but sufficient) amounts of B12 are absorbed passively by the small intestine. Hyperhomocysteinemia is now recognized as a cardiovascular risk factor and since optimal cobalamin levels in serum are not known, in high-risk patients with high blood levels of homocysteine, oral administration of crystalline B12 and folate is advocated. Various outcome trials with B12 and folate are being conducted to evaluate the role of replacement in preventing coronary heart disease.

See Also the Following Articles

Dietary Reference Intakes (DRI): Concepts and Implementation • Gastrectomy • Intrinsic Factor • Malabsorption • Pernicious Anemia • Vitamin B12: Absorption, Metabolism, and Deficiency

Further Reading

Babior, B. M., and Bunn, H. F. (2001). Megaloblastic anemia. *In* "Harrison's Principles of Internal Medicine" (E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and L. Jameson, eds.), 15th Ed., pp. 2245–2256. McGraw-Hill, New York.

Carmel, R. (2000). Current concepts in cobalamin deficiency. *Annu. Rev. Med.* **51**, 357–375.

Klee, G. G. (2000). Cobalamin and folate evaluation: Measurement of methylmalonic acid and homocysteine vs vitamin B(12) and folate. *Clin. Chem.* **46**, 1277–1283.

Marsh, E. N. (1999). Coenzyme B12 (cobalamin)-dependent enzymes. *Essays Biochem.* **34**, 139–154.

Selhub, J., and D'Angelo, A. (1998). Relationship between homocysteine and thrombotic disease. *Am. J. Med. Sci.* **316**, 129–141.

Seshadri, N. and Robinson, K. (2000). Homocysteine, B vitamins, and coronary artery disease. *Med. Clin. North Am.* **84**, 215–237.

Colectomy

ANTHONY RAMSANAHIE* AND RONALD BLEDAY*,†
*Brigham and Women's Hospital and †Harvard Medical School

anastomosis Surgical union of two hollow organs, for example, blood vessels or parts of the intestine, to ensure continuity of the passageway.
fistula Abnormal connection between two epithelial-lined surfaces.
ischemic Inadequate supply of blood to a part of the body, caused by partial or total blockage of an artery.
lithotomy Body position in which the patient lies on their back on the examining table, with hips and knees fully flexed; also called dorsosacral position. For obstetric procedures, the buttocks are at the edge of the table and the feet are held in stirrups.
megacolon Massive abnormal dilation of the colon.
pouch of Douglas A pouch formed by a fold of peritoneum between the rectum and the uterus.
volvulus Twisting or axial rotation of a portion of bowel about its mesentery.

Colectomy is the excision of the colon in a segment or in its entirety. It is a common procedure performed either electively or as emergency; performance by a general surgeon requires knowledge of technique and careful planning and execution.

INTRODUCTION

Reybard, of Lyon, France, performed the first recorded colectomy, an anastomosis, in 1844. Since then, of course, there have been many changes in the field of surgery. Principles of antiseptic surgery following Lister's treatise in 1867, antibiotic prophylaxis, progress in general anesthesia, and intravenous fluid and blood transfusion have all made intraabdominal surgery increasingly safer. Current technologies in the form of stapling devices and laparoscopic instruments as well as newer imaging modalities have added to the surgeon's armamentarium. Depending on the disease and its extent and site, most colectomies result in restoration of gastrointestinal continuity.

The initial principles of bowel anastomosis first reported by Halsted are still practiced. The importance of no tension on the anastomosis, incorporation of the subserosa by suture, inversion of the mucosa, and adequate blood supply are tenets still upheld today. Closure of the mesenteric defect and a two-layered anastomosis are usually practiced but are not essential. Exposure of the site to be anastomosed is important and removal of serosal fat and epiploic appendages at the site of anastomosis is helpful, provided the serosa is left intact. The use of stapling sutures does not negate these principles for a successful anastomosis.

INDICATIONS

Colectomy is commonly indicated in the treatment of primary resectable colorectal carcinoma and the precancerous condition familial adenomatous polpyposis. It is also required in inflammatory bowel diseases such as ulcerative colitis, Crohn' s disease, and diverticulitis that have become refractory to medical therapy or are life threatening due to their complications (bleeding, perforation, or sepsis). Other conditions that may require excision of the colon include intestinal obstructions, perforation of the colon wall, ischemic colon, toxic megacolon, volvulus, and fistulas between the colon and other viscera.

PREOPERATIVE PREPARATION

The principle of preoperative care is that the patient is made as physically and biochemically fit as possible; in cancerous cases, patients are accurately staged prior to surgery. Hence, nutritional and biochemical status are evaluated and corrected as necessary. This maybe possible in elective cases, but in the emergency setting, time may not permit. Mechanical cleansing of the colon is helpful in colon surgery and along with antibiotics can reduce the incidence of infection rates following colorectal surgery. In a survey of colon and rectal surgeons, two-thirds of surgeons preferred polyethylene glycol electrolyte solution for their patients due to the reliability of the cleansing results. Prophylactic antibiotics complement cleansing to decrease the incidence of postoperative septic complications, because the concentration of bacteria per millimeter of effluent is not affected by mechanical cleansing. The antibiotic regime used most commonly is the one advocated by Nichols and Condon, consisting of a neomycin and erythromycin base, 1 g each orally at 1, 2, and 10 pm, on the day prior to surgery. More recently, many surgeons have

 Copyright 2004, Elsevier (USA). All rights reserved.

substituted metronidazole, 500 g, for erythromycin, because metronidazole is bactericidal against a wider spectrum of anaerobic bacteria.

Intravenous antibiotics are also given perioperatively to supplement the preoperative regime. Double prophylaxis (oral and intravenous) is most beneficial for procedures below the peritoneal reflection, for operations lasting longer than 3.5–4 hours, in immunocompromised patients (e.g., patients on steroids), and in situations in which mechanical preparation is impossible. The intravenous antibiotics are directed against aerobic and anaerobic bowel bacteria. Cefazolin (1 g, for aerobic and skin flora) is usually given and metronidazole (500 g, for anaerobic bacteria) is used at induction of anesthesia, the aim being to achieve a high tissue level at the time of surgery. Coverage with antibiotics is usually for 24 hours, although the preoperative dose is most critical.

Patients with known valvular disease of the heart or with implanted vascular or recently placed orthopedic prostheses also require additional prophylactic antibiotics. Intravenous ampicillin (2 g) and gentamycin (1.5 mg/kg) are given at 30 minutes to 1 hour before the procedure and at least one postoperative dose is given in place of cefazolin. Metronidazole is given as usual. Vancomycin is substituted for ampicillin in patients who are allergic to penicillins and cephalosporins.

Patients who have had colon and rectal surgery, especially pelvic surgery, are at risk for deep venous thrombosis complications. Patients should be given either subcutaneous heparin (5000 units) or low-molecular-weight heparin, or provided with compression stockings. Patients on anticoagulants such as warfarin need to stop the medication 2–5 days prior to surgery, with timing of medication dependent on individual patient risk.

SURGICAL TECHNIQUES

Right Hemicolectomy

Right hemicolectomy is undertaken with the patient placed in a supine position. The surgeon stands on the left side of the patient. The procedure incorporates the removal of the last few centimeters of the ileum, the ascending colon, and the first few centimeters of the transverse colon. Intraabdominally, the colon is mobilized along the avascular plane—the white line of Toldt (which fixes the colon to the lateral abdominal wall)—from the cecum to beyond the hepatic flexure. In mobilizing, the surgeon must be aware of the right kidney, second part of duodenum, right ureter, and right gonadal vessels.

The colon is reflected medially to expose the origin of the ileocolic artery. The surgeon performs ligation and division of the right branches of the middle colic artery, right colic vessels, ileocolic vessels, and vessels to last 5–10 cm of the ileum. The major vascular trunks should be double ligated to avoid risk of knot slippage and hemorrhage. If clamps are used, they are placed at the site of resection of both ends of the bowel, and the bowel is excised flush with the clamps. Many surgeons use staplers to divide and occlude the ends of the bowel prior to anastomosis. After removal of the specimen, the ileum is anastomosed in a side-to-side (functional end-to-end) fashion to the transverse colon. The side-to-side arrangement is used because the caliber of the small bowel and large bowel is very different.

In cases of hepatic flexure tumors, the excision is extended along the gastrocolic ligament toward the spleen. The splenic flexure and left colon can be mobilized if necessary. The omentum is resected if involved with cancer, otherwise it can be preserved. For an "extended" right hemicolectomy, most of the transverse colon is resected. For cancer cases, the surgeon needs to take the mesenteric dissection near their origin. The right colic artery should be ligated close to the superior mesenteric artery. Ligation of the middle colic artery needs to be customized to each patient. Patients with an occluded inferior mesenteric artery will need to have the left branch of the middle colic artery preserved for adequate blood supply to be maintained to the distal transverse colon and left colon.

Transverse Colectomy

Transverse colectomy is the excision of the transverse colon, with transection at the distal ascending and proximal descending colons. The resection requires ligation of the middle colic vessels. The stomach is detached from the transverse colon by division of the gastrocolic omentum. Both flexures are completely divided and the right and left paracolic gutters are mobilized. The site of resection is identified and the transverse colon is resected with its mesentery, and the greater omentum is included if resection is for malignant disease. The transverse colon is reflected upward to reveal the origin of the middle colic vessels beneath the peritoneum. The middle colic artery and its vein are ligated and divided. The bowel is reconstituted without tension.

Left Hemicolectomy

Left hemicolectomy is performed with the patient placed in the either the supine or the lithotomy position.

The surgeon stands to the right. Mobilization of the left colon begins with an incision of the white line of Toldt and extends around the splenic flexure. The spleen, tail of the pancreas, left kidney, left ureter, and left gonadal vessels are vulnerable in this procedure. The left ureter is identified best at the pelvic brim as it crosses the junction of the internal and external iliac vessels. The colon is mobilized medially to reveal the left colic and inferior mesenteric vessels.

Either the left colic artery or the inferior mesenteric artery can be ligated. The inferior mesenteric vein or left colic vein is also ligated. The left branch of the middle colic artery, the marginal vessels at midtransverse colon, and, on occasion, the superior hemorrhoidal artery are preserved. Less extensive or segmental resections involving only the affected portion of colon and its accompanying mesentery and lymph drainage are carried out for benign disease.

Anterior Resection

Anterior resection is indicated for rectal carcinoma between 5 cm from the anus and the rectosigmoid junction. The patient is placed in the lithotomy or modified lithotomy position using stirrups. This position allows for access through the anus for a low rectal anastomosis using a circular stapling device. In an anterior resection, the sigmoid colon is mobilized, the ureter is identified, and the resection is started at the junction of the sigmoid colon and the left colon. A high ligation should be performed on the superior hemorrhoidal vessels after the branches of the sigmoid arteries are divided. The dissection at this point needs to separate and preserve the sympathetic nerves to the pelvis from the mesocolon and mesorectum. The dissection over the pelvic brim into the pelvis starts the "total mesorectal excision" portion of the resection. By staying in the plane that is in front of the sympathetic nerves but outside the mesorectum, the surgeon can then remove all potentially cancerous lymph nodes. For a malignancy, the dissection should continue up to 4–5 cm beyond the lower border of the tumor in order to resect the complete zone of downward spread that can exist in the mesorectum.

The patient habitus (a thin female with a wide pelvis, compared with a narrow male pelvis), adequacy of the anal sphincter, encroachment of the tumor on the anal sphincters, and inadequacy of the distal margins are factors in determining the feasibility of anastomosis with a sphincter-sparing operation. The advent of circular stapling devices has allowed patients with very low-lying rectal cancer in whom the distal margin is at the minimal acceptable level, yet adequate for cancer clearance, to have sphincter-saving operations and coloanal anastomosis.

For anastomosing the colon to the distal rectum, prior to rectal cancer resection the rectum is stapled with a straight stapling device transversely across the rectum below the cancer. The bowel proximal to the staple is then clamped and the rectum is divided just above the stapler, and the cancer is removed. The circular stapling device with a trocar in its center is then placed in the anus. The trocar pierces the stapled top end of the residual rectum. The proximal bowel with the other end of the circular stapler is then connected to the lower end in the rectal remnant. Both parts are then closed and fired; the two rows of staples anastomose the bowel ends in a turned-in fashion (serosa to serosa). The resulting rings of residual tissue from inside the staple lines ("donuts") are then removed, with the circular stapler, via the anus.

Total Mesorectal Excision

Total mesorectal excision (TME) by sharp dissection, in conjunction with a low anterior resection or an abdominal perineal resection, has been shown to decrease the incidence of positive radial margins from 25% in conventional surgery to 7% in cases resected by TME. The hypothesis is that conventional surgery violates the circumference of the mesorectum during blunt dissection along undefined planes, leaving residual mesorectum in the pelvis and thus residual cancer. This has been reflected in the high rate of pelvic cancer recurrence in conventional surgery. The rectal mesentery is removed sharply under direct visualization, emphasizing autonomic nerve preservation, complete hemostasis, and avoidance of violation of the mesorectal envelope. The meticulous dissection is not without consequence; there is prolonged operative time and an increased anastomotic leak rate is noted. Anastomoses 3–6 cm from the anal verge have led to anastomotic leak rates of up to 17%. Some centers are routinely fashioning a protective diverting colostomy or ileostomy with a low anastomosis. Other postoperative complications related to the damage to the pelvic autonomic parasympathetic and sympathetic nerves by blunt dissection, such as impotence and retrograde ejaculation, have been reported to be as low as 10–29% of cases, as compared to 25–75% seen in conventional rectal surgery. (see Figs. 1 and 2).

The type of surgery offered for low rectal cancers depends on an accurate staging of the tumor and the location of the tumor in relation to the surgical resection.

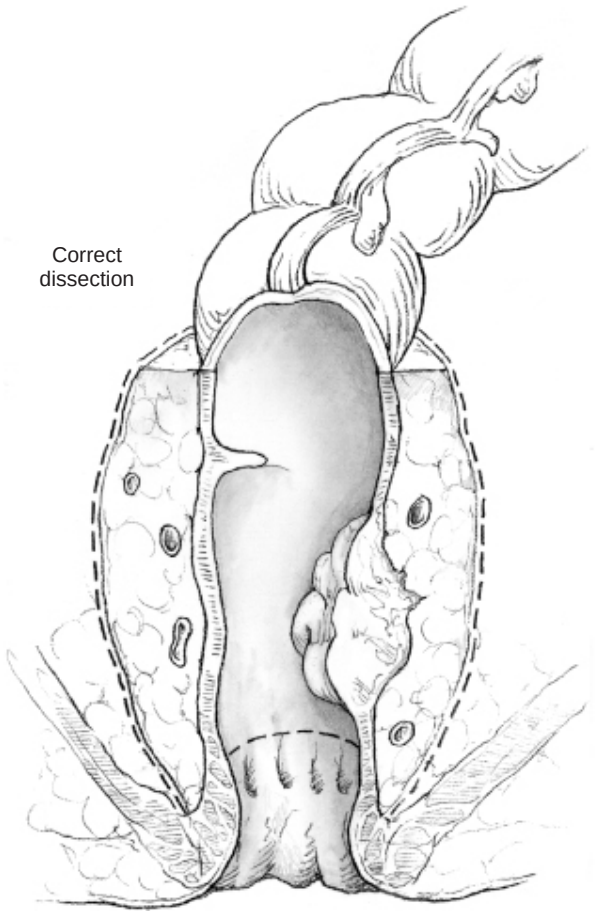


FIGURE 1 The proper plane for a total mesorectal excision is outside the mesorectal envelope, which would include all mesorectal fat, mesorectal lymph nodes, and any locoregional metastatic implants. However, the dissection is inside the pelvic autonomic nerves (both sympathetic and parasympathetic) and the pelvic side wall.

Clinically, the most reliable landmark is the dentate line where squamous mucosae of the anal skin merge into the columnar mucosa of the rectum. The dentate line is located in the middle of the anorectal ring, which is composed of the internal and external sphincters and the portions of the levators, the puborectalis. These muscles provide a high-pressure zone and are responsible for fecal continence. Commonly, the high-pressure zone proceeds 1–3 cm above the dentate line. But from a technical perspective, a tumor has to be located high enough above the top of the anorectal ring to allow for an adequate distal margin if sphincter preservation is to be achieved along with rectal cancer resection. However, a subset of rectal cancer patients are best treated by an abdominoperineal resection.

Abdominoperitoneal Excision

Abdominoperitoneal excision was first described in 1908 by Ernest Miles, though Czerny first performed the removal of the rectum via a combined abdominal and perineal approach in 1884. In this procedure, the patient is placed in a lithotomy with stirrups. There are two parts and two surgeons for this operation, with the abdominal operator standing on the left of the patient and the perineal operator sit facing the perineum. The abdominal resection is the same as for a low anterior tumor resection and its description is not repeated here, except to say that a preliminary examination is required to see whether the tumor is palpable above the pelvic floor and to assess the possibility of its removal.

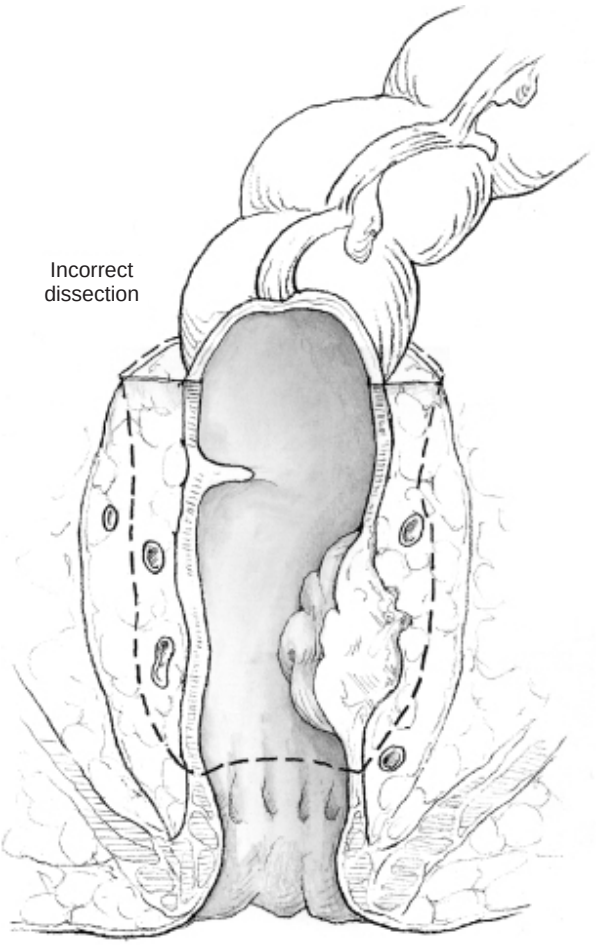


FIGURE 2 Blunt dissection of the mesorectum leads to a dissection plane that is within the mesorectal fat. Lymph nodes and/or implants can be left within the pelvis, leading to a higher recurrence rate.

Once the dissection has reached the level of the levator ani bundle, the peritoneum is divided down to the pouch of Douglas (in females) or to the Denonvilliers fascia at the seminal vesicles (in males). Anteriorly, the peritoneum is divided overlying the bladder, creating a plane of dissection anteriorly in males immediately posterior to the seminal vesicles and anterior to the rectum; in females, the plane lies posterior to the vagina. The posterior dissection is carried out using cautery for all of the mesorectum, from the sacrum proceeding as far as the coccyx.

The perineal dissection starts by placing a purse-string suture around the anus. A skin incision is then made in the form of an ellipse from the perineal body anteriorly, to the ischiorectal spines laterally, and extending posteriorly around each side to meet at the tip of the coccyx. The incision is deepened, exposing the coccyx. All vessels are ligated or cauterized. The perineal surgeon meets the abdominal surgeon behind the rectum but in front of the coccyx. After connecting the planes, the puborectalis muscle is then "hooked" by the hand of the perineal surgeon and divided with cautery. The anterior dissection is completed after the proximal portion of the specimen is inverted posteriorly in front of the coccyx out to the perineal surgeon. After carefully completing the dissection of the prostate or back wall of the vagina, the specimen is then removed. The perineum is closed with secure, interrupted sutures, ensuring hemostasis. A left-sided colostomy is fashioned but completed after the abdominal wound is closed.

Total Proctocolectomy with Ileal–Anal Anastomosis

Total proctocolectomy with ileal–anal anastomosis is a procedure that aims for the removal of all of the large bowel mucosa while preserving the anal fecal function.

POSTOPERATIVE COMPLICATIONS

Postoperative complications, as in many major operations, can generally occur early or late. Early complications can include hemorrhage, infections (wound, septicemia, peritonitis, and abscess formation), anastomotic leakage and stricture, deep venous thrombosis, and pulmonary embolism. Late complications are due to obstruction secondary to adhesions and local recurrence. Metastatic deposits can occur at a later time in the liver or peritoneum. Specific complications of the procedure, e.g., urinary dysfunction and impotence due to autonomic nerve damage in pelvic surgery, can occur.

Following resections of the colon and rectum, functional disturbances occur. Commonly, the frequency and consistency of the stool are altered after segmental colon resection. Usually, absorption of water and electrolytes occurs in the right colon, but following right colon resection and removal of the ileocecal valve, most patients experience only a mild increase in frequency with a looser stool. However, after a while, the remaining colonic mucosa adapts and efficiently absorbs water and electrolytes and so stool consistency and frequency return too normal. If treatment is required, it may take the form of stool bulk-forming agents in order to absorb excess stool water. Diarrhea secondary to bacterial overgrowth due to the degradation of primary bile salts to secondary bile salts occurs and can be treated with cholestyramine or low-dose antibiotics to reduce the bacterial population.

Constipation after segmental resection of the sigmoid is not unheard of and is most probably due to the disruption of the coordinated mass movement within the colon. Treatment initially should be with bulk-forming agents and a stool softener; however, most patients require stimulant laxatives to propel the fecal bolus into the rectum. Frequent small-volume stools, occasionally associated with seepage or frank fecal incontinence, occur after low anterior resection and coloanal anastomosis. The compliance of the proximal colonic segment above the anastomosis is less compared to that of the rectum, thus reducing the reservoir capacity of the colon. However, in most patients, the proximal bowel subsequently adapts and functions normally, with improved compliance and decreasing frequency of stools. In those patients who cannot adapt, bulk-forming agents are indicated.

LAPAROSCOPIC-ASSISTED COLORECTAL SURGERY

As surgeons become more comfortable with laparoscopic or minimally invasive techniques and more technologically advanced surgical instruments and staplers specifically designed for bowel work, it has become possible to perform laparoscopic and laparoscopic-assisted colon and rectal resections. The introduction of intestinal staplers has allowed milestones such as right hemicolectomy, sigmoid resection, low anterior resection, APR, and total abdominal and transverse colectomies to occur. This shift from conventional open colorectal surgery to minimally invasive surgery

is a continuation of the success derived from laparoscopic cholecystectomy and of the advantages of decreased morbidity and faster recovery due to smaller incision wounds.

The tenets of conventional surgery, exploration, resection, and anastomosis, can still occur via a smaller laparotomy incision, along with mini-incisions for trocars. Because the mesentery of the colon is a midline structure, the laparoscopic camera is first placed through a midline port; three to five more trocars, ranging in size from 5 to 12 mm, are then placed. Trocars are placed in all four quadrants of the abdomen. The colon is mobilized away from the left or right peritoneal reflections, up from the pelvis, or from its attachments to the omentum in exactly the same fashion as open surgery, but via laparoscopic instruments. The sigmoid colon, upper rectum, and right colon appear most amenable to laparoscopic techniques.

Various quality-of-life studies have reported that laparoscopic-assisted colectomy (LAC) results in a decrease in pain early after surgery, but that benefits are minimal. However, any benefit of LAC must also be shown to be efficacious in curing cancer. The final recommendation for the use of laparoscopic-assisted surgery in colon cancer awaits the outcome of ongoing prospective studies.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Diverticulosis • Familial Adenomatous Polyposis (FAP) • Laparoscopy • Toxic Megacolon • Volvulus

Further Reading

Block, G. E., and Moossa, A. R., eds. (1994). "Operative Colorectal Surgery." W. B. Saunders, Philadelphia.

Brief, D. K., Brener, B. J., Goldenkranz, R., *et al.* (1991). Defining the role of subtotal colectomy in the treatment of carcinoma of the colon. *Ann. Surg.* **213**, 248–252.

Corman, M. L. (1991). Principles of surgical technique and the treatment of carcinoma of the large bowel. *World J. Surg.* **15**, 592–596.

Jager, R. M., and Wexner, S. D. (1996). "Laparoscopic Colorectal Surgery" (R. M. Jager and S. D. Wexner, eds.). Churchill Livingstone, U. K.

MacFarlane, J. K., Rryall, R. D. H., and Heald, R. J. (1993). Mesorectal excision for rectal cancer. *Lancet* **341**, 457–460.

Monson, J. R. T., Darzi, A., Carey, P. D., and Guillou, P. H. (1992). Prospective evaluation of laparoscopic-assisted colectomy in an unselected group of patients. *Lancet* **340**, 831–833.

Pain, J., and Cahill, J. (1991). Surgical options for left-sided large bowel emergencies. *Ann. R. Col. l Surg. Engl.* **73**(6), 394–396.

Patterson, C. R., Bleday, R., and Steele, G. D., Jr. (1995). Malignant diseases of the colon. *In* "General Surgery" (W. P. Ritchie, Jr., G. Steele, Jr., and R. H. Dean, eds.), pp. 299–314. J. B. Lippincott, Philadelphia.

Solla, J. A., and Rothenberger, D. A. (1990). Preoperative bowel preparation: A survey of colon and rectal surgeons. *Dis. Colon Rectum* **33**(2), 154–159.

Weeks, J., Nelson, H., Gelber, S., Sargent, D., and Schroeder, G. (2002). Short-term quality of life outcomes following laparoscopic-assisted colectomy vs open colectomy for colon cancer. A randomized trial. *JAMA* **287**(3), 321–328.

CONCLUSION

Collagenous colitis and lymphocytic colitis, which comprise the syndrome of microscopic colitis, are unique forms of inflammatory bowel disease. Unlike Crohn's disease and ulcerative colitis, they do not carry an increased risk of cancer, and they rarely require surgery. Few controlled trials of therapy have been performed, but, in general, these disorders appear to respond to antiinflammatory or immune modulator therapy.

See Also the Following Articles

Anti-Diarrheal Drugs • Colitis, Ulcerative • Crohn's Disease • Diarrhea

Further Reading

Baert, F., Wouters, K., D'Haens, G., *et al.* (1999). Lymphocytic colitis: A distinct clinical entity? A clinicopathological confrontation of lymphocytic and collagenous colitis. *Gut* **45**(3), 375–381.

Fine, K. D., and Lee, E. L. (1998). Efficacy of open-label bismuth subsalicylate for the treatment of microscopic colitis. *Gastroenterology* **114**(1), 29–36.

Jarnerot, G., Hertervig, E., Granno, C., *et al.* (2001). Familial occurrence of microscopic colitis: A report on five families. *Scand. J. Gastroenterol.* **36**(9), 959–962.

Pardi, D. S., Loftus, E. V., Jr., Tremaine, W. J., and Sandborn, W. J. (2001). Treatment of refractory microscopic colitis with azathioprine and 6-mercaptopurine. *Gastroenterol.* **120**(6), 1483–1484.

Perner, A., Andresen, L., Normark, M., *et al.* (2001). Expression of nitric oxide synthases and effects of L-arginine and L-NMMA on nitric oxide production and fluid transport in collagenous colitis. *Gut* **49**(3), 387–394.

Tremaine, W. J. (2000). Collagenous colitis and lymphocytic colitis. *J. Clin. Gastroenterol.* **30**(3), 245–249.

Yusuf, T. E., Soemijarsih, M., Arpaia, A., Goldberg, S. L., and Sottile, V. M. (1999). Chronic microscopic enterocolitis with severe hypokalemia responding to subtotal colectomy. *J. Clin. Gastroenterol.* **29**(3), 284–288.

Zins, B. J., Sandborn, W. J., and Tremaine, W. J. (1995). Collagenous and lymphocytic colitis: subject review and therapeutic alternatives. *Am. J. Gastroenterol.* **90**(9), 1394–1400.

Colitis Cystica Profunda and Solitary Rectal Ulcer Syndrome

DAVID E. BECK

Ochsner Clinic Foundation, Louisiana

colitis cystica profunda A benign pathologic condition characterized by mucin-filled cysts located deep to the muscularis mucosae.

intussception A condition in which the rectum slips inside itself; this condition can cause obstructive symptoms and damage to the rectum.

perineal protectomy A surgical procedure, performed via the anus, that excises redundant rectum and reconnects the two remaining ends.

prolapse Rectal intussception continues to the point that portions of the rectum are outside the anus.

solitary rectal ulcer syndrome A clinical condition characterized by rectal bleeding, copious mucous discharge, anorectal pain, and difficult evacuation.

Colitis cystica profunda and solitary rectal ulcer syndrome are closely related diagnoses and some authors consider them interchangeable. The etiology of these conditions remains unclear, but a common feature is chronic inflammation and/or trauma. The inflammation may result from inflammatory bowel disease, resolving ischemia, or trauma associated with internal intussception or prolapse of the rectum, direct digital trauma, or the forces associated with evacuating a hard stool.

INTRODUCTION

Colitis cystica profunda (CCP) and solitary rectal ulcer syndrome (SRUS) are uncommon and controversial conditions. Cystica profunda is a benign condition characterized by mucin-filled cysts located deep to the muscularis mucosae. Although cysts can occur in any

 Copyright 2004, Elsevier (USA). All rights reserved.

segment of the digestive tract submucosa, they are most frequent in the colon and rectum. When these lesions are found in the colon or rectum, they are called colitis cystica profunda and appear as nodules or masses on the anterior rectal wall. Patients can be asymptomatic (with the lesions identified on screening endoscopy) or complain of rectal bleeding, mucous discharge, or anorectal discomfort. Most patients will admit to difficulty with bowel movements. CCP is a pathologic diagnosis whose most important aspect is to differentiate it from colorectal adenocarcinoma. This prevents unnecessary radical operations.

SRUS is a related clinical condition characterized by rectal bleeding, copious mucous discharge, anorectal pain, and difficult evacuation. Despite its name, patients with this condition can have single, multiple, or no rectal ulcers. When present, the ulcers usually occur on the anterior rectal wall just above the anorectal ring. Less commonly, they may occur from just above to 15 cm above the dentate line. Ulcers usually appear shallow with a "punched out" gray-white base surrounded by hyperemia.

DIAGNOSIS

In symptomatic patients, an endoscopic evaluation of the distal colon and rectum will reveal the lesions described above. Defecography documents intussusception in 45 to 80% of patients. The differential diagnosis of both CCP and SRUS includes polyps, endometriosis, inflammatory granulomas, infectious disorders, drug-induced colitidies, and mucus-producing adenocarcinoma. Differentiation among these entities is possible with an adequate biopsy. Biopsies obtained via a rigid proctoscope, or an endoscopic snare excision, may be necessary to obtain enough tissue for an accurate diagnosis. CCP is characterized pathologically by mucous cysts lined by normal columnar epithelium located deep to the muscularis mucosae. The overlying mucosa may be normal or ulcerated and the submucosa surrounding the cysts is fibrotic and contains a mixed inflammatory infiltrate. In adenocarcinoma, the epithelium is dysplastic and the surrounding stroma is reactive.

TREATMENT

Treatment is directed at reducing symptoms or preventing some of the proposed etiologic mechanisms. Conservative therapy (high-fiber diet and modifying bowel movements to avoid straining) will reduce symptoms in most patients and should be tried first. Patients without rectal intussusception should be offered biofeedback to retrain their bowel function. Pharmacologic therapy has had limited success, but it is reasonable to try this procedure before embarking on surgery. If symptoms persist, a localized resection may be considered in selected patients. Those suitable for localized resection should be significantly symptomatic, be good surgical risks, and have localized, accessible areas of disease. Patients with prolapse are considered for surgical treatment [abdominal rectopexy, segmental resection and rectal fixation, perineal protectomy (Altmeier), or a mucosal protectomy (Delorme)]. Those without prolapse may be offered excision, which varies from a transanal excision to a major resection with coloanal pull-through.

CONCLUSION

Colitis cystica profunda and solitary rectal ulcer syndrome are uncommon and related colorectal conditions. As benign conditions, efforts are directed to establishing the diagnosis, excluding malignancy, and treating symptoms. A directed history, physical examination, and endoscopic biopsy will confirm the diagnosis. Therapy to modify bowel movements and habits has had the most success. If these measures fail, surgical therapy to correct rectal prolapse or locally excise the lesions may be considered.

See Also the Following Articles

Intussusception • Solitary Rectal Ulcer Syndrome

Further Reading

Beck, D. E. (2002). Colitis cystica profunda and solitary rectal ulcer syndrome. *Curr. Treat. Opt. Gastroenterol.* **5**, 231–237.

Felt-Bersma, R. J., and Cuesta, M. A. (2001). Rectal prolapse, rectal intussusception, rectocele, and solitary rectal ulcer syndrome. *Gastroenterol. Clin. North Am.* **30**, 199–222.

Goodall, H. B., and Sinclair, I. S. (1957). Colitis custica profunda. *J. Pathol.* **73**, 33–42.

Gordon, P. H. (1999). Miscellaneous entities. *In* "Principles and Practice of Surgery for the Colon, Rectum, and Anus" (P. H. Gordon, and and S. Nivatvongs, ed.), pp. 1394–1396. Quality Medical Publishing, St. Louis, MO.

Green, G. I., Ramos, R., Bannayan, G. A., and McFee, A. S. (1974). Colitis cystica profunda. *Am. J. Surg.* **127**, 749–752.

Keighley, M. R. B., and Williams, N. S. (1993). Solitary rectal ulcer syndrome. *In* "Surgery of the Anus, Rectum, and Colon" (M. R. B. Keighley and N. S. Williams, eds.), pp. 720–738. W. B. Saunders, London.

Madoff, R. D. (1998). Rectal prolapse and intussusception. *In* "Fundamentals of Anorectal Surgery" (D. E. Beck and S. D. Wexner, ed.), 2nd Ed., pp. 99–114. W. B. Saunders, London.

Rutter, K. R. P. (1985). Solitary rectal ulcer syndrome of the rectum: Its relation to mucosal prolapse. *In* "Coloproproctology in the Pelvic Floor" (M. M. Henry and M. Swash, eds.), pp. 282–289. Butterworths, London.

Colitis Pseudomembranous

SUE ENG AND CHRISTINA M. SURAWICZ
University of Washington School of Medicine, Seattle

cytotoxin Substance that inhibits or prevents the function of the cells leading to its destruction.
enterotoxin Cytotoxin specific for the cells of the intestinal mucosa.
probiotics Live microorganisms that confer health benefits beyond inherent general nutrition when ingested.
toxic megacolon Acute nonobstructive dilation of the colon.

Pseudomembranous colitis is a pathological diagnosis of exudative plaques adherent to the colonic mucosa. This colitis is most commonly associated with antibiotic usage that leads to the disruption of the normal flora, allowing overgrowth of the bacterium *Clostridium difficile* and the elaboration of toxins. Although antibiotics are the most common cause, cancer chemotherapy also causes this colitis, and sporadic cases also occur. Pseudomembranous colitis is the most severe form in a wide spectrum of clinical diseases that result from *C. difficile* infection, with the associated self-limited diarrhea being more common and much less severe.

INTRODUCTION

Prior to the antibiotic era, pseudomembranous colitis (PMC) was a relatively rare disease and was associated with various risk factors, including surgery (especially of the gastrointestinal tract), shock, uremia, sepsis, ischemic cardiovascular disease, heavy metal poisoning, and colon cancer. However, in the past four decades, virtually all cases of PMC have been associated with *Clostridium difficile*, an anaerobic, gram-positive, spore-forming bacillus. More than 95% of *C. difficile* infections occur during or after antibiotic therapy, and almost every antibiotic has been implicated. In contrast, in antibiotic-associated diarrhea and antibiotic-associated colitis, only 10–30% and 60–75% cases, respectively, are associated with *C. difficile*. The increasing use of broad-spectrum antibiotics is reflected in the increasing incidence of *C. difficile* infections. The clinical manifestations of *C. difficile* infection range from asymptomatic carriage to *C. difficile*-associated diarrhea, to *C. difficile*-associated colitis, to the most severe form of PMC.

EPIDEMIOLOGY

Clostridium difficile is the most common organism isolated in nosocomial diarrhea. There are approximately 300,000 cases of *C. difficile* infection per year that occur in hospitals or long-term care facilities in the United States. Infection can occur sporadically or in outbreaks and can add an estimated cost of $2000 per patient to the cost of the hospitalization. Community-acquired *C. difficile* occurs less frequently and accounts for about 20,000 cases per year in the United States. *Clostridium difficile* can be found in up to 5% of healthy adults, and up to 20–30% of hospitalized patients will be colonized with *C. difficile* due to persistence of spores on contaminated surfaces. Transmission occurs via a fecal–oral route; the unwashed hands of the hospital staff are a common source.

Other risk factors for *C. difficile*-associated disease are age above 60 years, female gender, gastrointestinal manipulation with enemas or nasogastric tubes, inflammatory bowel disease, renal disease, chemotherapy, and HIV infection or AIDS.

PATHOPHYSIOLOGY

Exposure to antibiotics leads to alteration of the host's normal gut flora and presumably allows for the colonization of *C. difficile* acquired from the environment. The toxins produced by *C. difficile* in the intestinal lumen bind to the colonic mucosa and lead to the clinical manifestations of the disease. Although there are several toxic factors elaborated by *C. difficile*, only two are well studied. Toxin A is a 308-kDa enterotoxin and toxin B is a 269-kDa cytotoxin. Both toxins are endocytosed by the intestinal epithelium, where they can disrupt the actin cytoskeleton, leading to apoptosis. They can also induce macrophages and mast cells to elaborate inflammatory mediators such as tumor necrosis factor and interleukins. The major difference in the two toxins is that, in experimental animal models, toxin A leads to mucosal damage and the accumulation of viscous hemorrhagic fluid, whereas these enterotoxic activities do not occur with toxin B. However, toxin B cytopathic effects on cell culture lines are 1000 times more potent compared to

 Copyright 2004, Elsevier (USA). All rights reserved.

toxin A effects. Although *C. difficile* strains associated with clinical disease usually produce both toxin A and toxin B, there are reports of toxin A-negative/toxin B-positive strains that can also cause the full spectrum of disease, including PMC.

CLINICAL MANIFESTATION

Diarrheal symptoms usually occur within 1 week of starting antibiotics but can be delayed for as long as 8 weeks after discontinuing antibiotics, though the latter is uncommon. The reason for the range of clinical manifestations of *C. difficile* infection is not known, but the presence of immunoglobulin G (IgG) antibodies to toxin A may be one explanation; colonized asymptomatic patients are more likely to have IgG antibodies compared to patients who are symptomatic from *C. difficile*. Patients who are colonized with *C. difficile* are asymptomatic carriers and they continue to shed bacteria in the stool and represent a continued source of infection. In *C. difficile* diarrhea, there are usually only three to four loose watery stools and the patient may or may not experience lower abdominal crampy pain. Fever and leukocytosis are usually mild or absent. On endoscopy, the colonic mucosa is usually grossly normal. Patients with *C. difficile* colitis are clinically more ill. Constitutional symptoms of anorexia, low-grade fever, and nausea are usually present as well as profuse watery diarrhea with 5–15 stools per day, leading to dehydration and electrolyte abnormalities. Leukocytosis is typical and on endoscopic exam, there may be nonspecific patchy or diffuse erythematous colitis.

PMC will present with the same symptoms as *C. difficile* colitis, but on endoscopic evaluation, PMC will have the typical raised yellow or off-white plaques adherent to the colonic mucosa. Hypoalbuminemia occurs in a subset of PMC patients from protein-losing enteropathy; it may be secondary to the leakage of serum albumin from the affected colonic wall and can result in peripheral edema or anasarca. Although *C. difficile* colitis and PMC can diffusely affect the colon, the left colon and rectum are usually more severely affected. Some patients may present with colitis or PMC localized to only the right colon, but it is not known how commonly this isolated right-sided colitis occurs. In such cases, diarrhea may be mild or absent and the predominant symptoms are fever, right-sided abdominal pain, and tenderness with decreased intestinal motility.

Uncommonly, PMC can lead to toxic megacolon with associated ileus. At this point, patients may exhibit diffuse abdominal pain with high fever, chills, and minimal diarrhea secondary to the ileus. Laboratory evaluation may show a leukemoid reaction. Patient may require emergent laparotomy with subtotal colectomy if there is evidence of clinical deterioration due to imminent or actual perforation despite adequate antibiotic treatment.

Relapse of *C. difficile* is the recurrence of symptoms a few days to weeks or months after discontinuation of successful treatment; relapse may be due to the persistence of the original *C. difficile* or to reinfection by another strain. The frequency of relapses ranges from 5 to 50% and repeated episodes of recurrence are not common despite apparent successful treatment after each episode.

DIAGNOSIS

The gold standard for diagnosing *C. difficile* infection has long been the tissue culture test for toxin B. Filtrates of diarrheal stool are added to a monolayer of cultured fibroblasts; if toxin B is present, the cell culture will undergo a characteristic cytopathic effect. Preincubating the filtrate with antibodies against the toxin prevents the cytopathic effect and demonstrates that the cytotoxin present is toxin B from *C. difficile*. Sensitivity (94–100%) and specificity (99%) are both high and improve as the disease progresses along the clinical spectrum from diarrhea to colitis to PMC. The utility of the tissue culture test is limited by its expense and the time for testing, typically requiring 2–3 days. Thus, many hospital laboratories are using enzyme-linked immunosorbent assay (ELISA) technologies [also called enzyme immunoabsorbent assays (EIAs)]. These tests are rapid, less expensive, and widely used in the United States for the detection of toxin A or B. Although specificity is 99%, sensitivity ranges from 70 to 90% and may miss up to 10% of patients with *C. difficile* infection diagnosed either by endoscopy or by tissue culture test for toxin B. Thus, a negative EIA or ELISA test for either toxin A or B does not rule out *C. difficile*.

The latex agglutination test (LAT) detects the glutamate dehydrogenase produced by *C. difficile* rather than the actual toxins. The sensitivity of the LAT is equal to that of the EIA, but the specificity is lower because the protein is also produced by other colonic bacteria and nontoxigenic strains of *C. difficile*. Another diagnostic test is to culture stool for *C. difficile*. This test is limited by its cost, the time delay of 2–3 days for results, and its inability to distinguish between toxigenic and nontoxigenic strains; nontoxigenic strains would presumably not be capable of causing clinical disease. Asymptomatic carriers will have a positive culture.

Although endoscopy is the most rapid way to diagnose PMC, it is generally not recommended due to the high cost and the risk of perforation. It should be reserved for situations in which rapid diagnosis is needed, other test results are delayed and are not very sensitive, or if the patient has an ileus and no stool specimen is obtained.

TREATMENT

In mild disease, discontinuation of the antibiotic and supportive care lead to resolution in 15–20% of cases. Opiates and antidiarrheal medications should be avoided. Specific treatment should be initiated when supportive therapy fails after several days, if the culprit antibiotic cannot be discontinued, or if symptoms are severe. Specific treatment consists of either oral metronidazole at a dose of 250–500 mg four times a day or oral vancomycin at a dose of 125–500 mg four times a day, for 7 to 10 days. Although metronidazole can be administered intravenously at a dose of 500–700 mg three or four times a day if the patient is unable to take oral medication, it is not as effective as oral administration. Despite equivalent efficacy of metronidazole and vancomycin, metronidazole is the first-line therapy. Vancomycin is more expensive and there is the evidence that increased usage leads to the emergence of more vancomycin-resistant bacteria (such as enterococci) in hospital environments. Use of vancomycin should be reserved for patients who are unable to tolerate metronidazole or have failed to respond.

Recurrent infection with *C. difficile* is a difficult management problem, and no specific regimen is uniformly effective. One approach is a tapered dose regimen; a prolonged course of vancomycin is slowly tapered by reducing the dose and frequency over several weeks. Another approach is a pulse regimen; the antibiotic, with addition of cholestyramine, is given over 5- to 7-day periods, alternating with periods of no antibiotics. A combination of pulse or tapered regimen can also be effective. Another approach is the use of probiotics, such as *Lactobacillus GG* and *Saccharomyces boulardii*. *Saccharomyces boulardii* has been shown in controlled trials to decrease recurrences in combination with antibiotics, and especially with high-dose vancomycin. Probiotics in conjunction with antibiotics appear to decrease the frequency of recurrent episodes of *C. difficile*.

Surgical treatment is required in 4–5% of cases. The indications are clinical deterioration despite medical treatment, perforation, hemorrhage, sepsis, and multiorgan failure. The procedure of choice is total abdominal colectomy with ileostomy, because partial resections with diverting ileostomy or colostomy are associated with higher mortality rates.

See Also the Following Articles

Bacterial Toxins • Colitis, Radiation, Chemical, and Drug-Induced • Nosocomial Infections • NSAID-Induced Injury

Further Reading

Barbut, F., and Petit, J.-C. (2001). Epidemiology of *Clostridium difficile*-associated infections. *Clin. Microbiol. Infect. Dis.* **7**, 405–410.

Bartlett, J. (1998). Pseudomembranous enterocolitis and antibiotic-associated colitis. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed. Vol. 2., pp. 1633–1647. W. B. Saunders, Philadelphia.

Bliss, D., Johnson, S., Savik, K., Clabots, C., Willard, K., and Gerding, D. (1998). Acquisition of *Clostridium difficile* and *Clostridium difficile*-associated diarrhea in hospitalized patients receiving tube feeding. *Ann. Int. Med.* **129**, 1012–1018.

Borriello, S. (1998). Pathogenesis of *Clostridium difficile* infection. *J. Antimicrob. Chemother.* **41**(Suppl. C), 1319.

Brazier, J. (1998). The diagnosis of *Clostridium difficile*-associated disease. *J. Antimicrob. Chemother.* **41**(Suppl. C), 2940.

Brown, E., Talbot, G. H., Axelrod, P., Provencher, M., and Hoegg, C. (1990). Risk factors for *Clostridium difficile* toxin-associated diarrhea. *Infect. Control Hosp. Epidemiol.* **11**, 283–290.

Fekety, R. (1997). Guidelines for the diagnosis and management of *Clostridium difficile*-associated diarrhea and colitis. American College of Gastroenterology, Practice Parameters Committee. *Am. J. Gastroenterol.* **92**, 739–750.

Kreisel, D., Savel, T., Silver, A., and Cunningham, J. (1995). Surgical antibiotic prophylaxis and *Clostridium difficile* toxin positivity. *Arch. Surg.* **130**, 989–993.

Kyne, L., Warny, M., Qamar, A., and Kelly, C. (2000). Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibodies against toxin A. *N. Engl. J. Med.* **342**, 390–397.

McFarland, L., Surawicz, C., and Greenberg, R. (1994). A randomized placebo-controlled trial of *Saccharomyces boulardii* in combination with standard antibiotics for *Clostridium difficile* disease. *J. Am. Med. Assoc.* **271**, 1913–1918.

Colitis, Collagenous and Lymphocytic

WILLIAM J. TREMAINE
Mayo Clinic

collagenous colitis Syndrome of watery diarrhea with colonic mucosal abnormalities that include a thickened subepithelial collagen band and a chronic inflammatory cell infiltrate in the lamina propria.

lymphocytic colitis Syndrome of watery diarrhea with colonic mucosal abnormalities that include intraepithelial lymphocytes and a chronic inflammatory cell infiltrate in the lamina propria.

microscopic colitis Syndrome of watery diarrhea with colonic mucosal abnormalities that include the histologic features of either collagenous colitis or lymphocytic colitis.

Collagenous colitis and lymphocytic colitis are two similar disorders that fit under the heading of microscopic colitis, a syndrome of watery diarrhea with a chronic inflammatory infiltrate in the colonic mucosa, and without specific colonoscopic abnormalities. These disorders have been recognized only in the past 25 years. Because colonic mucosal biopsies are necessary to make the diagnosis, many patients with microscopic colitis were previously diagnosed as having diarrhea-predominant irritable bowel syndrome or chronic diarrhea of unknown origin. The clinical features of collagenous colitis and lymphocytic colitis are similar, and differentiation is made by the histological findings.

CLINICAL AND PATHOLOGIC FEATURES

Clinical Presentation

The usual presentation of collagenous colitis and lymphocytic colitis is chronic watery diarrhea with a few stools, or up to 10 or more stools a day. Nocturnal stools, cramping abdominal pain, and fecal urgency are typical. About 40% of patients have an abrupt onset of symptoms and the rest have a gradual onset. Symptoms may be mild and intermittent, without nocturnal stools and with clinical features indistinguishable from diarrhea-predominant irritable bowel syndrome. Rarely, the diarrhea is severe, with up to 5 liters of watery stool per 24 hours reported. On physical examination, patients may have mild to moderate abdominal tenderness, or the exam may be normal.

The course of these diseases is variable. Symptoms may be intermittent, with apparent spontaneous resolution for months followed by recurrent symptoms. In others, symptoms are chronic or progressive.

Histopathology

Collagenous Colitis

The diagnosis of collagenous colitis is made by finding specific colonic mucosal histologic abnormalities in a patient with watery diarrhea. A continuous or patchy collagen band is present between myofibroblasts and around capillaries in the upper lamina propria. The width of the collagen band may vary from 7 to 100 μm, and the presence of the collagen band is necessary to make the diagnosis. Intraepithelial lymphocytes may or may not be present in collagenous colitis. There is a chronic inflammatory cell infiltrate in the lamina propria, composed mainly of lymphocytes, but also including eosinophils, mast cells, and some neutrophils. Crypt distortion with branching may be found.

Lymphocytic Colitis

There is no increase in the subepithelial collagen in lymphocytic colitis. By definition, intraepithelial lymphocytes are present. There is a chronic inflammatory cell infiltrate in the lamina propria, just as is found in collagenous colitis.

EPIDEMIOLOGY

The reported incidence of collagenous colitis is 1–2 per 100,000 persons, and the prevalence is about 15 per 100,000. Incidence and prevalence for lymphocytic colitis are unknown. The disorders have been reported from countries throughout the world. Most series of patients have shown a female predominance for collagenous colitis, but the female:male ratio has been nearly equal for lymphocytic colitis. The mean age for the onset of collagenous colitis is reported to be 59 years; for lymphocytic colitis, the mean age at onset is reported

 Copyright 2004, Elsevier (USA). All rights reserved.

to be 51 years. Four cases of collagenous colitis in children aged 5–12 years have been reported. A familial occurrence of microscopic colitis has been reported in seven families, each with two affected sisters. In one family, the sister who smoked had collagenous colitis and the sister who did not smoke had lymphocytic colitis.

POSSIBLE CAUSES

The causes of these disorders are unknown, but they are likely the result of mucosal injury from unidentified toxins in the fecal stream. Several different observations support this hypothesis. Fecal extracts from patients with collagenous colitis are cytotoxic to McCoy cell cultures, and mixing cholestyramine with the fecal extract eliminates the cytotoxic effect. However, only a minority of patients with collagenous colitis who are treated with cholestyramine improve, so toxins other than those bound by cholestyramine appear to be important in most patients. Ileostomy with diversion of the fecal stream results in clinical and histological remission, and restoration of intestinal continuity causes clinical and histologic relapse.

A recent study suggests that the diarrhea in collagenous colitis is mediated by up-regulation of nitric oxide synthase, with a marked increase in the production of intraluminal nitric oxide, which in turn causes the secretory diarrhea. However, the agents that induce nitric oxide production in these disorders are unidentified. Originally, the collagen band was thought to cause diarrhea by creating a barrier effect to prevent resorption of fluids, but the collagen band is likely a consequence of the mucosal injury and not the cause of the diarrhea.

It is unclear whether collagenous colitis and lymphocytic colitis are two distinct entities or are different expressions of the same underlying process. Indeed, biopsies showing collagenous colitis and lymphocytic colitis can been found in the same individual at one time, or on sequential exams.

DISEASE ASSOCIATIONS

Celiac sprue has been reported in some patients with collagenous colitis. In a series from the Mayo Clinic, among 45 patients with collagenous colitis in whom intestinal malabsorption was suspected and proximal small intestinal biopsies were obtained, 1 out of the 45 patients had celiac sprue. There are case reports of ulcerative colitis and of Crohn's disease that developed in patients with previously diagnosed collagenous colitis. Collagenous colitis has been found coincidentally in two patients with adenocarcinoma of the colon. However, colorectal cancer did not develop in a series of 117 patients with collagenous colitis followed for 7 years. The relative risk of overall malignancy and mortality is no different from that of the general population. Because collagenous colitis and lymphocytic colitis do not have a premalignant potential, there is no need for increased surveillance for colorectal cancer in this group, above and beyond the recommendations for screening in average-risk patients.

TREATMENT

Medical Therapy

Bismuth subsalicylate in a dose of nine chewable tablets per day in three divided doses appears effective in some patients with either collagenous or lymphocytic colitis, based on the findings in a small, placebo-controlled trial. A recent placebo-controlled trial in 28 patients with collagenous colitis showed efficacy of delayed-release budesonide, a corticosteroid with a potent local topical effect in the proximal colon and terminal ileum, but with less systemic effect than prednisone. A number of other drugs have been used in uncontrolled studies and treatment has been largely empirical. A stepwise approach is used for therapy. Patients with mild symptoms may get relief with symptomatic therapy such as loperamide, diphenoxylate with atropine, or bulk agents. For more troublesome symptoms, chewable bismuth subsalicylate tablets, four to nine per day, may be effective. In those patients who do not respond to bismuth, sulfasalazine or mesalamine may be beneficial. Other agents reported to be beneficial in small, uncontrolled series include cholestyramine and antibiotics such as metronidazole, erythromycin, and penicillin. For more severe symptoms, a corticosteroid such as prednisone or budesonide may be used. Patients dependent on or refractory to corticosteroids have been successfully treated with the immune modulators, 6-mercaptopurine and azathioprine. Octreotide has been effective in some patients with severe diarrhea.

Surgery

In a small number of patients with collagenous and lymphocytic colitis, the diarrhea remains severe despite medical therapy. With a diverting ileostomy, the diarrhea resolves. Some patients with collagenous colitis and with lymphocytic colitis have undergone proctocolectomy with ileal pouch to anal anastomosis.

CONCLUSION

Collagenous colitis and lymphocytic colitis, which comprise the syndrome of microscopic colitis, are unique forms of inflammatory bowel disease. Unlike Crohn's disease and ulcerative colitis, they do not carry an increased risk of cancer, and they rarely require surgery. Few controlled trials of therapy have been performed, but, in general, these disorders appear to respond to antiinflammatory or immune modulator therapy.

See Also the Following Articles

Anti-Diarrheal Drugs • Colitis, Ulcerative • Crohn's Disease • Diarrhea

Further Reading

Baert, F., Wouters, K., D'Haens, G., *et al.* (1999). Lymphocytic colitis: A distinct clinical entity? A clinicopathological confrontation of lymphocytic and collagenous colitis. *Gut* **45**(3), 375–381.

Fine, K. D., and Lee, E. L. (1998). Efficacy of open-label bismuth subsalicylate for the treatment of microscopic colitis. *Gastroenterology* **114**(1), 29–36.

Jarnerot, G., Hertervig, E., Granno, C., *et al.* (2001). Familial occurrence of microscopic colitis: A report on five families. *Scand. J. Gastroenterol.* **36**(9), 959–962.

Pardi, D. S., Loftus, E. V., Jr., Tremaine, W. J., and Sandborn, W. J. (2001). Treatment of refractory microscopic colitis with azathioprine and 6-mercaptopurine. *Gastroenterol.* **120**(6), 1483–1484.

Perner, A., Andresen, L., Normark, M., *et al.* (2001). Expression of nitric oxide synthases and effects of L-arginine and L-NMMA on nitric oxide production and fluid transport in collagenous colitis. *Gut* **49**(3), 387–394.

Tremaine, W. J. (2000). Collagenous colitis and lymphocytic colitis. *J. Clin. Gastroenterol.* **30**(3), 245–249.

Yusuf, T. E., Soemijarsih, M., Arpaia, A., Goldberg, S. L., and Sottile, V. M. (1999). Chronic microscopic enterocolitis with severe hypokalemia responding to subtotal colectomy. *J. Clin. Gastroenterol.* **29**(3), 284–288.

Zins, B. J., Sandborn, W. J., and Tremaine, W. J. (1995). Collagenous and lymphocytic colitis: subject review and therapeutic alternatives. *Am. J. Gastroenterol.* **90**(9), 1394–1400.

Colitis Cystica Profunda and Solitary Rectal Ulcer Syndrome

DAVID E. BECK

Ochsner Clinic Foundation, Louisiana

colitis cystica profunda A benign pathologic condition characterized by mucin-filled cysts located deep to the muscularis mucosae.

intussception A condition in which the rectum slips inside itself; this condition can cause obstructive symptoms and damage to the rectum.

perineal protectomy A surgical procedure, performed via the anus, that excises redundant rectum and reconnects the two remaining ends.

prolapse Rectal intussception continues to the point that portions of the rectum are outside the anus.

solitary rectal ulcer syndrome A clinical condition characterized by rectal bleeding, copious mucous discharge, anorectal pain, and difficult evacuation.

Colitis cystica profunda and solitary rectal ulcer syndrome are closely related diagnoses and some authors consider them interchangeable. The etiology of these conditions remains unclear, but a common feature is chronic inflammation and/or trauma. The inflammation may result from inflammatory bowel disease, resolving ischemia, or trauma associated with internal intussception or prolapse of the rectum, direct digital trauma, or the forces associated with evacuating a hard stool.

INTRODUCTION

Colitis cystica profunda (CCP) and solitary rectal ulcer syndrome (SRUS) are uncommon and controversial conditions. Cystica profunda is a benign condition characterized by mucin-filled cysts located deep to the muscularis mucosae. Although cysts can occur in any

 Copyright 2004, Elsevier (USA). All rights reserved.

Colitis, Indeterminate

KONSTANTINOS A. PAPADAKIS
Cedars-Sinai Medical Center and University of California, Los Angeles

Crohn's disease Chronic idiopathic inflammatory bowel disease characterized by patchy, transmural, and often granulomatous inflammation of any part of the gastrointestinal tract; the inflammatory process more often involves the ileum and colon.

indeterminate colitis Chronic idiopathic inflammatory colitis that cannot be classified with certainty as ulcerative colitis or Crohn's disease based on endoscopic, radiographic, and histopathologic criteria.

ulcerative colitis Chronic idiopathic inflammatory colitis, characterized by continuous superficial inflammation extending to a varying degree in a continuous fashion from the rectum to the proximal colon.

Inflammatory bowel disease (IBD) has been traditionally classified into ulcerative colitis (UC) and Crohn's disease (CD). In UC, the mucosal inflammation is limited to the superficial layers of the colon (mucosa and submucosa), whereas in CD, which can affect any part of the gastrointestinal tract, the inflammatory process is patchy, transmural, and often associated with granuloma formation. The term "indeterminate colitis" (IC) was originally used by pathologists to describe colitis observed in surgical specimens that could not be accurately classified as UC or CD. This was therefore a provisional classification prior to establishing a definitive diagnosis. More recently, the use of the term IC has been extended in the precolectomy evaluation of patients with IBD to characterize the type of colitis (in the absence of small intestinal inflammation) that cannot be classified with certainty as UC or CD. It is still debated whether IC represents a distinct clinical entity or simply a problem of classification at the time of evaluation.

INTRODUCTION

Recent advances in our understanding of the pathogenesis of mucosal inflammation in animal models and humans suggest that inflammatory bowel disease (IBD), and particularly Crohn's disease (CD), may represent a heterogeneous group of diseases based on clinical, subclinical, and genetic characteristics. This notion has been supported by the development of numerous animal models of IBD with distinct phenotypes and immunologic features following selective manipulation of a variety of genes, including those of pro-inflammatory or immunoregulatory cytokines. Given these advances, indeterminate colitis (IC) may in fact be a distinct manifestation within the inflammatory bowel disease spectrum.

INCIDENCE, NATURAL HISTORY, AND DIAGNOSTIC EVALUATION

Overall, 5 to 23% of initial diagnoses of IBD are classified as IC and its incidence is approximately 2.4 per 100,000. It has been reported that IC is associated with a higher risk of colorectal cancer development and increased mortality compared to ulcerative colitis (UC). Also, the cumulative incidence of colectomy in patients with IC in a population-based study was four times higher than in patients with definite UC. Since the cancer risk is related to the extent of colonic inflammation, the higher risk of colorectal cancer in patients with IC may be due to the fact that most patients classified as having IC are more likely to have pancolitis. In addition, the higher mortality rate in patients with IC reflects a more severe and extensive inflammatory process in patients with IC compared to patients with UC, since the majority of the deaths have been attributed to complications of colitis. Recent studies have shown that most patients with IC will be reclassified as having CD or UC within 8 years of the initial diagnosis. The probability of having a diagnosis of Crohn's disease was increased in patients with fever at their initial presentation, segmental endoscopic lesions, or extraintestinal complications and in current smokers, whereas the probability of having a diagnosis of UC was increased in patients who had not undergone appendectomy before diagnosis. Serologic testing in patients with an initial diagnosis of IC may be helpful in categorizing the disease and predict the follow-up diagnosis. The most extensively studied serologic markers in IBD include perinuclear anti-neutrophil cytoplasmic antibody (pANCA) and anti-*Saccharomyces cerevisiae* antibodies (ASCA). As subclinical indicators of immune dysregulation, these markers can be used in IBD to stratify patients based on clinical, immunologic, and

 Copyright 2004, Elsevier (USA). All rights reserved.

genetic characteristics. pANCA is detected in the serum of 60 to 70% of patients with UC and in 10–20% of patients with CD. ASCA is present in 50 to 70% of patients with CD and 6 to 14% of patients with UC. ASCA is rarely expressed in individuals who do not have IBD and thus is highly specific for CD. In a recent prospective study, pANCA and ASCA were studied in 97 patients with IC. ASCA-positive/pANCA-negative IC predicted evolution to CD in 80% of the patients. ASCA-negative/pANCA-positive IC predicted UC in 64% of patients, whereas this combination was 100% predictive of UC or "UC-like CD." The latter represent cases of Crohn's disease with left-sided colonic involvement, which resemble UC both clinically and endoscopically. Interestingly, almost half of the patients with IC were negative for both of these serologic markers and most of these patients (85%) continued to carry the diagnosis of IC in follow-up evaluation.

MEDICAL THERAPY OF INDETERMINATE COLITIS

Unfortunately, evidence-based medical treatment strategies are lacking for IC simply because randomized controlled trials have included only well-documented cases of UC or CD. Since many patients with IC will eventually be found to have either UC or CD, any therapy that is effective in both diseases may prove useful in the treatment of IC. The goals of medical treatment in IC are the same as in the other forms of IBD, namely, induction and maintenance of remission (absence of symptoms related to colitis such as abdominal pain, diarrhea, rectal bleeding, or tenesmus). The mainstay of medical treatment for mild to moderately active IC includes the use of aminosalicylates. There are three delivery systems for oral aminosalicylates and several topical (rectal) formulations. The oral formulations include azo-bond conjugates (sulfasalazine, olsalazine, and balsalazide), pH-dependent mesalamine with varied Eudragit (USP) coatings (Asacol, Salofalk, and Claversal), and time/pH release formulations of mesalamine encapsulated into ethylcellulose beads (Pentasa). Since IC usually involves the entire colon, combined administration of oral and topical 5-ASA may have an additive therapeutic effect. In patients with IC who fail to respond to 5-ASA treatment or those with severe disease, corticosteroids will often be required to induce remission. However, corticosteroids have no maintenance benefit in patients with CD or UC and therefore probably should not be used long-term in patients with IC either. In cases where IC becomes steroid-dependent, immunomodulatory treatment with 6-mercaptopurine (6-MP) or azathioprine (AZA) should be initiated. In patients who are intolerant to 6-MP/AZA or fail to respond adequately methotrexate can be tried. This drug has been shown to be effective for the induction and maintenance of remission in steroid-dependent, chronic active CD but is less effective in UC. However, its efficacy in IC is unknown. Novel treatments that target inflammatory mediators [e.g., tumor necrosis factor α (TNFα)] thought to be central in the pathogenesis of mucosal inflammation, such as the human/chimeric monoclonal antibody directed against TNFα, infliximab (Remicade), have been found to be highly effective for the treatment of inflammatory and fistulizing CD, but less so for the treatment of UC. In a small series of patients with treatment-resistant IC, infliximab was effective in 67% of patients and several patients avoided colectomy. Cyclosporine (CsA) may also be effective in patients with IC who failed to respond to corticosteroid treatment similarly to patients with UC. Successful induction of remission with CsA, however, will invariably require the long-term use of 6-MP/AZA for maintenance of remission.

SURGICAL THERAPY OF INDETERMINATE COLITIS

As discussed in a previous section, the risk of major surgery is increased several fold in patients with IC compared to those with UC. Ileal pouch-anal anastomosis (IPAA) has been considered the gold standard surgical treatment for UC and is generally contraindicated in patients with CD due to the high incidence of complications, including pelvic abscess, fistula, and pouch failure, which may require pouch excision. In some cases of Crohn's disease, those with no small bowel or ano-perineal disease may do well with restorative proctocolectomy. This treatment, therefore, has been advocated in a select group of patients. The major decision regarding surgical treatment of IC is whether to perform restorative proctocolectomy (IPAA) or total colectomy with ileostomy or in cases of rectal sparing subtotal colectomy with ileorectal anastomosis. The outcome of patients operated on for IC with IPAA, who did not develop CD in long-term follow-up evaluation, is reportedly as good as those operated on for UC, although other studies have reported a higher incidence of complications in patients operated on for IC. This discrepancy may be related to the differences in the definition of IC used in different studies. Nevertheless, the development of CD following IPAA for either UC or IC is associated with poor long-term outcome. Prior to surgery, however, extensive

evaluation of patients with an upper gastrointestinal series/small bowel follow-through or even wireless capsule endoscopy should be performed to rule out Crohn's disease of the small bowel and avoid performing an IPAA in these patients. IBD serologies precolectomy may be helpful in predicting the potential development of complications following IPAA for IC, but long-term follow-up studies are needed to assess their utility in this setting.

CONCLUSION

Indeterminate colitis was originally reported in colectomy specimens to describe the inflammatory process that could not be adequately classified as UC or CD. Recently, the term has been extended in cases of chronic IBD, without small bowel inflammation and inconclusive diagnostic features of either UC or CD on endoscopic and histopathologic evaluation. Although long-term follow-up evaluation may change the diagnosis to either UC or CD, a significant number of patients will still have IC. This observation supports the idea that although IC may represent a problem of classification at the time of evaluation in some cases, other cases may truly represent a distinct disease entity. Medical therapies are similar to those used for the treatment of UC and sometimes CD. Surgical treatment may be needed in those patients with severe treatment-resistant disease.

See Also the Following Articles

Colectomy • Colitis, Ulcerative • Crohn's Disease • Ileoanal Pouch

Further Reading

Hanauer, S. B. (2001). Update on medical management of inflammatory bowel disease: Ulcerative colitis. *Rev. Gastroenterol. Dis.* **1**, 169–176.

Joossens, S., Reinisch, W., Vermeire, S., Sendid, B., Poulain, D., Peeters, M., Geboes, K., Bossuyt, X., Vandewalle, P., Oberhuber, G., Vogelsang, H., Rutgeerts, P., and Colombel, J. F. (2002). The value of serologic markers in indeterminate colitis: A prospective follow-up study. *Gastroenterology* **122**, 1242–1247.

Lindsey, I., Warren, B. F., and Mortensen, N. J. (2001). Indeterminate colitis: Surgical approaches. *In* "Advanced Therapy of Inflammatory Bowel Disease" (T. Bayless and S. Hanauer, eds.), pp. 241–244. B. C. Decker, Lewiston, NY.

MacDermott, R. P. (2001). Indeterminate colitis. *In* "Advanced Therapy of Inflammatory Bowel Disease" (T. Bayless and S. Hanauer, eds.), pp. 157–160. B. C. Decker, Lewiston, NY.

McIntyre, P. B., Pemberton, J. H., Wolff, B. G., Dozois, R. R., and Beart, R. W., Jr. (1995). Indeterminate colitis. Long-term outcome in patients after ileal pouch–anal anastomosis. *Dis. Colon Rectum* **38**, 51–54.

Meucci, G., Bortoli, A., Riccioli, F. A., Girelli, C. M., Radaelli, F., Rivolta, R., and Tatarella, M. (1999). Frequency and clinical evolution of indeterminate colitis: A retrospective multi-centre study in northern Italy. GSMII (Gruppo di Studio per le Malattie Infiammatorie Intestinali). *Eur. J. Gastroenterol. Hepatol.* **11**, 909–913.

Papadakis, K. A., and Targan, S. R. (1999). Serologic testing in inflammatory bowel disease: Its value in indeterminate colitis. *Curr. Gastroenterol. Rep.* **1**, 482–485.

Papadakis, K. A., Abreu, M. T., Targan, S. R., and Vasiliauskas, E. A. (2002). Infliximab in the treatment of indeterminate colitis. *Gastroenterology* **122**, W1356.

Price, A. B. (1978). Overlap in the spectrum of non-specific inflammatory bowel disease—"colitis indeterminate." *J. Clin. Pathol.* **31**, 567–577.

Stewenius, J., Adnerhill, I., Anderson, H., Ekelund, G. R., Floren, C. H., Fork, F. T., Janzon, L., Lindstrom, C., and Ogren, M. (1995). Incidence of colorectal cancer and all cause mortality in non-selected patients with ulcerative colitis and indeterminate colitis in Malmo, Sweden. *Int. J. Colorect. Dis.* **10**, 117–122.

Stewenius, J., Adnerhill, I., Ekelund, G., Floren, C. H., Fork, F. T., Janzon, L., Lindstrom, C., Mars, I., Nyman, M., and Rosengren, J. E. (1995). Ulcerative colitis and indeterminate colitis in the city of Malmo, Sweden. A 25-year incidence study. *Scand. J. Gastroenterol.* **30**, 38–43.

Wells, A. D., McMillan, I., Price, A. B., Ritchie, J. K., and Nicholls, R. J. (1991). Natural history of indeterminate colitis. *Br. J. Surg.* **78**, 179–181.

Colitis, Radiation, Chemical, and Drug-Induced

WALLACE K. MACNAUGHTON
University of Calgary, Canada

colitis Inflammation of the colon, which can be transmural or confined to the mucosa, which may be acute or chronic, and which may resolve, resolve and recur, or be persistent.

chemotherapeutics Anti-neoplastic drugs used to treat leukemic cancers, solid tumor cancers, or other cancers.

ionizing radiation Energetic particles, emitted by some isotopes of certain atoms, that interact with biological tissues to cause ionization, which leads to chemical, structural, or physiological changes in cells.

neuroleptics Drugs used to treat various psychiatric disorders including schizophrenia.

nonsteroidal anti-inflammatory drugs Medications used to treat a variety of inflammatory conditions and which are often prescribed for extended periods of time in the case of chronic inflammatory diseases such as rheumatoid arthritis.

Colonic inflammation, distinct from that which characterizes Crohn's disease or ulcerative colitis, is occasionally caused by external agents that are used to treat other diseases or which may gain access to the colon through ingestion or other routes. Many of the agents associated with noninflammatory bowel disease colitis, including radiation, chemicals, and drugs, are suspected, but not necessarily proven, causative agents. Causation has been implied in case studies, but often carefully controlled clinical trials are not available or ethically possible, resulting in uncertainty as to whether or not a given agent is responsible for the colitis.

INTRODUCTION

Colitis, or inflammation of the colon, is most commonly associated with the inflammatory bowel diseases Crohn's disease and ulcerative colitis. These diseases have an unknown etiology and are characterized by their "relapsing–remitting" nature. Less common, but no less deleterious to the patient, is colitis occurring in the setting of exposure to a defined external factor. In many cases, colitis is caused by a therapeutic modality that is being used for an indication not related to the colon per se. Examples are ionizing radiation, chemotherapeutic agents, other drugs, or other chemicals present in the environment and to which patients may be exposed. In most cases, when a specific agent is identified as causing colitis, cessation of that particular agent leads to the resolution of the colonic inflammation. In other cases, however, removal of the associated agent (such as cessation of abdominopelvic exposure to ionizing radiation in the case of radiotherapy) does not result in resolution of the colitis. In these cases, chronic colonic inflammatory disease may develop, which is often difficult to manage and is associated with significant morbidity.

RADIATION-INDUCED COLITIS AND RADIATION ENTEROPATHY

Abdominopelvic exposure to ionizing radiation, such as that received during cancer radiotherapy, can result in the development of an acute colonic inflammation that often progresses to a chronic radiation enteritis. Irradiation results in a rapid cellular burst of free radicals that can cause single- and double-stranded DNA breaks, as well as damage to proteins, lipids, carbohydrates, and other molecules. Cells that are rapidly dividing and undifferentiated (such as stem cells in the crypt region of the intestinal mucosa) are most radiosensitive, particularly when they are in the G2 or M phases of the cell cycle, and are induced to undergo p53-dependent apoptosis. Following irradiation, there is rapid activation of the coagulation cascade and apoptosis of endothelial cells within the radiation field. There is also an acute inflammatory response characterized by leukocyte infiltration. This inflammatory response can exacerbate the radiation-induced epithelial damage. The acute phase of the response to ionizing radiation will resolve, but during most fractionation protocols, another dose of radiation is delivered before this resolution can occur. Thus, there is an "accumulated injury," which can exacerbate the endothelial response and lead to increased production of cytokines such as

 Copyright 2004, Elsevier (USA). All rights reserved.

transforming growth factor-β (TGF-β), which can delay the repair process further.

Epithelial damage is associated with decreased barrier function. This is due to a combination of direct epithelial cellular damage, decreased responsiveness to secretagogues, and damage to the basement membrane. The impaired barrier function may lead to increased bacterial translocation, which has the potential to exacerbate or prolong the injury, although this remains controversial. Indeed, it has been observed that acute clinical symptoms (from the second to the sixth week postirradiation) are manifested even at times when endoscopic and histological changes have normalized. Nevertheless, impaired barrier function is thought to contribute to the progression to chronic disease. The chronic phase of radiation enteritis is characterized by progressive fibrosis and vascular sclerosis. The fibrosis can lead to bowel wall thickening and narrowing of the colonic lumen, leading to stricture formation and dilation of the segment of bowel proximal to the area of stricture. These regions are also characterized by an inflammatory cell infiltrate and, on occasion, fistula and abscess formation as a result of perforation. The chronic fibrosis is thought to be mediated by TGF-β_1, which is initially expressed following radiation-induced cytokine release. TGF-β_1 overexpression has been observed in irradiated tissues, including the colon.

CHEMICALLY INDUCED COLITIS

The incidence of colitis induced by chemicals distinct from therapeutic drugs is rare. The most common of these, however, is disinfectant colitis caused by inadvertent exposure of the colon to the 2% glutaraldehyde solutions used to clean endoscopes. Glutaraldehyde-containing disinfectant solutions can cause an acute, nonspecific colitis or proctitis characterized by bloody diarrhea, cramping, and fever, in the absence of *Clostridium difficile* infection. There has been a report of glutaraldehyde-induced colitis mimicking pseudomembranous colitis. The development of acute colitis within 24 h after endoscopy, particularly when endoscopic findings are normal, should be considered iatrogenic in nature.

There are several experimental animal models of chemically induced colitis, designed to mimic various clinically important inflammatory diseases of the gut. Intracolonic trinitrobenzene sulfonic acid or dinitrobenzenesulfonic acid, often administered in an ethanol vehicle, has been used in a number of rodent species to mimic Crohn's disease. Typically, instillation of these agents causes a T-helper 1 (T_H1) T-cell-driven colitis with acute and chronic phases that resolve within several weeks. Dextran sodium sulfate administered in the drinking water to mice or rats causes an acute colitis mimicking ulcerative colitis. This is a mixed T_H-1/T_H-2-driven inflammation characterized by bloody diarrhea and a neutrophilic mucosal inflammation. Oral administration of oxazolone causes a T_H-2-mediated colitis in mice and rats. Intralumenal administration of acetic acid causes a nonspecific colitis in the rat, guinea pig, and rabbit.

DRUG-INDUCED COLITIS

Chemotherapy Agents

Chemotherapeutic drugs are occasionally associated with the development of various colitides. In general, patients undergoing combination chemotherapy for leukemias or advanced solid tumors are at risk of developing necrotizing enteropathy or neutropenic enterocolitis. It is believed that the primary cause is due to a drug-induced decrease in mucosal defense, perhaps as a result of compromised epithelial integrity, with subsequent infection, often polymicrobial. The colon becomes edematous and hemorrhagic with transmural necrosis. Involved areas are often sharply demarcated from uninvolved areas. Unlike radiation-induced enteropathy, patients in most cases recover with supportive care and administration of broad-spectrum antibiotics. There have been case reports of severe early-onset colitis, occasionally causing death, associated with taxane-based drugs used in the treatment of breast cancer (e.g., docataxel, paclitaxel). These agents do not cause typical neutropenic colitis, but rather a pseudo-membranous or ischemic colitis.

Antibiotics

Although antibiotic use does not directly cause colitis, administration of some antibiotics is often associated with the development of a pseudo-membranous colitis. Antibiotics disrupt the normal microbial balance in the colon, allowing overgrowth of some pathogenic species of bacteria, predominantly *C. difficile*. Particularly susceptible are surgical and intensive care patients. The majority of patients who contract *C. difficile* infection will develop diarrhea and 10% of these will progress to pseudo-membranous colitis.

Other Drugs

Some neuroleptic drugs have been associated with the development of colitis. Clozapine, used in some schizophrenic patients, is associated with the development of neutropenia and has, in some instances, been

linked to the development of a neutropenic colitis. Others have associated clozapine administration with eosinophilia and still others associate its effect with its anticholinergic activity. Phenothiazines have also been linked with the development of colitis, perhaps due to reduced blood flow in the bowel. Hypersensitivity associated with anti-epileptic drugs may include colitis. In all of these cases, colitis resolves with cessation of drug intake.

There is a growing literature detailing the development of colitis with the use of pseudoephedrine as a nasal decongestant. Several reports of ischemic colitis have been made, with the vasoconstrictor activity of the drug proposed as the cause. In each case, colitis resolved spontaneously when the pseudoephedrine use was halted.

Sumatriptan, a serotonin type 1 receptor antagonist used to treat migraine headaches, is associated with several side effects, including ischemic colitis, although most case reports are confounded by coadministration of other drugs.

Nonsteroidal anti-inflammatory drugs (NSAIDs) are also associated with the development of colitis, presumably due to the fact that they decrease the production of prostaglandins involved in the preservation of gastrointestinal integrity. Several case reports demonstrate that NSAIDs that block both cyclooxygenase (COX)-1 and COX-2, or that selectively block COX-2, may induce nonspecific colitis or, more rarely, eosinophilic, pseudo-membranous, or collagenous colitis.

Effects of Drugs on Existing Colitis

Occasionally, patients with inflammatory bowel disease may experience relapse or exacerbation of their disease as a result of drugs they may be taking. For example, there are reports of exacerbation of inflammatory bowel disease (IBD) in patients taking NSAIDs. Of particular note is the fact that selective COX-2 inhibitors are capable of exacerbating colitis, despite the fact that they are touted as being "gastrointestinally safe." In addition, the immunosuppressant mycophenolate mofetil, which is used in IBD patients refractory to azathioprine and in renal transplant patients, has recently been associated with toxicity, including colitis, which has been described as "graft-versus-host disease-like."

See Also the Following Articles

Colitis, Collagenous and Lymphocytic • Colitis, Pseudomembranous • Colitis, Ulcerative • Colonic Ischemia • Crohn's Disease • Necrotizing Enterocolitis • Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) • Transforming Growth Factor-β (TFG-β)

Further Reading

Baerg, J., Murphy, J. J., Anderson, R., and Magee, J. F. (1999). Neutropenic enteropathy: A 10-year review. *J. Pediatr. Surg.* **34**, 1068–1071.

Cappel, M. S., and Simon, T. (1993). Colonic toxicity of administered medications and chemicals. *Am. J. Gastroenterol.* **88**, 1684–1699.

Denham, J. W., and Hauer-Jensen, M. (2002). The radiotherapeutic injury—A complex "wound". *Radiother. Oncol.* **63**, 129–145.

Elson, C. O., Sartor, R. B., Tennyson, G. S., and Riddell, R. H. (1995). Experimental models of inflammatory bowel disease. *Gastroenterology* **109**, 1344–1367.

Faucheron, J.-L., and Parc, R. (1996). Non-steroidal anti-inflammatory drug-induced colitis. *Int. J. Colorect. Dis.* **11**, 99–101.

MacNaughton, W. K. (2000). Review article: New insights into the pathogenesis of radiation-induced intestinal dysfunction. *Aliment. Pharmacol. Ther.* **14**, 523–528.

Neitlich, J. D., and Burrell, M. I. (1999). Drug-induced disorders of the colon. *Abdom. Imaging* **24**, 23–28.

Nguyen, N. P., Antoine, J. E., Dutta, S., Karlsson, U., and Sallah, S. (2002). Current concepts in radiation enteritis and implications for future clinical trials. *Cancer* **95**, 1151–1163.

Surawicz, C. M., and McFarland, L. V. (1999). Pseudomembranous colitis: Causes and cures. *Digestion.* **60**, 91–100.

Yassin, S. F., Young-Fadok, T. M., Zein, N. N., and Pardi, D. S. (2001). *Clostridium difficile*-associated diarrhea and colitis. *Mayo Clin. Proc.* **76**, 725–730.

Colitis, Ulcerative

LOUISE LANGMEAD AND DAVID S. RAMPTON
Barts and The London Queen Mary School of Medicine and Dentistry, United Kingdom

5-aminosalicylic acid Drug that reduces inflammation in UC.
ankylosing spondylitis Chronic inflammation of the spine and adjacent joints.
arthralgia Pain in the joints.
azathioprine Immunosuppressive drug used in refractory UC.
ciclosporin Immunosuppressive drug used to treat acute severe refractory UC.
colonoscopy Examination of the colon, rectum, and terminal ileum using a flexible videoendoscope introduced through the anus, usually after preparation of the patient with laxatives to clear the colon of feces.
Crohn's disease Idiopathic chronic inflammatory disease of the entire gastrointestinal tract; included with UC in the category "inflammatory bowel disease."
cytokines Group of peptides produced by inflammatory and other cells; control cell function and may have pro- or antiinflammatory effects.
dysplasia Alterations in the microscopic appearance of the cells lining the colon; may herald the development of cancer.
episcleritis Inflammation of sclera of the eye, sometimes occurring in patients with active UC.
erythema nodosum Red, tender swellings, usually on the legs; may occur in patients with active UC.
ileoanal pouch Reservoir created from the distal ileum to connect ileum to anus after total colectomy.
ileostomy Operation in which, after removal of the colon (colectomy), the cut end of the ileum is brought through an opening in the abdominal wall. Ileal effluent is collected thereafter into a bag fixed over the stoma.
6-mercaptopurine Metabolic product of azathioprine; also used in refractory UC.
mesalamine Approved name for 5-ASA products.
prednisolone Corticosteroid drug used to reduce inflammation in active UC.
proctitis Inflammation of the rectum; the least extensive variety of UC.
pseudopolyp Polypoid projection from the mucosa into the lumen of the colon, resulting from healing after inflammation.
pyoderma gangrenosum Chronic skin ulcer sometimes complicating UC.
sclerosing cholangitis Slowly progressive inflammatory scarring of bile ducts, leading eventually to liver failure; may occur in association with UC.
sulfasalazine Drug containing both 5-ASA and a sulfonamide; useful in the treatment of UC.
tenesmus Persistent urge to defecate in patients with rectal inflammation.
toxic megacolon Dilatation of the colon occurring rarely in patients with severe attacks of UC; may cause colonic perforation and usually requires urgent surgery.
ulcerative colitis Idiopathic chronic inflammatory disease of the colon and rectum.
uveitis Inflammation of the anterior chamber of the eye, sometimes complicating active UC.

Inflammatory bowel disease comprises two chronic relapsing and remitting inflammatory disorders of the gastrointestinal tract, ulcerative colitis and Crohn's disease. The cause of these diseases is unknown. Each is a lifelong disease, characterized by recurrent episodes of diarrhea, which is often bloody, with abdominal pain, malaise, and weight loss. Unlike in Crohn's disease, gut inflammation in ulcerative colitis affects only the colon and rectum.

INCIDENCE AND CAUSE

Ulcerative colitis (UC) is likely to result from a genetically determined inappropriately prolonged mucosal inflammatory response to as yet unidentified environmental factors, including, for example, gut microbial products.

Epidemiology

In North America and Europe, the incidence of UC is about 10 new cases per year per 100,000 population, and its prevalence 150/100,000. In Africa, Asia, and South America, UC is less common, although now increasing in incidence, probably as a result of improved hygiene. UC is slightly more common in women. It has a major peak of onset at 20–40 years of age and a lesser one at ages 60–80.

Cause

Genetic Factors

UC and Crohn's disease (CD) are likely to be related heterogeneous polygenic disorders with no single

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Genetic Factors in the Etiology of IBD

Factor[a]	Ulcerative colitis	Crohn's disease
Ethnic and familial incidence		
First-degree relatives	5%	10%
Identical twins	10%	60%
Genetic abnormalities		
HLA DR2	Increased (Japanese)	—
Susceptibility loci		
Chromosomes 5 and 16	—	Yes
Chromosomes 2 and 6	Yes	—
Chromosomes 1, 3, 4, 7, and 12	Yes	Yes
Gene products and markers		
Increased gut permeability	—	Yes
Defective colonic mucus	Yes	—
Abnormal immune regulation	Yes	Yes
DR3/DQ2 for extensive disease	Yes	—
pANCA	Yes	—
ASCA	—	Yes

[a] Abbreviations: pANCA, perinuclear antineutrophil cytoplasmic antibody; ASCA, anti-*Saccharomyces cerevisiae* antibody.

Mendelian pattern of inheritance. The contribution of genetic, as opposed to environmental, factors in UC is about 40%, and in CD, 60%. Epidemiological studies have shown that inflammatory bowel disease (IBD) is more common in Ashkenazi than in Sephardic Jews, and in North American Caucasians than in African-Americans. There is a 10-fold increased risk of IBD in the first-degree relatives of patients with UC. There is also high concordance for IBD in identical twins, particularly in CD (60%) as opposed to UC (10%).

The results of some genetic studies have varied according to the population under investigation. For example, the frequency of human leukocyte antigen (HLA) DR2 is increased in Japanese and Californians patients with UC, but not in British or European patients with UC. However, in several populations, susceptibility loci have been reported for UC on chromosomes 2 and 6, for CD on chromosomes 5 and 16, and for both on chromosomes 1, 3, 4, 7, and 12, suggesting that the two diseases are distinct, although sharing some susceptibility genes.

It is estimated that between 10 and 20 genes may be involved in the pathogenesis of IBD. Some of the pathophysiological abnormalities in IBD, such as abnormal immune regulation, increased gut permeability, defective colonic mucus, and disordered immune regulation (Table I), are probably genetically determined, but the clinical variability of IBD may also reflect genetic heterogeneity. In UC, for example, DR3/DQ2 is associated with extensive disease.

Genome scanning techniques using microsatellite markers have been employed to highlight areas of chromosomes linked to disease, including, for example, those on chromosomes 12 (*IBD2*) and 16 (*IBD1*). *IBD2* appears to make a major contribution to UC susceptibility but to have only a relatively minor effect in CD, for which there may be substantially more locus heterogeneity. The presence of perinuclear antineutrophil cytoplasmic antibodies (pANCA) in the serum of most patients with UC (but in only 5% of those with CD) is determined genetically. The genetics of IBD is currently under intense investigation; genetic clarification of this disease will have major implications for our understanding of its etiopathogenesis and for improving the treatment of IBD.

Environmental Factors

Several environmental factors appear to play a role in the causation of IBD.

Smoking Only about 10% of patients with UC smoke, compared with 30% of the normal population and with 40% of those with CD. UC commonly presents for the first time soon after a patient stops smoking, and nicotine patches have a modest therapeutic benefit. Nicotine and other constituents of tobacco have a variety of effects on the inflammatory response, but it is not known why these are beneficial in UC but harmful in CD.

Diet Up to 5% of patients with UC improve on avoidance of cow's milk; no other dietary influences are known.

Specific infection Despite its resemblance to, and occasional onset after, infectious diarrhea, there is no evidence that UC is due to a single infectious agent.

Enteric microflora The resident gut microflora are likely to be a major factor in the pathogenesis of IBD. Clinical and experimental evidence shows the importance of the fecal stream in driving mucosal inflammation. Patients with active IBD show loss of immunological tolerance to intestinal microflora. Antibiotics and, in preliminary studies, probiotics may have a therapeutic role in IBD.

Drugs Relapse of IBD may be precipitated by nonsteroidal antiinflammatory drugs (NSAIDs) and by antibiotics, the latter probably acting by altering enteric flora.

Appendectomy Appendectomy seems to protect patients from developing UC. In an inflamed appendix in genetically prone individuals, T lymphocytes may trigger inflammation (UC) in the more distal colon.

Stress Psychological stress is common in patients with IBD as a result of the unpleasantness, chronicity, and intractability of their illness. It is possible, however, that in some patients stress may trigger relapse of IBD (for example, by activation of leukocytes by enteric nerve endings in the gut wall) (Fig. 1).

Pathogenesis

Although the initiating factors in UC are unknown, altered immune regulation leads to mucosal inflammation, which is amplified and perpetuated by recruitment of leukocytes from the gut vasculature. Up-regulation of expression of nuclear transcription factors is likely to increase local release of cytokines and inflammatory mediators (Fig. 1). In UC, a non-T helper 1 (non-Th1) cytokine response generates a mainly humoral immune profile, whereas in CD, a Th1-induced cell-mediated response occurs (Table II). In UC, defective colonic mucus and abnormal mucosal permeability may facilitate access of bacterial and dietary products to the mucosa; furthermore, impaired availability of bacterially derived short-chain fatty acids may adversely affect colonocyte function.

Pathology

UC usually begins in the rectum, and either remains there or spreads proximally with time (Fig. 2). The colonic mucosa shows diffuse inflammation with hyperemia, granularity, and surface pus and blood, leading in severe cases to extensive ulceration (Fig. 3). This heals by granulation to form multiple pseudopolyps.

Microscopically, acute and chronic inflammatory cells infiltrate the lamina propria and crypts (producing crypt abscesses). Crypt architecture is distorted and goblet cells lose their mucin (Fig. 4). The mucosa is edematous with epithelial ulceration. Biopsies in long-standing total colitis may show dysplasia, in which epithelial cell nuclei are enlarged and crowded and lose their polarity; carcinoma may supervene.

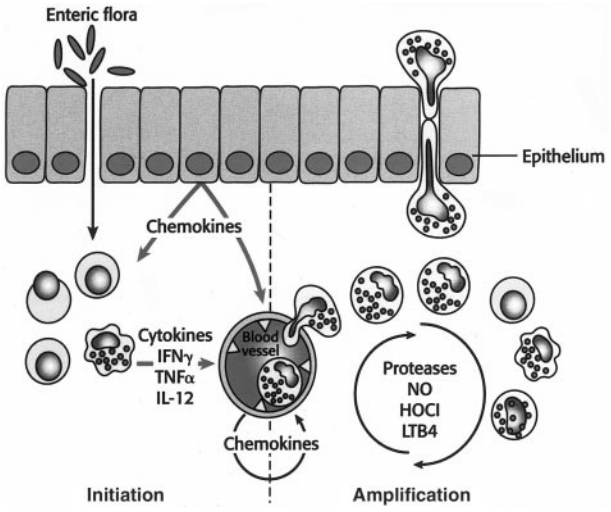


FIGURE 1 Mediators and mechanisms in the pathogenesis of inflammatory bowel disease. Although the initiating factors of the disease are uncertain, and may include a breakdown in tolerance to enteric flora, T cell and macrophage activation lead to production of cytokines acting at several levels. These include the local microvasculature, with generation of a chemokine gradient that causes transmigration of neutrophils, leading to tissue damage by metalloproteases and other reactive substances, augmentation of the inflammatory response, and disruption of the epithelial barrier, causing further ingress of enteric flora and their products. (IFNγ, interferon γ; TNFα, tumor necrosis factor α; IL-12, interleukin-12; NO, nitric oxide; HOCl, hypochlorite; LTB4, leukotriene B4). Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

CLINICAL FEATURES AND COMPLICATIONS

Clinical Features

The onset of UC is usually gradual and its natural history is chronic, with relapses and remissions over many years. The features of active disease depend on the extent as well as activity of the disease:

1. Acute severe UC, which most commonly occurs in patients with subtotal or total disease (Fig. 2), causes profuse, frequent diarrhea (>6 loose stools per day) with blood and mucus, abdominal pain, fever, malaise, anorexia, and weight loss. The patient is thin, anemic, fluid depleted, febrile, and tachycardic. In patients developing toxic megacolon and/or perforation, there

TABLE II Immune and Inflammatory Response in IBD[a]

Factor	Ulcerative colitis	Crohn's disease
Humoral immunity		
Association with autoimmune disease (thyroiditis, SLE)	Strong	Weak
Autoantibody production (anticolon antibody, pANCA, etc.)	Common	Rare
Cell-mediated immunity		
Mucosal infiltrate	Nongranulomatous; neutrophils prominent	Granulomatous; T lymphocytes prominent
T cell reactivity	Normal/decreased	Increased
Cytokine profile		
Th response	Non-Th1 (IL-10, IL-4, and IL-13)	Th1 (IL-2, IFN, IL-12, and TNFα)

[a] Abbreviations: SLE, systemic lupus erythematosus; pANCA, perinuclear antineutrophil cytoplasmic antibody; Th, T helper lymphocyte; IL, interleukin; IFN, interferon; TNFα, tumor necrosis factor α.

is sudden worsening of abdominal pain, distension, fever, tachycardia, sepsis, and shock.

2. Moderately active UC, usually left sided (Fig. 2), causes rectal bleeding and mucus discharge with diarrhea (<6 loose stools per day), urgency, and sometimes abdominal pain.

3. Active proctitis causes rectal bleeding and mucus discharge, often with tenesmus and pruritus ani. There may be increased stool frequency but some patients are constipated, particularly proximal to the inflamed rectum. General health is usually maintained.

Extraintestinal Associations and Complications

Systemic associations and complications of UC, as of CD most commonly affect the liver and biliary tree, joints, skin, and eyes (Table III). Some complications occur mainly in active disease. In some instances, the condition is a metabolic complication of IBD (gallstones and urinary stones). In others, there is a genetic and/or immunological (ankylosing spondylitis, uveitis, arthropathy) association with IBD. The most important complications are sclerosing cholangitis, joint disease, osteoporosis, skin conditions, ocular diseases, and thromboembolism.

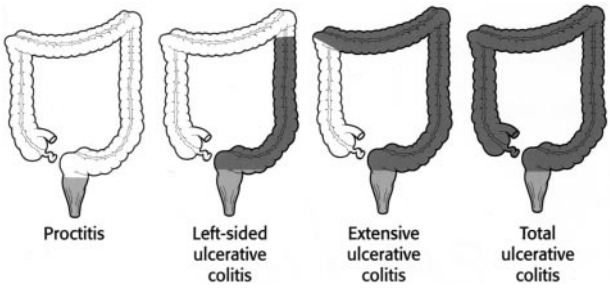


FIGURE 2 Distribution of UC. Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

Sclerosing Cholangitis

Sclerosing cholangitis occurs in about 5% of patients with UC. Its pathogenesis is unknown, but it may occur years before the onset of overt colitis. It is characterized by gradually progressive obliterative fibrosis of the biliary tree and is sometimes complicated by cholangiocarcinoma. The risk of colorectal cancer in patients with UC and sclerosing cholangitis exceeds that associated with UC alone. Patients usually present with obstructive jaundice, cholangitis, or abnormal liver function tests at routine screening. The diagnosis may be suggested by ultrasound, computer tomography (CT), magnetic resonance imaging, and/or liver biopsy; endoscopic retrograde choledochopancreatography (ERCP) is useful for diagnosis and stenting of strictures. Although widely used, ursodeoxycholic acid is of unproved benefit. Liver transplant is the only hope of long-term survival; median survival without transplant is about 15 years.

Joint Disease

IBD-related arthropathy occurs in up to 10% of patients with UC. Pauciarticular disease involves less than five joints; characteristically it affects one large joint (for example, the knee) and is most common in women. Attacks usually coincide with relapse of colitis and may be due to deposition in the affected joint of gut-derived immune complexes. Pauciarticular disease is neither progressive nor deforming. In most patients, the joint symptoms resolve on successful treatment of the active UC. Sulfasalazine may be more effective for the joints than other 5-aminosalicylic acids (5ASAs). NSAIDs may exacerbate UC, and should if possible be avoided. Joint aspiration with steroid installation may help.

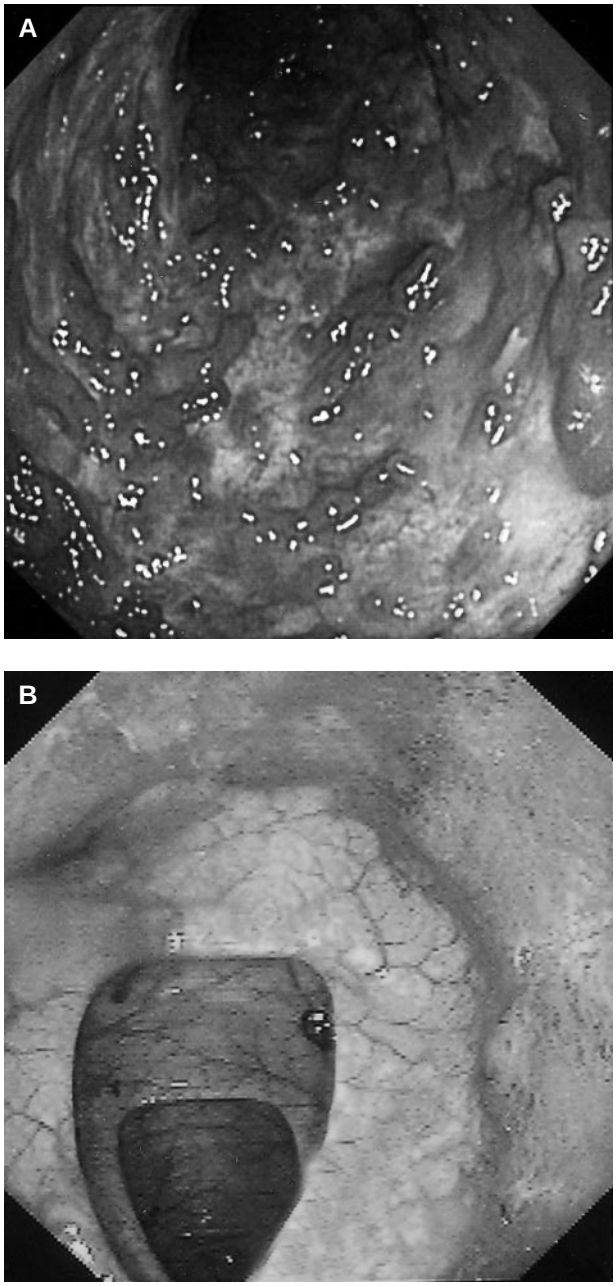


FIGURE 3 Colonoscopic appearance of (a) active UC, showing mucopurulent exudate, erythema, granularity, and superficial ulceration, and (b) inactive UC, showing erythema and edema, with a sharp "cutoff" to normal more proximal colonic mucosa.

Polyarticular IBD-related arthropathy affects more than five joints, particularly small joints in the hands. Symptoms are more common in women, chronic, and not clearly related to activity of UC. Management resembles that of pauciarticular disease, except that response of the arthropathy to treatment of UC is poor.

Ankylosing spondylitis is characterized by inflammation of the spinal column leading to formation of syndesmophytes between vertebra and later calcification and ossification of the interspinous ligaments. It affects about 5% of patients with UC and is usually accompanied by inflammation, erosion, and sclerosis of the sacroiliac joints (i.e., sacroiliitis). Although about 95% of patients without IBD who have ankylosing spondylitis are HLA-B27 positive, this is true of only 70% of patients with both diseases. Presentation is with back pain and stiffness, and diagnosis is confirmed by X ray. The course of ankylosing spondylitis is independent of the activity of UC and it may present years before the colitis becomes manifest. Treatment consists of vigorous physiotherapy, sulfasalazine, and, if tolerated, NSAIDs.

Osteoporosis

Osteoporosis is much less common in UC than in CD and is a consequence of chronic intestinal inflammation and its treatment with corticosteroids. The disease is asymptomatic for many years, presenting eventually with vertebral collapse or long bone fractures. Diagnosis is best made by bone densitometry. To prevent osteoporosis, patients should eat adequate amounts of dairy products and, if necessary, take calcium and vitamin D tablets. Patients should control their weight, stop smoking, and take regular exercise. Postmenopausal women are given hormone replacement therapy. Patients with proved osteoporosis are usually treated with cyclical bisphosphonates.

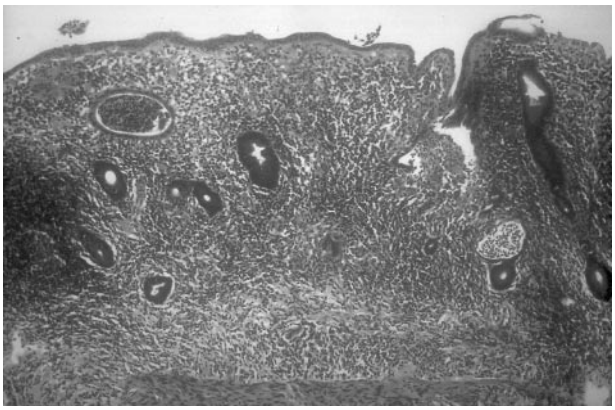


FIGURE 4 Microscopic appearance of active UC, showing intense inflammatory cell infiltration of the lamina propria, goblet cell depletion, and crypt abscesses. Courtesy of R. M. Feakins. Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

TABLE III Extraintestinal Associations and Complications of UC

Organ	Complication
Joints/bones	Enteropathic arthropathy[a]
	Sacroiliitis
	Ankylosing spondylitis
	Osteoporosis
Eyes	Episcleritis[a]
	Uveitis[a]
Skin	Erythema nodosum[a]
	Pyoderma gangrenosum
Liver	Fatty change
	Chronic active hepatitis
Biliary tract	Sclerosing cholangitis
	Cholangiocarcinoma
Kidneys	Uric acid stones
Lungs	Fibrosing alveolitis
Blood	Anemia[a]
	Arterial and venous thrombosis[a]
Constitutional	Weight loss[a]
	Growth retardation (children)[a]

[a] Worsens when UC is active.

Skin Associations

Erythema nodosum occurs in about 8% of patients with UC, usually when the disease is active. Hot, red, tender nodules appear, usually on extensor surfaces of the lower legs and arms (Fig. 5); these nodules gradually subside after a few days and leave brownish skin discoloration. Treatment is of the active associated UC.

Pyoderma gangrenosum occurs in about 2% of patients with UC. There is initially a discrete pustule with surrounding erythema; this develops into an indolent, enlarging ulcer. The pustules most commonly occur on the leg (Fig. 6), sometimes at sites of previous trauma. Biopsy shows a lymphocytic vasculitis with neutrophilic infiltration. Pyoderma is often refractory to treatment; options include intralesional, topical, and systemic corticosteroids, and immunosuppressive drugs. Colectomy does not reliably induce healing of the skin lesion.

Ocular Associations

The eye is involved in about 3% of patients with UC, often when the bowel disease is active. Episcleritis causes burning and itching accompanied by dilated blood vessels at the site of inflammation. Uveitis is more serious and often recurrent, causing headache, red eye, and blurred vision: slit lamp examination shows pus in the anterior chamber. Treatment of both disorders includes topical steroids, cycloplegics, and therapy of active UC. Prompt referral to an ophthalmologist of UC patients with ocular symptoms is advisable.

Thromboembolism

A hypercoagulable state exists in both forms of IBD, leading to an increased incidence of both venous and arterial thromboembolism. This involves all components of the clotting system with elevations of fibrinogen, factor V, and factor VIII, and reduction of antithrombin III concentrations. There is also increased activation of platelets. Patients particularly at risk of thromboembolic complications are those with active disease.

INVESTIGATION

The aims of evaluative tests in UC are to establish the diagnosis and the extent and activity of the disease, to check for complications of the disease, and to guide treatment. In patients aged less than 50 years, the main differential diagnoses include infection and irritable bowel syndrome; while in older people, the differential diagnoses are primarily neoplasia, diverticular disease, and ischemia.

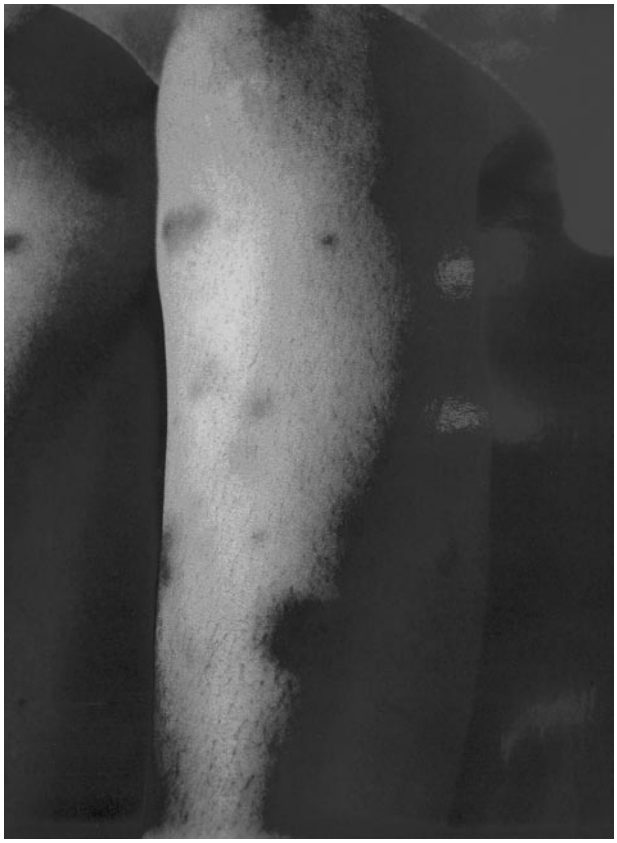


FIGURE 5 Erythema nodosum. Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

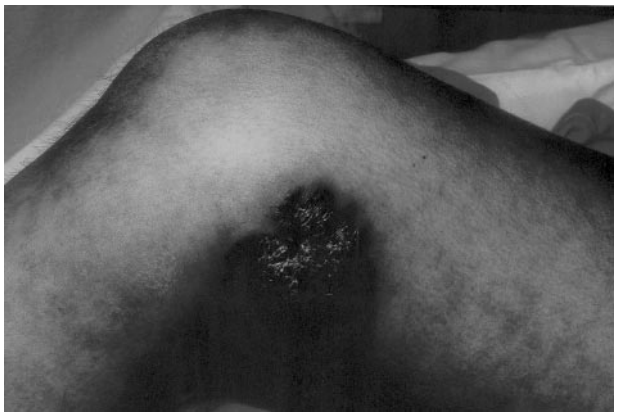


FIGURE 6 Pyoderma gangrenosum. Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

Blood Tests

In patients with abdominal pain and/or diarrhea, results indicating anemia, elevated platelet count, and elevated erythrocyte sedimentation rate (ESR) may suggest active UC, but are not diagnostic. Elevated C-reactive protein and low serum albumin levels suggest active disease in patients with established UC; prior to diagnosis, they are only suggestive but not diagnostic of UC.

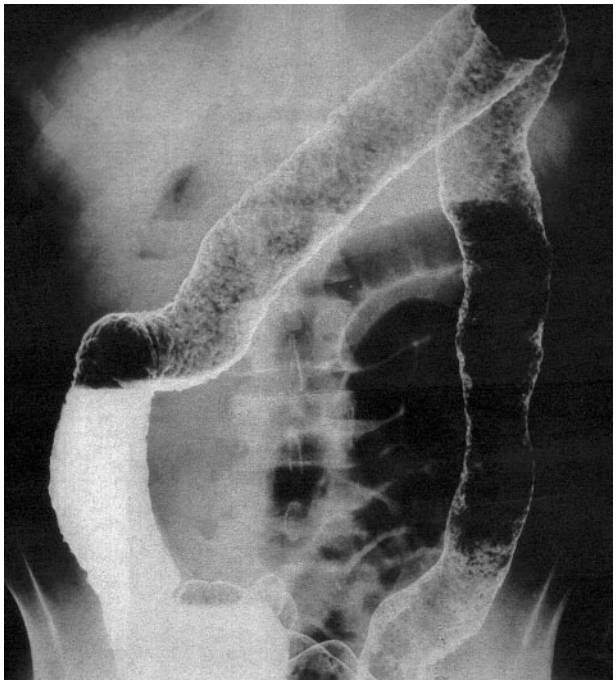


FIGURE 7 Barium enema, showing superficial ulceration in active total UC. This test has now been largely superseded by colonoscopy in patients with UC; both tests are potentially dangerous in active disease.

Liver function is abnormal in patients with hepatobiliary complications (Table III), and requires regular monitoring of patients on immunosuppressive therapy. Most patients with UC have circulating perinuclear antineutrophil antibodies, but the test for pANCAs is not sufficiently sensitive or specific to be of diagnostic value.

Regular blood checks are necessary to check for bone marrow depression and hepatitis in patients maintained on immunosuppressive drugs. Patients on sulfasalazine are, additionally, at risk for hemolytic anemia and folate deficiency.

Stool Microbiology

Stool microscopy shows red and white blood cells in active colitis, whether due to IBD or infection. Hot, fresh samples are essential to look for amebic trophozoites in stools of recent travelers. Even in patients with known UC, microbiological testing of stools may reveal associated infection when patients relapse.

Endoscopy and Histopathology

In patients with diarrhea, with or without rectal bleeding, rigid or flexible sigmoidoscopy with biopsy can provide immediate confirmation of colitis and its activity (Figs. 3 and 4). In patients who are not severely ill, colonoscopy is the best test for diagnosing UC and assessing its extent and activity. In acute severe UC, full colonoscopy may cause perforation and should be avoided. Inactive UC is characterized by mucosal edema, erythema, and granularity, whereas in active disease, there is also contact or spontaneous bleeding, mucopus, and ulceration (Fig. 3). In chronic UC, there are pseudopolyps and loss of the haustral pattern, with apparent shortening of the colon, and the mucosa may be atrophic. Colonoscopy is used for cancer surveillance in chronic extensive UC.

Radiology

In patients with active UC, a plain abdominal radiograph sometimes helps to assess disease extent, because fecal residue on the X ray may indicate sites of uninflamed colonic mucosa. Plain films are also used to exclude colonic dilatation (diameter more than 5.5 cm) in acute severe UC; in this setting, severe disease is also indicated by deep ulceration and coarse nodularity of the mucosa, or "mucosal islands," and linear gas tracking in the gut wall.

Double-contrast barium enema (Fig. 7) has largely been superseded by colonoscopy. In patients with active colitis, "air enemas," in which air is introduced into the

FIGURE 8 [^{99m}Tc]Hexamethylene propyleneamine oxime-labeled leukocyte scanning appearance of active UC, showing inflammation affecting colon up to hepatic flexure (extensive colitis).

unprepared rectum, are sometimes used to assess disease extent.

The intensity and extent of colonic uptake on an isotope scan taken 1 hour after injection of autologous leukocytes radiolabeled with [^{99m}Tc]hexamethylene propyleneamine oxime (HMPAO) provides information, noninvasively, about disease activity and extent (Fig. 8). Increased isotopic activity is not specific for UC and can also occur in other inflammatory gut diseases.

DRUGS USED IN THE TREATMENT OF UC

The mechanism of action, pharmacology, and side effects of drugs used as specific antiinflammatory agents in UC are discussed in the following sections.

Corticosteroids

By combining with intracellular glucocorticoid receptors, corticosteroids have multiple antiinflammatory actions (Table IV), but which of these is of predominant importance is unclear. Corticosteroids can be given intravenously, orally, or topically, the route selected depending on the severity and site of disease (Table IV). The side effects of conventional steroids have prompted a search for safer formulations. For topical therapy, two enema preparations (budesonide and prednisolone metasulfobenzoate) contain steroids that are poorly absorbed and/or undergo rapid first-pass hepatic metabolism, thereby producing fewer side effects and less adrenocortical suppression compared to other steroid enemas. Oral formulations are being developed to release prednisolone in the colon and thus reduce its systemic absorption and side effects. The many

TABLE IV Corticosteroids[a]

Preparations
- Intravenous
  - Hydrocortisone (300–400 mg/day)
  - Methylprednisolone (40–60 mg/day)
- Oral
  - Prednisolone, prednisolone (enteric coated), prednisone (< 60 mg/day)
- Enemas
  - Liquid: prednisolone metasulfobenzoate (Predenema), prednisolone sodium phosphate (Predsol), budesonide (Entocort)
  - Foam: prednisolone metasulfobenzoate (Predfoam), hydrocortisone (Colifoam)
- Suppositories
  - Hydrocortisone, prednisolone sodium phosphate (Predsol)

Indications
- Active UC

Side effects
- General: Cushingoid facies, weight gain, dysphoria
- Metabolic: Adrenocortical suppression, hyperglycemia, hypokalemia
- Cardiovascular: Hypertension, fluid retention
- Infection: Opportunistic infections, reactivation of tuberculosis, severe chickenpox
- Skin: Acne, bruising, striae, hirsuties
- Eyes: Cataracts, glaucoma
- Musculoskeletal: Osteoporosis, avascular osteonecrosis, myopathy
- Children: Growth retardation

Mechanisms of action
- Leukocytes
  - Reduced migration, activation, survival
  - Reduced activation of NF-κB
  - Phospholipase A_2 inhibition
  - Reduced induction of COX-2, iNOS
  - Reduced production of cytokines and lipid mediators
- Endothelial cells
  - Reduced expression of adhesion molecules
  - Reduced capillary permeability

[a] Abbreviations: NF-κB, nuclear (transcription) factor κB; COX-2, cyclooxygenase-2; iNOS, inducible nitric oxide synthase.

TABLE V Oral Formulations of Aminosalicylates

Drug	Formulation	Dose range (maintenance–conventional maximum)[a]
Prodrugs (5ASA azo-linked to carrier)		
Sulfasalazine	5ASA-sulfapyridine	1 g bd–2 g tds
Olsalazine	5ASA-5ASA	500 mg bd–1 g tds
Balsalazide	5ASA-aminobenzoylalanine	1.5 g bd–2.25 g tds
Mesalazine (5ASA alone)		
Delayed release		
Asacol	Eudragit S coating dissolving at pH > 7	400 mg tds–800 mg tds
Salofalk	Eudragit L coating dissolving at pH > 6	500 mg tds–1 g tds
Slow release		
Pentasa	Ethylcellulose microspheres	500 mg tds–2 g bd

[a] bd, Two times/day (*bis die*); tds, three times/day (*ter die sumendum*).

side-effects of steroids are related to dose and duration of treatment except in avascular osteonecrosis.

Aminosalicylates

Like corticosteroids, 5-aminosalicylates have numerous antiinflammatory effects. Aminosalicylates are available orally (Table V) and as enemas and suppositories (Table VI). The original compound, sulfasalazine (Fig. 9), consists of 5ASA linked by an azo bond to sulfapyridine (Table V). The sulfonamide acts as a carrier to deliver 5ASA, the active moiety, to the colon, where it is released by bacterial action. About 20% of patients cannot tolerate sulfasalazine because of side effects, which are mostly due to sulfapyridine (Table VI). The newer oral 5ASA formulations (Table V) are much better tolerated. The pH-dependent delayed-release and, particularly, slow-release mesalazine preparations release 5ASA more proximally in the gut, making them useful in small-bowel CD as well as in UC and Crohn's colitis. In contrast, olsalazine and balsalazide, like sulfasalazine, release 5ASA by bacterial azo reduction in the colon and are used only in UC.

Although better tolerated than sulfasalazine, the newer 5ASA formulations (Table V) may also cause rash, headache, nausea, diarrhea, exacerbation of UC, pancreatitis, and/or blood dyscrasias (Table VI). Interstitial nephritis occurs in about 0.2% of patients on mesalazine, and watery diarrhea occurs in about 5% of patients on olsalazine.

Azathioprine and 6-Mercaptopurine

Azathioprine is a prodrug that is rapidly converted to 6-mercaptopurine (6MP). Azathioprine and 6MP (Table VII) modify the immune response by inhibiting DNA synthesis in T lymphocytes. Both drugs are used only orally and take up to 4 months to reach maximum efficacy. Homozygous deficiency of 6-thiopurine methyltransferase (TPMT), an enzyme causing their degradation, occurs in about 0.2% of the population and accounts for some of the side effects of these drugs, particularly bone marrow suppression.

Up to 20% of patients cannot tolerate azathioprine because of nausea, rash, fever, arthralgia, abdominal pain, and headache; for some patients, a switch to 6MP averts these problems. Acute pancreatitis occurs in about 3% of patients. Their other potentially serious side effects, bone marrow depression (particularly in the first few weeks of treatment; 2% of patients) and cholestatic hepatitis (0.5% of patients), necessitate regular blood monitoring; if available, TPMT levels should be checked prior to therapy to exclude a deficiency of this enzyme. There is an increased risk of infections. Long-term use may increase the risk of malignancy, particularly lymphoma.

Ciclosporin

Ciclosporin is a fungus-derived cyclic undecapeptide that reduces T helper cell and cytotoxic T cell function by inhibiting interleukin-2 (IL-2) gene transcription. The variable absorption of the conventional oral preparation (Sandimmun) necessitates intravenous therapy in active disease. Close monitoring of blood concentrations is used to optimize ciclosporin dosage. Because the cytochrome P450 enzyme system metabolizes ciclosporin, grapefruit juice, St. John's wort, and drugs that modify ciclosporin activity should be avoided. The most serious side effects of ciclosporin (Table VIII) are opportunistic infections (20% of patients), renal impairment (20%), hypertension (30%), hepatotoxicity (20%), and epileptic fits (3%).

TABLE VI Aminosalicylates

Preparations
Oral (see Table V)
Enemas
Liquid: Pentasa, Salofalk, sulfasalazine
Foam: Asacolfoam
Suppositories
Asacol, Pentasa, Salofalk, sulfasalazine
Indications
Active and inactive UC
Side effects
General: Headache,[a] fever[a]
Gut: Nausea,[a] vomiting,[a] diarrhea, exacerbation of UC
Blood: Hemolysis,[a] folate deficiency,[a] agranulocytosis,[a] thrombocytopenia,[a] aplastic anemia,[a] methemoglobinemia[a]
Renal: Orange urine,[a] interstitial nephritis
Skin: Rashes, toxic epidermal necrolysis,[a] Stevens–Johnson syndrome,[a] hair loss
Other: Oligospermia,[a] pancreatitis, hepatitis, lupus syndrome, pulmonary fibrosis
Mechanisms of action[b]
Leukocytes
Reduced migration, cytotoxicity
Reduced activation of NF-κB
Reduced synthesis of interleukin-1 and lipid mediators
Antioxidant
TNF antagonist
Reduced FMLP receptor binding
Epithelium
Reduced MHC class II expression
Induction of heat-shock proteins
Reduced apoptosis

[a] Side effects usually due to sulfonamide component of sulfasalazine.

[b] Abbreviations: NF-κB, nuclear (transcription) factor-κB; TNF, tumour necrosis factor; FMLP, *f*-methionyl-leucyl-phenylalanine; MHC, major histocompatibility complex.

Long-term use, for which there is no clear indication in UC, may predispose to lymphoma.

Antibiotics

Limited data suggest that oral tobramycin may improve outcome in acute severe UC. However, the use of antibiotics is usually restricted to bacteremia and endotoxic shock complicating severe acute colitis.

Other New Therapeutic Approaches

Recognition of altered cytokine expression in IBD has prompted trials using a variety of cytokine-related therapies (Table IX), of which only antitumor necrosis factor α (anti-TNFα) antibody (infliximab) has reached clinical use for refractory CD. In the longer term, it is possible that gene transfer techniques will be used to induce intestinal mucosal production of antiinflammatory cytokines such as IL-4 and IL-10.

Progressive elucidation of the pathogenesis of UC has led to the evaluation of a number of further therapies aimed at specific pathophysiological targets (Table IX). Although none has yet reached routine application, there is currently major interest in the possible therapeutic role of orally administered preparations of bacteria, or "probiotics," thought to have a beneficial effect on mucosal immune function. Recent trials suggest that capsules containing *Lactobacillus*, *Bifidobacterium*, and/or nonpathogenic *Escherichia coli* may have a prophylactic effect in UC and pouchitis, but further studies are needed to clarify their efficacy and safety.

MEDICAL MANAGEMENT

Treatment in UC is aimed to induce and then maintain remission. Management of UC comprises general measures, supportive treatment, and specific pharmacological and surgical therapies (Table X).

General Measures

Explanation and Psychosocial Support

Patients with newly diagnosed UC need a full explanation about their disease. This process can be facilitated by written information provided by patient support groups such as the Crohn's and Colitis Foundation of America (United States) and the National Association for Colitis and Crohn's Disease (United Kingdom). Services offered by such groups include not only educational literature and web sites, but also

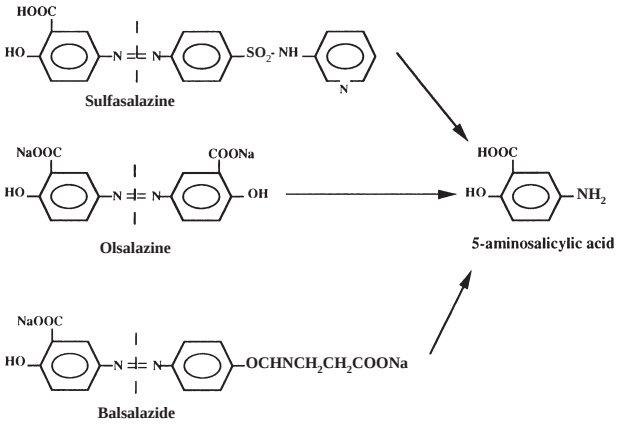


FIGURE 9 Chemistry of 5-aminosalicylic acid preparations.

TABLE VII Azathioprine and 6-Mercaptopurine

Preparations
Oral: Azathioprine (2–2.5 mg/kg/day), 6-mercaptopurine (1–1.5 mg/kg/day)
Indications
Steroid-dependent or -refractory UC
Side effects
General: Nausea, vomiting, abdominal pain, headache, arthralgia, fever, rash
Blood: Agranulocytosis, thrombocytopenia, macrocytosis
Infections: Opportunistic, including cytomegalovirus, herpes zoster
Hepatobiliary: Cholestatic hepatitis, acute pancreatitis
Malignancy: Lymphoma
Mechanism of action
Inhibition of T cell DNA synthesis

meetings at which patients and their families can share their problems, and counseling for individuals with difficulties relating to their illness. These groups can also direct patients to appropriate agencies in relation to employment and insurance problems. Importantly, support groups can exert political pressure to help maximize accessibility of health services and to generate funds for research.

Hospital Care

Patients with UC are best managed, whether as outpatients or during admission to hospital, by specialist gastroenterological medical, surgical, nursing, and dietetic staff working in collaboration, and with access to stoma therapists and counselors. Specialist IBD clinics are the best way of providing patients with joint medical/surgical consultations with open access for early review in the event of relapse (Table X).

TABLE VIII Ciclosporin

Preparations
Oral (5–8 mg/kg/day)
Intravenous (4 mg/kg/day)
Indication
Steroid-refractory acute severe UC (intravenous then oral)
Side effects
General: Nausea, vomiting, headache
Renal: Interstitial nephritis
Infection: Opportunistic, including *Pneumocystis carinii* pneumonia
Neurological: Epileptic fits, paresthesia, myopathy
Hypertension
Skin: Hypertrichosis, gingival hypertrophy
Metabolic: Hyperkalemia, hypomagnesemia, hyperuricemia
Liver: Cholestatic hepatitis
Malignancy: Lymphoma
Mechanism of action
Inhibition of T cell function and proliferation

Drugs

Iron and folic acid supplements are sometimes needed, as are drugs for osteoporosis. Antidiarrheal (loperamide, codeine phosphate, or diphenoxylate), opioid analgesic, antispasmodic, and anticholinergic drugs are avoided in active UC because they may provoke acute colonic dilatation. NSAIDs, and occasionally antibiotics, may provoke relapse of UC, and are not used unless essential.

Treatment of Active UC

Treatment in active UC is determined by the extent of disease and the severity of the attack. Knowledge of the extent of disease determines the feasibility of topical therapy, whereas the severity of the attack defines the optimal type and route of therapy, and whether patients can be treated as out-patients or need hospital admission.

In-Patient Management of Acute Severe UC

General measures These patients are admitted immediately to a gastroenterology ward for close joint medical, surgical, and nursing care. History and investigations help establish the diagnosis in patients presenting for the first time and, in those with established UC, exclude infection and assess disease extent and severity. Progress is monitored daily by clinical assessment, stool chart, 6-hourly measurement of temperature and pulse, blood count, ESR, C-reactive protein, routine biochemistry, and plain abdominal X ray.

Supportive treatment Intravenous fluids and electrolytes are required to replace diarrheal losses. Blood transfusion may be necessary. Patients can usually eat normally, with liquid protein and calorie supplements if necessary. Very sick patients may need total parenteral nutrition. To reduce the risk of venous and arterial thromboembolism, prophylactic subcutaneous heparin is given.

Specific medical treatment The cornerstone of specific medical treatment of acute severe UC remains corticosteroids. Ciclosporin is a useful adjunct, but thiopurines are too slow to work in patients with acute steroid-refractory attacks. 5ASAs (Tables V and VI) are continued in patients taking them on admission, but do not have a major role in acute severe UC.

Hydrocortisone (300–400 mg/day) or methylprednisolone (40–60 mg/day) (Table IV) given intravenously will improve about 70% of patients in 5–7 days.

TABLE IX Potential New Treatments for UC Aimed at Specific Pathophysiological Targets

Target	Agent[a]
Colonic bacterial flora	Probiotics (*Lactobacillus*, nonpathogenic *Escherichia coli*)
Epithelium	Short-chain fatty acid enemas, trefoil peptides
Leukocytes	
Reduce numbers	Apheresis, anti-CD4 antibodies
Reduce migration	Adhesion molecule antibodies or antisense oligonucleotides
Cytokines	
Reduce proinflammatory cytokines	NF-κB antisense oligonucleotide
Antagonize inflammatory cytokines	Anti-TNF antibodies, anti-IL-12 antibodies, IL-1, IL-2
Receptor antagonist	
Increase antiinflammatory cytokines	IL-10, IL-11, TGF-β, IL-4 gene therapy
Mediators	Cytoprotective prostaglandins
	Synthesis inhibitors and receptor antagonists of leukotrienes, thromboxanes
	Antioxidants
	Inducible NOS inhibitors
	Fish oil (eicosapentaenoic acid)
	metalloproteinase inhibitors
Vasculature	Heparin
Enteric nerves	Local anesthetics (lignocaine enemas)
Unknown targets	Nicotine

[a] Abbreviations: NF-κB, nuclear (transcription) factor κB; IL, interleukin; TGF-β, transforming growth factor-β; NOS, nitric oxide synthase.

Oral prednisolone (40–60 mg/day) is then used to induce complete remission, tapering the dose to zero over 2–3 months. Conventionally, failure to respond to intravenous steroids after 5–7 days has indicated urgent colectomy, but use of ciclosporin is now an alternative.

Intravenous ciclosporin (4 mg/kg/day for 5 days) followed by oral ciclosporin (5–8 mg/kg/day) (Table VIII), given with continued steroids, averts colectomy in about 70% of patients not responding to intravenous steroids alone. Enthusiasm for this treatment has to be tempered by the frequency of relapse necessitating colectomy (up to 50%) that follows later withdrawal of ciclosporin, and by its serious adverse effects (Table VIII).

Toxic megacolon In sick patients with clinical and/or radiological evidence of incipient colonic dilatation, rolling into the knee–elbow position for 15 minutes every 2 hours may help to evacuate gas per rectum and prevent toxic megacolon. If colonic dilatation does occur, immediate surgery is needed if patients do not improve after 24 hours of treatment with rolling, antibiotics, and a nasogastric tube to aspirate bowel contents.

Colonic perforation and massive hemorrhage Emergency surgery is required for these rare complications of severe UC. Even with immediate surgery, the mortality of colonic perforation approaches 30%.

Active Left-Sided or Extensive UC

The principles of management of moderate attacks of UC resemble those described above. These patients, however, do not usually need hospital admission. Stools are examined to exclude infection. In mild attacks, an oral 5ASA (Table V), with twice daily 5ASA or steroid enemas, may suffice. Prednisolone, 20–60 mg/day for 2–4 weeks, depending on the severity of the attack, may be needed to induce remission, followed by gradual reduction of the dose to zero. About 30% of patients fail to respond to these measures within 2 weeks, or deteriorate, and need more intensive management. An alternative treatment in steroid-refractory patients who are not acutely ill, and in whom a response taking up to 4 months is acceptable, is oral azathioprine

TABLE X Principles of Management of UC

General measures
- Explanation, psychosocial support
  - Patient support groups
- Specialist multidisciplinary hospital care
  - Monitoring disease activity, nutrition, therapy
  - Checking for extraintestinal complications
  - Colonoscopic cancer surveillance

Supportive treatment
- Dietary and nutritional advice
- Drugs
  - Antidiarrheals (not in active UC)
  - Hematinics
  - Vitamins, electrolytes
  - Osteoporosis prophylaxis and treatment
  - Subcutaneous heparin (in-patients with active UC)
- Drugs to avoid
  - Antidiarrheal drugs (in active UC)
  - Inessential nonsteroidal antiinflammatory drugs, antibiotics

Specific treatment (according to presentation)
- Drugs
  - Corticosteroids
  - Aminosalicylates
  - Immunomodulatory drugs (azathioprine/6-mercaptopurine, cyclosporine)
- Surgery (see text)

(2–2.5 mg/kg/day) or 6-mercaptopurine (1–1.5 mg/kg/day) (Table VII).

Active Proctitis

The principles of treating proctitis resemble those for more extensive UC. Any coexisting constipation is treated with a high-fiber diet and/or a stool-softening laxative. Topical treatment is usually used in order to obtain increased drug concentrations in the rectal mucosa, without the risk of systemic side effects. Depending on disease extent, suppositories or enemas of 5ASA (Table V) or corticosteroids (Table IV) are used; up to 80% of patients respond in 4 weeks. In patients with unresponsive proctitis, options include oral or intravenous corticosteroids, and a thiopurine. Arsenic-containing suppositories (Acetarsol) may help but, to avoid toxicity, should not be used for more than 4 weeks. Experimental possibilities include ciclosporin, short-chain fatty acid or lignocaine enemas, and nicotine patches (Table IX). If all medical treatments fail, patients require panproctocolectomy, because limited colectomy has too high a recurrence rate to be a practicable option.

Maintenance of Remission of UC

An oral regimen of 5ASA for life (Table V) reduces annual relapse rate of UC to 25%, compared with 75% with no treatment. Some patients with recurrent attacks of distal disease prefer prophylactic 5ASA therapy given as daily or alternate daily enemas or suppositories. Patients with disease of limited extent and having relapses less than once a year may decline maintenance therapy. In patients who relapse repeatedly despite a 5ASA drug regimen, and/or when steroids are withdrawn after acute episodes, oral azathioprine or 6-mercaptopurine, monitored and given for at least 2 years, is of proved benefit.

Surveillance for Colorectal Cancer

The increased risk of colorectal cancer in chronic extensive UC (about 20% at 30 years after diagnosis) has led to colonoscopic surveillance, in which multiple biopsies from throughout the colon, and from any raised lesions, are taken every 1–3 years starting 8–10 years after diagnosis. If the biopsies show the premalignant changes of high-grade epithelial dysplasia (Fig. 10), colectomy is necessary. If there is confirmed low-grade dysplasia, either colectomy or, in patients reluctant to have surgery, more frequent surveillance is needed. Unfortunately, such surveillance has not yet been shown to reduce mortality from colon cancer in UC, perhaps because 25% of cancers occur in patients without dysplasia. Molecular biological techniques (e.g., DNA aneuploidy, p53 heterozygosity) are likely soon to supersede detection of dysplasia for cancer prevention in UC.

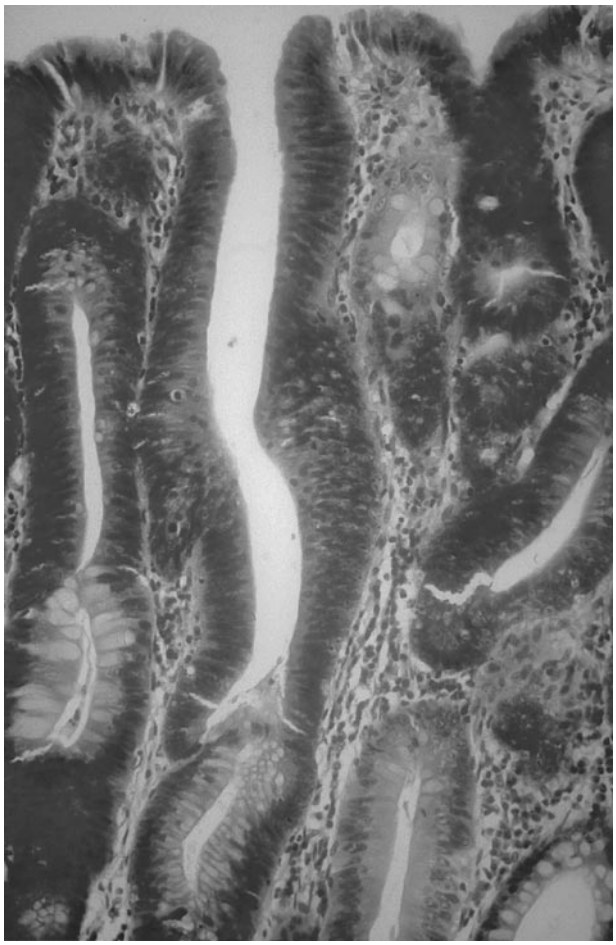


FIGURE 10 Severe epithelial dysplasia in UC. There is glandular distortion, stratification of the epithelium with heaping of nuclei, and nuclear polymorphism and hyperchromaticism. Note the normal crypts at the bottom of the picture. Courtesy of R. M. Feakins. Reproduced from D. S. Rampton and F. Shanahan (2000), *Fast Facts—Inflammatory Bowel Disease*, with permission of Health Press Ltd., Oxford.

SURGICAL MANAGEMENT

The management of UC requires close liaison between physician and surgeon. Specialist care from a stoma therapist is also necessary. About 30% of patients with total UC require colectomy by 15 years after its onset.

Indications for Surgery

Indications for surgery can be categorized as follows:

1. Emergency colectomy is necessary in patients with colonic perforation or massive hemorrhage.
2. Urgent colectomy is necessary in patients with acute severe UC who fail to respond to intensive medical treatment in 5–7 days, or who develop unresponsive toxic colonic dilatation.
3. Elective colectomy is indicated in refractory chronic active UC, dysplasia, or frank carcinoma. Colectomy may be necessary in children with chronically active disease to prevent growth retardation and, very rarely, in patients with intractable pyoderma gangrenosum.

Options

There are several options for surgical management of UC (Fig. 11):

1. Proctocolectomy with permanent ileostomy has the lowest morbidity and mortality of the surgical options, is technically the easiest, and involves only one operation.
2. Colectomy with ileorectal anastomosis is useful in older patients with rectal sparing who cannot cope with a stoma or are unsuitable for an ileoanal pouch because of frailty or a weak anal sphincter. This option should not be chosen for patients with marked rectal inflammation, or for young patients (because of the risk of later rectal cancer).
3. Restorative proctocolectomy with ileoanal pouch avoids the need for permanent ileostomy. It is the favored operation in patients less than 60 years old with normal anal sphincter function. Construction of an ileoanal pouch usually requires a temporary ileostomy that is closed at a second operation a few months later.

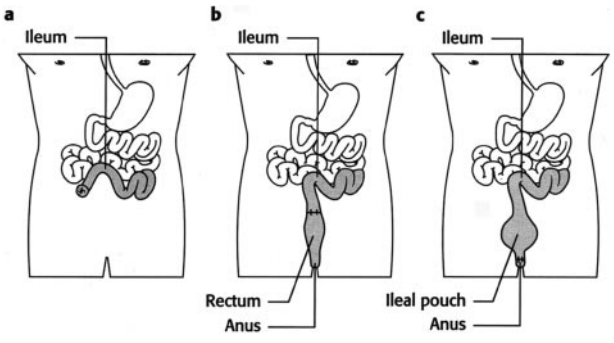


FIGURE 11 Surgical options in UC: (a) panproctocolectomy with ileostomy, (b) subtotal colectomy with ileorectal anastomosis, and (c) total colectomy with ileoanal pouch.

TABLE XI Complications of Ileostomy and Ileoanal Pouches

Ileostomy	Ileoanal pouch
Early	**Early**
Skin problems	Pelvic sepsis
Adhesive intestinal obstruction	Anastomotic leaks
Necrosis, fistulas, retraction, parastomal herniation	Adhesive intestinal obstruction
Excess stomal output (normal, approx. 500 ml/day)	
Late	**Late**
Sexual dysfunction	Poor function: diarrhea, urgency, incontinence
Uric acid renal stones	Sexual dysfunction
Pouchitis	
Vitamin B_{12} deficiency	

Complications

Complications of ileostomy and ileoanal pouch surgery are outlined in Table XI. Complications of ileoanal pouch surgery can lead to excision of the pouch and conversion to permanent ileostomy ("pouch failure") in about 10% of patients. Even after successful pouch surgery, daytime stool frequency is 4–7, urgency is common, and nocturnal incontinence is present in about 20% of patients.

About 30% of patients with pouches develop pouch mucosal inflammation of unknown etiology. The diagnosis of pouchitis is made in patients with worsening diarrhea, endoscopic signs of pouch inflammation, and acute inflammation histologically. Treatments include oral metronidazole and topical or oral steroids or 5ASAs (Tables IV and V). Probiotic therapy is a novel option, but a minority of patients with refractory pouchitis require pouch resection and a permanent ileostomy.

UC IN PREGNANCY AND CHILDHOOD

Fertility, Pregnancy, and Lactation

Female fertility is normal except in active UC. Male fertility in patients taking sulfasalazine is reduced as a result of azospermia, but this can be reversed within a few weeks by switching to a different 5ASA (Table V). Although the outcome of pregnancy is normal in quiescent UC, there is an increased rate of spontaneous abortion, premature delivery, and stillbirth in persistently active disease. Pregnancy has no consistent effect

on the activity of UC, although the disease occasionally flares in the puerperium.

Corticosteroids and 5ASAs can be used safely during pregnancy and lactation; withholding them exposes the mother and fetus unnecessarily to the risks of active disease. Although their teratogenic potential means that thiopurines should be avoided during pregnancy if possible, accidental pregnancies in patients taking these drugs have been uneventful. Surgery is occasionally necessary in very sick patients but has a high rate of fetal loss.

Childhood

UC may occur in children of any age. Allergy to protein in milk from cows produces a similar syndrome in babies after weaning and needs to be excluded. The diagnosis of UC in children is often delayed; it needs to be considered early not only in patients with diarrhea, but also in those with delayed growth and puberty. Prompt referral to a pediatric gastroenterology unit is advised for appropriate investigation.

The principles of treatment of UC in children resemble those in adults. However, the adverse effects of UC corticosteriods on growth and pubertal development mean that active UC should be suppressed promptly, undernutrition reversed, and prolonged courses of corticosteroids avoided. To maintain growth and development, prepubertal colectomy may be necessary. Azathioprine is a useful steroid-sparing agent in children for whom surgery is inappropriate or is declined.

PROGNOSIS

The risk of death in UC is highest in the first year of diagnosis and relates mainly to acute severe UC. In this setting, risk of death is now less than 2%. The overall mortality of patients with UC resembles that of the normal population.

FUTURE TRENDS

The intense research effort underway to clarify the genetics of inflammatory bowel disease is likely to have a major impact on its management. Identification of the genes involved, and the proteins they encode, will shed new light on the pathogenesis of the disease and is likely in turn to lead to new treatments. Such information will enable us to identify relatives of index patients who are at risk of developing IBD, to facilitate diagnosis, to predict the phenotype and natural history of the disease, and to determine the likely response of individuals to specific therapies. Advances in molecular biology are likely also to enable us to specify which patients with UC are at particular risk of developing colorectal cancer, such that colonoscopic screening may become obsolete.

Whatever advances are made in the coming years, the management of patients with UC will continue to require close collaboration between physicians, surgeons, nurses, dietitians, and counselors. Most importantly, the patient with UC must be viewed as a person rather than a case. As management becomes more complex, and the options more varied, it is essential that the patient remains at the center of the decision-making process: the individual with UC must be the final arbiter of the type of therapy he or she is to be given.

See Also the Following Articles

Cholangiocarcinoma • Cholangitis, Sclerosing • Colonoscopy • Colorectal Adenocarcinoma • Crohn's Disease • Ileoanal Pouch • Pouchitis • Proctitis and Proctopathy • Toxic Megacolon

Further Reading

Allan, R. N., Rhodes, J. M., Hanauer, S. B., Keighley, M. R. B., Alexander-Williams, J., and Fazio, V. W. (eds.) (1997). "Inflammatory Bowel Diseases," 3rd Ed. Churchill Livingstone, New York.

Kirsner, J. B., and Shorter, R. G. (eds.) (1995). "Inflammatory Bowel Disease," 4th Ed. Lea & Febiger, Philadelphia.

Rampton, D. S. (ed.) (2000). "Inflammatory Bowel Disease: Clinical Diagnosis and Management." Martin Dunitz, London.

Rampton, D. S., and Shanahan, F. (2000). "Inflammatory Bowel Disease." Fast Facts Series. Health Press, Oxford.

Stein. S. H., and Rood. R. P. (1998). "Inflammatory Bowel Disease—A Guide for Patients and Their Families," 2nd Ed. Crohn's & Colitis Foundation of America, New York.

Targan, S. R., and Shanahan, F. (eds.) (2003). "Inflammatory Bowel Disease: From Bench to Bedside." 2nd Ed. Williams and Wilkins, Baltimore.

Colitis, Ulcerative (Pediatric)

SEUNG-DAE PARK* AND JAMES F. MARKOWITZ*,†
*Schneider Children's Hospital, New York and †New York University

Ulcerative colitis (UC), a chronic inflammatory bowel disease, is an important pediatric gastrointestinal disorder. It is a condition that causes significant morbidity, and rarely mortality, in children and adolescents. Despite the increased understanding of its pathophysiology and novel medical therapy, UC remains medically incurable. However, current medical and surgical therapeutic options have improved the overall outlook for children with UC.

EPIDEMIOLOGY

The incidence of ulcerative colitis (UC) appears to have remained fairly stable over the past 30 years, with a reported incidence in children of 1.5 to 10 cases per 100,000 and a prevalence of 18 to 30 per 100,000. Males and females are equally affected. Although presentation during adolescence is common, as many as 40% of children present before the age of 10 years and a number of well-documented cases have developed before the first birthday.

Risk factors associated with the development of UC during childhood include ethnic background and a positive family history of inflammatory bowel disease (IBD). Ashkenazic Jews are affected more often than other Caucasians, and Caucasians are affected more often than African Americans. Overall, 10 to 15% of children with UC have first-degree relatives with IBD. Although cigarette smoking apparently protects against the development of UC in adult populations, the data regarding exposure to cigarette smoke as a risk factor for the development of UC in childhood are less clear-cut. Epidemiological studies have documented, however, that passive smoking during childhood protects against the development of UC in adulthood.

GENETICS

The search for genes associated with the development of ulcerative colitis has begun to uncover intriguing possibilities. Genome-wide screens have suggested that there may be significant genetic heterogeneity between different ethnic and racial populations with UC. Interest has focused on the human leukocyte antigens for many years. Studies in both the United States and Japan have suggested a significant association between specific HLA-DR2 alleles and the development of UC. More recent studies from Korea and Mexico suggest a similar association with human leukocyte antigen (HLA)-DRB1 alleles. Gene mutations in the intracellular adhesion molecule (ICAM)-1 gene and the tumor necrosis factor (TNF) (or another nearby) gene have also been identified more frequently in patients with UC than in healthy populations or those with Crohn's disease. Finally, a recent report has identified a genetic polymorphism in the multidrug resistance 1 (MDR1) gene, associated with decreased production of P-glycoprotein, which is found more frequently in patients with UC than in controls. This abnormality likely impairs colonic defenses against lumenal bacteria or toxins, resulting in a chronic immune response. This observation is of particular importance, as an MDR1 knockout mouse has been shown to develop a form of colitis similar to UC that can be prevented by antibiotics. The Crohn's disease susceptibility gene NOD2/CARD15 is not associated with UC.

Other serologic markers may identify specific genetic subsets of UC. The best defined is pANCA, a perinuclear anti-neutrophil cytoplasmic antibody that is present in approximately 70% of UC patients but in only 6% of those with Crohn's disease and 3% of healthy controls. The presence or absence of pANCA is concordant within families and also tends to correlate with the presence of HLA-DR2. UC phenotype correlates poorly with the presence or absence of pANCA, except for patients with sclerosing cholangitis and UC, who are almost always pANCA positive. The fact that not all UC patients manifest this marker strongly suggests that UC is a genetically heterogeneous disorder.

CLINICAL FEATURES

Anatomic Distribution

Pancolitis is common in children with UC, with reports suggesting frequencies of 40 to 60%. Left-sided colitis is seen in approximately one-third of children and proctitis in approximately one-quarter. Although UC is described as a disease restricted to

 Copyright 2004, Elsevier (USA). All rights reserved.

the colon, several recent reports have demonstrated that pathologic changes may also be found in the proximal gastrointestinal tract. Endoscopic studies have revealed histologic esophagitis in 15–50% of children and gastroduodenal inflammation in 25–69%. Although the presence of granulomas can support a diagnosis of Crohn's disease, inflammation in the proximal gastrointestinal tract is not a sufficient finding to exclude UC.

Symptoms and Signs

Diarrhea, rectal bleeding, and abdominal pain are almost universal. Frequent loose stool can contain either streaks of blood or clots. Children describe both tenesmus and urgency, although the former symptom is at times misinterpreted as constipation. Acute weight loss is common, but linear growth is usually maintained. When only proctitis is present, children may have stools of normal consistency, no systemic symptoms, and minimal hematochezia.

At the time of initial presentation, disease activity can be quite variable. Approximately 50% of children present with mild symptoms, characterized by fewer than four stools per day, intermittent hematochezia, and minimal if any systemic symptoms or weight loss. These children generally have normal physical examinations or only minimal tenderness on palpation of the lower abdomen. Another third of children are moderately ill, with weight loss, more frequent diarrhea, and systemic symptoms. Physical examination demonstrates more significant abdominal tenderness. An acute fulminant disease presentation, characterized by severe crampy abdominal pain, fever, more than six diarrheal stools per day, and, at times, copious rectal bleeding, is seen in the remaining 10–15% of cases. These children commonly manifest tachycardia, orthostatic hypotension, diffuse abdominal tenderness without peritoneal signs, and distension. Toxic megacolon represents the most dangerous extreme of acute, fulminant colitis, but is unusual in children.

Extraintestinal Manifestations

Extraintestinal manifestations of disease have been described in many organ systems of the body, most commonly in the hepatobiliary tree, joints, skin, and eyes. Many of these inflammatory manifestations develop during bouts of increased colitis activity. Although the cause of these manifestations remains unknown, an anti-colonocyte antibody detectable in the sera of patients with UC has been shown to cross-react with antigens present in the skin, ciliary body of the eye, bile duct, and joints. One hypothesis for the development of extraintestinal symptoms in these organs is therefore that they occur when autoantibodies capable of recognizing these nonintestinal tissues develop as part of the humoral response characteristic of UC.

Hepatobiliary Disease

Abnormal liver function tests are common in children with UC, often seen transiently during periods of increased disease activity or associated with the use of a number of treatments including corticosteroids, sulfasalazine, parenteral nutrition, azathioprine, and 6-mercaptopurine. Unfortunately, however, more significant hepatobiliary problems are also seen in association with UC. Primary sclerosing cholangitis (PSC) occurs in 3.5% and immune hepatitis in $<1\%$ of children and adolescents with UC. Both conditions may be present before or at the time of the initial diagnosis of UC or can develop at any time during the course of the illness. Although most children described in the literature appear to have mild liver disease, this may be a function of relatively short-term follow-up. In adults, PSC commonly progresses to end-stage liver disease, requiring transplantation or death from cholangiocarcinoma. Immune hepatitis may also progress to end-stage liver disease. Neither hepatobiliary disease appears to positively or negatively influence the activity of colitis, but the presence of PSC in UC has been shown to enhance the risk of colorectal aneuploidy, dysplasia, and cancer. Absolute cumulative risk for colorectal cancer in UC patients with PSC after 10, 20, and 25 years of disease is 9, 31, and 50%, respectively, compared with 2, 5, and 10%, respectively, in UC patients without PSC. Since the time to dysplasia may be accelerated, once the diagnosis of UC is made in the setting of PSC, more frequent colonoscopic surveillance may be indicated.

Joint Disorders

Arthralgias are common in children with UC and arthritis, either a peripheral migratory type affecting the large joints or a monoarticular, nondeforming type primarily affecting the knees or ankles, is reported in 10 to 20% of children. The presence and activity of arthritis usually (but not invariably) correlate with the activity of the bowel disease. Ankylosing spondylitis occurs in up to 6% of adults with UC but is rare during childhood.

Skin Disorder

Both erythema nodosum and pyoderma gangrenosum have been seen in children, generally during periods of enhanced colitis activity. Erythema nodosum lesions appear as raised, erythematous, painful circular nodules that usually occur over the tibia but may be

present on the lower leg, ankle, or extensor surface of the arm. Lesions persist for several days to as long as a few weeks, but usually remit following treatments directed at the enhanced colitis activity. Pyoderma gangrenosum usually appears as small, painful, sterile pustules that coalesce into a larger sterile abscess. This abscess usually drains, forming a deep, necrotic ulcer. Lesions usually occur on the lower extremities, although the upper extremities, trunk, and head are not spared. Although pyoderma lesions have traditionally been quite resistant to therapy, the anti-TNFα monoclonal antibody infliximab or the immune modifiers ciclosporin or tacrolimus can rapidly heal the lesions and should be considered the current treatments of choice.

Thromboembolic Disorders

There have been case reports of thromboembolic complications in children with UC. Sites of venous or arterial thrombosis include the extremities, portal or hepatic vein, lung, and central nervous system. Though the presence of factor V Leiden mutations may be associated with this hypercoaguable state, other prothrombotic inherited abnormalities and other factors, such as hyperhomocysteinemia or resistance to activated protein C, do not appear to be important.

Ocular Disorders

Severe eye disorders, such as optic neuritis, are rare in children with UC. Less serious conditions including episcleritis and asymptomatic uveitis have been described. Other ocular disorders, such as posterior subcapsular cataracts or increased ocular pressure, may result from corticosteroid therapy.

Complications

Toxic Megacolon

This complication develops rarely with current medical management, although the literature reports rates of up to 5% of children and adolescents with UC. Toxic megacolon represents a medical and potentially a surgical emergency with the potential for rapidly progressive deterioration complicated by severe electrolyte disturbances, hypoalbuminemia, hemorrhage, perforation, sepsis, and/or shock. Precipitating factors include the use of anti-diarrheal agents, such as anticholinergics or opiates, and excessive colonic distension during enema or colonoscopy.

Carcinoma

UC is a premalignant disease, with a well-documented proclivity for adenocarcinoma. A flat dysplastic mucosa rather than an adenomatous polyp represents the precursor lesion. The genetic alterations that precede the development of dysplasia occur multifocally in the colon, so that the resulting adenocarcinomas may be evenly distributed throughout the colon. Multifocal or synchronous tumors are present in 10 to 20% of patients.

Duration of disease and extent of colitis are the two most critical risk factors for cancer in UC. Because of this, children with UC have a particularly high lifetime risk of colorectal adenocarcinoma. Less well-characterized risk factors include concomitant PSC; an excluded, defunctionalized, or bypassed segment of bowel; and depressed red blood cell folate levels. Although patients as young as 16 years of age have been demonstrated to have colonic aneuploidy, dysplasia, or cancer, as in adults the risk for these changes does not appear to be significant until after the first decade of illness.

Given the high risks of colorectal cancer, surveillance colonoscopy is recommended. Studies have documented that undergoing colectomy following detection of low- or high-grade dysplasia at surveillance can reduce the risk of advanced adenocarcinoma, thereby improving mortality in populations undergoing regular surveillance. Although no prospective studies have assessed the optimal schedule of surveillance, a cost–benefit analysis suggests colonoscopies every 3 years for the first 10 years of surveillance, generally starting 7 to 10 years after diagnosis, with more frequent investigations as the duration of colitis increases. Although many advocate initiating surveillance only after 15 to 20 years of disease in adults with left-sided colitis or proctosigmoiditis, the frequent proximal extension of these disease distributions in children mandates that at least a screening colonoscopy be performed in all children 7 to 10 years after diagnosis. Subsequent frequency of surveillance can then be determined based on whether the extent of colitis has increased. Surveillance assessments require a panendoscopy to the cecum, with two to four biopsies obtained every 10 cm from the cecum to the sigmoid and every 5 cm in the sigmoid and rectum. Additional biopsies must be performed if a mass or other suspicious lesion is identified. Current recommendations for colectomy include any identification of dysplasia (low- or high-grade) confirmed by two independent, experienced pathologists. Repeat colonoscopy for confirmation of dysplasia on new biopsies is not recommended. If indefinite dysplasia is identified, aggressive medical management to reduce active inflammation, followed by repeat surveillance colonoscopy within 3 to 6 months, is indicated.

Growth and Development

In contrast to children with Crohn's disease, disturbances in growth are unusual in children with UC. Although weight loss during periods of active colitis and decreased dietary intake is common, significant impairments in linear growth occur in only approximately 10% of children with UC. Differences in the cytokine alterations and the absence of small bowel involvement in UC are postulated to explain the difference in growth patterns between these two diseases. When children with UC are growing poorly, control of active disease, avoidance of excessive corticosteroid use, and improvement in caloric intake generally result in resumption in linear growth.

DIAGNOSIS

History

Clinical symptoms, including abdominal pain, diarrhea, and rectal bleeding, are often obvious. However, in children with mild disease activity or inflammation restricted to only the rectum, symptoms may be less obvious and more difficult to elicit, especially in children or adolescents who resist discussing their bowel habits. Nocturnal bowel movements and symptoms such as weight loss, poor growth, arrested sexual development, or, in the postmenarchal adolescent, secondary amenorrhea suggest an organic rather than a functional condition. A positive family history for IBD also should suggest the need to evaluate for possible UC.

Physical Examination

Children with active UC often have mild to moderate abdominal tenderness, especially in the left lower quadrant or in the midepigastric area. With fulminant disease, marked tenderness can be present. Rebound tenderness and increasing abdominal distension are particularly ominous signs and should alert the physician to possible toxic megacolon. The perianal area is generally normal. Extraintestinal manifestations, such as arthritis, erythema nodosum, or pyoderma gangrenosum, are important clues to the autoimmune nature of the child's illness.

Laboratory Studies

Laboratory studies cannot be used to make a diagnosis of UC. Rather, recommended laboratory studies (see Table I) help exclude other illnesses and provide evidence to support proceeding to more diagnostic radiologic and endoscopic procedures. Serologic testing for pANCA may be useful, as the presence of this marker is unusual in children who do not have IBD. A negative serologic study does not, however, exclude the possibility of UC, and studies have not demonstrated significant advantage to the use of this serologic testing over more traditional laboratory screening tests. The presence of anti-*Saccharomyces cerevisiae* antibody may suggest a diagnosis of Crohn's disease rather than UC. Stool studies for enteric bacterial and parasitic pathogens must be performed to exclude infectious colitis. Given the frequency with which children are exposed to antibiotics, screening for *Clostridium difficile* toxin must also be performed. If pathogens are present, monitoring the child's response to treatment is necessary, as it is not unusual for children with UC to present initially with superimposed infection. Microcytic anemia, mild to moderate thrombocytosis, elevated erythrocyte sedimentation rate, and hypoalbuminemia are present in 40 to 80% of cases. The total leukocyte count is normal to only mildly elevated, unless the illness is complicated by acute fulminant colitis. Elevated serum aminotransferase levels are present in 3% of children at the time of initial diagnosis and reflect signs of potential serious concomitant liver disease (chronic active hepatitis or PSC) in approximately half of them. In a number of children, however, all laboratory studies are normal.

TABLE I Laboratory Evaluation of the Child with Suspected Ulcerative Colitis

Blood tests
Complete blood count, differential, reticulocyte count
Erythrocyte sedimentation rate, C-reactive protein
Electrolytes, serum chemistries (including total protein, albumin, liver functions)
Serum iron, total iron-binding capacity, ferritin
Stool tests
Stools for enteric pathogens (including *Salmonella, Shigella, Campylobacter jejuni, Yersinia enterocolitica, Aeromonas hydrophilia, Escherichia coli*)
Stool for *Clostridium difficile* toxin A and B
Direct microscopic examination of the stool for ova and parasites, Charcot-Leyden crystals, leukocytes
Stool for *Giardia lamblia* antigen
Serologic assessment
Anti-neutrophil cytoplasmic antibody
Anti-*Saccharomyces cerevisiae* antibody
Celiac serologies (tissue transglutaminase antibody, anti-endomysial antibody)

Radiography

Radiographic studies have a limited role in the evaluation of a child with suspected UC. Barium enema is only rarely indicated, having been replaced by

colonoscopy. Abdominal ultrasound and computed tomography (CT) offer little help to establish the diagnosis. Plain films of the abdomen are important in determining the degree of colonic distension and the presence of possible toxic megacolon or perforation. In most circumstances, however, the child with suspected UC should undergo an upper gastrointestinal series with small bowel follow-through to help exclude the possibility of Crohn's disease. CT may also be helpful in this regard.

Endoscopy and Histology

Colonoscopy provides the best means of diagnosing UC, as the procedure allows direct visualization and biopsy of the mucosa of the entire colon. This provides the physician with the ability to accurately determine the extent, distribution, and histologic characteristics of the disease. Typically, UC is characterized by diffuse inflammation that begins at the anal verge and progresses proximally to a variable degree. Whereas rectal sparing is generally thought to indicate Crohn's disease, untreated children with UC have been described with rectal sparing at initial evaluation, only to manifest findings of typical UC at a later date. Endoscopically, active UC is characterized by diffuse exudates, ulceration, and marked hemorrhage. With milder disease activity, the mucosa may only appear erythematous, with loss of the normal vascular markings and increased contact friability. All children who undergo endoscopy should be biopsied. Mucosal biopsies are characterized by neutrophilic infiltration of the crypts, crypt abscesses, goblet cell depletion, crypt distortion, and a papillary configuration to the surface epithelium. Although these findings are not pathognemonic for UC (as they can be seen in cases of severe Crohn's colitis), the biopsies often allow differentiation between UC and both self-limited colitis and most cases of Crohn's disease.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis for the child or adolescent with suspected UC is summarized in Table II. Though much of the differential diagnosis will be the same for children of all ages, as noted in Table II, a number of these conditions are likely to occur only in the very young. Similarly, a number of conditions in the differential diagnosis list can be excluded simply by ascertaining whether or not there is blood in the stool.

TABLE II Differential Diagnosis of Ulcerative Colitis in Children

Enteric infection
Salmonella
Shigella
Campylobacter jejuni
Aeromonas hydrophilia
Yersinia enterocolitica
Enterohemorrhagic *Escherichia coli*
Entomeba histolytica
Giardia lamblia[a]
Pseudo-membranous (postantibiotic) enterocolitis
Clostridium difficile
Carbohydrate intolerance[a]
Lactose
Sucrose
Nondigestible carbohydrates (sorbitol, xylitol, mannitol, maltitol, sucralose)
Vasculitis
Henoch-Schonlein purpura
Hemolytic–uremic syndrome
Allergic enterocolitis[b]
Hirschsprung's enterocolitis[b]
Eosinophilic gastroenteritis
Celiac disease[a]
Laxative abuse[a]
Neoplasms
Juvenile polyp[b]
Adenocarcinoma
Intestinal polyposis

[a] Watery, nonbloody diarrhea.
[b] Primarily in the young child.

MEDICAL THERAPIES

Despite extensive research, UC remains medically incurable. Treatment is therefore designed to minimize symptoms and improve a child's quality of life with the hope that complications can be avoided or controlled with a minimum of treatment-induced toxicity. Unfortunately, as pediatric trials are limited, much of the data supporting the currently recommended treatments for children with UC have been extrapolated from adult studies. The following discussion focuses on those aspects of treatment that have been shown to be particularly effective in or unique to the pediatric population.

Nutritional Therapy

No nutritional or dietary therapies have been shown to be effective as primary treatment for children with UC. Rare children may benefit from an exclusion diet, presumably because their particular illness is at least in part an allergy rather than ulcerative colitis. Such a

patient is the exception, however, rather than the rule. Possibly because the colonocyte derives energy from short-chain fatty acids in the fecal stream, bowel rest to reduce the frequency of bowel movements is not effective therapy. Although preliminary treatment trials using short-chain fatty acids have demonstrated modest benefits in adults with UC, no pediatric trials have been reported. Supplementation of the diet with omega-3 fatty acids (fish oils) has been shown to offer some protection against early relapse in adults with UC, but again there are no comparable studies in children. Nutritional interventions in UC are therefore generally adjunctive to other treatments. Assurance of an adequate dietary intake promotes normal growth and prevents catabolism, thereby enhancing the effect of other treatment modalities. Nutritional support can be accomplished successfully by a number of approaches, including dietary supplementation and enteral or parenteral nutrition.

Corticosteroids

Corticosteroids remain the gold standard for inducing remission in children with moderate to severe colitis. These agents are not, however, effective or safe for maintaining remission. Their use in children with UC has largely been extrapolated from trials in adults, with current recommendations evolving from empiric use and clinical experience rather than controlled clinical trials. Systemically acting agents, including prednisone, methylprednisone, and hydrocortisone, are most frequently prescribed and result in response rates of 65–90% in children with moderate to severe disease. Intravenous and rectal formulations in addition to oral preparations make these medications suitable for a wide variety of treatment scenarios. Rectal preparations, including enemas, foams, and suppositories, are particularly suitable as adjunctive therapy for the child with severe tenesmus and urgency and may be the only treatment required in cases of limited proctitis. The use of corticosteroids must be weighed against their potential adverse effects (see Table III). Systemically active corticosteroids can interfere with linear bone growth even in the face of adequate dietary intake. Alternate-day dosing can minimize these effects while maintaining reduced disease activity and appears to have only limited deleterious effects on bone mineralization in children. However, in most cases, the deleterious effects of an extended course of steroids as well as their limited effectiveness for maintenance of remission make the long-term use of a corticosteroid inadvisable. Newer topically active corticosteroids with high first-pass metabolism, such as budesonide, have the potential to provide anti-inflammatory activity to the gut without systemic toxicity. Such an agent may offer particular advantages for the treatment of children, but only limited pediatric trials have been reported. The enema formulation of budesonide is as effective as rectal mesalamine and rectal hydrocortisone in the treatment of left-sided and distal colitis. Multiple courses of rectal budesonide are safe and effective for recurrent flares of UC. By contrast, the oral formulation that is currently available is a controlled ileal release capsule, which may not deliver adequate medication to the distal colon for effective treatment of UC. Studies of this formulation in children with UC have not, however, been performed.

TABLE III Adverse Effects of Corticosteroids

Cosmetic
Moon facies
Acne
Hirsutism
Striae
Central obesity
Metabolic
Hypokalemia
Hyperglycemia
Hyperlipidemia
Endocrinologic
Suppression of linear growth
Adrenal suppression
Musculoskeletal
Osteopenia
Osteoporosis
Vertebral collapse
Aseptic necrosis of bone
Myopathy
Psychological
Mood swings
Psychosis
Ocular
Cataracts
Increased ocular pressure
Other
Systemic hypertension
Pseudotumor cerebri
Immunosuppression

5-Aminosalicylates

Clinical experience suggests that the 5-aminosalicylate (5-ASA) drugs (sulfasalazine, mesalamine, olsalazine, balsalazide) effectively induce and maintain remission in 50–90% of children with mild to moderate UC. These agents exert local anti-inflammatory effects through a number of different mechanisms including inhibition of 5-lipoxygenase with resulting decreased

production of leukotriene B4, scavenging of reactive oxygen metabolites, prevention of the up-regulation of leukocyte adhesion molecules, and inhibition of interleukin-1 (IL-1) synthesis. 5-ASA is rapidly absorbed from the upper intestinal tract on oral ingestion and various delivery systems have been employed to prevent absorption until the active drug can be delivered to the distal small bowel and colon. Sulfasalazine (Azulfidine) links 5-ASA via an azo bond to sulfapyridine. Bacterial enzymes in the colon break the azo linkage, releasing 5-ASA to exert its anti-inflammatory effect in the colon. Because the sulfapyridine moiety causes most of the untoward reactions to sulfasalazine and is thought to have no therapeutic activity, newer agents have been designed to deliver 5-ASA without sulfapyridine. Olsalazine (Dipentum) links two molecules of 5-ASA via an azo bond, whereas balsalazide (Colazal) links 5-ASA via an azo bond to an inert, nonabsorbable carrier. A number of other delayed-release preparations (Asacol, Claversal, Mesasal, Salofalk, Pentasa) prevent rapid absorption of 5-ASA (generically called mesalamine) by coating tablets or microspheres with different materials designed to dissolve and release medication in a time- or pH-dependent manner. Uncoated mesalamine is also available as a rectal suppository (Canasa/Salofalk) or enema formulation (Rowasa), which can be of significant value in the treatment of a child with UC or proctitis. Adverse reactions to all of the 5-ASA preparations have been described and have required discontinuation of treatment in 5 to 15% of cases. Serious complications reported in pediatric patients include pancreatitis, nephritis, exacerbation of UC, and sulfa- or 5-ASA-induced allergic reactions.

Antibiotics

There is little role for antibiotics in the primary therapy of active UC. Based on experience in adults, metronidazole is occasionally used for the treatment of mild to moderate UC or for maintenance of remission in the 5-ASA-intolerant or allergic patient. A controlled trial of ciprofloxacin as an adjunct to corticosteroids in adults with active UC demonstrated no benefit compared with placebo.

Immunomodulators and Biologicals

Although colectomy "cures" UC, many parents and physicians are reluctant to perform such an operation in children with even severely active UC. As a consequence, immunomodulators and new biologic agents are increasingly being used therapeutically when chronic or intractable symptoms are present. The most commonly prescribed treatments in this category are discussed below.

6-Mercaptopurine and Azathiopurine

6-Mercaptopurine and its prodrug azathioprine are purine analogues that inhibit RNA and DNA synthesis, thereby down-regulating cytotoxic T-cell activity and delayed hypersensitivity reactions. Both medications have steroid-sparing effects and effectively maintain remission in approximately two-thirds of children with UC. Onset of action is delayed (mean time to response of 4.5 ± 3.0 months), limiting the role of these agents for the induction of remission. Adverse reactions requiring discontinuation of treatment, such as allergic reactions, pancreatitis, abnormal liver function tests, or severe leukopenia, occur in less than 5−15% of pediatric patients. The recent availability of assays to measure the active metabolites of these agents offers the prospect of further enhancing treatment efficacy while reducing toxicity.

Ciclosporin and Tacrolimus

Ciclosporin and tacrolimus (FK506) are potent inhibitors of cell-mediated immunity that have been demonstrated to rapidly and markedly reduce severe colitis activity in 20 to 80% of children with fulminant UC who might otherwise require imminent colectomy. However, relapses necessitating colectomy occur within 1 year in 70 to 100% of initial responders during or after discontinuation of therapy. By contrast, if 6-mercaptopurine or azathiopurine is added to the regimen once ciclosporin or tacrolimus has induced remission, 60 to 90% of treated patients maintain long-term remission. These agents bind to intracellular receptors, forming complexes that ultimately down-regulate the production of cytokines IL-2 and IL-4. As a consequence, T-cell function and, to a lesser extent, B-cell function are impaired.

The use of these agents has been limited by drug-induced toxicity. Tremors, hirsutism, decreased renal function, and systemic hypertension are the most common side effects that have been described in children with IBD. However, isolated reports of pneumocystis carinii pneumonia (PCP), lymphoproliferative disease, and serious bacterial and fungal infections in ciclosporin-treated patients merit careful monitoring of all children treated with either drug, especially those treated in combination with corticosteroids and 6-mercaptopurine or azathiopurine. When these agents are used, PCP prophylaxis and anti-candida therapy are indicated.

Other Immunomodulators

Limited studies in adults have suggested that methotrexate may be an effective treatment for UC in patients who are intolerant of or resistant to azathioprine. Clinical trials in children have not been performed and clinical experience with this agent in children with UC is very limited.

Infliximab

The chimeric anti-TNFα antibody, infliximab, is a potent inhibitor of TNFα. Infliximab rapidly downregulates cytokine activity in Crohn's disease. This results in a rapid and significant decrease of disease activity. A few open-label trials in adults with UC and a single small open-label trial in children have demonstrated a meaningful clinical improvement in patients with chronic intractable and severe UC. Response to a single infusion lasts 4 to 12 weeks and repeat infusions are necessary to maintain response in virtually all responsive patients. Adverse reactions are primarily related to minor infusion reactions, but only a very small number of children with UC have been treated. More severe reactions, such as delayed hypersensitivity reactions, anaphylaxis, and reactivation of latent tuberculosis, have been seen in adults and children receiving infliximab for Crohn's disease or rheumatoid arthritis.

Surgery

Curative surgery requires total proctocolectomy with a reconstructive procedure (see below). This is indicated because of intractable or fulminant disease or because of the development of dysplasia or cancer. In fact, approximately 20% of children and adolescents require colectomy within 5 years of diagnosis because of intractable disease or fulminant symptoms. Although proctocolectomy and ileostomy result in a healthy patient with no risk of future recurrence, few children or parents readily accept the option of a permanent ileostomy. Most instead opt for restorative surgery [the ileal pouch–anal anastomosis (IPAA)], which allows the child to continue to defecate by the normal route. Although this can be performed as a one-stage operation, it is often advisable to carry out an IPAA as a two- or three-stage procedure, especially when there is any preoperative suggestion that a patient might have Crohn's rather than ulcerative colitis or when large doses of steroids have been used chronically. Reviews of pediatric surgical experience document that IPAA utilizing an ileal J-pouch (or less commonly a W- or an S-pouch) results in fewer daytime and nocturnal bowel movements and less fecal soiling than an ileoanal anastomosis without a pouch. Anorectal function is well preserved in children and postoperative fecal soiling is unusual. The most common early postoperative complication of IPAA is small bowel obstruction. Pouchitis, a chronic inflammation of the ileal pouch, is the most common late complication described, occurring in 19% of children and adolescents. Pouchitis generally responds to treatment with metronidazole, ciprofloxacin, 5-ASA, or corticosteroids.

COURSE AND PROGNOSIS

Clinical experience amassed since 1975 suggests that 70% of children with UC can be expected to enter remission within 3 months of initial diagnosis, irrespective of the severity of their initial attack. Though symptoms remain inactive in 45–58% over the first year after diagnosis, 10% of those initially presenting with moderate to severe colitis can be expected to remain continuously symptomatic. In any given year over the next decade, approximately 55% of patients have inactive disease, 40% have chronic intermittent symptoms, and 5–10% have continuous symptoms. Approximately 5% of all children require colectomy within the first year after diagnosis and 19–23% require surgery within 5 years after diagnosis. These rates rise to 9 and 26%, respectively, in the subgroup of children initially presenting with moderate to severe symptoms.

Children with proctitis or proctosigmoiditis have less morbidity, as more than 90% are asymptomatic within 6 months of diagnosis and less than 5% have continuously active disease. In contrast to adults, however, proximal extension of disease may occur in up to 70% of children over the course of follow-up. Colectomy may eventually be required in 5% of these patients.

See Also the Following Articles

Colorectal Adenocarcinoma • Crohn's Disease, Pediatric • Diarrhea, Pediatric • Toxic Megacolon

Further Reading

Ahsgren, L., Jonsson, B., Stenling, R., and Rutegard, J. (1993). Prognosis after early onset of ulcerative colitis: A study from an unselected patient population. *Hepatogastroenterology* **40**, 467–470.

Bonen, D. K., and Cho, J. H. (2003). The genetics of inflammatory bowel disease. *Gastroenterology* **124**, 521–536.

Ekbom, A., Helmick, C., Zack, M., and Adami, H.-O. (1990). Ulcerative colitis and colorectal cancer: A population-based study. *N. Engl. J. Med.* **323**, 1228–1233.

Griffiths, A. M., and Sherman, P. M. (1997). Colonoscopic surveillance for cancer in ulcerative colitis: A critical review. *J. Pediatr. Gastroenterol. Nutr.* **24**, 202–210.

Gryboski, J. D. (1993). Ulcerative colitis in children 10 years old or younger. *J. Pediatr. Gastroenterol. Nutr.* **17**, 24–31.
Hildebrand, H., Fredrikzon, B., Holmquist, L., Kristiansson, B., and Lindquist, B. (1991). Chronic inflammatory bowel disease in children and adolescents in Sweden. *J. Pediatr. Gastroenterol. Nutr.* **13**, 293–297.
Hyams, J., Davis, P., Lerer, T., Colletti, R., Bousvaros, A., Leichtner, A., Benkov, K., Justinich, C., and Markowitz, J. (1997). Clinical outcome of ulcerative proctitis in children. *J. Pediatr. Gastroenterol. Nutr.* **25**, 149–152.
Hyams, J., Markowitz, J., Treem, W., Davis, P., Grancher, K., and Daum, F. (1995). Characterization of hepatic abnormalities in children with inflammatory bowel disease. *Inflamm. Bowel Dis.* **1**, 27–33.
Hyams, J. S., Davis, P., Grancher, K., Lerer, T., Justinich, C. J., and Markowitz, J. (1996). Clinical outcome of ulcerative colitis in children. *J. Pediatr.* **129**, 81–88.
Langholz, E., Munkholm, P., Krasilnikoff, P. A., and Binder, V. (1997). Inflammatory bowel diseases with onset in childhood: Clinical features, morbidity, and mortality in a regional cohort. *Scand. J. Gastroenterol.* **32**, 139–147.
Markowitz, J. F. (2002). Medical management of IBD: Children versus adults. *In* "Falk Symposium 122, Inflammatory Bowel Disease: A Clinical Case Approach to Pathophysiology, Diagnosis, and Treatment" (M. Campieri, C. Fiocchi, S. B, Hanauer, D. P. Jewell, D. Rachmilewitz, and J. Scholmerich, eds.), pp. 71–84. Kluwer Academic, Dordrecht.
Ringheanu, M., and Markowitz, J. (2002). Inflammatory bowel disease in children. *Curr. Treat. Options Gastroenterol.* **5**, 181–196.
Sarigol, S., Caulfield, M., Wyllie, R., Alexander, F., Lavery, I., Steffen, R., Kay, M., and Michener, W. (1996). Ileal pouch–anal anastomosis in children with ulcerative colitis. *Inflamm. Bowel Dis.* **2**, 82–87.

Colon, Anatomy

LUCA MAZZUCCHELLI* AND CHRISTOPH MAURER†
**University of Bern, Switzerland and †Kantonspital Liestal, Switzerland*

intestinal barrier The border formed by the intestinal mucosa and in particular the epithelial cell layer between the external world, i.e., the intestinal lumen with its contents and the internal environment.
intestinal intrinsic nervous system A complex system of interconnecting nerve fascicles and ganglia occurring within the intestinal wall and able to function autonomously.

The etiology, pathogenesis, and symptoms caused by several pathologic processes occurring in the large intestine are strictly related to its anatomy. The success of numerous surgical procedures performed on the colon largely depends on profound knowledge of its anatomical relationships with adjacent organs and structures as well as its vascular supply and innervation.

INTRODUCTION

The large intestine, approximately 1.5 m in length, extends from the distal end of the ileum to the anus. It includes the cecum with the appendix vermiformis and the ileocecal valve, the colon, and the rectum with the anus. In sequence, the segments of the colon are designated as the ascending, the transverse, the descending, and the sigmoid colon. Clinicians refer to the large intestine proximal to the middle of the transversum as the right colon and to the distal portions as the left colon. The colon initiates in the right iliac region, arches around the small intestine, and finally forms a sinuous loop in the left iliacal region before entering the lesser pelvis. The shape and course of the large intestine, which can be easily recognized on a barium enema radiograph, are not firmly fixed, but may vary according to the position of adjacent peritoneal and retroperitoneal organs. For instance, the hepatic flexure descends with the liver with each inspiration, the middle portion of the transverse colon, which is normally found in an approximately horizontal position, descends by dilation of the stomach, and the course of the ascending and descending colon may be affected by extension of the coils of the small intestine. The diameter of the large intestine progressively diminishes from the cecum to the sigmoid colon, with the narrowest portion being located in

Copyright 2004, Elsevier (USA). All rights reserved.

Gryboski, J. D. (1993). Ulcerative colitis in children 10 years old or younger. *J. Pediatr. Gastroenterol. Nutr.* **17**, 24–31.
Hildebrand, H., Fredrikzon, B., Holmquist, L., Kristiansson, B., and Lindquist, B. (1991). Chronic inflammatory bowel disease in children and adolescents in Sweden. *J. Pediatr. Gastroenterol. Nutr.* **13**, 293–297.
Hyams, J., Davis, P., Lerer, T., Colletti, R., Bousvaros, A., Leichtner, A., Benkov, K., Justinich, C., and Markowitz, J. (1997). Clinical outcome of ulcerative proctitis in children. *J. Pediatr. Gastroenterol. Nutr.* **25**, 149–152.
Hyams, J., Markowitz, J., Treem, W., Davis, P., Grancher, K., and Daum, F. (1995). Characterization of hepatic abnormalities in children with inflammatory bowel disease. *Inflamm. Bowel Dis.* **1**, 27–33.
Hyams, J. S., Davis, P., Grancher, K., Lerer, T., Justinich, C. J., and Markowitz, J. (1996). Clinical outcome of ulcerative colitis in children. *J. Pediatr.* **129**, 81–88.
Langholz, E., Munkholm, P., Krasilnikoff, P. A., and Binder, V. (1997). Inflammatory bowel diseases with onset in childhood: Clinical features, morbidity, and mortality in a regional cohort. *Scand. J. Gastroenterol.* **32**, 139–147.
Markowitz, J. F. (2002). Medical management of IBD: Children versus adults. *In* "Falk Symposium 122, Inflammatory Bowel Disease: A Clinical Case Approach to Pathophysiology, Diagnosis, and Treatment" (M. Campieri, C. Fiocchi, S. B, Hanauer, D. P. Jewell, D. Rachmilewitz, and J. Scholmerich, eds.), pp. 71–84. Kluwer Academic, Dordrecht.
Ringheanu, M., and Markowitz, J. (2002). Inflammatory bowel disease in children. *Curr. Treat. Options Gastroenterol.* **5**, 181–196.
Sarigol, S., Caulfield, M., Wyllie, R., Alexander, F., Lavery, I., Steffen, R., Kay, M., and Michener, W. (1996). Ileal pouch–anal anastomosis in children with ulcerative colitis. *Inflamm. Bowel Dis.* **2**, 82–87.

Colon, Anatomy

LUCA MAZZUCCHELLI* AND CHRISTOPH MAURER†
**University of Bern, Switzerland and †Kantonspital Liestal, Switzerland*

intestinal barrier The border formed by the intestinal mucosa and in particular the epithelial cell layer between the external world, i.e., the intestinal lumen with its contents and the internal environment.
intestinal intrinsic nervous system A complex system of interconnecting nerve fascicles and ganglia occurring within the intestinal wall and able to function autonomously.

The etiology, pathogenesis, and symptoms caused by several pathologic processes occurring in the large intestine are strictly related to its anatomy. The success of numerous surgical procedures performed on the colon largely depends on profound knowledge of its anatomical relationships with adjacent organs and structures as well as its vascular supply and innervation.

INTRODUCTION

The large intestine, approximately 1.5 m in length, extends from the distal end of the ileum to the anus. It includes the cecum with the appendix vermiformis and the ileocecal valve, the colon, and the rectum with the anus. In sequence, the segments of the colon are designated as the ascending, the transverse, the descending, and the sigmoid colon. Clinicians refer to the large intestine proximal to the middle of the transversum as the right colon and to the distal portions as the left colon. The colon initiates in the right iliac region, arches around the small intestine, and finally forms a sinuous loop in the left iliacal region before entering the lesser pelvis. The shape and course of the large intestine, which can be easily recognized on a barium enema radiograph, are not firmly fixed, but may vary according to the position of adjacent peritoneal and retroperitoneal organs. For instance, the hepatic flexure descends with the liver with each inspiration, the middle portion of the transverse colon, which is normally found in an approximately horizontal position, descends by dilation of the stomach, and the course of the ascending and descending colon may be affected by extension of the coils of the small intestine. The diameter of the large intestine progressively diminishes from the cecum to the sigmoid colon, with the narrowest portion being located in

Copyright 2004, Elsevier (USA). All rights reserved.

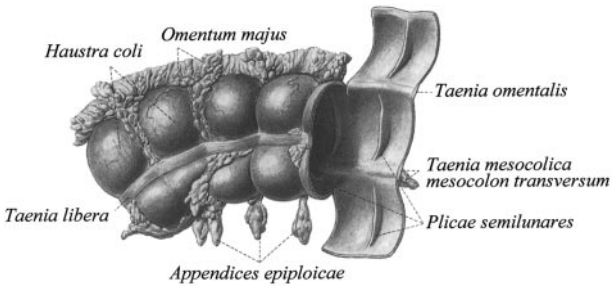


FIGURE 1 Macroscopic appearance of the colon. Reprinted from Benninghoff, A., and Goerttler, K., "Lehrbuch der Anatomie des Menschen," Vol. 2, 12th Ed., Urban & Schwarzenberg, 1979, with permission. Copyright Urban & Fischer Verlag.

the colon sigmoideum. Conversely, the thickness of the muscle layers of the colonic wall gradually increases from the cecum to the sigmoid colon. The characteristic macroscopic appearance of the colon is given by the three separate longitudinal "teniae coli," which are formed by a concentration of fascicles of the outer muscle layer, by "haustra," i.e., protrusions of the bowel wall between the teniae and the plicae semilunares, and finally, by appendages of adipose tissue, the appendices epiploicae, scattered on the serosal surface of the colon but absent in the cecum and rectum (Fig. 1).

VERMIFORM APPENDIX

The vermiform appendix represents embryologically an extension of the cecum. It consists of a blind tube with a diameter usually measuring 0.5 to 0.8 cm and with lengths ranging from 2.5 cm to more than 20 cm (average $\sim$8 cm). It arises from the posteromedial wall of the cecum approximately 1 to 3 cm below the ileocecal junction, where the three teniae coli converge and merge into the longitudinal muscle layer of the appendix. The strong circular and longitudinal muscle layers of the vermiform appendix prevent any substantial dilation of the lumen. Although the location of the base of the appendix, with respect to the cecum, is constant, the position of the tip is extremely variable. Mainly, the vermiform appendix is found in a retrocecal and retrocolic (posterior) or in a pelvic descending (anterior) position. Other positions, such as in front of and behind the terminal ileum, or descending to the right, occur more rarely. The mesentery of the appendix consists of a peritoneal fold contiguous with the mesentery of the terminal ileum. It extends along the entire appendix length and contains the appendical vascular supply. The functions of the vermiform appendix remain a matter of debate. The presence of numerous and hyperplastic lymphoid follicles in young individuals suggests specialized functions in the context of the mucosal immune system.

CECUM AND ILEOCECAL VALVE

The cecum is generally located in the right iliacal fossa. Commonly, it is entirely covered by peritoneum and in only a small percentage of individuals is it fixed posteriorly. Grossly, the cecum is wider than it is long and, due to large haustrae located medially to the tenia libera, an asymmetric shape characterizes it. The lack of a true mesentery accounts for its increased motility, which eventually can lead to cecal volvulus or even to cecal herniation into the groin channel. The large diameter of the cecum and the characteristic thin muscular layer in this portion of the large intestine also have important clinical implications. In fact, cecal cancers may reach a considerable size before causing symptoms, and obstructing processes of the left colon or rectum can cause perforation of the cecum, the site of lowest resistance within the colon.

The ileocecal valve is characterized by two rather horizontally oriented semilunar lips protruding into the posterior medial aspect of the cecum. At their end, the two lips merge into the frenula, which extends into transverse mucosal folds dividing the cecum from the ascendens colon. Whether the ileocecal junction is a competent valve is still uncertain. In normal individuals, reflux into the terminal ileum has been observed, especially when the cecum is empty. However, by increasing intralumenal pressure in the proximal colon, for instance, by obstruction beyond the cecum, the two semilunar lips forming the ileocecal valve tend to collapse together, preventing reflux. Conversely, although the ileocecal valve is formed by an extension of the enteric musculature and not by a true sphincter, it is usually assumed that this well-innervated enterocolic junction is also important in regulating the passage of ileal contents into the cecum.

COLON

The ascending colon is mainly involved in absorptive processes that eventually lead to feces formation. The ventral aspect is covered by the peritoneum, whereas the posterior surface is retroperitoneally fixed to the iliac fascia. A poorly formed mesocolon can occasionally be observed. The anatomic relationships of the ascending colon include the right kidney and ureter, the duodenum, the liver, and the gallbladder. The right colic flexure, also called the hepatic flexure, consists of a curve at the junction with the transverse colon. A peritoneal fold

extending from the hepatorenal ligament most likely supports this portion of the large intestine. The transverse colon is entirely intraperitoneal with the exception of its right end, which is in close contact with the duodenum and the pancreas. It courses diagonally across the upper abdomen from the ventrally located hepatic flexure to the left colic (splenic) flexure, which is located in the posterior plane of the abdominal cavity in a more cranial position than its counterpart on the right (Fig. 2). The left flexure forms an acute angle fixed to the diaphragm by the phrenicocolic ligament, which also sustains the lower pole of the spleen. The position of the middle part of the transverse colon is variable. In asthenic individuals, it may reach the pelvis. The descending colon first passes the lateral border of the left kidney and then curves medially, reaching the pelvic rim. The anterior surface is covered by peritoneum, whereas the posterior face lies retroperitoneally or is occasionally sustained by a mesentery. The sigmoid colon begins at the pelvic entrance. It is characterized by an S shape of variable length, a short mesentery, and therefore good mobility. The sigmoid colon may be in close contact with the urinary bladder, the ventral surface of the upper rectum, or, in females, the uterus and annexes. The junction between the sigmoid colon and the rectum is located 15 cm from the anal verge. It is characterized by fusion of the teniae into the complete longitudinal muscle layer of the rectum and by an abrupt loss of appendices epiploicae.

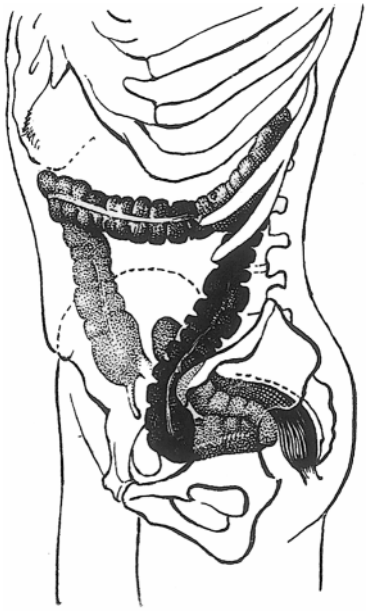


FIGURE 2 Course of the large intestine as seen from the side. Reprinted from Benninghoff, A., and Goerttler, K., "Lehrbuch der Anatomie des Menschen," Vol. 2, 12th Ed., Urban & Schwarzenberg, 1979, with permission. Copyright Urban & Fischer Verlag.

HISTOLOGY

The mucosa, the submcosa, the muscle coat, and the serosa constitute the wall of the large intestine. The mucosa has no villi and is characterized by crypts arranged in parallel. It consists of a single-cell layer of tall columnar epithelium covering the surface including the crypts, smooth muscle cells of the muscularis mucosae at the base, and, in between, a stromal compartment called the lamina propria. The columnar epithelium is supported by a thin collagenous basement membrane and represents an important border between the lumenal environment of the intestine and the inner environment. It is formed by highly specialized polarized absorptive cells located mainly at the surface and at the neck of the crypts, and by goblet cells, which secrete mucus granules and are particularly abundant in the crypts. Immature precursor cells with marked proliferative activity as well as specialized endocrine cells with multiple and still not fully understood regulatory functions are found toward the crypt base. As in the small intestine, Paneth cells are regularly found in the mucosa of the cecum and ascending colon. Numerous T lymphocytes are normally present within the colonic epithelium. Occasionally, eosinophilic granulocytes may also be observed within the epithelial layer. In contrast, neutrophilic granulocytes migrate into and through the epithelial cell layer only during inflammatory processes. The lamina propria contains a heterogeneous cell population consisting mainly of leukocytes, such as plasma cells, T and B lymphocytes, macrophages, and occasionally eosinophilic granulocytes or mast cells. In addition, scattered smooth muscle cells, blood capillaries, lymphatic vessels, and nerve twiglets, but not nerve ganglia, are arranged in its loose connective tissue. Lymphoid follicles are scattered throughout the entire mucosa of the large intestine but appear to be slightly more predominant in the cecum and rectum. Morphologically, two types may be identified: one type displays a pit-like follicle-associated epithelium. Many of these lymphoid follicles extend into the submucosa. The second type is entirely intramucosal, has a flat follicle-associated epithelium, and is observed mainly in the rectum.

The submucosa of the large intestine is very wide, constituting up to approximately half of the total wall thickness. It consists mainly of loosely organized strands of collagen and elastic fibers, vascular structures including arterioles, venules, and lymphatic vessels, few leukocytes, and two neural plexuses. The plexus of Meissner is located beneath the muscularis mucosae where nerve cells and glia cells may be grouped in intramural ganglia arranged in an irregular pattern, and with declining density along the distal segments. The second submucosal

plexus, Henle' s plexus, lies adjacent to the muscularis propria. Both plexuses are connected by interganglionic fascicles. A circular inner layer and a longitudinal outer muscle layer form the muscle coat of the colon. The plexus of Auerbach, also called the plexus myentericus, lies between these two muscle layers. It is morphologically related to the submucosal plexuses, but the ganglia are generally larger and they are arranged in a more regular pattern. The interstitial cells of Cajal form a complex cell network within the gastrointestinal tract wall. They are intercalated to the autonomic nerves of the plexus myentericus and to the smooth muscle cells of the inner portion of the circular muscle layer. These cells may act as a pacemaker to regulate smooth muscle contraction. Recently, Cajal cells have been implicated as possible cells of origin of human gastrointestinal tumors. The mesothelial cells of the serosa cover the external colon surface. A thin layer of loose connective tissue, commonly called the subserosa, separates the serosa from the external muscle layer.

VASCULAR SUPPLY AND LYMPHATIC DRAINAGE

The superior and the inferior mesenteric arteries provide the arterial blood supply of the colon (Fig. 3). The superior mesenteric artery branches into the ileocolic, the right colic, and the middle colic arteries. They supply the terminal portion of the ileum, the vermiform appendix, the colon ascendens, and part of the transverse colon. It is noteworthy that the right colic artery is missing in approximately 10% of individuals and that an accessory middle colic artery is present in approximately 10–20%. The inferior mesenteric artery branches in the superior left colic, the sigmoid, and the superior rectal arteries. All these major branches anastomose to form a series of arterial arches constituting the marginal artery of Drummond, which lies a few centimeters from the colonic wall and extends from the ileocecal region to the rectosigmoid junction. The connection between the middle colic artery and the left colic artery is called the arc of Riolan. It is located approximately at the junction between the transverse and the descending colon and connects the domains of the superior and inferior mesenteric artery. The arc of Riolan is absent in 0.1–1% of individuals. From the marginal artery, multiple small arteries, the arteriae rectae, enter the colonic wall along the line of the mesenteric insertion, where they branch to arterioles that extend symmetrically to the anti-mesenteric side and penetrate the muscular coat to reach the mucosa. The venous return reflects the arterial supply and thus bears the same names. The superior mesenteric and the splenic vein converge to the portal vein. The inferior mesenteric vein leads mostly into the splenic vein, rarely directly into the portal vein, and extremely rarely into the superior mesenteric vein. A peculiarity of the portal venous system is the absence of valves, which may lead to severe intestinal blood congestion by portal hypertension.

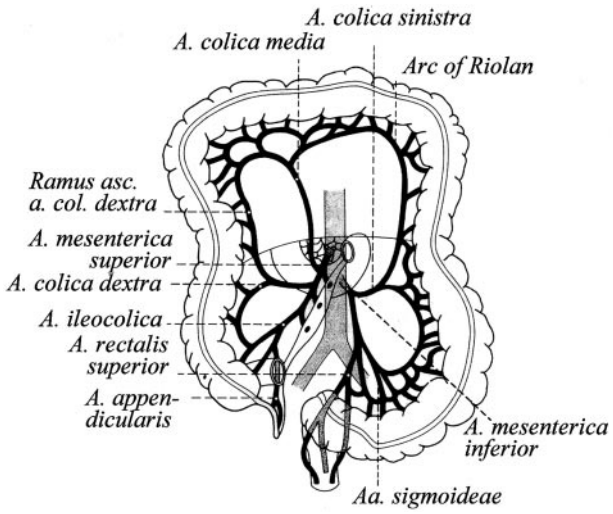


FIGURE 3 Arterial supply of the large intestine. Reprinted from Benninghoff, A., and Goerttler, K., "Lehrbuch der Anatomie des Menschen," Vol. 2, 12th Ed., Urban & Schwarzenberg, 1979, with permission. Copyright Urban & Fischer Verlag.

The lymphatic drainage follows in general the mesenteric vessels and converges to the cisterna chyli. The lymph channels pass several lymph nodes, which may be grossly gathered in four groups: the epicolic lymph nodes adjacent to the colonic wall; the paracolic lymph nodes, situated along the marginal arteries; the intermediate lymph nodes, alongside the main mesenteric vessels; and the central lymph nodes, at the mesenteric roots.

INNERVATION

The motor and secretory activities of the colon are regulated by the autonomic nervous system and by the neurohormonal system, which includes a complex network of modulators, such as serotonin, dopamine, somatostatin, bombesin, Substance P, and vasoactive intestinal peptides. The autonomic nervous system includes the sympathetic and parasympathetic pathways and it may be divided into components intrinsic or extrinsic to the colon, i.e., lying within or outside the colon wall, respectively. The intrinsic component is formed mainly by the neural plexuses described

above. It has a great functional significance since it is capable of working independently of the extrinsic component.

The extrinsic component of the parasympathetic innervation is formed by fibers originating either from the medulla oblongata or from the sacral segments of the spinal cord. The fibers from the medulla oblongata are gathered mainly in the vagus nerve, which descends along the esophagus, ramifies in the preaortal celiac regions, and extends approximately to the left third of the transverse colon. Synapses are formed in the ganglia of the submucosal plexus and the myenteric plexus. The sacral parasympathetic nerves form the pelvic nerves, which are located within the hypogastric plexus and contribute to the innervation of the distal colon segments. Conversely, the afferent sympathetic innervation originates from thoracic and lumbar segments of the spinal cord as part of the anterior spinal nerves. It then transverse the sympathetic ganglionic chain, mostly without making synapses, and forms the splanchnic nerves, which meet synapses located in preaortic ganglia, namely, the celiac, superior mesenteric, inferior mesenteric, and pelvic ganglia. The postganglionic adrenergic fibers eventually reach the colon along the adventitia of the arteries. Connections between the preaortic ganglia are numerous and as a consequence the respective innervation zones of the large intestine overlap to a considerable degree.

See Also the Following Articles

Anal Canal • Autonomic Innervation • Gastrointestinal Matrix, Organization and Significance • Gastrointestinal Tract Anatomy, Overview • Interstitial Cells of Cajal • Parasympathetic Innervation • Rectum, Anatomy • Sympathetic Innervation

Further Reading

Christensen, J. (1991). Gross and microscopic anatomy of the large intestine. *In* "The Large Intestine" (S. F. Phillips, J. H. Pemberton, and R. G. Shorter, eds.), pp. 13–35. Raven Press, New York.

Fenoglio-Preiser, C., Noffsinger, A. E., Stemmermann, G. N., Lantz, P. E., Listrom, M. B., and Rilke, F. O. (eds.) (1999). The normal anatomy of the colon. *In* "Gastrointestinal Pathology: An Atlas and Text," 2nd Ed., pp. 747–761. Lippincott-Raven, Philadelphia/New York.

Haubrich, W. S. (1995). Anatomy of the colon. *In* "Gastroenterology" (W. S. Haubrich, F. Schaffner, and J. E. Berk, eds.), 5th Ed., Vol. 2, pp. 1573–1591. W. B. Saunders Co., Philadelphia.

Levine, D. S., and Haggitt, R. C. (1992). Colon. *In* "Histology for Pathologists" (S. S. Sternberg, ed.), pp. 573–605. Raven Press, New York.

Williams, P. L., Warwick, R., Dyson, M., and Bannister, L. H. (eds.). (1989). The large intestine. *In* "Gray's Anatomy," 37th Ed., pp. 1365–1369. Churchill Livingstone, London.

Colonic Absorption and Secretion

ROGER T. WORRELL, JOHN CUPPOLETTI, AND JEFFREY B. MATTHEWS
University of Cincinnati

absorption Active or passive movement of a substance from the lumen of the gut (mucosal side) to the blood side of the gut (serosal).
apical membrane Side of an epithelial cell that faces toward the lumen, outside of the body; the mucosal side.
basolateral membrane Side of an epithelial cell that faces toward the blood or inside of the body; the serosal side.
conductance Electrogenic movement of a charged particle.
constipation Condition in which there is infrequent or difficult defecation or hard and dry stool.
cotransporter Membrane protein with which one or more substances are moved in concert. Typically, the energy gradient of one substance (almost always Na^+) is used to establish a gradient for the other(s).
crypts of Lieberkuhn Invaginations in the surface of the colonic mucosa, predominant location of fluid secretory processes. The base of the crypt contains the progenitor cells for the entire crypt as well as surface cells.
diarrhea Condition in which there is a two- to threefold or greater increase in the normal amount of fecal water content (approximately 300–450+ ml/day).
ion channel Membrane protein that forms a selective and regulated pore for the diffusion of a particular ion or ions. The direction of movement is dependent on the electrochemical driving force for the given ion, which is determined by the concentration of ion on each side of the membrane and the electrical potential across the membrane.
ion exchangers Subclass of cotransporters that move substances in opposite directions; antiporters.
ion pump Membrane protein in which ATP hydrolysis is used as an energy source to move one or more ions across the cell membrane.
secretion Active or passive movement of a substance from the blood side of the gut (serosal) to the lumen of the gut (mucosal side).
tight junction Complex of proteins near the apical side of epithelial cells; acts as a barrier to the movement of molecules between the cells.

Fluid and electrolytes are secreted as well as absorbed by the colonic epithelium. The net transport, in the normal state, is absorption of sodium, chloride, water, ammonium ions, and short-chain fatty acids, and net secretion of potassium, bicarbonate, and mucus. Similar to fluid processing in the kidney, both the secretory and the absorptive processes can be fine-tuned by selective regulation, such that homeostasis is maintained. Abnormalities in individual transport processes or abnormalities in their regulation can lead to excessive fluid loss (diarrhea) or excessive fluid retention (constipation), both of which can cause substantial symptomology and, if extreme, can be life threatening.

INTRODUCTION

For the average individual, the colon receives approximately 1.5 liters of fluid/solid load per day. On average, approximately 150 ml of water and 100–200 g of solid waste are excreted from the colon. The balance between these amounts constitutes the colonic processing of luminal contents. The colonic epithelium is capable of both secretion and absorption; the difference between these processes determines the final ionic and water composition of the fecal material. Under normal conditions, the integrated outcome of bidirectional fluid movement is absorption. Although the colon normally receives 1.5 liters of input per day, it is capable of processing approximately 2.5 liters per day. Delivery volumes in excess of 2.5 liters per day by the small bowel may exceed the absorptive capacity of the colon and lead to an overflow diarrheal state, despite otherwise normal colonic function. The colonic absorptive processes predominantly involve movement of sodium (Na^+), chloride (Cl^-), and water, such that 80–90% of the ileal fluid is removed from the gut lumen. The ionic composition of the ileal fluid delivered to the colon is essentially isosmotic to plasma, with 140 mEq/liter Na^+, 60 mEq/liter Cl^-, 70 mEq/liter HCO_3^-, and 7 mEq/liter K^+. The fecal ionic composition is reduced in Na^+ and Cl^- and enriched in K^+ and HCO_3^-. In addition to the net absorption of Na^+ and Cl^- and the net secretion of K^+ and HCO_3^-, short-chain fatty acids (SCFAs) and NH_4^+, both produced by colonic bacteria within the colon, are major constituents of colonic luminal fluid and are also absorbed. Mucus, secreted by the colonic mucosa, is hydrated and propelled along the crypt by fluid secretion driven by K^+ and/or Cl^- secretion within the crypts. The focus in this article is on the mechanisms whereby the colonic content is

 Copyright 2004, Elsevier (USA). All rights reserved.

modified, the regulation of those mechanisms, and the pathology associated with alterations in these processes.

CRYPT–SURFACE AXIS

The human colon contains millions of crypts of Lieberkuhn (Fig. 1a), with each crypt being 30–60 μm in diameter and 200–400 μm in length, composed of approximately 2000 cells, termed "enterocytes." The epithelial cells of the colon are generated from progenitor cells at the base of the crypts. Each crypt contains at least one stem cell, thus each individual crypt is considered to represent a clonal or quasi-clonal population of cells, based on DNA methylation patterns. Stem cells divide in an asymmetric manner, producing another stem cell and a cell destined for differentiation. The cells destined for differentiation migrate along the crypt axis and eventually form the surface cell layer. During the migratory process, the cells become more differentiated and less proliferative. The fully differentiated surface cells subsequently undergo apoptosis and slough off into the lumen. This process of enterocyte production, migration, and terminal differentiation occurs over a 4- to 8-day period, thus the stem cells of the crypt are highly proliferative and represent a major locale for tumorigenesis.

During the migration of enterocytes from the base of the crypt to the surface, the differentiation process results in functionally distinct cell populations. The three general cell types produced are goblet cells, columnar epithelial cells, and enteroendocrine or enterochromaffin cells. Of these, the latter represent only 5%. The enterochromaffin cells act as mechanical sensors within the epithelium and provide the initial signal for neuronal reflex regulation of transport. Goblet cells are mucus-secreting cells and occur along the length of the crypt as well as on the surface, varying little in function along the crypt–surface axis. The columnar cells are responsible for fluid and electrolyte transport but also secrete a mucus distinct from that produced by goblet cells. Columnar cell function varies greatly along the crypt–surface axis, with the early and midcrypt cells having a predominantly secretory function and the surface cells having a predominantly absorptive function. Columnar cells at the neck of the crypt display a mixed phenotype. Although the secretory function appears uniform along the length of the colon, the absorptive process (the functional fate of terminal differentiation) differs. Absorption in the proximal colon (ascending and transverse segments) is electroneutral, involving apical membrane cotransporters, whereas the process of absorption in the distal colon (descending and sigmoid segments) is electrogenic and involves apical ion channels.

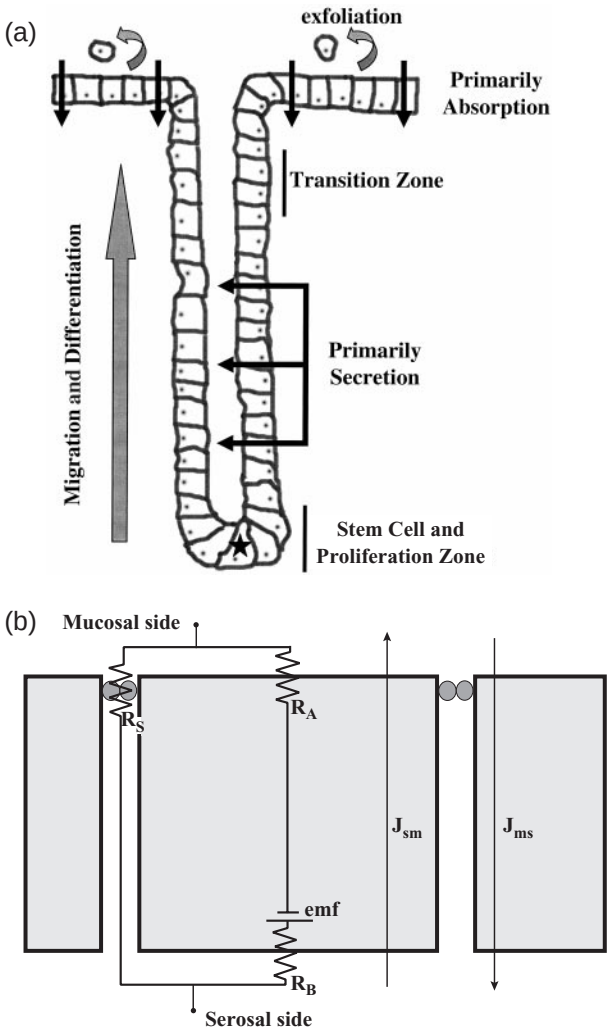


FIGURE 1 General aspects of colonic transport. (a) Crypt of Lieberkuhn functional model. Stem cells (★) at the base of each crypt proliferate and cells differentiate as they migrate toward the lumen surface. Surface epithelial cells undergo apoptosis and slough off. Cells in the lower crypt primarily secrete whereas surface cells are primarily absorptive. (b) Minimal equivalent circuit model for epithelial transport: R_S, the barrier to paracellular ion movement (tight junctions, ◯◯); R_A, the cell apical membrane barrier; R_B, the cell basolateral membrane barrier; emf, the electromotive force (energy source, Na^+,K^+-ATPase). Net transepithelial transport, or flux (J_{Net}), is determined by the difference between the serosal to mucosal flux (J_{sm}) and the mucosal to serosal flux (J_{ms}).

MECHANISMS OF FLUID AND ELECTROLYTE TRANSPORT

The consensus model of fluid and electrolyte transport by colonic epithelium was established in the 1970s and 1980s. The unidirectional movement of ions and thus

water by osmotic flow results from the polarized distribution of specific membrane proteins in epithelial cells and the occurrence of a barrier to flow between the cells. The net movement, or flux, of a substance across an epithelium (Fig. 1b), J_{Net}, is defined as the difference between the rate of absorption (mucosal to serosal flux), J_{ms}, and the rate of secretion (serosal to mucosal flux), J_{sm}. For passive diffusion, J_{Net} of a substance occurs down its electrochemical gradient. When J_{Net} cannot be explained by passive diffusion, an active transport system must be considered. The energy source (electromotive force, emf) for directed ion transport is principally supplied by the Na^+,K^+-ATPase, which pumps three Na^+ ions out of the cell per two K^+ ions into the cell and thus establishes an ion gradient for both K^+ and Na^+ across the cell membrane. The established electrical potential and the Na^+ concentration gradient then provide the energy for active transport of other ions. The epithelium represents two barriers or resistances to flux—the paracellular pathway and the transcellular pathway.

The paracellular space between adjacent epithelial cells is bridged by an intercellular functional complex that consists of three zones, the zonula occludins (tight junction), the zonula adherens (adherens junction), and the desmosomes. The tight junction is the anatomical locus of epithelial barrier function and determines the leakiness of the paracellular pathway. The relative permeability of the tight junction to specific ions, the ion selectivity, is in part determined by the isoform type and state of claudin proteins within the tight junctional complex. This leakiness of the tight junction can be described in electrical terms as the paracellular resistance, R_S (Fig. 1b). The paracellular resistance is higher in the colon than in the jejunum and ileum, increasing along the length of the colon from approximately $100\ \Omega \cdot cm^2$ in the proximal segments to approximately $300-400\ \Omega \cdot cm^2$ in the distal segments. Thus, the ability to establish and maintain transepithelial gradients increases along the length of the colon.

The flux across the epithelial cell is further determined by two components, the apical membrane resistance, R_A, and the basolateral membrane resistance, R_B (Fig. 1b). Thus, the net flux of a substance across the cell is determined by the net flux across the apical membrane, $J_{Net}{}^A$, in combination with the net flux across the basolateral membrane, $J_{Net}{}^B$. In colonic epithelial cells, $J_{Net}{}^A$ is considered to be rate limiting (e.g., $R_A \gg R_B$), although it is now appreciated that changes in $J_{Net}{}^B$ can have a regulatory effect on the magnitude of the cellular flux. Although poorly understood at present, regulatory changes in the paracellular resistance or flux pathway are also thought to occur and likely involve posttranslational modification of claudins, which comprise part of the tight junction complex. Increased paracellular permeability, particularly when ion selective, can enhance the rate of directed transport, or, when nonselective, can decrease the ability of the epithelium to maintain an ionic or osmotic gradient.

The functional phenotype of a transporting epithelial cell is determined by the identity of specific membrane transport pathways in the apical and basolateral membrane. The asymmetric distribution of proteins in the apical versus basolateral membrane produces cell polarity. The direction and magnitude of transport, that is, the conductance of a particular ion or molecule, are set by the relative activity of these apical and basolateral transport processes. Crypt cells are similar along the length of the colon, with the key features being apical K^+ and Cl^- ion channels in combination with basolateral Na^+,K^+-ATPase—the $Na^+/K^+/2Cl^-$ cotransporter 1 (NKCC1)—and K^+ ion channels. The surface cells of the proximal colon contain the apical Na^+/H^+ exchanger (NHE), the Cl^-/HCO_3^- exchanger (anion exchanger), and the Cl^-/OH^- exchanger [originally designated "down-regulated in adenoma" (*DRA*)], in combination with basolateral K^+ and Cl^- ion channels, the K^+/Cl^- cotransporter 1 (KCC1), the anion exchanger (AE), and Na^+,K^+-ATPase. The surface cells of the distal colon differ in that the apical membrane contains K^+ and Na^+ ion channels and two different isoforms of H^+,K^+-ATPase.

Absorption

The majority of NaCl and fluid absorption occurs in the proximal colon via an electroneutral process depicted in Fig. 2a. Na^+ enters across the apical membrane in exchange for H^+ via two isoforms of the Na^+/H^+ exchanger (NHE2 and NHE3), which have distinct regulatory profiles, although the functional significance of their redundancy remains unclear. *DRA* was originally described as a sulfate/oxalate exchanger and was later shown to have Cl^-/OH^- exchange as well as Cl^-/HCO_3^- exchange activity. Although apical AE isoform 1 (AE1) is present in rat colon cells, it has not been identified in human colon cells, thus *DRA* may serve as the apical anion exchanger (Cl^-/HCO_3^-) in humans. Cl^- enters via *DRA* in exchange for OH^- and via the AE and/or *DRA* in exchange for HCO_3^-. The driving force for apical Na^+ entry is maintained by active removal via the basolateral Na^+,K^+-ATPase. Cl^- exits the cell via basolateral KCC1, AE2 and AE3, and Cl^- channels. AE2 and AE3 provide a pathway for the basolateral entry of HCO_3^-. K^+ is recycled across the basolateral membrane by basolateral K^+ channels. The combined result of these transport processes is the net transport of NaCl

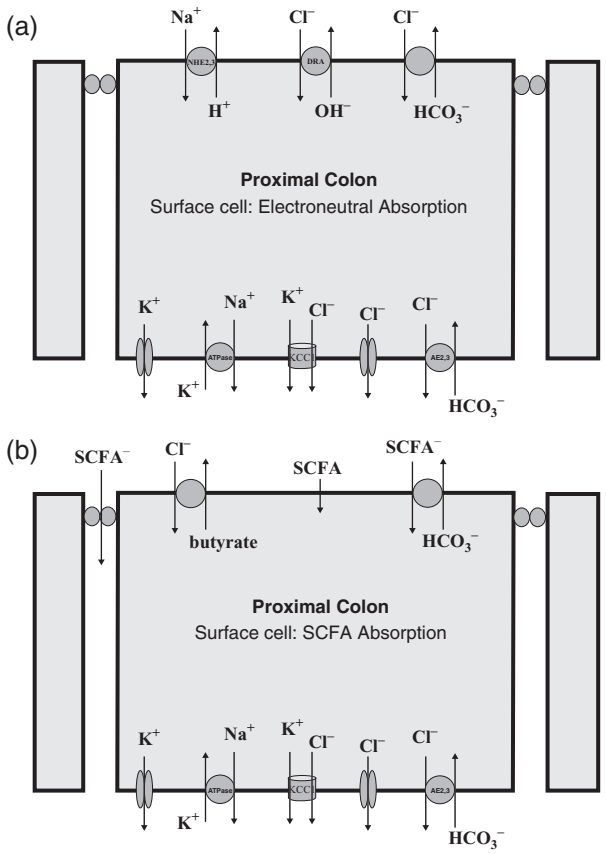


FIGURE 2 Cellular models for surface cell transport in the proximal colon. (a) Electroneutral absorption. Na^+ and Cl^- are taken up across the apical membrane by Na^+/H^+, Cl^-/OH^-, and Cl^-/HCO_3^- exchangers. Intracellular Na^+ is kept low by basolateral Na^+,K^+-ATPase. Intracellular Cl^- is kept low by basolateral K^+ and Cl^- cotransport, Cl^-/HCO_3^- exchange, and Cl^- channels. Note that HCO_3^- secretion occurs in conjunction with NaCl absorption, but does not result in a significant luminal pH change because H^+ is also secreted. NHE2,3, Na^+ and H^+ exchangers, isoforms 2 and 3; AE2,3, anion exchangers, isoforms 2 and 3. (b) Short-chain fatty acid (SCFA) absorption. Uncharged SCFA passively diffuses across the cellular membrane. $SCFA^-$ is absorbed across the apical membrane in exchange for HCO_3^-. Via a butyrate/Cl^- exchanger, Cl^- absorption can occur. $SCFA^-$ may also permeate the paracellular pathway driven by a negative lumen-to-serosa potential. The processes depicted here and in (a) occur within the same cells but are shown separately for clarity.

from lumen to serosa and the net movement of HCO_3^- from serosa to lumen. Activity of NHE2 and NHE3 does not significantly contribute to luminal acidification because the secreted H^+ is buffered by the OH^- and HCO_3^- secreted by the *DRA* and AE.

The surface cells of the proximal colon also have the capacity to absorb short-chain fatty acids (SCFAs) (Fig. 2b). Absorption occurs predominantly by nonionic diffusion; however, the charged forms can also be transported across the apical membrane via a $SCFA^-/HCO_3^-$ exchanger. There also exists an exchange process for butyrate and Cl^- in the apical membrane, thus, via $SCFA^-/HCO_3^-$ exchange coupled to Cl^-/butyrate exchange, net Cl^- absorption and HCO_3^- secretion can occur.

The cell model for electrogenic NaCl absorption occurring in the surface cells of the distal colon is depicted in Fig. 3a. Intracellular Na^+ is kept low by the basolateral Na^+,K^+-ATPase and provides the driving force for the electrogenic diffusion of Na^+ through an apical epithelial Na^+ channel (ENaC). The electrogenic movement of Na^+ from the lumen to the serosal side creates a negative lumen potential, which facilitates the cellular and/or paracellular diffusion of Cl^-. An apical K^+ channel provides a pathway for K^+ secretion. The driving force for K^+ secretion is dependent on the relative permeability of the apical membrane to Cl^-; that is, Na^+ entry can be "charge balanced" by either apical K^+ exit or apical Cl^- entry. At the basolateral membrane, K^+ channels provide a pathway to recycle K^+. The net movement of K^+ also depends on the relative activity of the apical K^+ channel with respect to that of the basolateral K^+ channel and the driving force across each respective membrane. The colon contributes to body K^+ homeostasis and can function both to secrete and to absorb K^+. The cells responsible for K^+ absorption are the same as involved in electrogenic Na^+ absorption. The cellular model for K^+ absorption, for the sake of simplicity, is shown separately in Fig. 3b. Two H^+,K^+-ATPases are responsible for the active exchange of luminal K^+ for intracellular H^+, the colonic H^+,K^+-ATPase (HKc) being sensitive to omeprazole and the other sensitive to ouabain. Accumulated intracellular K^+ exits the cell across the basolateral membrane either by a K^+ channel or by KCC1.

It is well appreciated that the colon absorbs ammonium from the lumen, because portal vein ammonium levels are higher than those of the systemic circulation. However, the exact mechanism of absorption is not understood. It is likely that, similar to SCFA absorption, both the passive diffusion of uncharged ammonia and the directed transport of ammonium ions contribute to the process.

Secretion

Within the crypts, NaCl, KCl, and HCO_3^- secretion can occur. The cellular model for these secretory processes is depicted in Fig. 4. In KCl secretion, both apical membrane Cl^- and K^+ ion channels are active, with a relatively less active basolateral K^+ channel. Intracellular Cl^- is accumulated above electrochemical

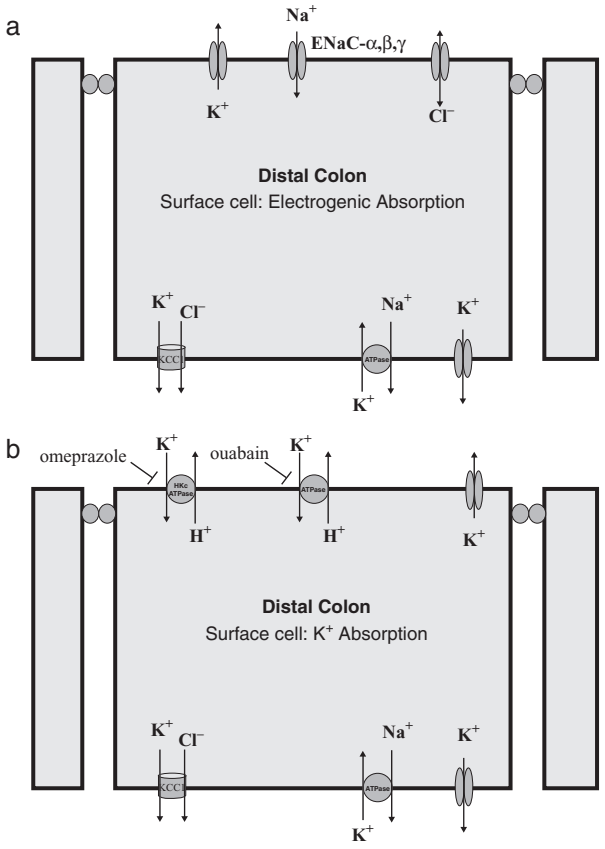


FIGURE 3 Cellular models for surface cell transport in the distal colon. (a) Electrogenic absorption of Na^+. Intracellular Na^+ is kept low by basolateral Na^+,K^+-ATPase. Na^+ absorption occurs via the passive diffusion of Na^+ through an apical α, β, and γ epithelial Na^+ channels (ENaC-α,β,γ). Apical K^+ and Cl^- channels provide a pathway for charge balance across the apical membrane. (b) K^+ absorption. Two distinct isoforms of H^+,K^+-ATPase provide a pathway for apical entry of K^+. The exit of K^+ across the basolateral membrane occurs by K^+,Cl^- cotransport and/or K^+ channels. The processes depicted here and in (a) occur within the same cells but are shown separately for clarity.

equilibrium by basolateral NKCC1. A basolateral Na^+,K^+-ATPase maintains the driving force for NKCC1. In NaCl secretion, the apical K^+ channel activity is relatively lower than that of basolateral K^+ channels such that K^+ preferentially recycles across the basolateral membrane. The electrogenic movement of Cl^- across the apical membrane creates a lumen negative potential, which acts to drive the paracellular movement of Na^+. Numerous apical Cl^- channels have been described, the two most prominent being the Ca^{2+}-activated Cl^- channel (CaCC) and the cAMP-activated Cl^- channel; the latter, designated the cystic fibrosis transmembrane conductance regulator (CFTR), was originally identified by genetic screening as the protein that is defective in cystic fibrosis. Subsequently, by electrophysiological studies, CFTR was identified as being a cAMP-activated Cl^- channel. Electrogenic HCO_3^- secretion can also occur via these apical Cl^- channels.

Mucus secretion by goblet cells serves to protect the epithelium from abrasion and to create a protective unstirred layer near the apical membrane. Goblet cell mucus secretion occurs via exocytosis of apically located mucus-containing granules. Movement of the granules is microtubule dependent and restricted by the actin cytoskeleton at the apical membrane. On stimulation of goblet cells by Ca^{2+} agonists such as acetylcholine (ACh) and histamine, but not by cAMP agonists, the granule contents are released into the lumen of the crypt or on the surface lumen. Columnar cells in the crypt can also secrete mucus contained within apical vesicles. In contrast to goblet cells, columnar cells secret mucus in response to cAMP agonists but not to Ca^{2+} agonists. Mucus within the granules is compact and in a somewhat dehydrated state; on exocytosis, hydration of the mucus results in an approximately twofold expansion in volume. Crypt fluid secretion is necessary both to hydrate the luminal mucus and to provide a flow for expulsion of the mucus from the crypt lumen.

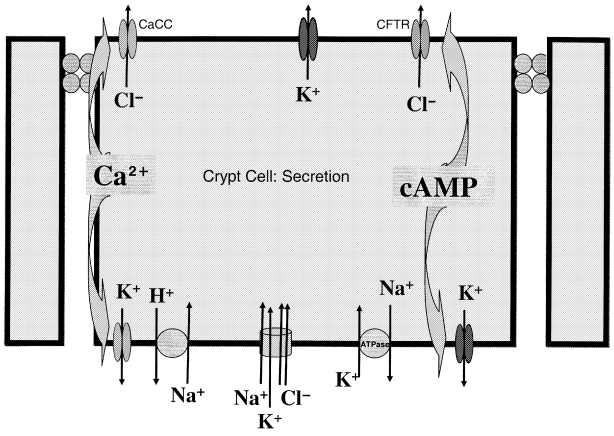


FIGURE 4 Cellular model for crypt cell secretion. Apical Cl^- channels provide an exit pathway for Cl^- across the apical membrane. Intracellular Cl^- is maintained by functional basolateral $Na^+/K^+/2Cl^-$ cotransport. Agonists that increase intracellular Ca^{2+} increase the activity of the apical Ca^{2+}-activated Cl^-channel (CaCC) and a Ca^{2+}-sensitive basolateral K^+ channel. Agonists that increase intracellular cAMP increase the activity of the apical cystic fibrosis transmembrane conductance regulator (CFTR) and, depending on the level of cAMP, either an apical K^+ channel or a basolateral K^+ channel. The directional flow of K^+ (e.g., secretion or basolateral recycling) is determined by the relative activity of the apical and basolateral conductive pathways. The cAMP agonists also up-regulate NKCC1 in the basolateral membrane.

REGULATION OF ABSORPTION

The normal default function of the colon is absorption, thus effects on net absorption are more typically associated with an inhibition of secretion. Electrogenic absorption in the distal colon is under more significant regulation than is the electroneutral absorption in the proximal colon. Nonetheless, electroneutral absorption is affected by a number of extracellular factors. NHE2 and NHE3 expression is up-regulated with Na^+ depletion and, in a nongenomic manner, with low-dose glucocorticoids and aldosterone. Stimulation of $\alpha 1$ or $\beta 2$ receptors increases NHE3 activity and thus increases absorption. Agonists that stimulate protein kinase C (PKC), calmodulin-dependent kinase II (CaMKII), or cyclic adenosine monophosphate (cAMP) inhibit NHE3 function; interestingly these second-messenger systems are also strong activators of secretion. Norepinephrine, somatostatin, and neuropeptide Y (NPY) are able to increase electroneutral absorption; however, their major influence on net transport is via inhibition of secretion. Electrogenic Na^+ absorption in the distal colon is strongly influenced by circulating glucocorticoids and mineralocorticoids. Both exert genomic early- and late-phase stimulation of apical Na^+ entry by effects on the apical Na^+ channel, ENaC. The early-phase response, which requires protein expression, is the up-regulation of ENaC activity in the apical membrane, whereas in the late phase, ENaC protein expression is increased along with an increase in Na^+,K^+-ATPase. Increases in intracellular cAMP increase ENaC activity, whereas increases in intracellular Ca^{2+} or PKC activity tend to diminish activity. ENaC is also subject to feedback inhibition by increases in intracellular Na^+. A number of studies have demonstrated that the expression of CFTR and CFTR activity has a negative effect on ENaC activity. Although CFTR abundance is low in the surface cells, it may play a significant role in the regulation of ENaC as well as in pathological conditions that increase CFTR activity, or in cystic fibrosis, in which ENaC-mediated Na^+ absorption is enhanced. Dietary Na^+ and K^+ depletion as well as aldosterone cause increased K^+ absorption by increasing the expression and activity of the H^+,K^+-ATPase.

REGULATION OF SECRETION

Homeostatic regulation of colonic function is primarily directed at inhibition of the absorptive pathway or stimulation of the secretory pathway. Regulation of secretion occurs by paracrine, autocrine, endocrine, and neuroendocrine mechanisms. Although a plethora of agents have been demonstrated to affect Cl^- and K^+ secretion, the majority of these do so by effects on intracellular cAMP, cGMP, Ca^{2+}, or PKC. Discussion of all the extracellular and intracellular regulatory pathways involved in the regulation of secretion is beyond the scope of this article, but the two most significant physiological agonists for secretion are acetylcholine (ACh) and vasoactive intestinal polypeptide (VIP), which cause increases in intracellular Ca^{2+} and cAMP, respectively. Both ACh and VIP are locally released by intrinsic primary afferents within the submucosal plexus in response to mechanical stimulus of enterochromaffin cells and subsequent 5-hydroxytryptamine (5-HT) release. Whereas the secretory response to cAMP agonists is long term, the secretory response to Ca^{2+} agonists is more transient.

A number of peptide and nonpeptide factors act on the secretory cells to produce an increase in intracellular cAMP, including VIP, adenosine, and prostaglandins. Increases in intracellular cAMP lead to the activation of protein kinase A (PKA), which in turn leads to the activation of the apical membrane Cl^- channel CFTR, activation of NKCC1, and activation of either or both apical and basolateral K^+ channels. Low levels of cAMP tend preferentially to activate the apical K^+ conductance and support KCl secretion, whereas high levels preferentially activate the basolateral K^+ conductance, and thus support NaCl secretion. Guanylin acting via apical receptors leads to an increase in intracellular cGMP, which activates CFTR via protein kinase G (PKG) and increases cAMP levels by inhibiting phosphodiesterase (PDE) breakdown of cAMP. cAMP may also, acutely increase the number of CFTR proteins in the apical membrane and thus further promote secretion.

ACh and histamine stimulate secretion by causing an increase in intracellular Ca^{2+} levels. For ACh and histamine, Ca^{2+} is increased in an inositol 1,4,5-trisphosphate (IP_3)-dependent manner. The increased Ca^{2+} leads to activation of apically located Ca^{2+}-dependent Cl^- channels and the stimulation of Ca^{2+}-sensitive basolateral K^+ conductance. The transient nature of this response has been attributed to the generation of inositol 3,4,5,6-tetrakisphosphate (IP_4) as well as phosphatidylinositol 3-kinase (PI 3-kinase)-dependent activation of PKC-ϵ and their respective negative effect on CaCC. Growth factors—epidermal growth factor (EGF), insulin, and insulin-like growth factor (IGF)—also exert a negative effect of Ca^{2+}-dependent Cl^- secretion by inhibition of the basolateral Ca^{2+}-dependent K^+ conductance. Enkephalins, neuropeptide Y, and somatostatin act to inhibit secretion by inhibition of ACh release; these substances also increase

Na^+ reabsorption. Bile acids, in part, stimulate secretion by an IP_3-dependent increase in intracellular Ca^{2+} and by increasing prostaglandin E_2 (PGE_2) release from colonic fibroblasts.

PATHOLOGICAL INFLUENCES ON SECRETION AND ABSORPTION

Dysregulation of colonic secretion and absorption underlies a number of digestive disorders. These conditions may arise from congenital or acquired defects. In addition, the presence of unabsorbed osmolytes in the lumen of the colon can result in an osmotic diarrhea. These osmolytes may either be voluntarily ingested or result from small intestinal malabsorption.

Congenital chloride diarrhea, or congenital chloridorrhea, is the result of a mutation in *DRA*, a condition known as Pendred syndrome that also results in deafness. Although *DRA* acts as a Cl^-/OH^- exchanger, there is some evidence that it can also effect Cl^-/HCO_3^- exchange. In addition, defective *DRA* leads to an increased intracellular pH, which in turn reduces the activity of apical anion exchange. Congenital chloridorrhea is characterized by an acidic watery diarrhea with chloride concentrations in excess of 90 mEq/liter (due to decreased Cl^- absorption) and metabolic alkalosis (due to decreased HCO_3^- and OH^- secretion). Defects in NHE3 cause congenital sodium secretory diarrhea, in which stool sodium and HCO_3^- are elevated, with only slight diarrhea and metabolic acidosis. Cystic fibrosis (CF) is the result of a defect in the cAMP-stimulated Cl^- channel, CFTR. CF results in a decrease in colonic secretion or an antidiarrheal state, which in 5–10% of CF newborns results in colonic obstruction (meconium ileus). Diarrhea is seen in some patients with CF due to steatorrhea caused by pancreatic exocrine insufficiency. Interestingly, the high prevalence CF (1 in 2500 births) may relate to a heterozygote protective effect against the development of acquired secretory diarrheas, particularly in cholera.

The major cause of acquired secretory and inflammatory diarrheas is pathogenic colonization within the gastrointestinal (GI) tract. Bacterial species that commonly affect colonic transport are *Vibrio cholerae, Escherichia coli, Clostridium difficile*, and *Salmonella typhimurium*. In children, rotavirus is a common cause of secretory diarrhea. Toxins produced by these organisms produce diarrhea (1) by increasing the chloride secretory pathway, (2) by indirectly or directly decreasing electrogenic Na^+ absorption, (3) by decreasing the tight junction permeability, thus making the epithelia more leaky, and (4) by initiating an inflammatory response, which in turn can activate the secretory pathway via the enteric nervous system. It should be noted that these toxins also increase fluid secretion in the small bowel, which can in and of itself produce an overflow diarrhea due to the increased fluid volume delivery to the colon.

Vibrio cholerae produces a number of toxins. Cholera toxin causes a covalent and irreversible activation of the G protein, G_s, which in turn activates adenylate cyclase and raises intracellular cAMP levels, activating CFTR. Zonula occludens toxin (ZOT) acts to increase the ionic permeability of the tight junctions, whereas other toxins result in the release of serotonin from enterochromaffin cells. *Escherichia coli* produces a heat-labile toxin (ST_b) that acts to increase intracellular cAMP and a heat-stable toxin (ST_a) that acts to reversibly increase intracellular cGMP, resulting in stimulation of CFTR. ST_a increases cGMP by hijacking the endogenous guanylin receptor, which, when active, acts as a guanylate cyclase. With cholera toxin and ST, activation of CFTR results in increased secretion of Cl^- as well as HCO_3^- via CFTR and the subsequent development of metabolic acidosis accompanying the diarrhea. *Clostridium difficile* produces two toxins, A and B, that increase tight junction permeability via glucosylation of Rho protein. Toxin A has also been implicated to increase intracellular Ca^{2+} and activate CaCC. Rotavirus and *C. difficile* act similarly, rotavirus producing a nonstructural glycoprotein, NSP4, which acts both to increase intracellular Ca^{2+} and thus stimulate CaCC as well as to decrease tight junction permeability.

The chronic diarrhea observed in colonic carcinoma is likely due to increased proliferation and decreased differentiation within the crypt cells. As mentioned earlier, a protein (*DRA*) found to be down-regulated in carcinoma is expressed predominantly on the fully differentiated surface cells, with little expression within the undifferentiated crypt cells. Thus, the loss of *DRA* in colon carcinoma is indicative of decreased colonocyte differentiation, which suggests an increase in the secretory cell population at the expense of a decrease in the absorptive cell population. A number of neoplastic disorders can lead to humoral-mediated chronic diarrhea. The VIP-secreting tumors (VIPomas), due to high plasma levels of VIP, present with profuse chronic watery diarrhea associated with hypokalemia and achlorhydria (WDHA syndrome).

SUMMARY

The overall state of colonic transport determines the water and electrolyte content of the feces and is the

integrated result of both absorptive and secretory processes. The normal balance is such that the net effect is absorption of salt and water and thus conservation of water. The absorptive process predominantly occurs at the surface cells and is electroneutral in the proximal colon and electrogenic in the distal colon, with the latter being more tightly regulated. The secretory process predominantly occurs in the crypts and is more uniform throughout the length of the colon. In contrast to the absorptive processes, the secretory processes are much more tightly regulated. Pathological disruption of the balance between the absorptive and secretory processes generally involves an increase in secretion and thus diarrhea.

Acknowledgments

The work reported here was supported by NIH grants DK048010 and DK051630.

See Also the Following Articles

Cholera • Constipation • Cystic Fibrosis • Diarrhea • Epithelial Barrier Function • Epithelium, Proliferation of • Small Intestine, Absorption and Secretion

Further Reading

Barrett, K. E., and Keely, S. J. (2000). Chloride secretion by the intestinal epithelium: Molecular basis and regulatory aspects. *Annu. Rev. Physiol.* **62**, 535–572.

Cooke, H. J. (2000). Neurotransmitters in neuronal reflexes regulating intestinal secretion. *Ann. N.Y. Acad. Sci.* **915**, 77–80.

Dawson, D. C. (1991). Ion channels and colonic salt transport. *Annu. Rev. Physiol.* **53**, 321–339.

Field, M., and Semrad, C. E. (1993). Toxigenic diarrheas, congenital diarrheas, and cystic fibrosis: Disorders of intestinal ion transport. *Annu. Rev. Physiol.* **55**, 631–655.

Greger, R. (2000). Role of CFTR in the colon. *Annu. Rev. Physiol.* **62**, 467–491.

Haas, M., and Forbush III, B. (2000). The Na-K-Cl cotransporter of secretory epithelia. *Annu. Rev. Physiol.* **62**, 515–534.

Halm, D. R., and Frizzell, R. A. (1991). Ion transport across the large intestine. *In* "Handbook of Physiology. The Gastrointestinal System. Intestinal Absorption and Secretion," Sect. 6, Vol. IV, pp. 257–274. American Physiology Society, Bethesda, MD.

Halm, D. R., and Halm, S. T. (2000). Secretagogue response of goblet cells and columnar cells in human colonic crypts. *Am. J. Physiol. Cell Physiol.* **278**, C212–C233.

Kim, K.-M., and Shibata, D. (2002). Methylation reveals a niche: Stem cell succession in human colonic crypts. *Oncogene* **21**, 5441–5449.

Kunzelmann, K., and Mall, M. (2002). Electrolyte transport in the mammalian colon: Mechanisms and implications for disease. *Physiol. Rev.* **82**, 245–289.

Masyuk, A. I., Marinelli, R. A., and LaRusso, N. F. (2002). Water transport by epithelia of the digestive tract. *Gastroenterology* **122**, 545–562.

Rajendran, V. M., and Binder, H. J. (2000). Characterization and molecular localization of anion transporters in colonic epithelial cells. *Ann. N.Y. Acad. Sci.* **915**, 15–29.

Stephen, J. (2001). Pathogenesis of infectious diarrhea. *Can. J. Gastroenterol.* **15**, 669–683.

Turnberg, L. A. (1970). Electrolyte absorption from the colon. *Gut* **11**, 1049–1054.

Colonic Ischemia

JÖRN KARHAUSEN* AND CORMAC T. TAYLOR†
*Brigham and Women's Hospital, Boston and †University College, Dublin

hypoxia Condition that arises when the cellular demand for molecular oxygen necessary to maintain physiologic function exceeds supply.
ischemia Pathologic reduction in blood supply to a tissue.

Colonic ischemia is associated with a diverse array of pathologies wherein the epithelial barrier is compromised, including acute mesenteric ischemia, ischemic colitis, necrotizing enterocolitis, and inflammatory bowel disease. Although the underlying disease processes leading to colonic ischemia are numerous, a major consequence for the mucosal tissue is that oxygen demand exceeds supply, resulting in tissue hypoxia. Because the chemical reduction of molecular oxygen is the primary source of metabolic energy for all eukaryotic cells, hypoxia represents a severe threat to tissue function and survival. Hypoxia has an impact on intestinal cell function, which relates to disease processes. This entry reviews the molecular mechanisms underlying such events.

INTRODUCTION

Tissue ischemia and subsequent hypoxia are common physiologic and pathophysiologic occurrences. Because of its juxtaposition with the anoxic lumen of the gut, the colonic mucosa relies heavily on the rich mucosal vascular bed for oxygen supply and is consequently highly susceptible to ischemic/hypoxic insult (Fig. 1). Colonic ischemia can occur as a result of frank vascular occlusion or vasoconstriction, such as in cases of occlusive or nonocclusive acute mesenteric ischemia, respectively. Alternatively, at the tissue level, hypoxia can originate from fibrosis and microvascular breakdown associated with chronic inflammation. Furthermore, oxygen supply can be compromised as a function of a general decrease in perfusion in shock and hypotension, or may be due to whole body hypoxia (e.g., carbon monoxide poisoning, altitude sickness). Although other factors such as acidosis, insufficient tissue detoxification, and hypercapnia play a role in ischemic disease, a primary consequence common to all the pathophysiologic mechanisms is that cellular oxygen demand exceeds supply, leading to tissue hypoxia.

The intestinal epithelium represents an important dynamic barrier, preventing the free mixing of luminal antigenic material with the mucosal immune system, while at the same time facilitating antigen presentation and the development of oral tolerance. The epithelial barrier is maintained through the existence of tight and adherens junctions between cells. These junctions are dynamic structures, the organization and function of which are highly regulated processes. In both health and disease, intercellular tight junction permeability is under the regulation of a number of autocrine and paracrine factors that signal through basolaterally localized receptors. Though the mechanisms involved are still under investigation, it is clear that hypoxia, either directly or through ATP depletion, disrupts assembly of both tight and adherens junctions, crucially impairing barrier integrity. As a consequence, epithelial barrier dysfunction leads to dramatically increased transepithelial paracellular permeability, as has been documented both *in vitro* and *in vivo*. Under such conditions, luminal antigenic material gains unregulated access to the lamina propria, which houses the mucosal immune system. Resulting interactions between luminal antigens and cells of the mucosal immune system lead to the activation and development of inflammatory episodes, which actively contribute to ongoing disease processes.

Because of the challenge that hypoxia represents to physiologic cell function, an exquisite oxygen-sensing mechanism has evolved that allows a rapid, adaptive molecular response to occur. Activation of the adaptive response to hypoxia results in the expression of specific genes encoding proteins that increase tissue blood flow/oxygenation and promote cell survival in oxygen-depleted environments. Temporally downstream of the adaptive phenotype, and through independent signaling mechanisms, hypoxia also initiates an inflammatory response that can lead to epithelial dysfunction. The resulting increase in transepithelial permeability is an event that is of pivotal importance in the propagation of disease states in the large intestine, where inflammation and hypoxia are intimately associated.

 Copyright 2004, Elsevier (USA). All rights reserved.

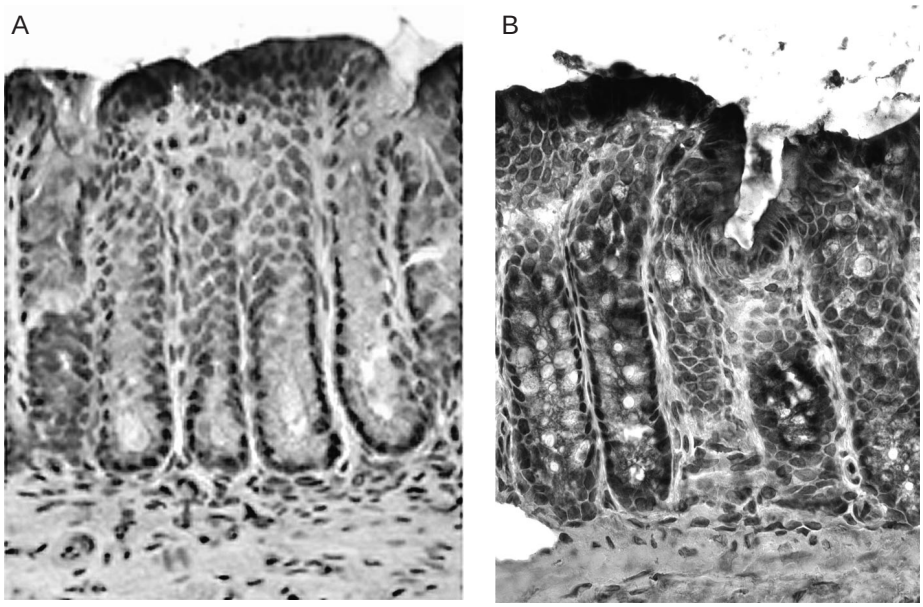


FIGURE 1 Immunhistological localization of retention of EF5 (a pentafluorinated derivative of etanidazole) as a marker of tissue hypoxia. (A) In control mice (room air), colonic sections demonstrate predominantly apical EF5 retention, indicating the presence of significant hypoxia under physiological conditions. (B) In contrast, animals exposed to 8% ambient oxygen for 4 hours display a profound increase in staining of the whole crypt. Under both conditions, staining is almost exclusively mucosal, supporting the concept that the epithelial barrier is specially challenged by oxygen deficiency.

MOLECULAR MECHANISMS IN THE INTESTINAL EPITHELIAL HYPOXIC RESPONSE

Adaptive

Due to the imminent threat that hypoxia represents to cell, tissue, and whole animal survival, mammalian cells have evolved a highly effective mechanism that allows a rapid molecular response through the induction of an adaptive phenotype. This response is mediated primarily by the heterodimeric transcription factor hypoxia inducible factor-1 (HIF-1), a master regulator of oxygen-dependent gene transcription. Whereas the β subunit is constitutively expressed and dimerizes with several other target proteins, HIF-1α is the specific and O_2-regulated subunit that determines biological activity. Under normoxic conditions, HIF-1α has a high rate of turnover, with the protein being targeted for immediate degradation by the ubiquitin/proteasome pathway in a manner dependent on the oxygen-dependent hydroxylation of proline residues. Hypoxia-dependent stabilization of HIF-1α results in its association with HIF-1β and the formation of a transcriptionally active complex that activates the expression of a number of genes with specific binding sites known as hypoxia response elements (HREs) in their promoter or enhancer regions. These genes encode angiogenic factors such as vascular endothelial growth factor (VEGF), vasodilatatory factors such as the inducible form of nitric oxide synthase (iNOS), and other factors such as erythropoietin (EPO), all of which can optimize tissue oxygenation. Furthermore, hypoxia induces the HIF-1-dependent expression of genes encoding glycolytic enzymes that facilitate the generation of cellular ATP through oxygen-independent pathways.

Apart from these general adaptive mechanisms, which are responses to the risk of metabolic demise during decreased oxygen levels, organ-specific mechanisms exist that preserve tissue function and constitute part of the specific hypoxic phenotype of that tissue. A resistance to impairment of barrier function by colonic epithelial cells in response to hypoxia is not found in barrier-forming cells derived from other tissues (endothelia, oral epithelia). Pivotal for this intestinal epithelial hypoxic phenotype is the HIF-1-dependent up-regulation of a number of barrier-protective genes, such as intestinal trefoil factor (ITF) and ecto-5′-nucleotidase (CD73). ITF is a small protease-resistant peptide that is widely expressed in the mammalian intestinal tract, especially in proximity to sites of mucosal injury.

ITF promotes epithelial barrier function, protects against injury, and facilitates repair following damage. One possible mechanism by which this is achieved is via the ITF-regulated expression of E-cadherin and α- and β-catenin, crucial components of the adherens junction. Here, ITF adopts a role as a motogenic factor that facilitates wound healing. ITF also supports barrier integrity, likely through epidermal growth factor (EGF)-dependent and -independent activation of the mitogen-activated protein (MAP) kinase pathway, which leads to reorganization of tight junctional components such as zonula occludens-1 (ZO-1) and claudin-2.

The potential relevance of CD73 in the regulation of barrier function was revealed following observations that neutrophil-derived adenosine monophosphate (AMP) promotes endothelial barrier function through CD73-dependent conversion into adenosine. Though the exact mechanisms remain unclear, the effect is most likely conveyed via an increase of intracellular cAMP that results from adenosine activation of the A_{2b} receptor. Inflammatory cells accumulate at hypoxic/ischemic tissue sites and represent a significant source of extracellular AMP. Furthermore, hypoxic epithelia generate extracellular AMP, allowing it to contribute to this pathway in either an autocrine or paracrine manner.

Both ITF and CD73 are dramatically induced by hypoxia in colonic epithelial cells. Functional HRE consensus sequences that exist within the regulatory domains of these genes mediate transcriptional activation in hypoxia. Furthermore, *in vivo* data clearly support the concept that these two genes constitute part of an innate protective phenotype of intestinal epithelia. Treatment with the specific CD73 inhibitor, α,β-methylene-ADP, leads to augmentation of the intestinal barrier injury induced by hypoxia. Similarly, ITF knockout mice display increased susceptibility in their intestinal permeability response to hypoxia. Thus, intestinal epithelial cells are equipped with adaptive mechanisms that promote cell and tissue survival and enhance epithelial barrier function in hypoxia.

Inflammatory

In a number of pathologies, colonic ischemia is associated with the induction of inappropriate inflammatory episodes that may actively contribute to ongoing disease processes, particularly through the disruption of epithelial barrier function. Exposure of intestinal epithelial cells to hypoxia results in the induction of an inflammatory phenotype, typified by the expression of proinflammatory markers, decreased capacity of the epithelium to restitute following immune cell transmigration, and decreased ion and fluid transport. A primary factor in hypoxia-elicited epithelial cell dysfunction is mediated through the basolaterally polarized release of epithelial tumor necrosis factor α (TNFα), which synergizes with immune-derived cytokines such as interferon γ (IFNγ; present at physiologically high levels in the colonic mucosa) to cause dramatically increased transepithelial permeability to antigenic material and subsequently to promote mucosal inflammation. Inhibition of hypoxia-elicited TNFα using thalidomide or neutralizing antibodies reverses hypoxia-elicited increases in IFNγ-induced permeability. Because TNFα is a major therapeutic target in inflammatory disease of the intestine, the molecular mechanisms underlying hypoxia-elicited TNFα have been investigated. The hypoxia-elicited epithelial expression of TNFα appears to be HIF-1 independent and is dependent instead on the activation of the proinflammatory transcription factor nuclear factor κB (NF-κB) and coincidental degradation of the cyclic AMP response-element binding protein (CREB). NF-κB activation and CREB depletion rely on the phosphorylation-dependent targeting of inhibitor κBα (IκBα) and CREB, respectively, to ubiquitination and subsequent proteasomal degradation. Interestingly, IκBα and CREB share similar phosphorylation sites that act as targeting sequences for ubiquitination and subsequent degradation, implicating a common signaling pathway, leading to hypoxia-elicited activation of these distinct pathways. The net result of NF-κB activation and coincidental CREB degradation is the activation of inflammatory gene expression.

CLINICAL RELEVANCE AND FUTURE PERSPECTIVES

Mucosal hypoxia is an important contributing factor in ischemic disease of the large intestine. Furthermore, it has become increasingly clear that disease processes not classically associated with hypoxia are significantly influenced by hypoxia-related pathways. For example, significant signaling cross-talk between hypoxia and inflammatory pathways exists. This is exemplified by TNFα-dependent induction of HIF-1α in normoxia. Traditionally, inappropriate activity of the mucosal immune system has been deemed responsible for much of the pathophysiology of inflammatory bowel disease. However, mounting evidence indicates that intestinal epithelial hypoxia, due to microvascular breakdown resulting from either tissue fibrosis or mesenteric vasculitis, might be an important factor in mediating

barrier disturbance associated with the disease. Evidence for a role of epithelial hypoxia in inflammatory disease processes in the gut stems from observations both in human disease and animal models. For example, in experimental colitis, the degree of intestinal hypoxia has been directly related to the severity of disease; conversely, knockout of hypoxia-induced adaptive genes such as ITF increases susceptibility. Thus, it is likely that hypoxia significantly contributes as a pathological stimulus in chronic inflammatory disease in the gut.

Bacterial translocation and endotoxemia are devastating complications in intensive care patients. Following intraoperative hypovolemia and/or vasoconstriction associated with surgery, intestinal permeability increases and the degree of endotoxemia is related to this increase. Endotoxemia not only may arise from disturbances of intestinal perfusion, but may also contribute to it by further decreasing intestinal blood supply and increasing mucosal permeability through the induction of inflammatory episodes. These events may constitute a self-propagating series of events that raise challenges to effective therapeutic intervention. These two distinct clinical entities exemplify the crucial importance of intestinal barrier integrity and the profound impact that ischemia/hypoxia exerts on this system. Consequently, innate mechanisms that protect against hypoxia are of fundamental importance. One such mechanism could be the HIF-1-regulated expression of barrier protective genes. Efforts to better understand this pathway as a therapeutic modality could prove beneficial to diseases that are characterized by barrier disruption. Taking into account the significant progress made in immunomodulatory therapy in Crohn's disease, this could constitute a profoundly new approach in the sense of an "epithelial therapeutic." One such approach already under investigation is the use of recombinant trefoil factor, which has proved beneficial in the reduction of inflammation and in acceleration of healing in a rat model of colitis.

Due to its role as a master regulator of hypoxia adaptive responses, enhancement of the HIF-1 pathway, e.g., through stabilization of HIF-1α, could consequently lead to a broad barrier protective response involving multiple downstream effectors. In cancer therapy, suppression of the adaptive HIF-1 response is required, but in the case of ischemic disease in the colon, support of barrier function through induction of the same pathway would be a desirable consequence. Data from studies utilizing induction of HIF-1α by pharmacologic intervention (peptide inhibitor of HIF-1α degradation) in models of ischemic injury have provided promising preliminary data. Although these or other approaches constitute significant advances, a better understanding of the complexity of the hypoxia/inflammation interrelationship is required to permit useful clinical applications.

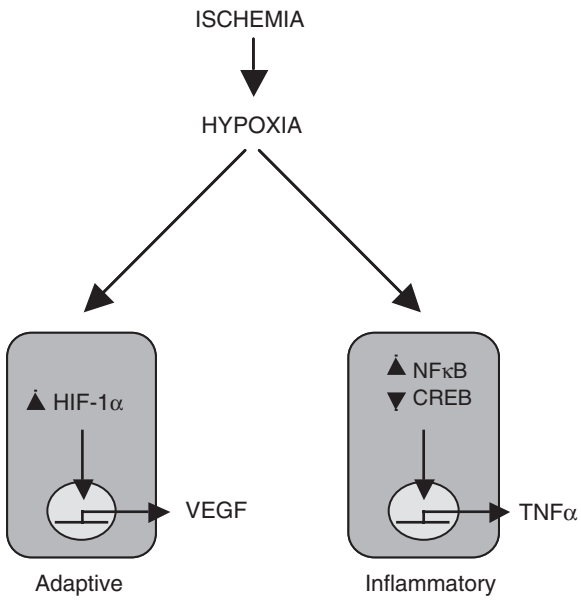


FIGURE 2 Distinct molecular mechanisms underlie the induction of adaptive and inflammatory responses of colonic epithelial cells to hypoxia. Mucosal ischemia leads to intestinal epithelial cell hypoxia. A rapid adaptive response is induced through the accumulation of hypoxia inducible factor-1α (HIF-1α), an event that leads to the induction of transcription of a number of adaptive genes, typified by vascular endothelial growth factor (VEGF). Temporally downstream, an inflammatory response to hypoxia occurs that is mediated through the activation and suppression of nuclear factor κB (NF-κB) and cyclic AMP response-element binding protein (CREB), respectively. TNFα, Tumor necrosis factor α.

SUMMARY

Colonic ischemia is a relatively frequent pathophysiologic occurrence that leads to colonic epithelial cell hypoxia and the activation of distinct molecular pathways. Adaptive and inflammatory responses to hypoxia are mediated through HIF-1α and NF-κB/CREB (Fig. 2). A greater understanding of the oxygen-sensing and signaling mechanisms leading to the activation of these pathways will lead to a more rational definition of novel therapeutic targets in ischemic disease in the large intestine.

Acknowledgments

This work was supported by grants from the National Institutes of Health (United States), The Wellcome Trust, The Science Foundation of Ireland and the Health Research Board of Ireland, and the Crohn's and Colitis Foundation of America.

See Also the Following Articles

Colitus, Radiation, Chemical, and Drug-Induced • Colitis, Ulcerative • Crohn's Disease • Necrotizing Enterocolitis

Further Reading

Furuta, G. T., Turner, J. R., Taylor, C. T., Hershberg, R. M., Comerford, K., Narravula, S., Podolsky, D. K., and Colgan, S. P. (2001). Hypoxia-inducible factor 1-dependent induction of intestinal trefoil factor protects barrier function during hypoxia. *J. Exp. Med.* **193**, 1027–1034.

Haglund, U., and Bergqvist, D. (1999). Intestinal ischemia—The basics. *Langenbecks Arch. Surg.* **384**, 233–238.

Hauser, C. J., Locke, R. R., Kao, H. W., Patterson, J., and Zipser, R. D. (1989). Visceral surface oxygen tension in experimental colitis in the rabbit. *J. Lab. Clin. Med.* **112**, 68–71.

Madara, J. L. (1990). Pathobiology of the intestinal epithelial barrier. *Am. J. Pathol.* **137**, 1273–1278.

Podolsky, D. K. (1999). Mucosal immunity and inflammation. V. Innate mechanisms of mucosal defense and repair: The best offense is a good defense. *Am. J. Physiol.* **277**, G495–G499.

Sartor, R. B. (1997). Pathogenesis and immune mechanisms of chronic inflammatory bowel diseases. *Am. J. Gastroenterol.* **92**, 5S–11S.

Semenza, G. L. (2001). HIF-1, O(2), and the 3 PHDs: How animal cells signal hypoxia to the nucleus. *Cell* **107**, 1–3.

Synnestvedt, K., Furuta, G. T., Comerford, K. M., Louis, N., Karhausen, J., Eltzschig, H. K., Hansen, K. R., Thompson, L. F., and Colgan, S. P. (2002). Ecto-5′-nucleotidase (CD73) regulation by hypoxia-inducible factor-1 mediates permeability changes in intestinal epithelia. *J. Clin. Invest.* **110**, 993–1002.

Taylor, C. T., and Colgan, S. P. (1999). Therapeutic targets for hypoxia-elicited pathways. *Pharm. Res.* **16**, 1498–1505.

Taylor, C. T., Dzus, A. L., and Colgan, S. P. (1998). Autocrine regulation of intestinal epithelial permeability induced by hypoxia: Role for basolateral release of tumor necrosis factor-α. *Gastroenterology* **114**, 657–668.

Wakefield, A. J., Sawyerr, A. M., Dhillon, A. P., Pittilo, R. M., Rowles, P. M., Lewis, A. A., and Pounder, R. E. (1989). Pathogenesis of Crohn's disease: Multifocal gastrointestinal infarction. *Lancet* **2**, 1057–1062.

Colonic Motility

SUSHIL K. SARNA

University of Texas Medical Branch, Galveston

excitation–contraction coupling Signaling cascades in smooth muscle cells that transform the binding of neurotransmitters to cell surface receptors into the phosphorylation of contractile proteins, resulting in cell contraction.

giant migrating contraction A large-amplitude, long-duration, lumen-occluding contraction that rapidly propagates over long distances. It causes mass movement and produces descending inhibition of contractions and relaxation of tone to facilitate rapid propulsion.

interneurons Neurons that relay signals among ganglia in a three-dimensional structure.

interstitial cells of Cajal Specialized nonneuronal non-smooth-muscle cells found at specific locations in the gut wall.

intrinsic primary afferent neurons Neurons with cell bodies located in the colonic wall and axons projecting to mucosa to detect chemical and mechanical stimuli in the lumen.

mass movement Ultrarapid propulsion of a large bolus of digesta over a sizable length of the small bowel or colon.

motor neurons Neurons with cell bodies located in the myenteric plexus and axons projecting to longitudinal and circular muscle layers.

rhythmic phasic contractions Regularly occurring contractions of colon musculature that cause mixing and slow net distal propulsion of digesta.

slow waves Periodic spontaneous depolarization of smooth muscle cells that regulate the maximum frequency, direction of propagation, and timing of occurrence of contractions on receiving an excitatory signal from the motor neurons.

The motility requirements of the colon are threefold: first, to cause extensive turning over and mixing movements; second, to cause slow net distal propulsion of colonic contents; and third, to produce infrequent mass movements. This article discusses the types of colonic contractions and their spatiotemporal organization and

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Colitus, Radiation, Chemical, and Drug-Induced • Colitis, Ulcerative • Crohn's Disease • Necrotizing Enterocolitis

Further Reading

Furuta, G. T., Turner, J. R., Taylor, C. T., Hershberg, R. M., Comerford, K., Narravula, S., Podolsky, D. K., and Colgan, S. P. (2001). Hypoxia-inducible factor 1-dependent induction of intestinal trefoil factor protects barrier function during hypoxia. *J. Exp. Med.* **193**, 1027–1034.

Haglund, U., and Bergqvist, D. (1999). Intestinal ischemia—The basics. *Langenbecks Arch. Surg.* **384**, 233–238.

Hauser, C. J., Locke, R. R., Kao, H. W., Patterson, J., and Zipser, R. D. (1989). Visceral surface oxygen tension in experimental colitis in the rabbit. *J. Lab. Clin. Med.* **112**, 68–71.

Madara, J. L. (1990). Pathobiology of the intestinal epithelial barrier. *Am. J. Pathol.* **137**, 1273–1278.

Podolsky, D. K. (1999). Mucosal immunity and inflammation. V. Innate mechanisms of mucosal defense and repair: The best offense is a good defense. *Am. J. Physiol.* **277**, G495–G499.

Sartor, R. B. (1997). Pathogenesis and immune mechanisms of chronic inflammatory bowel diseases. *Am. J. Gastroenterol.* **92**, 5S–11S.

Semenza, G. L. (2001). HIF-1, O(2), and the 3 PHDs: How animal cells signal hypoxia to the nucleus. *Cell* **107**, 1–3.

Synnestvedt, K., Furuta, G. T., Comerford, K. M., Louis, N., Karhausen, J., Eltzschig, H. K., Hansen, K. R., Thompson, L. F., and Colgan, S. P. (2002). Ecto-5′-nucleotidase (CD73) regulation by hypoxia-inducible factor-1 mediates permeability changes in intestinal epithelia. *J. Clin. Invest.* **110**, 993–1002.

Taylor, C. T., and Colgan, S. P. (1999). Therapeutic targets for hypoxia-elicited pathways. *Pharm. Res.* **16**, 1498–1505.

Taylor, C. T., Dzus, A. L., and Colgan, S. P. (1998). Autocrine regulation of intestinal epithelial permeability induced by hypoxia: Role for basolateral release of tumor necrosis factor-α. *Gastroenterology* **114**, 657–668.

Wakefield, A. J., Sawyerr, A. M., Dhillon, A. P., Pittilo, R. M., Rowles, P. M., Lewis, A. A., and Pounder, R. E. (1989). Pathogenesis of Crohn's disease: Multifocal gastrointestinal infarction. *Lancet* **2**, 1057–1062.

Colonic Motility

SUSHIL K. SARNA

University of Texas Medical Branch, Galveston

excitation–contraction coupling Signaling cascades in smooth muscle cells that transform the binding of neurotransmitters to cell surface receptors into the phosphorylation of contractile proteins, resulting in cell contraction.

giant migrating contraction A large-amplitude, long-duration, lumen-occluding contraction that rapidly propagates over long distances. It causes mass movement and produces descending inhibition of contractions and relaxation of tone to facilitate rapid propulsion.

interneurons Neurons that relay signals among ganglia in a three-dimensional structure.

interstitial cells of Cajal Specialized nonneuronal non-smooth-muscle cells found at specific locations in the gut wall.

intrinsic primary afferent neurons Neurons with cell bodies located in the colonic wall and axons projecting to mucosa to detect chemical and mechanical stimuli in the lumen.

mass movement Ultrarapid propulsion of a large bolus of digesta over a sizable length of the small bowel or colon.

motor neurons Neurons with cell bodies located in the myenteric plexus and axons projecting to longitudinal and circular muscle layers.

rhythmic phasic contractions Regularly occurring contractions of colon musculature that cause mixing and slow net distal propulsion of digesta.

slow waves Periodic spontaneous depolarization of smooth muscle cells that regulate the maximum frequency, direction of propagation, and timing of occurrence of contractions on receiving an excitatory signal from the motor neurons.

The motility requirements of the colon are threefold: first, to cause extensive turning over and mixing movements; second, to cause slow net distal propulsion of colonic contents; and third, to produce infrequent mass movements. This article discusses the types of colonic contractions and their spatiotemporal organization and

Copyright 2004, Elsevier (USA). All rights reserved.

regulation, as well as disordered motility in conditions such as diarrhea, constipation, irritable bowel syndrome, and inflammatory bowel disease.

REQUIREMENTS OF COLONIC MOTILITY FUNCTION

The colon is the last major organ in the gastrointestinal tract. As such, it plays a critical role in regulating the frequency of defecation and consistency of stools. These two parameters are determined by the rates of absorption and secretion, the rate of propulsion, and the degree of mixing and turning over of lumenal contents in the colon. The colon absorbs primarily water and some electrolytes to maintain homeostasis. The colonic mucosa is tight and, therefore, its absorption rate is slower than that in the small bowel. Consequently, for adequate absorption an important requirement of colonic motility is to mix and turn over the lumenal contents frequently and regularly and at the same time propel them slowly for their uniform and prolonged exposure to the mucosa.

Another major requirement of colonic motility is rapid propulsion of the feces during defecation. Such rapid propulsion is also required infrequently over the rest of the colon, which, otherwise, has sluggish movements to accommodate the slow absorption rate. These rapid propulsions are called "mass movements." In health, mass movements occur only a few times a day in the colon. However, these rapid propulsions play an important role in colonic motor function since a significant decrease in their frequency causes constipation and a significant increase in their frequency causes motor diarrhea. Motor diarrhea results from frequent mass movements as opposed to secretory diarrhea, which results from excessive secretions. The softer stools in motor diarrhea are due to the lack of absorption of water due to rapid propulsion.

TYPES OF COLONIC CONTRACTIONS

The colonic circular smooth muscle cells generate three types of contractions: (1) rhythmic phasic contractions (RPCs); (2) giant migrating contractions (GMCs); and (3) tone. The longitudinal muscle cells also generate similar contractions but the role of these contractions in colonic motor function is minimal due to their axis of contraction. The longitudinal muscle contractions would reduce the length of a colonic segment without directly shortening the lumen size and, hence, will be minimally ineffective in propulsion. On the other hand, contractions of circular smooth muscle cells partially or completely occlude the lumen, causing mixing and/or propulsion. An underlying requirement for propulsion in the gut is that the circular muscle contractions must propagate. Additionally, the larger amplitude contractions occlude the lumen strongly and are more effective in propulsion. Similarly, a higher frequency of propagating contractions propels lumenal contents more frequently. Nonpropagating contractions, on the other hand, cause basically the mixing and turning over of lumenal contents with minor propulsion due to the to-and-fro movement of contents during contractions.

Rhythmic Phasic Contractions

Rhythmic phasic contractions are responsible for the mixing and slow net distal prolusion in the colon. The extensive mixing and slow distal propulsion in the colon are achieved because the RPCs in the colon are largely nonpropagating and small in amplitude so that they only partially occlude the lumen (Fig. 1).

Giant Migrating Contractions

The mass movements in the colon and rapid expulsion of feces during defecation are achieved by GMCs. These are large-amplitude, lumen-occluding contractions that rapidly propagate over fairly long distances in the colon (Fig. 2). In higher species, such as humans and dogs, these contractions occur infrequently (a few times a day). However, in rodents, these contractions occur regularly and are primarily responsible for their colonic motor function. One major difference between rodent and human GMCs is that, in rodents, most of them do not propagate or propagate only over short distances. The reason that GMCs occur regularly in rodents may be that they pass fecal pellets, which may require stronger propulsive forces. The phasic contractions do occur in the rodent colon, but they are very small in amplitude compared with the amplitude of GMCs and, therefore, have only a minor effect on colonic motility. The GMCs have the characteristics of peristaltic contractions, as descending relaxation precedes their distal propagation. By contrast, the RPCs do not produce descending relaxation. This distinction is often missed in the literature.

The GMCs not only produce mass movements, but they are also important physiological stimuli to produce the sensation of intermittent abdominal cramping (see Fig. 3). Strong contractile forces generated by GMCs can exceed the nociceptive threshold of the gut wall receptors and activate sensory neurons. In addition, if rapid propulsion due to a GMC is not accompanied by descending relaxation to accommodate the large bolus of digesta being propelled, or if a lumenal obstruction exists ahead of a GMC, distension may occur in the

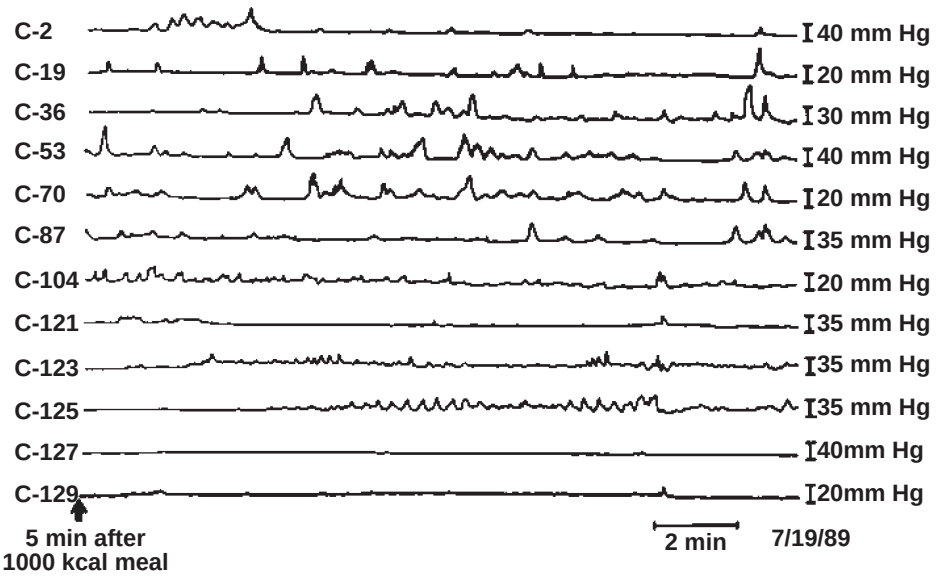


FIGURE 1 Postprandial RPCs recorded from human colon by a 12-channel manometric device. Note that the RPCs show little propagation.

descending segment, which may act as a stimulus to activate stretch sensory receptors.

Tone

The colonic circular smooth muscle cells also generate tonic contractions, which can last from several minutes to hours. Typically, the tone in the circular muscle is increased following a meal. However, an increase in tone by itself does not cause any significant mixing or propulsion. It may, however, enhance the mixing and propulsive functions of phasic contractions and GMCs by reducing lumenal diameter.

REGULATION OF COLONIC CONTRACTIONS

The three types of colonic contractions, discussed above, and their spatial and temporal organizations are regulated at multiple levels to achieve the desired motility functions. The basic scheme is as follows: (1) the mechano- and chemoreceptors in the mucosa sense the presence and characteristics of lumenal contents and send appropriate signals to the myenteric ganglia either directly or via interneurons; (2) the excitatory and inhibitory motor neurons release their respective neurotransmitters when excited by the above inputs; (3) the neurotransmitters bind to membrane receptors on smooth muscle cells or diffuse through the membrane to initiate cascades of signaling pathways to phosphorylate or dephosphorylate contractile proteins; (4) an important condition to initiate these smooth muscle signaling cascades in response to ligand–receptor binding is that the smooth muscle must be depolarized; (5) the myenteric neurons also have the ability to contract smooth muscle cells during their depolarization, independent of any input from mucosal sensory neurons, and colon contractions can occur even when the lumen is completely empty. This suggests that some neurons in the enteric ganglia may spontaneously and periodically release excitatory neurotransmitters to stimulate excitatory motor neurons (6). The gut is an intelligent organ. The mixing and propulsion at any given location depend not only on the lumenal contents at that location, but also on the conditions orad and aborad to that location. This modulation of motility function is achieved by reflexes mediated by orad- and aborad-projecting interneurons in the gut wall, as well as by reflexes mediated by vagal and sympathetic neurons and circulating hormones. Both the extrinsic neurons and hormones eventually work through the enteric ganglia and smooth muscle cells described in (2), (3), and (4) to stimulate or inhibit contractions. (7) Finally, the colon also responds to emotions and stress generated in higher centers. These effects are mediated by extrinsic neurons described in (6). Some salient features of multiple mechanisms in the regulation of colonic motility are discussed below.

Sensory Regulation

The intrinsic primary afferent neurons (IPANS) have their nerve endings in the mucosa close to

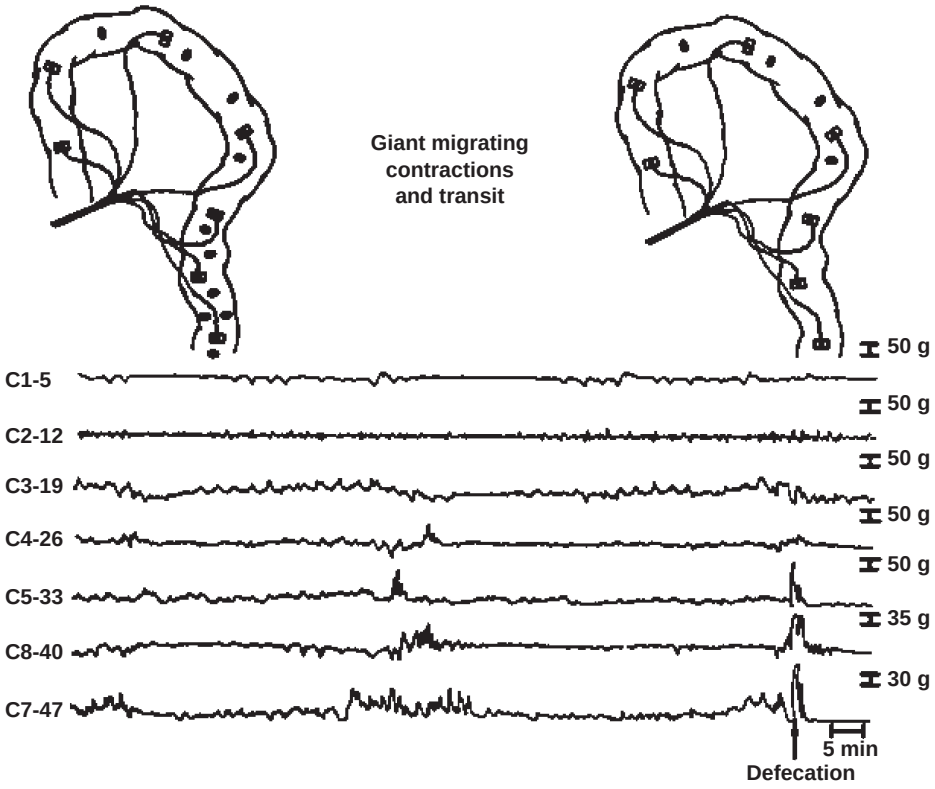


FIGURE 2 Colonic motor activity recorded from canine colon by surgically implanted strain gauge transducers. A set of 10 radio-opaque markers was positioned in the colon. There was little propulsion of the markers during the occurrence of RPCs. However, when a GMC occurred in the distal colon, all markers ahead of its starting point were expelled during defecation. The numbers after C1 to C7 indicate the distances of the transducers from the ileocolonic junction. Reprinted from Sethi, A. K., and Sarna, S. K. (1995). Contractile mechanisms of canine colonic propulsion. *Am. J. Physiol.* **268**, G530–G538.

enterochromaffin (EC) cells containing serotonin. Mucosal stimulation releases 5-hydroxytryptamine (5-HT) from these cells, which acts on 5-HT type 4/5-HT type 1p receptors on IPANS to send a signal to the enteric ganglia either directly or via interneurons. The neurotransmitter for IPANS is acetylcholine, acting on nicotinic receptors, and calcitonin gene-related peptide (CGRP), acting on CGRP receptors. In the small bowel, at least, CGRP administered close intra-arterially stimulates both RPCs and GMCs. The 5-HT released from EC cells may also act on 5-HT type 3 receptors of first-order extrinsic sensory neurons to send signals to higher centers. It has been hypothesized that this route may transmit nociceptive signals, but a complete understanding of this afferent sensory limb is lacking.

Balloon distension is sometimes used as an experimental tool to elicit stretch-induced reflexes from the colon. However, distension is not the primary stimulus for spontaneous colonic contractions in the fasting and postprandial states. Unless there is a lumenal obstruction, the colon may rarely distend to thresholds required for stretch-induced reflexes. The lumenal volume of the colon is normally enough to accommodate the fecal material before any significant stretch can occur under healthy conditions. The primary physiological sensory stimulus is the gentle contact of digesta to stimulate the chemo- and mechanoreceptors in the mucosa. These receptors can be activated experimentally by applying chemicals to the mucosal surface or by stroking the mucosa. Stretch and contractile receptors are, of course, present in the colon wall, but they are more likely to be utilized under abnormal conditions.

Motor Neurons

The enteric ganglia in the myenteric plexus project excitatory and inhibitory motor neurons to smooth muscle cells. Acetylcholine (ACh) is the established physiological neurotransmitter of excitatory motor neurons to stimulate all three types of contractions,

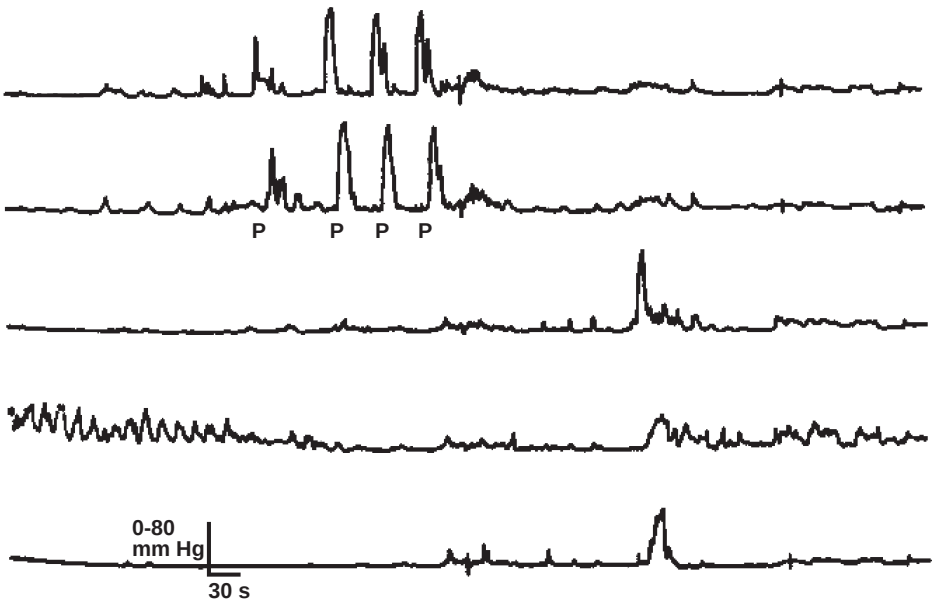


FIGURE 3 Colonic motor activity recorded by a manometric device from a patient with idiopathic chronic constipation. Each occurrence of a GMC was associated with the sensation of abdominal cramping. Reprinted from Bassotti, G., Gaburri, M., Imbimbo, B. P., Rossi, L., Farroni, F., Pelli, M. A., and Morelli, A. (1988). Colonic mass movements in idiopathic chronic constipation. *Gut* **29**, 1173–1179, with permission from the BMJ Publishing Group.

i.e., RPCs, GMCs, and tone. The primary inhibitory neurotransmitters are nitric oxide (NO) and vasoactive intestinal polypeptide (VIP). Of these, NO seems to play a more prominent role for inhibitory control in colonic motility. The smooth muscle cells are in their basal resting state in the absence of any excitatory or inhibitory input from the motor neurons. They contract in response to the release of ACh, subject to the concurrent occurrence of a slow-wave depolarization. Under some conditions, there may be a constitutive release of NO in the basal resting state of cells, but the excitatory input dominates to stimulate ACh-induced contractions. The hypothesis that spontaneous colon contractions occur by the inhibition of inhibitory input to smooth muscle cells *in vivo* lacks evidence. The sudden block of neural conduction, *in vivo* or *in vitro*, sometimes stimulates colon contractions, but they are transient, lasting only a few minutes. The major role of inhibitory motor neurons is in response to specific inputs, such as those that produce descending relaxation or inputs from higher centers in which case the increased inhibitory excitation may overcome the cholinergic excitatory input and prevent the occurrence of contractions, even when the sensory signals to stimulate them are present.

Tachykinins, particularly substance P (SP), are also putative neurotransmitters of excitatory motor neurons and are co-localized with ACh in the myenteric neurons. However, their precise physiological role is not understood. Exogenously administered substance P stimulates GMCs as well as RPCs in the canine colon and it is also released in the ascending limb of the peristaltic reflex in an organ bath, which suggests that it may participate in the stimulation of GMCs, along with ACh. Substance P inhibitors, however, do not block the spontaneous phasic contractions in intact conscious dogs, indicating that SP may not mediate their spontaneous stimulation.

Interneurons

As noted above, the enteric motor neurons and smooth muscle cells are the primary regulators of gut contractions at any given location, in response to sensory input from the same location. However, in the context of overall gut function, these two regulatory mechanisms need integrated information from proximal and distal locations to fine-tune the mixing and propulsion rates in the segment under their direct control. This information may come from proximal and distal segments within the same organ, such as occurs in an "ileal brake," or it may come from adjacent organs, such as duodeno-gastric and colo-ileal reflexes to regulate the rate of emptying of the proximal organ

into the next gut organ. The information from the distal locations is used to sense the state of digestion and the readiness of the distal segments to receive more digesta. The information from the proximal segments is used to prepare the distal segments for the arrival of new digesta. This coded information is transmitted by interneurons. The extrinsic neurons and hormones also perform a part of this function, but these will not be discussed here.

Peristaltic reflex that manifests as a GMC in the intact conscious state is the most studied reflex involving interneurons. But, as noted above, GMCs occur only infrequently. The rest of the time, the orad or aborad transmission of information in the colonic interneurons is used to fine-tune the spatiotemporal patterns of RPCs for optimal digestion. Note also that the descending inhibition occurs only with GMCs; the RPCs do not produce descending relaxation.

A majority of interneurons in the human colon are immunoreactive for choline acetyltransferase (ChAT) and nitric oxide synthase (NOS). Specifically, approximately 90% of the orad-projecting neurons in the human colon contain ChAT, whereas no ascending neurons contain NOS. On the contrary, approximately 46% of the aborad-projecting neurons contain NOS immunoreactivity alone, 29% contain both ChAT and NOS immunoreactivities, and approximately 20% contain ChAT activity alone. In agreement with the orad or aborad projections of ChAT-containing neurons, both the orad and aborad reflexes are blocked by nicotinic receptor blockade with hexamethonium. The average length of orally and aborally projecting neurons is approximately 10 mm. This suggests that a chain of interneurons, where each neuron innervates the lateral excitatory or inhibitory motor neurons in the ganglia, as well as orad- or aborad-projecting interneurons in the same ganglia conveys the information along the entire length of the colon. The interneurons in the human colon display the Dogiel type 1 morphology. Although a physiological role for NO as one of the inhibitory neurotransmitters in the gut is well established, a physiological role for NO in the interneurons has not yet been identified.

The human colonic interneurons also display immunoreactivity for other neurotransmitters, including tachykinins, VIP, calretinin, and 5-HT. These neurotransmitters may perform specialized roles, which have largely not yet been identified. The roles of neurotransmitters, other than ACh, are likely to be supportive rather than primary. It must be noted, though, that the neurochemical coding of interneurons in the colon of animals, such as guinea pig, differs substantially from that in the human colon.

Electrophysiological Regulation (Slow Waves)

Periodic depolarization of smooth muscle cells, called slow waves, is an important mechanism of the regulation of gut contractions. The slow waves regulate the maximum frequency of RPCs, their timing of occurrence, and the direction and distance of their propagation. A phasic contraction can occur only once during cell depolarization. Therefore, the maximum frequency of contractions at a given location cannot exceed the frequency of slow waves. The slow waves are omnipresent but contractions occur only under the conditions of concurrent depolarization and excitatory neurotransmitter release from motor neurons. The frequency of slow waves in the intact human colon is highly irregular and varies from approximately 2 to 13 cycles/min. Since a local contraction occurs only once during a slow wave depolarization, it will propagate distally only if the slow waves are phase-locked in that direction. In the colon, the slow waves exhibit very poor coupling, which is the reason that the phasic contractions do not propagate or propagate only over very short distances. This is the underlying cause of slow distal propulsion of colonic contents, as noted above.

The importance of coordination of the actions of slow waves and excitatory neural input to regulate contractions is sometimes overlooked in the literature. This confusion arises because, in other organs, the muscle cells alone regulate spontaneous contractions, such as the cardiac muscle that contracts with every membrane depolarization. The colonic smooth muscle cells do not contract with every depolarization. In other organs, such as those containing only skeletal muscle, the neurons alone regulate contractions. On the other hand, the excitatory input from neurons alone cannot contract colonic smooth muscle cells. Also, the colonic smooth muscle cells are not like the vascular smooth muscle cells, which generate predominantly a tone to regulate blood flow. The colonic muscle generates three distinct types of contractions, RPCs, GMCs, and tone, to regulate motility function.

Unlike the small bowel and stomach, the colon of most species, such as human and dog, also generates long-duration phasic contractions. These contractions are not regulated by the omnipresent slow waves as discussed above; they are regulated by contractile electrical complexes (CECs) that occur intermittently. The frequency of CECs and their corresponding long-duration contractions in the human colon varies from approximately 0.5 to 2 per minute. The short- and long-duration phasic contractions together are responsible for the extensive mixing and slow net distal propulsion in the colon.

The gut of most species exhibits three types of interstitial cells of Cajal (ICC): one layered network in the myenteric plexus (ICC-MP), one in the submucosal plexus (ICC-SP), and one in the circular muscle layer (ICC-IM). The hypothesis is that the ICC (MP) drive the slow waves in the small intestine, whereas ICC (SP) drive them in the colon. The ICC (IM) may mediate the inhibitory neuronal input to muscle cells. The majority of the studies in support of this hypothesis have been carried out in the rat and mouse guts. The dissection of submucosa from the rat colon muscle strips obliterates phasic contractions regulated by slow waves. The presence and distribution of ICC in the human colon, however, differ substantially from those in rodents. The highest density of ICC in the human colon has been reported to be in the myenteric plexus. The ICC (SM) at the inner layer of the circular muscle are sparse and may not form a continuous network as has been reported in rodents. The density of ICC (IM) is several-fold less than that of ICC (SM) or ICC (MP) in the entire human colon. The differences in the distributions of ICC between humans and rodents leave open the question whether there are differences in their roles in the two species. Several studies have reported a decrease in the density of ICC or damage to their processes in colonic motility disorders, such as constipation, but in most of these reports the concurrent absence of slow waves or impaired inhibitory motor function related to the defects in ICC was not investigated.

Excitation–Contraction Coupling through Cell Signaling

On ligand binding to surface receptors, a number of signaling cascades, starting with the activation of G-proteins, are brought into play for smooth muscle cells to contract. An immediate-early event is the increase of cytosolic [Ca^{2+}]. This event utilizes Ca^{2+} influx through L-type Ca^{2+} channels, as well as calcium release from the endoplasmic reticulum. The gut smooth muscle cells express primarily the M_2 and M_3 receptors. M_3 receptors are coupled to G_{q11} protein and M_2 to $G_{\alpha i3}$. The activation of G_{q11} by M_3 receptors activates phospholipase C to hydrolyze inositol phospholipids. The hydrolysis of phosphatidylinositol 4,5-biphosphate generates two second messengers, inositol 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG). IP_3 acts on its receptors on the endoplasmic reticulum to release calcium from intracellular stores. Ca^{2+} release from these stores also occurs in response to ryanodine. DAG, on the other hand, activates protein kinase C (PKC), which translocates it from the cytosol to the plasma membrane. The activation of G-protein and PKC is also linked to the opening of L-type calcium channels for Ca^{2+} influx. The activation of PKC and increase of cytosolic [Ca^{2+}] are key steps in cell contraction, but several other steps, such as the release of arachidonic acid from membrane phospholipids and activation of mitogen-activated protein kinases, also play important roles in excitation–contraction coupling in colonic smooth muscle cells. The expression of some of these signaling molecules is altered in inflammation, leading to motility abnormalities.

COLONIC MOTILITY DYSFUNCTION IN INFLAMMATION AND IRRITABLE BOWEL SYNDROME

Colonic Inflammation

The classic concept of inflammatory bowel disease has been that an autoimmune disorder results in uncontrolled inflammation in which the rest of the cells in the gut wall are "innocent bystanders" and get hurt. Accumulating evidence over the past decade or so shows, however, that although an immune disorder is a major component of this disease, the rest of the cells, including smooth muscle cells, enteric neurons, epithelial cells, myofibroblasts, and glia, play an active rather than a passive role in both acute and chronic inflammations. The role of smooth muscle cells and enteric neurons is twofold: (1) to generate abnormal motility patterns that produce the symptoms of diarrhea, urgency of defecation, and abdominal cramping; and (2) to secrete cytokines and chemokines and express cell adhesion molecules on their surface to amplify the inflammatory response triggered by the activation of resident and infiltrating immunocytes in the muscularis externa.

In both acute and chronic inflammation, rhythmic phasic contractions and tone are suppressed in the colon, whereas the frequency of GMCs is significantly increased. The absence of phasic contractions and decrease of tone result in the lack of mixing and slow propulsive movements, thus decreasing the uniform and prolonged exposure of the lumenal contents to the mucosa for absorption. The absence of phasic contractions also allows rapid propulsion of colonic contents, which otherwise move slowly to and fro. At the same time, frequent mass movements produced by the increased frequency of GMCs propel the fecal contents rapidly to further decrease their mucosal contact. When the GMCs occur in the distal colon or propagate from the proximal colon to the rectum, they produce a sense of urgency by forcefully propelling the bolus ahead of them against the anal sphincters. In addition, the GMCs, by their strong contractile force and/or by distending the

descending segment beyond the nociceptive threshold, may produce the sensation of abdominal cramping. The nociceptive threshold in the inflamed colon is reduced due to hyperalgesia and allodynia, further increasing the incidence of abdominal cramping. In health, the GMCs are rarely felt to be painful because of the lack of hyperalgesia and because of intact descending relaxation that accommodates the large bolus that is propelled rapidly by a GMC.

Primary cultures of human colonic smooth muscle cells, as well as longitudinal muscle/myenteric plexus strips, secrete a panel of cytokines and chemokines and express cell adhesion molecules in inflammation and on exposure to cytokines, such as tumor necrosis factor α and interleukin-1β. The recruitment of smooth muscle cells and enteric neurons, which are in much greater numbers than the immunocytes, may be essential for full-blown inflammation in the muscularis. Thus, the colonic smooth muscle cells and enteric neurons may serve as secretory cells, although their normal function is to contract and release neurotransmitters, respectively.

Both of the above changes in cell behavior, i.e., altered motility patterns and secretions of inflammatory mediators, are mediated by altered gene expression of key signaling molecules through the activation of transcription factors, such as nuclear factor κB (NF-κB). Calcium influx through L-type Ca^{2+} channels plays a critical role in the contraction of smooth muscle cells. Recent findings show that calcium influx through these channels is reduced in inflammation. Other findings also show that inflammation, as well as oxidative stress, activates NF-κB and, thereby, suppresses contractility. The inhibition of NF-κB in single cells or in intact animals reverses the suppression of contractility. Regulation of gene expression in colonic smooth muscle cells to affect their contractility, as well as in enteric neurons to affect neurotransmitter release, may be a potent target of molecular medicine to normalize motility disorders.

Irritable Bowel Syndrome

Irritable bowel syndrome results from motility dysfunction and visceral hypersensitivity. It is characterized by visceral pain accompanied by diarrhea (IBS-diarrhea; IBS-d), constipation (IBS-constipation; IBS-c), or alternating diarrhea and constipation (IBS-diarrhea/constipation; IBS-d/c). The symptoms of diarrhea and constipation result directly from enteric neuromuscular disorders. Since mental stress exacerbates the symptoms of IBS or can even precipitate them, abnormal input from the central nervous system and/or spinal cord may also play an important role in producing motility dysfunction in these patients.

Unlike in IBD patients, RPCs and tone are not generally suppressed severely in IBS patients. However, the precise alterations in the spatiotemporal patterns of RPCs that may lead to altered transit in IBS-d or IBS-c have not been identified for several reasons: (1) As noted above, the colonic RPCs are highly disorganized in space and are irregular in frequency, which makes it almost impossible to analyze them visually. (2) The manometric or solid-state transducer devices record primarily the lumen-occluding contractions. Most colonic RPCs do not occlude the lumen. Therefore, these devices do not record all RPCs. (3) Cleansing the colon prior to motility recordings by intralumenal devices may significantly alter its motility patterns. Under normal conditions, the colon is seldom empty. (4) The colonic motility is neither uniform nor consistent throughout its length, which reflects different mixing and propulsion rates in different segments of the colon. Reliable conclusions on colonic motility can be made only by concurrent recordings from the entire colon. This is technically challenging. (5) The normal transit time from the appearance of digesta at the ileocolonic junction to defecation is on the order of 36 to 48 h in the human colon. During this period, the motility patterns change almost continuously throughout the colon. Shorter periods of recording are unable to provide a complete understanding of motility patterns in either health or disease. These difficulties have led some investigators to speculate that there may be no abnormal motility in IBS patients. This hypothesis does not explain the rapid or slow transits in the colon in IBS-d and IBS-c patients, respectively, as well as intermittent occurrences of abdominal cramping.

GMCs play a major role in motility dysfunction in IBS patients as they can be related to all three major symptoms of this disease: rapid transit, slow transit, and visceral pain. Patients with constipation have been reported to have a significantly lower frequency of colonic GMCs than healthy subjects. The smaller frequency of GMCs would not only reduce mass movements, which are essential in an otherwise sluggish motility environment, but they may also affect the propulsive force required for defecation and the strong contractile stimulus required in the rectum to relax the anal sphincters. On the other hand, an increase in the frequency of GMCs causes motor diarrhea by producing excessive mass movements. Hyperalgesia in IBS patients coupled with the occurrence of GMCs can produce the sensation of abdominal cramping. In comparison with rhythmic phasic contractions, the GMCs are easier to record and analyze because they are mostly lumen

occluding. However, long-term recordings from the entire length of the colon may still be a prerequisite for a complete understanding of their contribution to colonic motor function in both health and disease.

Acknowledgments

This work was supported in part by National Institute of Digestive, Diabetes, and Kidney Diseases grant DK 32346.

See Also the Following Articles

Basic Electrical Rhythm • Colitis, Ulcerative • Colonic Obstruction • Constipation • Crohn's Disease • Defecation • Diarrhea • Gastro-colic Reflex • Haustra • Interstitial Cells of Cajal • Irritable Bowel Syndrome • Postprandial Motility • Power Propulsion

Further Reading

Bassotti, G., Chiarioni, G., Vantini, I., Betti, C., Fusaro, C., Pelli, M. A., and Morelli, A. (1994). Anorectal manometric abnormalities and colonic propulsive impairment in patients with severe chronic idiopathic constipation. *Digest. Dis Sci.* **39**, 1558–1564.

Chey, W. Y., Jin, H. O., Lee, M. H., Sun, S. W., and Lee, K. Y. (2001). Colonic motility abnormality in patients with irritable bowel syndrome exhibiting abdominal pain and diarrhea. *Am. J. Gastroenterol.* **96**, 1499–1506.

Choi, M. G., Camilleri, M., O'Brien, M. D., Kammer, P. P., and Hanson, R. B. (1997). A pilot study of motility and tone of the left colon in patients with diarrhea due to functional disorders and dysautonomia. *Am. J. Gastroenterol.* **92**, 297–302.

Foxx-Orenstein, A. E., and Grider, J. R. (1996). Regulation of colonic propulsion by enteric excitatory and inhibitory neurotransmitters. *Am. J. Physiol.* **271**, G433–G437.

Gonzalez, A., and Sarna, S. K. (2001). Different types of contractions in rat colon and their modulation by oxidative stress. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G546–G554.

Hagger, R., Gharaie, S., Finlayson, C., and Kumar, D. (1998). Regional and transmural density of interstitial cells of Cajal in human colon and rectum. *Am. J. Physiol.* **275**, G1309–G1316.

Karaus, M., and Sarna, S. K. (1987). Giant migrating contractions during defecation in the dog colon. *Gastroenterology* **92**, 925–933.

Li, M., Johnson, C. P., Adams, M. B., and Sarna, S. K. (2002). Cholinergic and nitrergic regulation of *in vivo* giant migrating contractions in rat colon. *Am. J. Physiol. Gastrointest. Liver Physiol.* **283**, G544–G552.

Liu, X., Rusch, N. J., Striessnig, J., and Sarna, S. K. (2001). Down-regulation of L-type calcium channels in inflamed circular smooth muscle cells of the canine colon. *Gastroenterology* **120**, 480–489.

Lyford, G. L., He, C. L., Soffer, E., Hull, T. L., Strong, S. A., Senagore, A. J., Burgart, L. J., Young-Fadok, T., Szurszewski, J. H., and Farrugia, G. (1994). Pan-colonic decrease in interstitial cells of Cajal in patients with slow transit constipation. *Gut* **51**, 496–501.

Mawe, G. M., Branchek, T. A., and Gershon, M. D. (1986). Peripheral neural serotonin receptors: Identification and characterization with specific antagonists and agonists. *Proc. Natl. Acad. Sci. USA* **83**, 9799–9803.

Porter, A. J., Wattchow, D. A., Brookes, S. J., and Costa, M. (2002). Cholinergic and nitrergic interneurones in the myenteric plexus of the human colon. *Gut* **51**, 70–75.

Sanders, K. M., Ordog, T., and Ward, S. M. (2002). Physiology and pathophysiology of the interstitial cells of Cajal: From bench to bedside. IV. Genetic and animal models of GI motility disorders caused by loss of interstitial cells of Cajal. *Am. J. Physiol. Gastrointest. Liver Physiol.* **282**, G747–G756.

Sarna, S. K. (2003). Neuronal locus and cellular signaling for the stimulation of ileal giant migrating and phasic contractions. *Am. J. Physiol. Gastrointest. Liver Physiol.* **284**, G789–G797.

Sarna, S. K. (1991). Physiology and pathophysiology of colonic motor activity (1). *Digest. Dis. Sci.* **36**, 827–862.

Sethi, A. K., and Sarna, S. K. (1991). Colonic motor activity in acute colitis in conscious dogs. *Gastroenterology* **100**, 954–963.

Shi, X.-Z., Lindholm, P. F., and Sarna, S. K. (2003). NF-κB activation by oxidative stress and inflammation suppresses contractility in canine colonic circular smooth muscle cells. *Gastroenterology* **124**, 1369–1380.

Shi, X. Z., and Sarna, S. K. (2000). Impairment of Ca(2+) mobilization in circular muscle cells of the inflamed colon. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G234–G242.

Wade, P. R., and Wood, J. D. (1988). Actions of serotonin and substance P on myenteric neurons of guinea-pig distal colon. *Eur. J. Pharmacol.* **148**, 1–8.

Wardell, C. F., Bornstein, J. C., and Furness, J. B. (1994). Projections of 5-hydroxytryptamine-immunoreactive neurons in guinea-pig distal colon. *Cell Tissue Res.* **278**, 379–387.

Colonic Obstruction

SCOTT W. ARLIN
Ohio State University College of Medicine

adynamic ileus Absence of intestinal motility, often occurring for several days following surgery.
diverticulosis Herniations of the mucosa and submucosa through the muscular wall of the colon.
intussusception Invagination or telescoping of proximal bowel into the adjacent distal bowel, often resulting in bowel obstruction.
volvulus Abnormal twisting of a segment of bowel on itself in the longitudinal axis, often resulting in bowel obstruction.

Colonic obstruction refers to blockage of fecal or other luminal contents from movement in the anal direction within the colon. There are many causes of colonic obstruction, most of which can be subdivided into mechanical or nonmechanical categories; nonmechanical obstruction may be congenital or acquired.

INTRODUCTION

Mechanical colonic obstructions are caused by a tumor or other lesion that results in narrowing or total stricture of the bowel lumen, thereby preventing intestinal contents from moving toward the anus. Nonmechanical colonic obstruction results from failure of normal propulsive motility. Commonly, this is due to a condition called adynamic ileus, which usually has a reversible cause. A more severe and prolonged form of ileus is termed colonic pseudo-obstruction, which is a syndrome with all the signs and symptoms of mechanical obstruction in the absence of an occluding lesion. Patients with this syndrome may have an underlying congenital defect involving absence of the nerve supply to the intestinal wall or lack of circular smooth muscle, which causes severe dysmotility that can lead to megacolon. An example of congenital aganglionosis is Hirschsprung's disease. Colonic pseudo-obstruction that is acquired later in life after trauma or surgery is referred to as Ogilvie's syndrome.

ETIOLOGY

By far the most common cause of mechanical colonic obstruction is cancer, which accounts for 60–70% of all cases. Most colorectal cancers occur in the distal third of the colon and up to 10% of these patients will require emergent surgery for decompression. Up to 20% of cases of colonic obstruction are caused by a volvulus. The most frequent site of colonic volvulus is the sigmoid colon. Other sites, in order of decreasing frequency, are the cecum, transverse colon, and splenic flexure. Diverticular stricture is the third most common cause of mechanical colonic obstruction. Diverticulosis is common in the United States and Western world in general, with a prevalence approaching 40% in some case series. Diverticulitis refers to inflammation in the involved segment of bowel, causing a perforation of the colonic diverticulum. Progression of this process can lead to complications that include development of a localized abscess, fistula formation, stricture, and complete bowel obstruction. Less common causes of mechanical colonic obstruction include intussusception, stricture secondary to inflammatory bowel disease, fecal impaction, and extrinsic compression (i.e., from intraabdominal adhesions, masses, or hernia).

Nonmechanical colonic obstruction likewise has many causes and can be categorized as either congenital or acquired. Congenital absence of the enteric nervous system (aganglionosis), lack of circular smooth muscle, or loss of extrinsic nervous supply to the bowel can all lead to ineffective colonic motility and ultimately functional obstruction of the lumen. Hirschsprung's disease is the classic example of congenital colonic aganglionosis. Due to lack of enteric inhibitory motor neurons, the involved segment of bowel fails to relax, resulting in narrowing of the lumen and obstruction to the passage of feces. This usually occurs in distal colonic segments or in the rectum and leads to megacolon in severe cases. Familial visceral myopathies, or progressive dystrophy of intestinal smooth muscle, can be the underlying cause of failure of propulsive motility resulting in nonmechanical (i.e., functional) obstruction. Finally, systemic neurologic disorders such as Parkinson's disease and brain stem lesions can lead to colonic dysmotility and functional obstruction via

 Copyright 2004, Elsevier (USA). All rights reserved.

compromise of nervous signaling from the brain and spinal cord to the bowel.

The most common acquired nonmechanical cause of acute colonic obstruction is Ogilvie's syndrome, which involves gross dilation of the cecum and proximal colon in the absence of an anatomic lesion. The etiology of this process is unknown; a postulated mechanism involves impairment of the autonomic nervous supply to the colon, usually following trauma or surgery. Metabolic disturbances (i.e., hypokalemia, hypomagnesemia, or hypocalcemia) or use of narcotic drugs may occur in conjunction with Ogilvie's syndrome and contribute to colonic dysmotility or may occur independently as a cause for transient colonic obstruction. Other drugs implicated as a cause for nonmechanical intestinal obstruction include calcium channel blockers, tricyclic antidepressants, antihistamines, phenothiazines, and clonidine. Endocrine disorders, such as hypothyroidism, also remain an important underlying factor in functional colonic obstruction.

Acquired colonic obstruction may be associated with a variety of degenerative changes to the musculature and nervous supply of the bowel that occur in other systemic diseases. Infiltrative diseases such as amyloidosis involve disruption of the contractile proteins in intestinal smooth muscle due to deposition of amyloid in the muscle. Systemic sclerosis causes similar disruption that is secondary to fibrosis and smooth muscle atrophy in the involved segments of bowel. Some visceral neuropathies reflect autoimmune attack on the enteric nervous system and can be the cause of intestinal pseudo-obstruction. Neuropathies of this nature occur in Chagas' disease and paraneoplastic syndrome secondary to small cell carcinoma of the lung. Circulating antineuronal antibodies directed to neural elements of the enteric nervous system occur in both Chagas disease and paraneoplastic syndrome and, when present, can lead to functional dysmotility.

SYMPTOMS

Whether the cause is mechanical or nonmechanical, most patients with colonic obstruction develop similar symptoms. Abdominal pain, distension, and bloating are the most common presenting symptoms. The degree of abdominal pain varies considerably among affected patients and is usually more severe if total obstruction or peritonitis is present. If the pain is severe and persistent, then a complication such as bowel perforation or strangulation must be considered. Depending on the degree to which the obstruction impedes the forward propulsion of the luminal contents, the patient may experience constipation or diarrhea. Change in stool caliber may reflect the location of the obstruction. Patients often report pencil-thin stools when a partially obstructing lesion is present in the distal colon or rectum. Vomiting can be a late symptom but is uncommon in colonic obstruction. Other symptoms, when present, may suggest the underlying cause of the obstruction. For example, weight loss, blood in the stools and poor appetite are common in patients with colorectal cancer.

DIAGNOSIS

Making the diagnosis of colonic obstruction always starts with taking a history and performing a physical exam. The most frequently reported symptoms are abdominal pain, distension, and bloating; however, it is important to consider that lack of pain does not exclude the diagnosis. This is especially true in elderly patients or postoperative patients being treated with narcotic medications. Duration of symptoms may also provide a clue to the cause of the obstruction. Acute onset of symptoms is common with volvulus or intussusception; whereas insidious onset tends to occur in patients who have colorectal cancer. The physical exam usually reveals a tympanitic, distended abdomen with tenderness to palpation that is often more severe below the level of the umbilicus. Even with complete mechanical obstruction, bowel sounds are usually present and may sound like "rushes and tinkles." Nevertheless, there are no pathognomonic physical exam signs or symptoms for colonic obstruction.

Plain radiographs should be taken with the patient both in the supine and the upright positions. The characteristic finding will be dilated loops of bowel proximal to the site of obstruction with little or no gas in the distal colon or rectum. Air/fluid levels in the proximal colon or small bowel may also be present and suggestive of an obstruction. Computed tomography (CT) scan of the abdomen is helpful in distinguishing between partial and complete obstruction and can identify segments of bowel that may be strangulated.

Ultimately, most patients with suspected colonic obstruction will undergo a proctosigmoidoscopy to identify the site and potential cause of the obstruction. If a neoplastic process is discovered, histologic evaluation of biopsy specimens can provide a definitive diagnosis. Exceptions are patients with signs of peritonitis, strangulation, or bowel perforation. Water-soluble contrast enema may be performed to confirm the diagnosis and define the site of the anatomic obstruction. Any

colonoscopic examination must be done with caution to avoid causing perforation secondary to air insufflation during the procedure.

There are no reliable blood tests that confirm the diagnosis of colonic obstruction. Nevertheless, it is useful to obtain blood work for measurement of serum electrolytes, kidney function, and complete blood count. These tests can reveal whether a patient is dehydrated, losing blood, or showing signs of infection. A serum lactate level is usually obtained and, if elevated, suggests compromise of blood supply to the affected bowel.

TREATMENT

The treatment of colonic obstruction depends on the cause and whether it is partial or complete. Complications secondary to the obstruction, such as bowel strangulation or perforation, can also change management. As a general rule, treatment for partial obstruction begins with supportive care. Intravenous fluids are given to restore blood volume and prevent dehydration and electrolyte imbalances are corrected. A nasogastric tube can be placed to decompress the stomach and proximal small bowel. Oral intake should be avoided. If there are signs of peritonitis, antibiotics are administered. Air or contrast enemas are sometimes useful in relieving obstructions, especially if caused by intussusception. Possible precipitants such as narcotic or anticholinergic medications are used sparingly, if at all, and promotility drugs such as neostigmine are often helpful. Colonoscopic placement of a rectal tube may be done to decompress the colon when supportive measures have failed.

Surgery is usually needed if a partial obstruction does not resolve or if the colonic obstruction is complete. It is also reserved as an option for patients who develop a complication, such as bowel strangulation or perforation. The type of surgery will depend on the cause of the obstruction, the patient's clinical condition, and the findings during the operation.

See Also the Following Articles

Autonomic Innervation • Chagas' Disease • Colonic Motility • Colorectal Adenocarcinoma • Defecation • Disinhibitory Motor Disorder • Diverticulous • Enteric Nervous System • Hirschsprung's Disease (Congential Megacolon) • Intestinal Pseudoobstruction • Intussusception • Paraneoplastic Syndrome • Volvulus

Further Reading

Turnage, R. H., and Bergen, P. (2002). Intestinal obstruction and ileus. In "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 2113–2150. Saunders, New York.

Colonic Ulcers

RHONDA K. YANTISS* AND ROBERT D. ODZE†

University of Massachusetts Memorial Health Care, Worcester and †Brigham and Women's Hospital, Boston

active colonic injury Colonic mucosal disease characterized by neutrophilic inflammation of the mucosa, cryptitis, and crypt abscesses, with or without ulcers.

chronic colonic injury Colonic mucosal disease characterized by alterations in the cellular components and architecture, suggestive of ongoing injury.

colitis Inflammation of the colonic mucosa characterized by neutrophilic and/or mononuclear cell inflammation in the lamina propria, with associated injury to the colonic epithelium.

inflammatory bowel disease Chronic, idiopathic inflammatory disease of the colon, specifically Crohn's disease and ulcerative colitis.

ischemia Pattern of disease injury characterized by fibrin and hemorrhage in the lamina propria, crypt loss and regeneration, and necrosis of the mucosa. This pattern of injury may be seen in association with a variety of vascular insults, infections, and medications.

Colonoscopy has assumed a major role in the evaluation of patients with lower gastrointestinal bleeding. As a result, clinicians increasingly encounter the presence of colonic ulcers and, thus, are faced with the problem of diagnosing these lesions based on the anatomic distribution, endoscopic appearance, and pathologic features of the affected areas of mucosa. A variety of different disease processes may lead to mucosal ulceration. Some are associated with specific pathologic features that may help clarify a specific etiology. Diagnosis requires knowledge of the more common types of colonic ulcers and information regarding methods to identify the etiology based on the clinical, endoscopic, and pathologic features of these lesions.

VASCULAR DISORDERS

Ischemic Colitis

Ischemic colitis is classified as occlusive or nonocclusive disease (Table I). Acute ischemic colitis results from a sudden reduction of splanchnic blood flow and may develop in association with a variety of systemic diseases, including atherosclerosis, diabetes mellitus, cardiovascular disease, vasculitis, hypercoagulable states, and hypertension; low-flow states, including cardiac failure, systemic hypotension, and sepsis; and certain drugs, including oral contraceptives and cocaine. Acute ischemic colitis may also occur as a result of mechanical occlusion of blood flow due to an incarcerated hernia, volvulus, intussusception, or adhesions.

TABLE I Causes of Ischemic Colitis

Etiologic agent
Occlusive disease
Arterial occlusion
Atheroemboli
Thromboemboli
Mechanical disorders
Incarcerated hernia
Volvulus
Intussusception
Extrinsic compression
Nonocclusive disease
Hypotension
Sepsis
Vasculitis

Ischemic colitis may develop at any age, but is most common in older individuals. The clinical manifestations vary according to the underlying etiology, the severity of the injury, and the distribution of disease. Patients typically complain of rapidly progressive crampy abdominal pain associated with abdominal distension, bloody diarrhea, leukocytosis, and fever. Extensive ischemic colitis may be associated with peritoneal signs, ileus, shock, and mental status changes.

Endoscopically, affected areas often demonstrate mucosal ulceration, friability, and erythema, or even pseudomembranes composed of sloughed epithelium admixed with fibrin and inflammatory cells. The distribution of disease tends to reflect the vascular supply to the colon, with a predilection to occur in "watershed" areas that lie within anastomosing branches of the arterial tree. Thus, the splenic flexure and rectosigmoid colon are particularly common areas affected by ischemic injury.

 Copyright 2004, Elsevier (USA). All rights reserved.

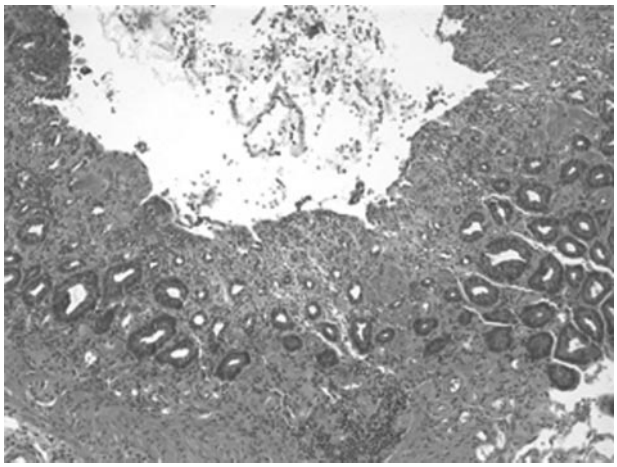


FIGURE 1 Ischemic colitis. The mucosa is ulcerated and the crypts appear small and regenerative. The lamina propria is filled with fibrin and extravasated red blood cells.

Pathologically, the mucosa may show a variety of changes, depending on the stage of disease. Early histologic features of acute ischemic colitis include crypt necrosis, edema, hemorrhage, and fibrin deposition in the lamina propria (Fig. 1). With progression, the mucosa may become entirely necrotic, resulting in the formation of ulcers. Extensive sloughing of mucosa and the development of pseudomembranes may also be apparent. With progression, the deeper layers of the bowel wall may become affected, resulting in transmural necrosis, serositis, and perforation. If the ischemic insult is reversed, or is only mild, the mucosa will undergo healing and repair, which, in some cases, may be complete. However, repeated injury may result in the development of chronic changes such as crypt architectural abnormalities, atrophy, fibrosis, and chronic inflammation, features that can mimic inflammatory bowel disease. Injury to the submucosa and muscularis propria usually heals with scar tissue.

Milder forms of ischemic colitis that are confined to the mucosa are frequently reversible and may be managed expectantly. However, transmural necrosis usually represents a surgical emergency and necessitates surgical resection of the involved areas. Complications of ischemic colitis include bacterial superinfection, particularly by *Clostridium difficile*, stricture, and the development of colonic obstruction.

Vasculitis

Many forms of systemic vasculitic illnesses, such as polyarteritis nodosa, Churg–Strauss syndrome, Henoch–Schönlein purpura, systemic lupus erythematosus, rheumatoid disease, Wegener's granulomatosis, hemolytic uremia syndrome, and Behçet's disease, may involve the gastrointestinal tract, resulting in colonic ulceration. These diseases commonly affect small arteries, capillaries, and venules, and thus may be encountered anywhere in the colon.

Clinical manifestations of vasculitic injury to the colon include abdominal pain and distension, bloody diarrhea, and fever. The endoscopic appearance of the affected colon is variable, depending on the severity of the disease. Mucosal erythema associated with shallow ulcers may be present in milder disease, whereas more extensive disease shows marked mucosal erythema and ulcers, epithelial necrosis, and, occasionally, transmural injury with perforation.

Histologic features that suggest the presence of an underlying vasculitis involve mucosal changes of ischemic colitis, including regeneration of colonic crypts, hemorrhage and fibrin deposition in the lamina propria, and ulcers; combined with the presence of leukocytoclasis and fibrin deposition within blood vessel walls. Also, rare vasculitic syndromes, such as enterocolic lymphocytic phlebitis, are encountered exclusively in the gastrointestinal tract. This unusual form of vasculitis is confined to the intestinal tract and is characterized by the presence of lymphocytic inflammation of medium and small veins.

The management of patients with vasculitis-induced colitis involves the use of immunosuppressive therapy. Patients who develop necrosis are at risk for the development of bacterial superinfection, bleeding, and colonic perforation in the acute setting and may require surgical intervention. Chronic complications, such as colonic strictures or obstruction, may also require surgical resection of the affected areas.

Radiation-Induced Injury

Radiation injury may produce acute or chronic colonic ulcers via two different mechanisms. Radiation is directly toxic to the colonic mucosa and causes mucosal necrosis and ulcers, which develop shortly after the initiation of therapy. However, radiation also causes progressive obliteration of blood vessels and may produce ischemic-type colonic ulcers in the months to years following therapy. Colonic disease usually occurs as a result of radiotherapy for a variety of tumors, such as those that arise from the prostate, urinary bladder, uterus, distal colon, and pelvic bones and soft tissues. The distal colon, particularly the rectosigmoid colon, is most frequently affected. Acute radiation injury, which is manifest approximately 7–14 days after initiation of therapy, is heralded by

the onset of lower abdominal pain, abdominal distension, rectal bleeding, diarrhea, mucoid discharge, and tenesmus. The endoscopic findings include edema, erythema, erosions, and friability. Vascular telangiectasias may present as well.

Pathologically, acute radiation injury is characterized by submucosal edema. However, there may also be active colitis with crypt abscesses, increased neutrophilic inflammation in the lamina propria, and ulceration associated with epithelial cell necrosis. The disease is largely confined to the mucosa, with sparing of the deeper layers of the bowel wall. In most cases, the symptoms of acute radiation colitis are self-limited and resolve with cessation of radiotherapy. Therapy consists predominantly of supportive care. In the months to years following radiotherapy, patients may develop chronic sequelae due to the effects on blood vessels and soft tissues. Radiation-induced vascular ectasia may develop in thin-walled venules and capillaries and lead to gastrointestinal bleeding. Chronic radiation injury results in progressive intimal and fibrous obliteration of arteries, which ultimately compromises the vascular supply to the affected colon and produces ischemic ulceration of the mucosa. In this setting, patients usually present with lower gastrointestinal bleeding due to vascular telangiectasias. If there is an ischemic component to the injury, they may also complain of crampy lower abdominal pain, bloody diarrhea, tenesmus, and/or rectal bleeding.

The endoscopic appearance of chronic radiation injury is characteristic. In many instances, the mucosa is atrophic and vascular ectasias are present. The mucosa may be friable and erythematous with superficial ulcerations as a result of vascular obliteration and ischemic injury. The histologic hallmark of chronic radiation injury is manifest by damage to endothelial cells and blood vessels, particularly arteries. Increased subepithelial collagen deposition may also be seen, in addition to crypt architectural distortion and atrophy of the mucosa (Fig. 2). Affected arteries show luminal obliteration, intimal hyperplasia, and abundant "foam" cells within the muscular wall (Fig. 3). "Atypical" fibroblasts, with abundant eosinophilic cytoplasm and hyperchromatic nuclei, may also be present in the stroma.

Chronic complications of radiation injury include stricture formation and serosal adhesions. These complications may become severe enough to warrant surgical resection. Lower gastrointestinal bleeding from mucosal and submucosal telangiectasias is a frequent occurrence and may be treated with argon plasma coagulation laser therapy. Post-irradiation sarcomas and carcinomas have been reported to develop after a long latency period as well.

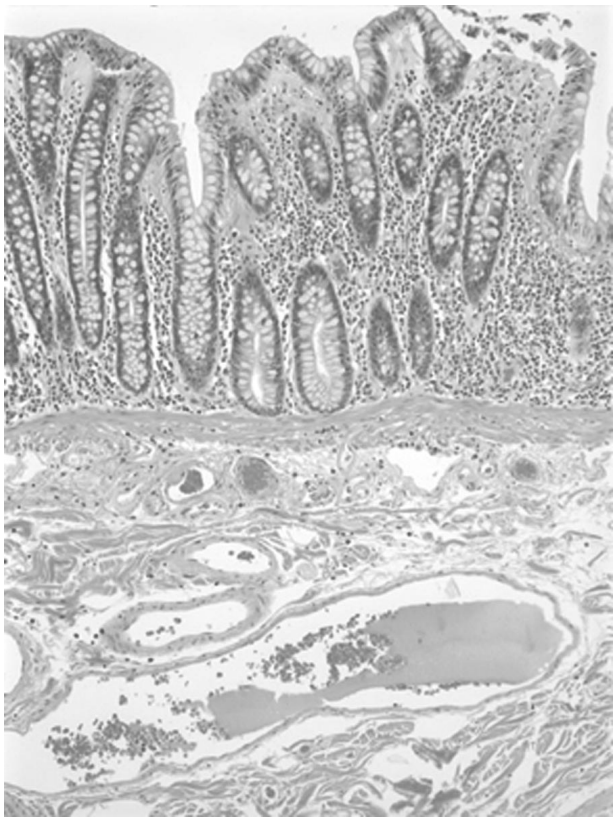


FIGURE 2 Post-irradiation changes include subepithelial collagen deposition and vascular ectasia in the submucosa.

INFECTIOUS COLITIS

Bacterial Colitis

Acute Self-Limited Colitis

Acute self-limited colitis is one of the most common causes of colonic ulcers. The majority of cases develop as a result of exposure to an infectious agent or toxic substance. Unfortunately, an etiologic agent is not identified in up to one-half of cases of acute self-limited colitis. Patients typically present acutely with crampy abdominal pain, bloody diarrhea, and fever. Laboratory tests often reveal leukocytosis and the presence of fecal leukocytes on stool examination.

The endoscopic features of acute self-limited colitis are usually evident within the first 4 days of onset of symptoms. The colonic mucosa appears erythematous, edematous, and friable, with small ulcers and superficial erosions that do not demonstrate a predilection for any specific site within the colon. Pathologically, the features of acute self-limited colitis are quite consistent. Early changes include edema, increased active (neutrophilic) inflammation in the lamina propria and epithelium, and crypt abscesses and ulcers. With

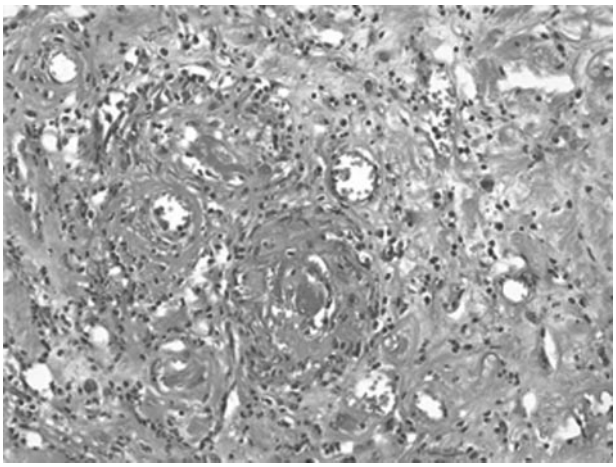


FIGURE 3 Radiation injury induces intimal hyperplasia and luminal narrowing of the submucosal blood vessels. In this example, mural foam cells are associated with inflammation. The submucosa is fibrotic, and scattered atypical stromal fibroblasts are present.

progression, edema becomes less prominent and the neutrophilic inflammation regresses. During the final stages of resolution, the epithelium regenerates to restore normal architecture to the mucosa. Architectural distortion and marked mononuclear cell inflammation of the lamina propria are usually conspicuously absent in cases of acute self-limited colitis. Generally, the disease is confined to the mucosa, with sparing of the deeper layers of the bowel wall. In most instances, acute self-limited colitis resolves spontaneously or following antibacterial therapy. On resolution, there is normally complete restoration of the mucosa to the pre-disease state. Because the disease is usually limited to the mucosa, it generally heals without any significant complications.

Clostridium difficile-Induced Pseudomembranous Colitis

Clostridium difficile is a toxin-secreting gram-positive rod that damages colonic epithelial cells, resulting in sloughing of the mucosa and the development of pseudomembranes (pseudomembranous colitis). In most cases, *C. difficile*-associated pseudomembranous colitis occurs following antibiotic therapy, presumably as a result of alterations of the colonic bacterial flora.

Patients with pseudomembranous colitis develop watery diarrhea, abdominal pain, distension, and leukocytosis. In severe cases, some individuals may show systemic manifestations such as hypotension, fever, and a change in mental status. Serologic measurement of *C. difficile* toxin is a highly sensitive method of detecting infection with this organism. In most cases, the diagnosis of pseudomembranous colitis is evident on endoscopic examination. Punctate or confluent plaques of yellow–tan fibrinous pseudomembranes adherent to an edematous, erythematous, and ulcerated mucosa are often seen (Fig. 4). The right colon is typically more severely affected than the left, but fulminant disease may involve the entire colon.

The pathologic changes are characteristic. The pseudomembranes consist of an admixture of denuded epithelial cells and inflammatory cells enmeshed within fibrinous and mucinous debris, which appear to "erupt" from the colonic crypts (Fig. 5). Initially, the pseudomembranes form discrete plaques; however, in severe disease, the mucosa may become extensively ulcerated and associated with confluent pseudomembranes, signaling a possible evolution toward toxic megacolon. Importantly, a number of infectious agents, including some strains of *Escherichia coli* and *Shigella*, may also cause a pseudomembranous colitis indistinguishable from that due to *C. difficile*.

The management of affected patients is determined by the severity of their illness. The mainstay of medical therapy is antibiotics, such as Flagyl or, less commonly, oral vancomycin. Surgical resection of the colon is reserved for patients who do not respond to antibiotic therapy or develop systemic toxicity despite medical therapy, or develop toxic megacolon. In addition to toxic megacolon, other complications may occur, including colonic perforation, which is associated with a high mortality rate.

Viral Colitis

Although a variety of different viruses may cause disease in the gastrointestinal tract, most are not pathogenic in the colon, and, with rare exception, the development of colonic ulcers occurs as a result of infection by only a few different viruses.

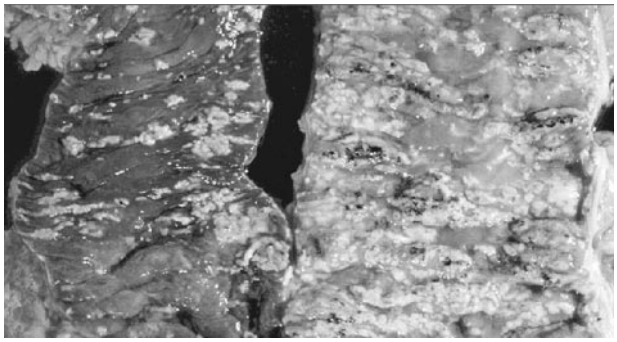


FIGURE 4 Pseudomembranous colitis. The mucosa is covered by numerous pseudomembranes. The background mucosa is edematous and erythematous.

FIGURE 5 Pseudomembranous colitis. Aggregates of cellular debris appear to "erupt" from the damaged mucosa. These aggregates are composed of necrotic epithelial cells enmeshed in mucin and fibrinopurulent material.

Cytomegalovirus Colitis

Cytomegalovirus (CMV) may cause a severe, life-threatening form of colitis associated with ulcers. Although CMV colitis has been reported to occur rarely in immunocompetent individuals, it is much more commonly encountered among immunocompromised patients. In this setting, the infection usually represents reactivation of latent disease that becomes apparent following immunosuppression. CMV demonstrates a predilection to infect rapidly proliferating fibroblasts and endothelial cells present in granulation tissue of ulcers, so it may complicate pre-existing colonic diseases such as inflammatory bowel disease and ischemic colitis.

Most patients with CMV colitis complain of fever, bloody diarrhea, and abdominal pain. In some cases, CMV colitis develops as a complication of immunosuppressive therapy for ulcerative colitis or Crohn's disease, such that the clinical picture is one of progressively severe colitis that persists despite increasing levels of immunosuppression. Serologic assays for CMV, including reverse transcription polymerase chain reaction (RT-PCR), may be extremely useful in establishing the correct diagnosis in these cases.

The endoscopic appearance of CMV colitis is nonspecific. Generally, there is a predilection for the right colon, but this is not always the case. The distribution of disease is usually segmental, characterized by the presence of severe colitis with edema and erythema of the mucosa, as well as deep ulcers. CMV induces a nonspecific histologic pattern of injury to the colonic mucosa. Frequently, the mucosa is ulcerated and the affected biopsy fragments consist entirely of granulation tissue and fibrinopurulent debris. Biopsy fragments obtained from the non-ulcerated mucosa often show active colitis with cryptitis and crypt abscesses. The diagnosis of CMV colitis rests on identification of pathognomonic viral inclusions in the cytoplasm and nuclei of infected cells by routine stains or by immunohistochemistry. The virus typically infects endothelial cells and fibroblasts, producing granular, brightly eosinophilic cytoplasmic inclusions as well as large eosinophilic "owl's eye" intranuclear inclusions (Fig. 6). Occasionally, viral inclusions may be encountered in the glandular epithelium as well, but these are much less common.

Treatment of CMV colitis includes the use of antiviral therapy with ganciclovir. Complications include systemic manifestations of infection, superinfection with other organisms, and colonic perforation in the acute setting.

Herpes Simplex Virus Colitis

Herpes simplex viruses (HSV I and HSV II) may cause colonic ulcers and are an important cause of proctitis in some patient populations. Importantly, HSV may cause disease in both immunocompetent and immunocompromised individuals. In immunocompetent patients, the disease usually occurs as a result of direct contact between the colonic mucosa and infected secretions or mucosal lesions. As a result, HSV-associated ulcers are almost exclusively encountered in the distal rectum and anus. In immunocompromised individuals, however, the disease may occur as a result of either direct contact or as a manifestation of systemic disease in otherwise severely ill patients.

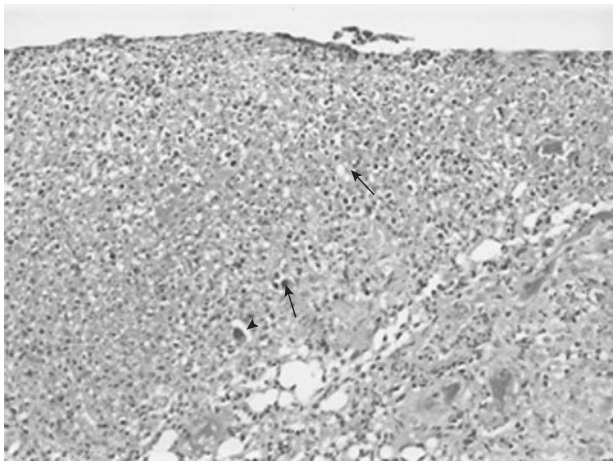


FIGURE 6 Cytomegalovirus colitis. This biopsy fragment consists entirely of inflamed granulation tissue. Many of the stromal and endothelial cells contain both cytoplasmic (arrow) and intranuclear (arrowhead) inclusions.

Patients present with anorectal pain, constipation, tenesmus, pain on defecation, and gastrointestinal bleeding or bloody diarrhea. Characteristic vesicular lesions may be present on the anorectal mucosa or perianal skin. Endoscopically, the viral-infected lesions are often located circumferentially within the affected colon and usually appear as small vesicles on a background of erythematous mucosa. As the vesicles enlarge, they frequently rupture, resulting in the formation of shallow, well-demarcated ulcers.

The histologic features of the ulcers are not specific. They are often associated with a marked inflammatory infiltrate rich in neutrophils, as well as a fibrinopurulent exudate. Diagnostic multinucleated giant cells with "ground glass" or amphophilic intranuclear inclusions and nuclear molding are most commonly encountered in the stromal cells of the submucosa and can be identified on routine stains or by immunohistochemistry.

Treatment of herpetic colitis/proctitis includes the management of symptoms in the acute setting as well as the use of antiviral medications. In most instances, the lesions resolve completely with therapy and are not associated with complications.

Fungal Colitis

Fungal colitis is extremely rare and is encountered almost exclusively as a terminal event in seriously ill individuals who are also immunocompromised. Some of the more common organisms associated with fungal colitis include *Aspergillus* and *Candida* species. Both cause a severe, ulcerative form of colitis characterized by the presence of neutrophilic abscesses as well as invasive yeast forms within the lamina propria.

Protozoan-Induced Colitis

Amebiasis

Amebiasis, particularly that due to *Entamoeba histolytica*, is a major cause of ulcerative colonic disease and dysentery worldwide. The trophozoites cause disease by adhering to epithelial cells and invading the underlying mucosa, destroying the tissue via elaboration of a variety of proteolytic enzymes and cytolytic agents.

Patients typically complain of acute-onset abdominal pain, bloody diarrhea with mucus, tenesmus, and fever. The correct diagnosis is often suspected on endoscopic examination because the organisms demonstrate a predilection for the cecum and right colon. However, the entire colon may be involved in severe cases. The ulcers are usually deep, flask shaped, and exudative, but may be confluent in severe cases. The adjacent mucosa may be erythematous, but is usually relatively unaffected.

The pathologic features of amebic colitis include deep, excavating ulcers that are typically associated with a marked inflammatory infiltrate (Fig. 7). Abundant fibrinopurulent material is often present overlying the ulcers and, although the organisms may be found within the mucosa, they are more frequently identified in the overlying necroinflammatory debris. *Entamoeba histolytica* trophozoites also ingest red blood cells, thus the presence of red cell fragments within the organisms confirms the diagnosis.The treatment of amebic colitis consists of antiprotozoal therapy and supportive care. Although complete resolution and restoration of the colonic mucosa may occur following treatment,

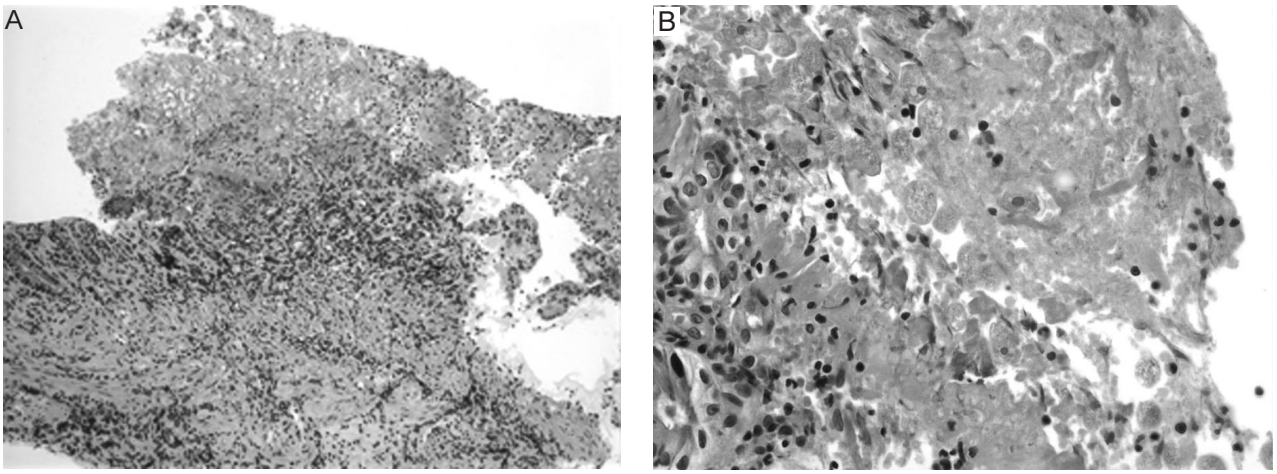


FIGURE 7 Amebiasis of the colon results in the development of deeply ulcerating lesions associated with a fibrinous exudate and mucosal necrosis (A). The organisms are usually apparent within necrotic debris on the mucosal surface (B).

complications, including the development of colonic strictures secondary to healing of deep ulcers, are not uncommon.

DRUG-INDUCED COLONIC ULCERS

Nonsteroidal Antiinflammatory Drugs

It is well-known that nonsteroidal antiinflammatory drugs (NSAIDs) cause ulcers in the gastrointestinal tract by inhibiting prostaglandin synthesis in the mucosa and altering local blood flow. In fact, NSAID use is one of the most common causes of gastrointestinal bleeding. Although NSAID use is more commonly associated with upper gastrointestinal disease, it has recently become clear that inflammation and ulceration may involve the colon as well. The clinical manifestations of NSAID-induced colonic injury are variable. In some cases, patients are asymptomatic and lesions are identified during endoscopic colon evaluation performed for other reasons. Symptomatic patients may complain of rectal bleeding, abdominal pain, and/or bloody diarrhea.

Typically, NSAIDs cause an acute self-limited type of colitis that is endoscopically indistinguishable from that due to infections. Ulceration occurs more often in the cecum and ascending colon. The mucosa may show erythema and edema with or without superficial ulcers. However, deeper ulcers associated with perforation may rarely occur. The use of NSAID-containing suppositories has also been associated with severe ulcerating proctitis. The pathologic features of NSAID-induced colitis are essentially non-specific. The mucosa is infiltrated with a mixed inflammatory cell population rich in neutrophils. Cryptitis and crypt abscesses, as well as mucosal ulcers, are variably present. The use of NSAIDs has also been linked to the development of collagenous and lymphocytic colitis, which are characterized by the presence of increased stromal and intraepithelial lymphocytes, eosinophils, and a thickened subepithelial collagen layer in the former.

The clinical and pathologic features of NSAID-induced colonic injury usually resolve with cessation of drug intake and are not normally associated with chronic mucosal injury, except in rare and unusual circumstances, such as the development of "diaphragm disease." This condition is characterized by the development of colonic obstruction due to the presence of web-like strictures composed of mucosa and submucosa, which cause luminal stenosis.

Oral Contraceptives

Estrogen and progesterone compounds may induce ischemic lesions in the colon due to decreased splanchnic blood flow and the development of small thrombi in the mesenteric veins and small venules. Patients may be asymptomatic or may present with lower abdominal pain, rectal bleeding, or bloody stools. In most instances, lesions involve the colon in a segmental fashion, resulting in mucosal erythema. Erosions and superficial ulcers as a result of epithelial cell necrosis may also be identified.

The pathologic features are indistinguishable from those of ischemic colitis due to other causes and consist of mucosal ulceration, fibrin deposition in the lamina propria, and regeneration of the colonic crypts. In most instances, the mucosal changes are completely reversible on cessation of contraceptives. However, severe, transmural disease may result in perforation or scarring.

Vasoconstrictive Drugs

Other medications that decrease splanchnic blood flow, such as cocaine and vasopressive agents, may result in ischemic colitis and ulceration. These medications may produce ischemic-type ulcers in the colon, resulting in gastrointestinal bleeding. The endoscopic appearance is that of a segmental area of colitis with mucosal erythema, edema, and ulcers. The histologic features include mucosal hemorrhage, fibrin deposition, and regeneration of injured colonic crypts associated with ulceration. Treatment is supportive and most of the changes are reversible on cessation of the drug.

Enemas

A number of different bowel preparations and enema solutions may cause colonic injury and mucosal ulceration. Mildly irritating enema solutions produce mucosal edema, but hypertonic solutions [Fleet PhosphoSoda (CB Fleet Co, Lynchburg, VA), bisacodyl, or hydrogen peroxide] may cause more severe colonic injury, such as mucosal erosions. The histologic changes are generally mild and consist of a variable amount of epithelial cell mucin depletion and acute inflammation.

In contrast, oral preparations and enemas that contain Kayexelate have been associated with the development of severe intestinal ulceration, which may be life threatening. Histologically, the lesions are characterized by necrosis, ulceration, and granulation tissue, which may contain refractile basophilic (Kayexelate) particles (Fig. 8). Interestingly, this complication occurs in association with Kayexelate preparations that contain a sorbitol carrier (Sanofi Winthrop Pharmaceuticals, NY), but does not develop with preparations that contain a lactulose carrier. Thus, most believe that the injury is the result of sorbitol, because similar lesions

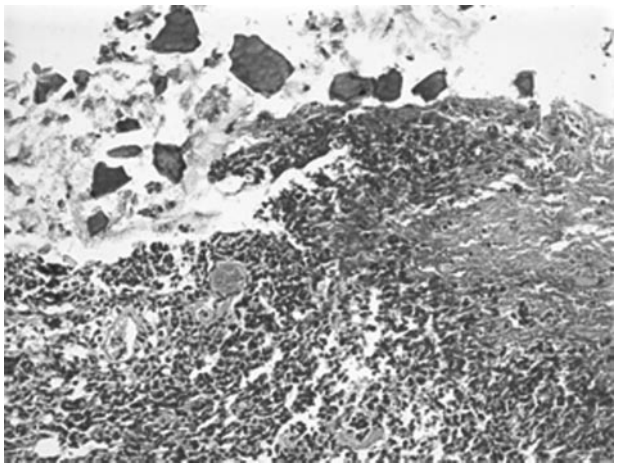


FIGURE 8 Kayexelate-induced colonic injury. Kayexelate preparations that contain a sorbitol carrier may induce ulceration. Kayexelate crystals may be identified in the mucosa or in ulcer debris and appear as rhomboid-shaped structures.

have been induced in animals using sorbitol enemas without Kayexelate.

MECHANICAL DISORDERS

Diverticulitis

Colonic ulceration may occur uncommonly in the setting of acute diverticulitis when the inflammation involves the mucosal surface. However, in some cases, an unusual form of chronic colitis, "diverticular disease associated colitis," may develop in patients with diverticulosis. This disease shows segmental mucosal erythema, friability, and ulceration, which is more pronounced at the orifices of the diverticula but involves the interdiverticular mucosa as well. The pathologic features include mucosal ulcerations, increased mononuclear cell inflammation of the lamina propria and deeper layers of the bowel wall, transmural lymphoid aggregates, and crypt architectural changes, such as crypt atrophy and/or branching. Treatment includes medical treatment of the diverticulitis as well as immunosuppressive therapy. Surgical intervention may be necessary in some cases that are refractory to medical management.

Trauma

Traumatic causes of colonic ulceration are uncommon and are generally limited to the distal rectum. Most cases result from foreign bodies, which may produce mucosal ulceration or pressure-induced ischemic injury. The changes may vary from mild to severe proctitis. Chronic injury may result in the development of inflammatory polyps and lumenal stenosis.

Obstruction and Pseudo-obstruction

Mechanical obstruction and pseudo-obstruction may produce colonic ulceration as a result of stasis of fecal contents ("obstructive colitis"). Initially, mucosal changes such as patchy erythema and ulceration may be identified. However, as the condition persists, histologic evidence of ischemic colitis may develop.

Solitary Rectal Ulcer Syndrome (Mucosal Prolapse)

Solitary rectal ulcer syndrome is an inflammatory disorder of the distal colon; it results from ischemia, torsion, and twisting of the mucosa due to chronic constipation and straining on defecation. It is associated with a primary malfunction of the internal anal sphincter and/or rectal musculature, and is more commonly encountered in older females. Affected patients may present with a variety of clinical manifestations, including rectal bleeding, pain on defecation, and incomplete rectal emptying. Some cases are entirely asymptomatic and are discovered only incidentally on routine examination.

The endoscopic features of solitary rectal ulcer syndrome are variable because they are temporally related to the stage of disease. Features include ulceration, erythema, and inflammatory polyps. Most lesions occur on the anterior wall of the distal two-thirds of the rectum. However, other areas of the rectum and sigmoid colon may be affected as well. At the time of endoscopic evaluation, mucosal prolapse may be demonstrated if the patient is asked to "bear down" during the procedure.

Histologically, the disease is characterized by the presence of ulceration, hemorrhage in the lamina propria, increased active inflammation, and regeneration and hyperplasia of the crypts. Inflammatory polyps typically have a smooth surface, but may be irregularly shaped and/or ulcerated. The polyps may contain regenerative crypts that are cystically dilated, resulting in a localized form of colitis cystica profunda. The mucosa of both ulcerating lesions and inflammatory polyps demonstrates fibromuscular hyperplasia of the lamina propria, which represents a proliferation of smooth muscle fibers arising from the muscularis mucosae. Fibromuscularization of the lamina propria results from chronic prolapse of the mucosa into the lumen, where the increased pressure and trauma induce a localized form of ischemic injury (Fig. 9).

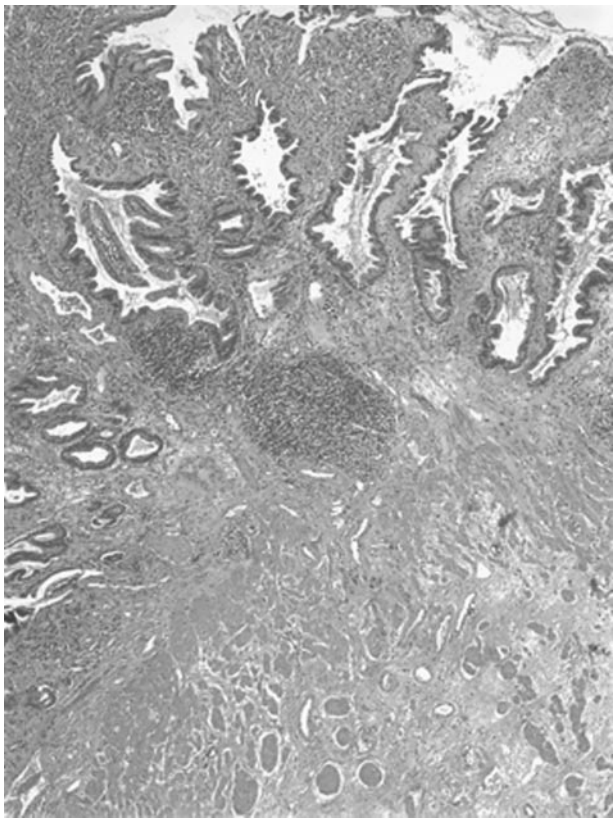


FIGURE 9 Solitary rectal ulcer syndrome (mucosal prolapse syndrome) may produce localized ischemic changes and ulceration in the mucosa. This polypoid lesion contains numerous cystically dilated hyperplastic colonic crypts. Acute and chronic inflammation and fibromuscular hyperplasia are present in the lamina propria.

Because solitary rectal ulcer syndrome/mucosal prolapse results from a mechanical disorder of sphincter and pelvic floor function, the treatment of these patients depends on dietary modification, stool softeners, and behavior modification, as well as medical management of symptoms. Although the polypoid lesions of solitary rectal ulcer syndrome may be resected or biopsied endoscopically, these lesions are generally not responsive to surgical management.

Stercoral Ulcers

Stercoral ulcers are defined as lesions that occur due to fecal impaction; they are therefore more commonly encountered in the distal colon. Ulcers typically develop following periods of severe constipation. Patients complain of severe constipation, abdominal pain, rectal pain, pain on defecation, and bleeding. The endoscopic appearance of stercoral ulcers is characteristic. Most lesions are small, range in size from a few millimeters to a few centimeters in diameter, are normally well demarcated, and are superficial in nature. The adjacent mucosa is frequently edematous and congested. Within the ulcerated area, there is tissue necrosis with loss of the epithelium and fibrin deposition in the lamina propria. Hemorrhage and entrapped fecal material may be present as well. With time, granulomatous inflammation and foreign-body-type giant cells develop in areas of entrapped fecal material.

The treatment of stercoral ulcers is mainly supportive. This includes the use of stool softeners and dietary modification. Most cases heal without sequelae. However, rare cases may be associated with colonic perforation, strictures, and obstruction.

IDIOPATHIC INFLAMMATORY BOWEL DISEASE

Crohn's Disease

Colonic involvement in Crohn's disease is common. It occurs in approximately 50% of patients with Crohn's disease, affecting other sites in the gastrointestinal tract, and the colon is the only site of disease in approximately 30% of patients. Affected patients complain of lower abdominal pain that may be crampy in nature, abdominal bloating, and bloody diarrhea. Many patients are anemic at the time of presentation and may also have a history of weight loss, particularly if they suffer from small intestinal disease as well.

The endoscopic manifestations of the disease include several different patterns of injury. Generally, colonic involvement in Crohn's disease is segmental in distribution. However, diffuse colonic involvement may occur as well. In milder cases, the affected mucosa is erythematous and edematous, with multiple punctate or aphthous ulcers. However, deep fissuring ulcers arranged in a longitudinal fashion are often present in patients with more severe disease. Luminal stenosis as a result of transmural injury and scarring may occur in chronic disease as well.

Microscopically, Crohn's disease is characterized by acute and chronic changes. Nonnecrotic granulomas are present in approximately 30% of cases. However, their presence is merely a marker of the disease and does not necessarily imply the presence of disease activity. Acute changes include segmental inflammation of the colonic mucosa with crypt abscesses and associated ulcers. The ulcers that occur in the setting of Crohn's disease are morphologically variable and include aphthous, flask-shaped, linear, and fissuring-type lesions. Most commonly encountered are aphthous ulcers, which are shallow, well-delineated

lesions that are confined to mucosa or superficial submucosa and tend to occur over lymphoid aggregates. Deeper ulcers may be flask shaped with a broad base centered in the submucosa, or fissuring, knifelike cracks lined by granulation tissue extending into the submucosa or muscularis propria (Fig. 10). Linear ulcers that course through the submucosa parallel to the mucosal surface are also a common feature and are typically lined by granulation tissue and marked acute and chronic inflammation. Chronic changes that develop in Crohn's disease include mucosal abnormalities, such as crypt architectural changes, Paneth cell metaplasia, increased mononuclear cell inflammation, and basal plasmacytosis. Crohn's disease may also cause chronic changes in the colonic wall, including transmural lymphoid aggregates, hypertrophy of nerves in the muscularis propria and submucosa, mural fibrosis, strictures, and fistulous tracts.

Immunosuppression of disease activity is the mainstay of therapy for Crohn's disease. Unfortunately, patients with Crohn's disease frequently develop complications of the disease, such as fistulas, sinus tracts, fissures, and luminal obstruction due to strictures. These complications usually require surgical intervention. Additionally, patients with colonic Crohn's disease should undergo regular surveillance for the possibility of epithelial dysplasia and carcinoma.

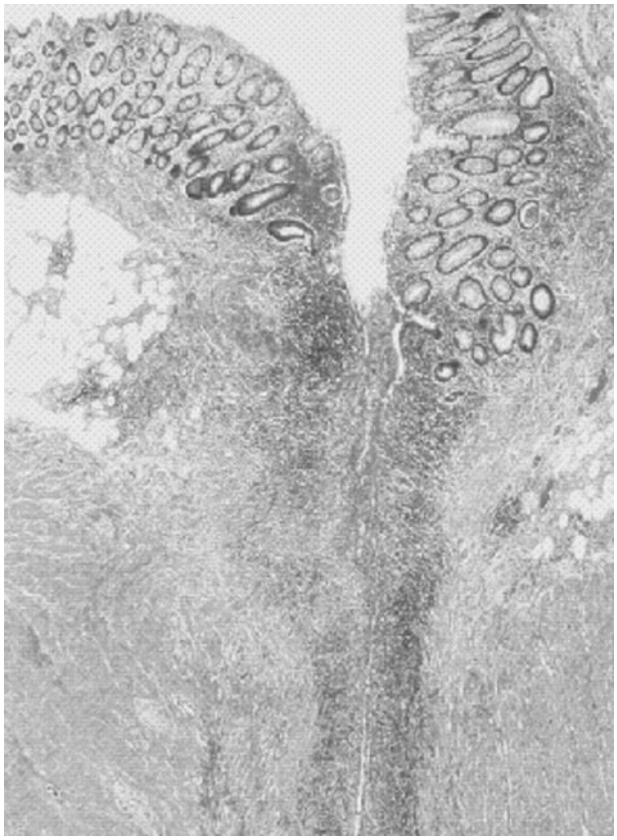


FIGURE 10 Crohn's disease. Sharply delineated fissuring ulcer lined by granulation tissue may involve all layers of the bowel wall.

Ulcerative Colitis

Ulcerative colitis is a chronic ulcerating disease of the colon that demonstrates a bimodal age distribution, with most cases occurring in young adults and a second peak in late adulthood. Although the mechanisms underlying its development are not entirely clear, it appears to be a multifactorial immune-mediated process that results from the complex interplay of a variety of genetic and environmental factors.

Patients typically present with crampy abdominal pain, bloody diarrhea, mucus discharge, and fever. Extraintestinal manifestations of the disease are not uncommon and include large-joint arthritis, primary sclerosing cholangitis, skin manifestations (erythema nodosum and pyoderma gangrenosum), sacroileitis, and ocular disease. Endoscopically, the mucosa is friable, erythematous, and frequently eroded or frankly ulcerated. Although ulcerative colitis classically involves the colonic mucosa in a contiguous retrograde fashion, this is not always the case. The distribution of disease may be discontinuous, involving the left side of the colon, sparing the middle and ascending colon, and involving the cecum and appendix. Such a pattern of disease has been termed a "cecal patch" of ulcerative colitis and does not imply the presence of Crohn's disease. Additionally, endoscopically and pathologically evident "skip areas" may be present early in the course of the disease. However, the process usually becomes confluent over time. Finally, skip areas of uninflamed colonic mucosa are frequently encountered in patients who have received oral or topical therapy for ulcerative colitis.

The pathologic features of ulcerative colitis include the presence of active inflammation on a background of chronic colitis. Depending on the severity of the disease, there may be a variable amount of cryptitis, crypt abscesses, or mucosal ulceration associated with an acute and chronic inflammatory exudate. Chronic changes include the presence of a lymphoplasmacytic inflammatory cell infiltrate, crypt architectural distortion, crypt atrophy, Paneth cell metaplasia, and basal plasmacytosis (Fig. 11). As the disease progresses, the ulcers become confluent, resulting in the formation of mucosal pseudopolyps

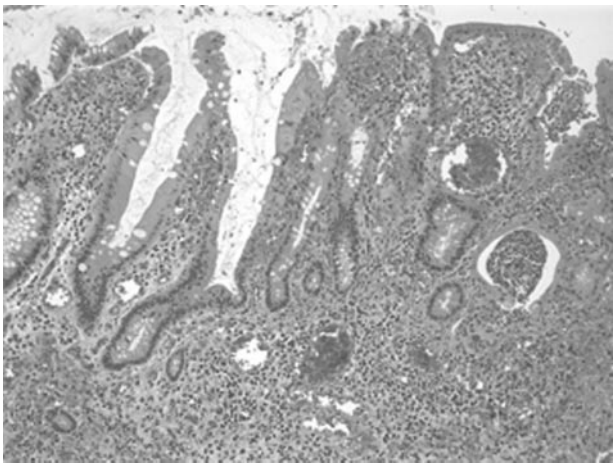


FIGURE 11 Ulcerative colitis. Active changes include the presence of crypt abscesses and superficial mucosal ulcerations. Chronic changes, including crypt atrophy and architectural distortion, increased mononuclear cell inflammation of the lamina propria, and basal plasmacytosis, are also present.

that represent islands of residual intact mucosa surrounded by ulcerated tissue.

Most cases of ulcerative colitis are managed with medical immunosuppression. Surgical intervention may be required in cases refractory to medical therapy or in the setting of toxic megacolon. Patients with ulcerative colitis are at risk for the development toxic megacolon and colonic perforation in the acute setting, as well as several complications of long-standing disease. Chronic immunosuppression predisposes patients to colonic superinfection with cytomegalovirus and other opportunistic organisms. Chronic complications of ulcerative colitis also include glandular dysplasia and the development of carcinoma.

OTHER COLITIDIES

Allergic Colitis

Allergic colitis may result in colonic ulceration as a result of hypersensitivity to foodstuffs. This disease most commonly affects infants and demonstrates a predilection for males. Patients generally present with bloody diarrhea that promptly resolves on withdrawal of the offending agent, which, in most cases, is cow's milk or soy protein. Rarely, breast milk may be the inciting agent. Peripheral eosinophilia may also be present, but this abnormality is not necessary in order to establish a diagnosis. Endoscopically, areas of mucosal erythema, occasionally associated with small ulcerations, may be observed.

The histologic changes of allergic colitis consist of an increase in eosinophils in the lamina propria and epithelium associated with eosinophilic cryptitis and rarely, mucosal erosions or ulceration. Mild lesions demonstrate focal infiltration of the crypts, and multiple sections may be necessary to demonstrate the lesions. More severe disease results in necrosis and ulceration of the surface epithelium. Treatment of the disorder requires discontinuation of the offending agent. In most instances, the mucosal injury heals completely and, thus, the disease is not associated with any significant sequelae.

Endometriosis

Colonic endometriosis typically results in mural fibrosis, inflammation, and stricture formation. However, it rarely involves the colonic mucosa, but, in these cases, may induce ulceration and lower gastrointestinal bleeding. The endoscopic appearance varies from mucosal erythema and ulceration to the presence of polypoid luminal masses (inflammatory polyps) or even ulcerative lesions that may resemble malignancy.

The pathologic features of endometriosis are characteristic. Aggregates of endometrial glands enmeshed in a myxoid or cellular stroma are typically present in the wall of the colon. Frequently, these aggregates are associated with marked hypertrophy of the muscularis propria as well as increased fibrosis of the muscularis propria and submucosa. Endometriosis involving the colonic mucosa may induce changes reminiscent of active colitis with cryptitis, increased neutrophilic inflammation of the lamina propria, and ulceration. However, endometrioid glands composed of mucin-depleted cells, some of which are ciliated, as well as endometrioid stroma, are features distinct from those of the native lamina propria.

The management of colonic endometriosis is variable and dependent on the colonic manifestations of the disease. Endometriosis associated with mucosal ulceration and colitis may be medically managed with hormonal therapy, whereas mass lesions may necessitate surgical resection due to luminal obstruction.

Idiopathic Ulcers and Isolated Cecal Ulcers

In some instances, the etiologic agent responsible for colonic ulceration may not be readily identifiable. These "idiopathic ulcers" demonstrate a predilection for the antimesenteric border of the cecum and proximal ascending colon and, less commonly, the sigmoid colon. As a result, they have occasionally been termed "isolated cecal ulcers." The anatomic distribution of disease

suggests that these lesions may occur as a result of a preexisting lesion, such as congenital diverticula or angiodysplasias in older individuals. However, many are due to an unidentified toxic metabolic agent or drug, such as NSAIDs.

Idiopathic or isolated cecal ulcers may occur in all age and gender groups. Right-sided lesions often present with occult blood loss, frank bleeding, or bloody diarrhea, whereas left-sided ulcers usually present with hematochezia, colonic obstruction, tenesmus, rectal pain, or pain on defecation.

The endoscopic appearance of these lesions is nonspecific. In many cases, they are solitary and range from one to several centimeters in greatest diameter. In most cases, they are superficial. However, they may penetrate deep into the colonic wall. The pathologic features of ulcers are also entirely nonspecific. Increased acute and chronic inflammation, hemorrhage, granulation tissue, and fibrosis are seen in the base of the ulcer. The adjacent, unaffected mucosa may be normal or may demonstrate mild changes with increased mixed cellular inflammation, edema, and hemorrhage.

The treatment of these lesions is variable and is determined by the severity and extent of the ulcers. Small, superficial ulcers may be managed expectantly. Large or deep ulcers may require surgical resection. Complications of these lesions include massive lower gastrointestinal bleeding and colonic perforation.

TUMORS

A variety of malignant tumors of the colon may present as endoscopically apparent ulcerated masses. Most commonly, these lesions are conventional adenocarcinomas. However, other lesions, such as lymphomas, gastrointestinal stromal tumors, unusual sarcomas, and metastases to the colon, may also infiltrate the colonic mucosa, resulting in mucosal ulceration.

See Also the Following Articles

Colitis, Pseudomembranous • Colitis, Radiation, Chemical, and Drug-Induced • Colitis, Ulcerative • Colonoscopy • Crohn's Disease • Diverticulosis • NSAID-Induced Injury • Rectal Ulcers • Solitary Rectal Ulcer Syndrome

Further Reading

Baird, D. B., and Norris, H. T. (1998). Vascular disorders. *In* "Pathology of the Gastrointestinal Tract" (S. Ming, and H. Goldman, eds.), pp. 283–289. Williams & Wilkins, Baltimore.

Berthrong, M., and Fajardo, L. F. (1981). Radiation injury in surgical pathology. Part II. Alimentary tract. *Am. J. Surg. Pathol.* **5**, 153–178.

Bjarnason, I., and Price, A. B. (1994). The small and large intestinal pathologies of non-steroidal anti-inflammatory drug ingestion. *Ann. Pathol.* **14**(5), 326–332.

du Boulay, C. E., Fairbrother, J., and Isaacson, P. G. (1983). Mucosal prolapse syndrome—A unifying concept for solitary ulcer syndrome and related disorders. *J. Clin. Pathol.* **36**(11), 1264–1268.

Ford, M. J., Anderson, J. R., and Gilmour, H. M. (1983). Clinical spectrum of "solitary ulcer" of the rectum. *Gastroenterology* **84**(6), 1533–1540.

Gekas, P., and Schuster, M. M. (1981). Stercoral perforation of the colon: Case report and review of the literature. *Gastroenterology* **80**(5, Pt. 1), 1054–1058.

Goldman, H. (1998). Allergic disorders. *In* "Pathology of the Gastrointestinal Tract" (S. Ming, and H. Goldman, eds.), pp. 235. Williams & Wilkins, Baltimore.

Goldman, H. (1998). Other inflammatory disorders of the intestines. *In* "Pathology of the Gastrointestinal Tract" (S. Ming, and H. Goldman, eds.), pp. 732–738. Williams & Wilkins, Baltimore.

Halter, F., Weber, B., and Huber, T. (1993). Diaphragm disease of the ascending colon. Association with sustained-release diclofenac. *J. Clin. Gastroenterol.* **16**(1), 74–80.

Hardin, R. D., and Tedesco, F. J. (1986). Colitis after hibiclens enema. *J. Clin. Gastroenterol.* **8**(5), 572–575.

Kang, Y. S., Kamm, M. A., and Engel, A. F. (1996). Pathology of the rectal wall in solitary rectal ulcer syndrome and complete rectal prolapse. *Gut* **38**(4), 587–590.

Langlois, N. E., Park, K. G., and Keenan, R. A. (1994). Mucosal changes in the large bowel with endometriosis: A possible cause of misdiagnosis of colitis? *Hum. Pathol.* **25**(10), 1030–1034.

Meisel, J. L., Bergman, D., and Graney, D. (1977). Human rectal mucosa: Proctoscopic and morphological changes caused by laxatives. *Gastroenterology* **72**(6), 1274–1279.

Meyer, C. T., Brand, M., and DeLuca, V. A. (1981). Hydrogen peroxide colitis: A report of three patients. *J. Clin. Gastroenterol.* **3**(1), 31–35.

Nostrant, T. T., Kumar, N. B., and Appelman, H. D. (1987). Histopathology differentiates acute self-limited colitis from ulcerative colitis. *Gastroenterology* **92**(2), 318–328.

Rashid, A., and Hamilton, S. R. (1997). Necrosis of the gastrointestinal tract in uremic patients as a result of sodium polystyrene sulfonate (kayexalate) in sorbitol: An underrecognized condition. *Am. J. Surg. Pathol.* **21**(1), 60–69.

Rutter, K. R., and Riddell, R. H. (1975). The solitary ulcer syndrome of the rectum. *Clin. Gastroenterol.* **4**(3), 505–530.

Scott, T. R., Graham, S. M., and Schweitzer, E. J. (1993). Colonic necrosis following sodium polystyrene sulfonate (kayexalate)-sorbitol enema in a renal transplant patient. Report of a case and review of the literature. *Dis. Colon Rectum* **36**(6), 607–609.

Surawicz, C. M., Haggitt, R. C., and Husseman, M. (1994). Mucosal biopsy diagnosis of colitis: Acute self-limited colitis and idiopathic inflammatory bowel disease. *Gastroenterology* **107**(3), 755–763.

Yantiss, R. K., Clement, P. B., and Young, R. H. (2001). Endometriosis of the intestinal tract: A study of 44 cases of a disease that may cause diverse challenges in clinical and pathologic evaluation. *Am. J. Surg. Pathol.* **25**(4), 445–454.

Colonoscopy

TROY D. SCHMIDT
Duke University Medical Center

argon plasma photocoagulation Use of ionized argon gas to produce a high-frequency current for the purpose of thermally coagulating tissues.
conscious sedation Level of sedation in which the patient is relaxed, comfortable, and perhaps in a state of light sleep. The patient is easily arousable and able to protect his or her airway.
digital rectal examination Use of the finger to manually inspect the anus, anal canal, and lower rectum for palpable or visual lesions.
heparin window Use of intravenous or subcutaneous heparin in between the time a patient, requiring anticoagulation, stops taking a long-acting oral anticoagulant (e.g., warfarin) and the time the procedure is performed. This practice is aimed to reduce thromboembolic risk while off warfarin for an invasive procedure such as colonoscopy.
informed consent Agreement and acknowledgment of understanding by the patient for a procedure or intervention following an explanation of the risks, benefits, and alternatives.
insufflation Installation of air into the colon for the purpose of distending the walls for examination.
looping A suboptimal curvature of the colonoscope within the colon, making it difficult to advance the colonoscope.
pneumatosis Gas within the submucosa or subserosa of the bowel wall, indicative of a loss of mucosal integrity. This may represent partial or impending perforation.

The first report of a completed colonoscopy was in 1969. Since that time, colonoscopes have evolved and colonoscopy has proven to be a very important tool for gastroenterologists. Colonoscopy allows for a variety of diagnostic and therapeutic options. For example, physicians use colonoscopy to diagnose or exclude inflammatory bowel disease and other forms of colitis. Physicians can diagnose and treat bleeding sources such as ulcers, angioectasiae, and polypectomy sites. Furthermore, the procedure allows for other therapeutic options such as stricture dilation/stenting, foreign body removal, and colon decompression. Because it gives such detailed views of the colonic mucosa, colonoscopy is the most sensitive examination for diagnosing colon polyps and cancer. As a result, it is now a commonly used screening tool for colorectal cancer. This article reviews the colonoscopy procedure, proper indications, complications, and patient preparation.

COLONOSCOPES

The modern colonoscope (Fig. 1) is a flexible instrument with a light and lens at its tip for visualization. It has four-way tip deflection and is typically 160–170 cm long. Control cables run the length of the shaft and are manipulated using two dials at the head of the instrument (Fig. 2). The light seen at the tip of the shaft is connected to the head with fiber-optic cables and these cables are further connected to an imaging console through what is frequently called "the umbilical cord." Fiber-optic connections allow the display of images onto a color monitor and this monitor is placed in front of the endoscopist for visualization. Having images displayed in such a manner allows both the endoscopist and the assistant to see lesions and coordinate the use of therapeutic instruments. It also allows for teaching during the examination because multiple people can watch the procedure in real time.

Colonoscopes have two buttons near the control dials (Fig. 2). When depressed, one button results in suction of colonic contents through the colonoscope and into a waste canister. The endoscopist insufflates

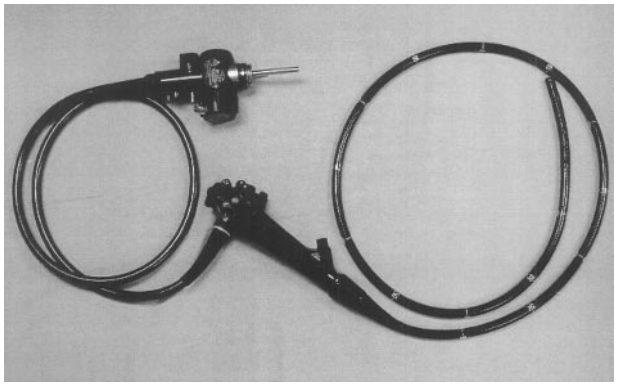


FIGURE 1 The modern colonoscope.

 Copyright 2004, Elsevier (USA). All rights reserved.

the colon by covering a small hole in the center of the second button. When fully depressed, this button cleans the lens with a spray of water.

Colonoscopes have either one or two channels located just below the instrument head (Fig. 2). These are often referred to as either suction or instrument channels because they are used for both. The channels allow endoscopists to flush the colon with water and then suction out residual stool or blood. The endoscopists can also suction out air or small, resected polyps. Polyps are collected in a trapping device before being suctioned into the main waste canister. All available therapeutic instruments are passed through the same channels.

Finally, some manufacturers have recently produced colonoscopes with adjustable stiffness capabilities. This allows the endoscopist to make the colonoscope more rigid as it is passed through the colon. Some studies have reported faster cecal intubation rates and improved patient satisfaction with these newer colonoscopes. An example of this type of colonoscope is found in Fig. 2.

THE PROCEDURE

Colonoscopy requires informed consent from the patient. The physician discusses the reasons for the examination, explains the procedure, and then lists the options, potential benefits, and potential risks. After the patient and physician sign the appropriate documentation, the patient typically receives conscious sedation while having careful cardiopulmonary monitoring. Although sedation is not medically required, most patients in the United States are sedated for colonoscopy. Surveys have indicated that patients prefer sedation.

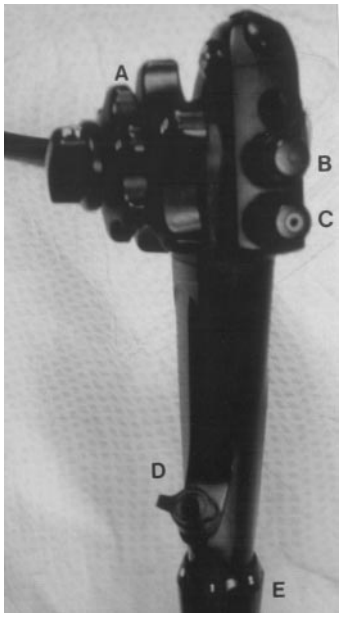


FIGURE 2 The instrument head: (A) control dials; (B) suction button; (C) air/water button; (D) instrument/suction channel; (E) colonoscope stiffening control knob.

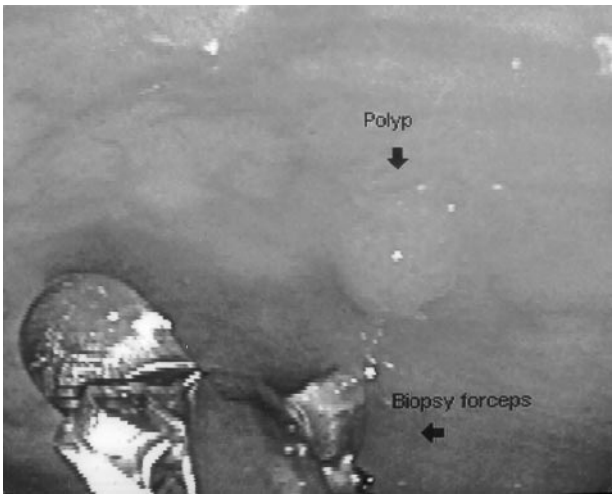


FIGURE 3 Biopsy forceps moving toward a polyp.

Prior to colonoscopy, the physician performs a digital rectal examination to exclude obvious lesions at the anus and lower rectum. Thereafter, the shaft of the colonoscope is lubricated with surgical jelly and inserted into the rectum. The operator carefully maneuvers the instrument to the cecum and/or terminal ileum. This is accomplished by using the control dials and torque to manipulate the shaft. The cecum is readily identified by the presence of the appendiceal orifice and the ileocecal valve. In certain cases, it may be necessary to examine the ileum before withdrawing the colonoscope (e.g., to assess inflammation as with inflammatory bowel disease).

Once the cecum and (or) terminal ileum have been properly examined, the colonoscope is withdrawn slowly. The mucosa is surveyed in a circumferential manner. During polyp screening or surveillance, small polyps (<5 mm) are removed with forceps (Fig. 3). Larger polyps are removed with a snare or with a combination of snare and forceps. Polyps too large to be removed endoscopically are biopsied and then tattooed with India ink for surgical removal.

If the procedure is performed for bleeding treatment and a source is discovered, several options exist. Bleeding angioectasiae may be treated with an electrocautery probe or with argon plasma coagulation. Bleeding vessels, ulcers, diverticula, or polypectomy sites can be treated with injection of epinephrine (1 : 10,000 dilution) and/or electrocautery. Appropriate therapy is

based on the nature of the lesion and the skill of the endoscopist. Other therapeutic issues are discussed below.

After the colon has been fully examined and appropriate therapies have been performed, the colonoscope is removed from the patient. The patient is then taken to a recovery area where vital signs are monitored until the patient awakens.

INDICATIONS

Common indications for diagnostic colonoscopy are summarized in Table I. These indications correspond to recommendations of the American Society for Gastrointestinal Endoscopy (ASGE) with the addition of screening colonoscopy for persons over the age of 50. Screening colonoscopy for the detection of colon cancer is rapidly becoming more widely accepted in the medical community. It is likely that medical societies and insurance carriers will endorse screening examinations for persons within that age group. The effectiveness of flexible sigmoidoscopy for screening purposes is limited because it does not involve examination of the entire colon.

Common indications for therapeutic colonoscopy are listed in Table II. As noted previously, a variety of instruments are available for therapeutic purposes. In addition to polypectomy and bleeding techniques, endoscopists are able to remove foreign bodies with snares, baskets, or forceps. Mesh stents are available for palliative placement in patients who have obstructing colon lesions and are not surgical candidates. Balloon dilation can be performed to open stenotic surgical anastamoses. Colonic decompression can be performed with placement of a decompression tube when the use of neostigmine is contraindicated or ineffective. Sigmoid volvulus can also be reduced when radiologic decompression with gastrografin enema fails. It is likely that therapeutic capabilities will continue to expand as technology improves and as needs arise.

CONTRAINDICATIONS

There are times when colonoscopy is contraindicated (Table III). It is unwise to perform colonoscopy in situations of suspected perforation, toxic megacolon, severe diverticulitis, and fulminant colitis because there is considerable risk for perforation during insufflation. Perforation can be the result of direct distension of mucosa already weakened by underlying disease or it can result from ensuing ischemia that occurs as the colonic wall becomes more distended.

LOW-YIELD EXAMINATIONS

Just as there are indications and contraindications for colonoscopy, there are certain situations in which colonoscopy has a low diagnostic yield or is not indicated. An example of this is when colonoscopy is considered to exclude colon malignancy as the source for metastatic adenocarcinoma of unknown primary type. Previous studies demonstrate that colonoscopy is not

TABLE I Routine Indications for Diagnostic Colonoscopy

1. Abnormal adequate barium enema (e.g., filling defect)
2. Exclusion of right-sided colon polyps when polyps are found during flexible sigmoidoscopy
3. Hematochezia not thought to originate from the rectum or perineum
4. Melena of unknown origin (negative upper gastrointestinal evaluation)
5. Positive fecal occult blood test
6. Unexplained iron deficiency
7. Evaluation of the entire colon to exclude synchronous cancer or neoplastic polyps in a patient with a treatable colon cancer or neoplastic polyp
8. Follow-up at 1 year and at 3- to 5-year intervals after colonic resection of cancer or neoplastic polyp
9. Surveillance exams every 3–5 years for persons with established adenomatous polyps
10. Screening every 2 years beginning at age 25, or 5 years younger than earliest age of a relative with colorectal cancer, in patients at risk for hereditary nonpolyposis colorectal cancer; yearly colonoscopy at age 40
11. Screening every 5 years, beginning 10 years earlier than the age of an affected relative when that relative developed colorectal cancer before age 60; surveillance every 3 years if adenomas found
12. Screening for colon polyps in average-risk patients over age 50
13. Exclusion of cancer or dysplasia at 1- to 2-year intervals in patients with chronic ulcerative colitis if they have pancolitis for greater than 7 years or left-sided colitis for greater than 15 years
14. Chronic inflammatory disease of the colon when precise *diagnosis* will influence immediate management
15. Chronic inflammatory disease of the colon when determination of *disease activity* will influence immediate management
16. Clinically significant diarrhea of unexplained etiology
17. Intraoperative identification of the site of a lesion that cannot be found by gross inspection or palpation

TABLE II Routine Indications for Therapeutic Colonoscopy

1. Active lower intestinal bleeding secondary to lesions such as ulcers, polypectomy sites, vascular malformations, and diverticula
2. Foreign body removal
3. Decompression of acute nontoxic megacolon or sigmoid volvulus
4. Palliative stent placement across a stenosing neoplasm
5. India-ink tattooing a cancer or neoplastic polyp for surgical resection
6. Dilation of stenotic lesions such as anastamotic strictures
7. Snare polypectomy of polyps too large for safe resection during flexible sigmoidoscopy

cost-effective or of high utility in such cases, particularly when patients do not have lower intestinal symptoms. Other examples are included in Table IV.

BOWEL PREPARATION

General Instructions

When preparing for colonoscopy, the patient should take only clear liquids by mouth beginning the day before elective examination and should have nothing but medications and/or preparation solution on the day of the exam. Physicians should instruct patients to avoid red food coloring or red liquids so as not to mistake passage of these for blood. Further instructions include discontinuation of iron supplements 1 week before examination to reduce surface epithelial discoloration.

Polyethylene Glycol–Electrolyte Solutions

A variety of bowel preparations have been used over the years to cleanse the colon for examination. Many of these have proved less than ideal. Currently, the two most widely accepted and useful preparations include electrolyte solutions of polyethylene glycol (PEG–ELS) or oral sodium phosphate (Table V). The available PEG–ELS agents are iso-osmotic electrolyte (mainly sodium) solutions, which are not readily absorbed by the intestinal mucosa. Assuming that the patient does not take dietary sugar with the preparation (which facilitates the absorption of sodium), all electrolytes and water remain within the bowel lumen. When taken over 2–3 h, there is rapid washout of the entire bowel. The volume required for adequate preparation ranges from 2 to 4 liters or more. Typically, the bowel preparation begins early in the afternoon on the day before elective colonoscopy. This schedule minimizes patient bowel movements late into the night.

TABLE III Contraindications for Colonoscopy

1. Established or suspected perforated viscus
2. Toxic megacolon (e.g., *C. difficile* colitis)
3. Severe acute diverticulitis
4. Fulminant colitis

Oral Phosphasoda Solutions

Recently, oral sodium phosphate solutions have been used for colon preparation. Results from comparative studies demonstrate that they are as effective as the PEG–ELS solutions. Phosphasoda solutions are highly osmotic and require much less volume than PEG–ELS solutions. The patient takes 1.5 oz (45 ml) with approximately 32 oz of water on the afternoon before colonoscopy. The same preparation is repeated the next morning. Although complications of preparations are discussed below, it is worth noting that this preparation can cause mild mucosal inflammatory reactions and may not be the best preparation for patients who are having examinations to exclude inflammatory bowel disease or other forms of colitis.

SEDATION

Just as bowel preparation regimens have changed through the years, so have sedation techniques. Whereas meperidine was widely used for many years, it was recently discontinued in many units because of its duration of action and the potential for seizures related to drug metabolites. Fentanyl has replaced meperidine for both its shorter half-life and reduced side effect profile. Patients tend to recover from colonoscopy more quickly after fentanyl sedation.

Another change has been the increased use of midazolam rather than diazepam. As with fentanyl, midazolam has the benefit of a shorter duration of action. The combination of fentanyl and midazolam has become a useful combination for adequate conscious

TABLE IV Situations in which Colonoscopy Is Not Indicated or Has a Low Diagnostic Yield

1. Chronic, stable irritable bowel syndrome or abdominal pain
2. Acute, self-limited diarrhea
3. Metastatic adenocarcinoma of unknown primary when the patient has no lower intestinal symptoms and when diagnosis will not alter management
4. Routine follow-up of inflammatory bowel disease
5. Preoperative assessment for patients undergoing elective abdominal surgery for noncolonic disease
6. Melena or other gastrointestinal bleeding when an upper source is identified

TABLE V PEG–ELS and Oral Phosphasoda Bowel Preparation Solutions

PEG–ELS	Oral phosphasoda solution
2–4 liters over a 2 to 3 h period the day before colonoscopy; repeat preparation if not clear; may be delivered via nasogastric tube for emergent, rapid preparation	1.5 oz of solution in 8 oz of water, followed by another 32 oz of water; taken the afternoon before the exam and on the morning of examination
May be given to most any patient	Avoid in patients with congestive heart failure, cirrhosis, or creatinine >1.5
Not known to cause mucosal inflammation	May cause mild inflammatory mucosal reactions that endoscopically resemble forms of colitis

sedation during most colonoscopies. Most patients report good outcomes with this regimen.

When fentanyl and midazolam do not provide adequate sedation, it may be necessary to pursue other options. One option is propofol, which has a very short half-life. Sedation lasts only as long as the drug is being administered. Patients are generally well sedated for the examination and tolerate the procedure without difficulty. However, there are issues regarding propofol that make it less than ideal for routine colonoscopy. First, it often provides a level of sedation beyond that of conscious sedation. In many units, an anesthesiologist and/or respiratory therapist must be available during administration should the patient become oversedated and require intubation. Second, the medication is more expensive than fentanyl and midazolam, making the approach less cost effective.

In the very rare case that conscious sedation or propofol is inadequate, the endoscopist can consider general anesthesia. This requires coordination with an anesthesiologist and is typically scheduled in advance. Patients often go for a preprocedure anesthesiology clinic evaluation. This form of sedation adds considerable cost to the procedure but should be used when necessary.

RISKS AND POTENTIAL COMPLICATIONS

Bleeding

The potential complications of colonoscopy are listed in Table VI. One of the more common complications is bleeding. Although the risk of bleeding from a diagnostic exam is only 0.02–0.03%, the rate can be as high as 1.6–5.4% among patients who have undergone polypectomy. Higher rates are reflected by resection of lesions greater than 2 cm. Postpolypectomy bleeding is most common within the first 12 days of the exam, but there have been reports of bleeding as late as 29 days. Management of a postpolypectomy bleed involves hemodynamic stabilization, transfusion as required, and reversal of anticoagulation in certain cases. If bleeding does not stop spontaneously, then repeat examination may be indicated. Bleeding polypectomy sites can often be treated with injection of epinephrine (1 : 10,000 dilution) and/or electrocautery, depending on the nature of the lesion. Occasionally, interventional radiologic or surgical intervention is required when endoscopic therapies fail.

Perforation

Although perforation is likely the most feared complication of colonoscopy, the rate of occurrence secondary to a diagnostic exam is approximately 0.045%. The risk can be higher following polypectomy, with some older studies reporting an incidence as high

TABLE VI Potential Complications of Colonoscopy and Bowel Preparations

1. Bleeding
   - Diagnostic exam 0.02–0.03%
   - Therapeutic exam 1.6–5.4%
2. Perforation
   - Diagnostic exam 0.045%
   - Therapeutic exam $<1\%$
3. Infection $<1\%$
4. Pain
5. Allergic reaction to sedatives
6. Hypotension or hypoxemia secondary to oversedation
7. Thrombophlebitis at intravenous entry site
8. Cardiac ischemia or arrhythmia
9. Postpolypectomy distension syndrome
10. Postpolypectomy coagulation syndrome
11. Hyperphosphatemia and/or hypocalcemia secondary to phosphasoda bowel preparations
12. Volume overload secondary to phosphasoda bowel preparations
13. Intravascular volume depletion secondary to phosphasoda bowel preparations

as 1%. Perforation can result from mechanical force, colonic distension, and effects of polypectomy. A looped colonoscope may create excessive mechanical stress and perforation as the colon wall is stretched outward. The bowel can also rupture if the colonoscope is pushed blindly through a lumenal segment or if an instrument tip (e.g., biopsy forceps) pierces the wall. Pneumatic dilation causes perforation when the lumen is overdistended during insufflation. Finally, polypectomy may cause perforation because of full-thickness cuts or burns and there may be delayed rupture of necrotic tissue in some instances.

Factors that increase the risk of perforation include poor visualization secondary to inadequate bowel preparation, acute bleeding, and an uncooperative patient. There is also considerable perforation risk in cases of fulminant colitis, severe diverticulitis, toxic megacolon, or in cases where pneumatosis has been visualized on diagnostic imaging. Care should be taken to recognize and/or reduce perforation risks before a colonoscopy is begun.

Perforation can manifest in several ways. Direct visualization of the peritoneum is the most obvious sign. However, perforation may not be immediately apparent during colonoscopy. Patients may not develop abdominal pain (beyond that expected for gaseous distension), rebound tenderness, hypotension, fever, or tachycardia until after transfer to the recovery area. Others may not have symptoms for several days after the procedure.

Diagnosis of suspected perforation requires careful physical examination followed by radiographic imaging. An abdominal X ray will show pneumoperitoneum when a considerable amount of air has escaped the bowel lumen. In some cases, diagnosis requires an enema with water-soluble contrast (gastrograffin). Leakage of contrast is diagnostic of perforation.

Whenever there is pneumoperitoneum or contrast extravasation on imaging, surgical intervention is immediately required. When there is no obvious evidence for gross perforation, treatment might involve intravenous antibiotics and supportive care. In either case, a surgical consultation should be made as soon as the problem is recognized or suspected.

Infection

Systemic Infection

The risk of developing systemic infection from colonoscopy is very low. In several studies looking at bacteremia following colonoscopy, the combined incidence was 0.01%. In these studies, blood cultures were drawn within 5 min of colonoscopy completion and none of the study patients developed an acute illness afterward. Follow-up cultures were persistently negative in these cases.

Endocarditis

Although endocarditis can be a concern for patients with increased generalized risk, there are no well-documented reports of endocarditis following colonoscopy. Still, some authors have implicated colonoscopy as the cause in a few cases. Antibiotic prophylaxis to prevent endocarditis is not recommended for most patients and will be discussed below.

Arrhythmia and Cardiac Ischemia

Generally, it is recommended that patients have continuous cardiac monitoring during colonoscopy because a number of studies have demonstrated transient and variable electrocardiogram abnormalities during colonoscopy. Although rare, there have been reports of myocardial infarction occurring during colonoscopy. Care should be taken to reduce cardiac ischemic risk in patients undergoing colonoscopy whenever possible.

Postpolypectomy Distension Syndrome

Postpolypectomy distension syndrome (PDS) refers to pain caused by excessive insufflation of the colon. The pain of this syndrome can be quite impressive. In fact, it can be difficult to distinguish from perforation in some cases. The key differentiating factors are that patients with PDS do not develop peritoneal signs, fever, tachycardia, hypotension, or other common signs of perforation. The treatment for this problem is to turn the patient onto his or her right side so that air moves up to the splenic flexure and then through the left colon for release. It is expected that this problem will resolve spontaneously with patient movement and expulsion of gas.

Postpolypectomy Coagulation Syndrome

Polypectomy coagulation syndrome (PCS) results from deep tissue injury caused by electrical currents in a polypectomy snare, "hot" biopsy forceps, or cautery probe. Whereas the operator intends to apply thermal energy only to a polyp or other lesion at the mucosa or submucosa, PCS results when tissue damage extends into the muscularis propria and/or serosa. PCS is most often caused by prolonged application of electrical current. Although the colon is not perforated, there can be transmural necrosis, which sometimes leads to perforation.

Clinical evidence of PCS typically manifests within hours to 5 days after colonoscopy and symptoms can mimic perforation. Patients may present with peritoneal signs, fever, leukocytosis, tachycardia, or hypotension. Abdominal imaging should be used to exclude perforation in such cases. Thereafter, treatment consists of bowel rest, intravenous fluid hydration, antibiotics, and supportive care. If the patient's condition worsens, repeated radiography should be used to exclude new perforation.

Complications of Bowel Preparations

As previously discussed, the most common solutions used for bowel preparation are PEG−ELS and oral sodium phosphate. The advantage of PEG−ELS solutions is that they have no real toxicity. Volume overload is avoided because sodium and fluids are not readily absorbed during preparation. This is especially important for patients with renal disease, congestive heart failure, or cirrhosis. The disadvantages of PEG−ELS preparations are that patients are required to drink large volumes (4 liters or more) and the taste is often poorly tolerated. Because of the large ingested volumes, patients may experience abdominal distension, nausea, vomiting, and even aspiration. This can lead to poor compliance with the preparation, a poorly cleansed colon, or more serious health problems.

Oral sodium phosphate solutions have become popular because of their smaller preparation volumes, improved taste, and lower cost. However, these solutions can cause significant electrolyte and fluid shift abnormalities as phosphate and salts move in and out of the vascular space. One risk is hyperphosphatemia. Although most documented cases have been asymptomatic, there have been several deaths related to phosphate overload. There can be also arrhythmic complications resulting from hypocalcemia and hypokalemia. Furthermore, there is a risk for transient volume overload related to phosphate absorption and fluid shifts into the intravascular space. These factors have led to the recommendation that phosphasoda preparations be avoided in patients with renal insufficiency (creatinine >1.5), congestive heart failure, and cirrhosis.

Although there can be transient volume overload with oral phosphasoda preparations, there is also the risk of volume depletion. This occurs secondary to fluid and electrolyte losses in the stool. These preparations should be prescribed only for the appropriate patient and the patient should be counseled to drink large volumes of water after the preparative solution has been consumed.

Combustion of Colonic Gases

A good bowel preparation is important for reducing the chance of a colonic explosion. Colonic bacteria release hydrogen and methane gas during normal metabolism. If bacteria and stool are not adequately removed, there is a chance of explosion when using electrocautery. This is more than a theoretical risk as there have been published reports of explosions occurring in poorly prepared colons.

ANTIBIOTIC PROPHYLAXIS

Infectious complications from colonoscopy are uncommon. Although there is a risk of transient bacteremia during colonoscopy, only a few published reports have directly attributed infective endocarditis to colonoscopy. Furthermore, there are no prospective, controlled trials proving that antibiotic prophylaxis prevents the development of endocarditis in this setting. Because of these facts, and because bacteremia rates are very low during colonoscopy, the ASGE has not made formal recommendations for antibiotic prophylaxis in any individual undergoing colonoscopy. The ASGE has left it to the practitioner's discretion whether to administer prophylactic antibiotics to patients considered "high risk" for the development of infective endocarditis. These patients include those with a history of endocarditis and those with prosthetic heart valves or surgically constructed systemic−pulmonary shunts or conduits. There are insufficient data to recommend antibiotic prophylaxis for all other patients, including those with cirrhosis, prosthetic joints, or synthetic vascular grafts and other immunocompromised patients. As with the high-risk patients described above, the decision of whether to give prophylaxis to a patient with a synthetic graft less than 1 year old is left up to the practitioner.

When antibiotic prophylaxis is administered, current recommendations are for the use of preprocedure intravenous ampicillin and gentamicin. Thirty minutes before colonoscopy, the patient is given 2 g of ampicillin and 1.5 mg/kg (up to 80 mg) of gentamicin. Six hours after the colonoscopy, the patient is given a 1.5 g oral dose of ampicillin. In the case of penicillin allergy, 1 g of vancomycin is given in place of the ampicillin.

ANTICOAGULATION ISSUES

Many patients take anticoagulants for reasons such as mechanical heart valves, previous stroke, myocardial infarction, and deep venous thrombosis. When making decisions about withholding anticoagulants for

TABLE VII ASGE Recommendations for Anticoagulation Management among Patients Undergoing Colonoscopy, Based on Their Thromboembolic Risk when Off Medication

Procedure risk+ thromboembolic risk	Recommendation
Low+Low	Do not change anticoagulation; delay elective procedures if INR supratherapeutic
Low+High	Do not change anticoagulation; delay elective procedures if INR supratherapeutic
High+Low	Discontinue warfarin 3–5 days prior to the procedure and restart after the procedure
High+High	Discontinue warfarin 3–5 days prior to the procedure and consider using a "heparin window"

colonoscopy, physicians must weigh procedural bleeding risks against the thromboembolic risks of stopping anticoagulation. It is useful to divide bleeding and thromboembolic risks into low- and high-risk categories when making these decisions. For example, diagnostic colonoscopy (with or without biopsy) is of low risk for causing bleeding (<1%). However, colonoscopy with snare polypectomy, coagulation therapy, or pneumatic dilation carries a much higher risk (1–5%). Likewise, persons with a low risk for thromboembolic complications are those with previous deep venous thrombosis, chronic/paroxysmal atrial fibrillation (without valvular disease), bioprosthetic heart valves, and mechanical aortic valves. High-risk patients include those who have mechanical mitral valves, atrial fibrillation plus valvular heart disease or mechanical valves, and mechanical valves in the setting of prior thromboembolic disease. ASGE recommendations for anticoagulant management are found in Table VII.

With regard to aspirin and other nonsteroidal anti-inflammatory agents, the ASGE has recommended continued usage through most endoscopic procedures. However, many physicians will ask patients to discontinue aspirin for therapeutic colonoscopy because of the inherent bleeding risks. Although there have been insufficient data for formal recommendations regarding use of dipyridimol and ticlodipine, many physicians follow the same recommendations as for aspirin.

See Also the Following Articles

Colonic Ulcers • Colorectal Adenocarcinoma • Colorectal Adenomas • Colorectal Cancer Screening • Lower Gastrointestinal Bleeding and Severe Hermatochezia • Sigmoidoscopy • Virtual Colonoscopy

Further Reading

American Society for Gastrointestinal Endoscopy (1998). Guideline on the management of anticoagulation and antiplatelet therapy for endoscopic procedures. *Gastrointest. Endosc.* **48**, 672–675.

American Society for Gastrointestinal Endoscopy (2003). "Antibiotic Prophylaxis for Gastrointestinal Endoscopy." Available at http://www.asge.org/gui/resources/manual/pe_anti.asp.

Basson, M., Etter, L., and Panzini, L. (1998). Rates of colonoscopic perforation in current practice. *Gastroenterology* **114**, 1115.

Clarkston, W., Tsen, T., Dies, D., *et al.* (1996). Oral sodium phosphate versus sulfate-free polyethylene glycol electrolyte solution in outpatient preparation for colonoscopy: A prospective comparison. *Gastrointest. Endosc.* **43**, 42–48.

Hsu, C., and Imperiale, T. (1998). Meta-analysis and cost comparison of polyethylene glycol lavage versus sodium phosphate for colonoscopy preparation. *Gastrointest. Endosc.* **48**, 276–282.

Huang, E., and Marks, J. (2001). The diagnostic and therapeutic roles of colonoscopy: A review. *Surg. Endosc.* **15**, 1373–1380.

Leslie, A., and Steele, R. (2002). How to do it: Colonoscopy. *J. R. Coll. Surg. Edinburgh* **47**, 502–509.

Rex, D. (2000). Colonoscopy. *Gastrointest. Endosc.* **10**, 135–160.

Ristikankare, M., Julkunen, R., Mattila, M., *et al.* (2000). Conscious sedation and cardiorespiratory safety during colonoscopy. *Gastrointest. Endosc.* **52**, 48–54.

Swaroop, V., and Larson, M. (2002). Colonoscopy as a screening test for colorectal cancer in average-risk individuals. *Mayo Clin. Proc.* **77**, 951–956.

Waye, J., Kahn, O., and Auerbach, M. (1996). Complications of colonoscopy and flexible sigmoidoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 343–377.

Colorectal Adenocarcinoma

PAUL J. LIMBURG AND DAVID A. AHLQUIST
Mayo Clinic and Foundation

aberrant crypt foci Clusters of colonic crypts that typically appear larger and thicker than normal; lumenal shape may also be altered; thought to represent an early precursor of adenocarcinoma, especially when dysplasia is present.

adenocarcinoma A malignant neoplasm derived from glandular epithelium.

adenomatous polyp (or adenoma) A premalignant neoplasm arising from glandular epithelium.

case-control study An epidemiological study design wherein exposures are compared between a group of individuals with the disease of interest (cases) and a group of individuals without the disease of interest (controls); also called a retrospective study, since the exposure information is collected after a diagnosis has been established.

chemoprevention The use of chemical compounds to prevent, inhibit, or reverse carcinogenesis before dysplastic epithelial cells invade across the basement membrane; may refer to either nutritional or pharmaceutical agents.

cohort study An epidemiological study design wherein a group of exposed individuals and a group of non-exposed individuals are followed forward in time to compare incidence rates for the disease of interest; also called a prospective study, since the exposure information is collected before a diagnosis has been established.

distal colon Anatomic subsite of the colorectum that includes the descending colon, sigmoid colon, and rectosigmoid colon; may sometimes include the rectum as well.

meta-analysis The analysis of combined data from multiple epidemiological studies, for the purpose of integrating the findings; often used to increase statistical power and/or provide a broader perspective when individual study results are inconsistent.

metachronous neoplasia Tumors (adenomas, adenocarcinomas, or a combination of both) that are diagnosed at different points in time; in clinical practice, metachronous neoplasms may refer to newly formed and newly discovered tumors.

proximal colon Anatomic subsite of the colorectum that includes the cecum, ascending colon, and transverse colon.

Surveillance, Epidemiology, and End Results (SEER) program An authoritative source of information on cancer incidence and survival in the United States; coordinated by the National Cancer Institute; currently collects and publishes cancer incidence and survival data from 11 population-based cancer registries and 3 supplemental registries, which cover approximately 14% of the U. S. population in total.

synchronous neoplasia Tumors (adenomas, adenocarcinomas, or a combination of both) that are diagnosed at the same point in time.

TNM staging system A universal cancer classification system that is based on tumor extent (T), lymph node status (N), and the presence or absence of distant metastases (M); provides a standardized measure for planning treatment and predicting prognosis.

Among all cancers, colorectal cancer (CRC) has the third highest incidence rate in the world, trailing only lung and breast cancers. According to recent estimates, 944,700 new CRC cases are diagnosed annually and 492,400 deaths are attributed to this disease each year. In the United States, approximately 1 in every 18 persons will be afflicted with CRC over the course of a lifetime. Despite these sobering statistics, emerging data suggest that the large majority of CRC cases may be prevented through the design of (and compliance with) effective early detection programs. Advances in CRC treatment and control strategies should further reduce the societal burden imposed by this relatively common disease.

CLINICAL ASPECTS

Ninety-eight percent of all malignancies that develop within the large intestine arise from glandular mucosa and are thus classified as adenocarcinomas by histology. Less frequently encountered histologic subtypes include lymphomas, carcinoid tumors, and leiomyosarcomas. Metastatic lesions rarely present within the colorectum, but can include adenocarcinomas arising from the breast, ovary, prostate, lung, or stomach. Malignant melanomas may occasionally spread to the colorectum as well. Because nearly all lower gastrointestinal tract neoplasms are primary adenocarcinomas, the term colorectal cancer (CRC) generally refers to this tumor subtype. This convention is used throughout the remainder of the text.

 Copyright 2004, Elsevier (USA). All rights reserved.

Carcinogenic Process

Colorectal carcinogenesis is a multistep process that begins with the clonal expansion of genetically altered epithelial cells (Fig. 1). Clustering of these abnormal cells results in the formation of aberrant crypt foci. Aberrant crypt foci (ACF) represent the earliest stage of dysplasia that can be recognized using current technology. In response to poorly understood molecular signals, a subset of ACF advance to become adenomatous polyps, which are often referred to simply as adenomas. Adenomas can further progress in terms of size (diminutive to small to large), glandular distortion (tubular to tubulovillous to villous), and/or dysplasia grade (low to high). When dysplastic cells invade across the basement membrane, the requisite criterion for transformation from a benign adenoma to a malignant adenocarcinoma is achieved. This process is thought to occur in a small fraction (8–12%) of all adenomas.

Although less fully characterized, a minority of CRCs are thought to arise *de novo* or from "flat" colorectal adenomas. These lesions tend to be found in the proximal colon and likely progress through an alternate pathway of carcinogenesis. Data regarding the prevalence and natural history of less-understood precursor lesions are currently being gathered. Ongoing research in this area may reveal new insights regarding the clinical characteristics of nonpolypoid CRCs during the next several years.

Presentation and Staging

The clinical manifestations of CRC relate in part to tumor location. Typical signs and symptoms of proximal (cecum, ascending, transverse) colon malignancies include ill-defined abdominal pain, weight loss, and occult bleeding. Distal (descending, sigmoid) colon and rectal cancers commonly present with altered bowel habits, decreased stool caliber, and hematochezia. Regardless of anatomic subsite, CRCs often remain asymptomatic until relatively late in the disease course. Accordingly, patients who delay seeking medical attention until after symptoms develop are more likely to be diagnosed with advanced-stage disease.

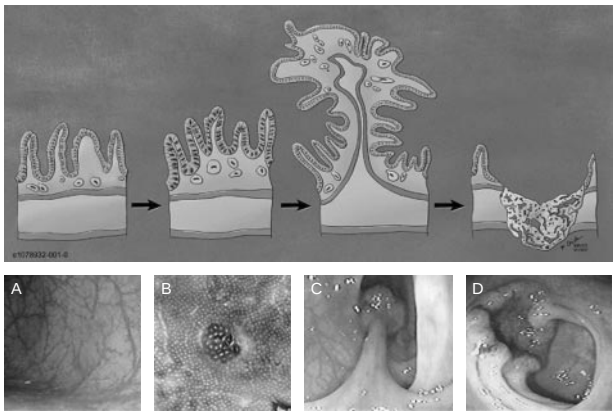


FIGURE 1 The process of colorectal carcinogenesis, as represented by schematic and endoscopic images. (A) Normal colorectal mucosa. (B) Aberrant crypt foci. (C) Adenomatous polyp. (D) Adenocarcinoma. Copyright Mayo Foundation.

Colonoscopy is the test of choice for establishing the diagnosis of CRC. Once cancer has been histologically confirmed, a staging work-up is performed among potential surgical candidates. Computerized tomography (CT) of the abdomen and pelvis is used as the primary tool for detecting both regional and distant metastases. In the setting of a lumenally obstructing CRC, CT colonography can be used to rule out synchronous neoplasia in the proximal colon, while simultaneously providing an abdominopelvic survey. For rectal cancers, endoscopic ultrasound helps to determine the locoregional extent of disease, which might influence the timing of adjuvant chemotherapy and/or radiation therapy. Positron emission tomography (PET) affords a whole body screen for malignant disease. However, because of its high cost and limited accessibility, PET scanning is often not included in the initial CRC staging evaluation.

The most definitive assessment of CRC stage is made from surgical pathology specimens. Tumor size, depth of invasion, lymph node involvement, and distant metastases represent core elements of the TNM staging system (Table I). TNM stages I–IV have been shown to correlate with 5-year survival rates. In general, stage I disease is associated with a very favorable prognosis, whereas stage IV disease portends a rapidly fatal outcome. Within TNM stages, 5-year survival rates are lower for rectal cancers than for colon cancers. This observation may be related, at least in part, to differences in anatomic structure (i.e., lack of a serosal lining in the rectum) and physiologic function between these organ subsites.

HOST FACTORS

Similar to many other malignancies, CRC risk increases with advancing age. Based on data from the Surveillance, Epidemiology, and End Results (SEER) program, CRC incidence rates begin to rise rapidly after the age of 50 years and increase steadily until the age of approximately 80 years. Fewer than 5% of all CRC cases occur among persons younger than 45 years of age. With respect to premalignant colorectal neoplasia, adenoma prevalence rates also increase

TABLE I Colorectal Cancer Survival by TNM Stage

TNM Stage	Description	Approximate 5-year survival rate (%)
0	Confined to the epithelium or invading into the lamina propria (Tis); no lymph node involvement (N0); no distant metastases (M0)	100
I	Invasion of the submucosa (T1); no lymph node involvement (N0); no distant metastases (M0)	95
	Invasion of the muscularis propria (T2) no lymph node involvement (N0); no distant metastases (M0)	90
II	Invasion through the muscularis propria into the subserosa or nonperitonealized pericolic/perirectal tissues (T3); no lymph node involvement (N0); no distant metastases (M0)	80
	Perforation of visceral peritoneum or direct invasion into adjacent organs or tissues (T4); no lymph node involvement (N0); no distant metastases (M0)	75
III	Any depth of invasion (Tis−T4); metastases in 1−3 pericolic or perirectal lymph nodes (N1); no distant metastases (M0)	72
	Any depth of invasion (Tis−T4); metastases in 4 or more pericolic or perirectal lymph nodes (N2); no distant metastases (M0)	60
	Any depth of invasion (Tis−T4); metastases in lymph nodes along a named vascular trunk or tumor invasion of adjacent organs (N3); no distant metastases (M0)	40
IV	Any depth of invasion (Tis−T4); any nodal status (N0−N3); distant metastases present (M1)	5

with age, with estimates of 30% at 50 years, 40−50% at 60 years, and 50−65% at 70 years of age. Gender is not strongly associated with overall CRC risk, but CRC distributions by anatomic subsite are different between men and women. Most notably, the rate ratio of proximal : distal CRCs is much higher among women (1.23) than among men (0.86) after the age of 65 years.

For reasons that remain incompletely defined, racial and ethnic subgroups of the U. S. population experience fairly striking dissimilarities in CRC incidence, mortality, and 5-year survival rates. For example, among the six race/ethnicity subgroups recognized by the SEER program, African Americans have the highest CRC incidence rate, highest CRC mortality rate, and the second lowest 5-year survival rate. Further exploration of CRC risk associations within and across racial, ethnic, and culturally diverse subject groups is needed to determine the predominant mechanisms of, and appropriate interventions to disrupt, carcinogenesis in these population subsets.

In addition to age, gender, and race/ethnicity, several host and environmental factors have been convincingly associated with CRC risk. Key points regarding these risk associations are highlighted below.

Personal History of Colorectal Neoplasia

Patients with a history of colorectal adenomas are three to six times more likely to develop metachronous neoplasms (adenomas or CRC diagnosed at a later point in time) than are persons of the same age, gender, and race/ethnicity in the general population. Adenomas that are large in diameter ($\geq$1 cm), multiple in number ($\geq$3 total), or "aggressive" by histology (villous glandular formation or high-grade dysplasia) portend the highest risks. In a retrospective study of patients with adenomas $\geq$1 cm in diameter that were managed by observation only (prior to the widespread availability of colonoscopy), CRC risks were found to increase progressively over time. After 5, 10, and 20 years of follow-

up, 2.5, 8, and 24% of these patients developed CRC. Data from the National Polyp Study additionally showed that recurrent adenomas were significantly more common at the first surveillance exam among patients with large adenomas (<1 cm in diameter; odds ratio 1.6; 95% confidence interval 1.1−2.5) or multiple adenomas (≥ 3 total; odds ratio 2.4; 95% confidence interval 1.7−3.5) at baseline, compared to patients with smaller and fewer adenomas, respectively. Importantly, neither diminutive hyperplastic polyps nor diminutive or small tubular adenomas ≤ 2 in total number appear to be associated with increased CRC risk.

Patients with a history of CRC are prone to developing recurrent primary cancers, second primary cancers, and metachronous adenomas. The majority of recurrent CRCs are diagnosed within 3 years (85%) of the initial operation. Median time to detection of metachronous adenomas ranges from 19 to 32 months in this patient group.

Family History of Colorectal Neoplasia

Familial clustering can be observed in up to 20% of all CRC cases. Large epidemiologic studies have shown that CRC risk is increased by 1.5- to 2-fold among first-degree relatives of patients with colorectal neoplasia. Although less thoroughly investigated, having second- or third-degree relatives with CRC also appears to confer a modestly increased risk. In a minority of families, multiple individuals are found to have documented histories of CRC or other cancers. Heritable cancer syndromes should be strongly considered in this context, particularly when a kindred includes one or more family members diagnosed with malignant disease at an early age (see Heritable Syndromes).

Inflammatory Bowel Disease

Idiopathic inflammatory bowel disease (IBD) is an established risk factor for CRC. Among patients with ulcerative colitis, cumulative CRC incidence rates range from 1.8% after 20 years to 43% after 35 years of disease. The extent of colonic inflammation is also thought to modify CRC risk (i.e., CRC risk is higher among patients with pancolitis than among patients with distal colitis or proctitis). In contrast, colitis severity and colitis activity appear to have little influence on CRC risk. Relatively few studies have specifically addressed an association between CRC and Crohn's disease. However, existing data suggest that CRC risks are similarly elevated among patients with either Crohn's disease or ulcerative colitis after adjusting for the duration of IBD. Other chronic inflammatory conditions such as microscopic colitis and lymphocytic colitis have not been convincingly associated with CRC risk.

Other Medical Conditions

Type II diabetes mellitus (DM) has been positively associated with CRC risk in some, but not all, epidemiologic investigations. In the largest prospective study reported to date, men with DM were 30% more likely to die from CRC than men without DM. CRC mortality rates were also higher among diabetic women, but this risk estimate was slightly lower (16%) and was not statistically significant. Acromegalics may be metabolically and/or anatomically predisposed to higher CRC risks. However, due to the relative rarity of this condition, most observational studies have lacked adequate statistical power to confirm or exclude a true risk association. Cholecystectomy results in altered fecal bile acid composition. Two recent meta-analyses found that CRC risk was increased by up to 34% after gallbladder removal. Yet, because of limitations in the design of some previous studies, this association remains controversial.

ENVIRONMENTAL FACTORS

Diet and Nutrition

An estimated 30% of all cancers may be avoidable through dietary modification. Unfortunately, identifying the food components that are most important to colorectal carcinogenesis remains an ongoing challenge. Select macro- and micronutrients with a putative influence on CRC risk are discussed below.

Macronutrients

Dietary fats induce the excretion of primary bile acids, which can then be converted into procarcinogenic secondary bile acids by the colonic bacteria. Although ecological studies have repeatedly shown a strong association between per capita fat consumption and CRC incidence rates (with correlation coefficients ranging from 0.8 to 0.9), data from case-control and cohort studies have been less consistent. In a comprehensive report published jointly by the World Cancer Research Fund and the American Institute for Cancer Research in 1997, total fat intake and saturated fat intake were both concluded to be possible, but not definite, CRC risk factors. Further exploration of hypothesized positive and negative associations with specific fatty acid subtypes (such as ω-6 and ω-3 polyunsaturated fatty acids, respectively) is needed to clarify the potential role of dietary fats in CRC risk modulation.

Fiber enhances stool bulk, stimulates intestinal transit, decreases secondary bile acid concentrations (possibly pro-carcinogenic), and increases short-chain fatty acid concentrations (possibly anti-carcinogenic). Numerous epidemiological studies have observed inverse associations between high fiber intake and CRC risk. However, several large clinical trials using various fiber-based interventions have been completed with little, if any, beneficial effects observed against the surrogate end-point of adenoma recurrence.

Micronutrients

Calcium binds to intralumenal toxins and may also decrease cellular proliferation within the colorectum. In one recent clinical trial, calcium supplementation was associated with a statistically significant 19% reduction in the adenoma recurrence rate among postpolypectomy patients after 4 years. Antioxidants (including retinoids, carotenoids, ascorbic acid, α-tocopherol, and selenium) neutralize free radical compounds, which might stimulate tumorigenesis within the colorectum. Observational data referent to antioxidant intake and CRC risk have been generally unimpressive. Similarly, mixed results have been obtained from five clinical trials, with the largest and most rigorously designed study finding no statistically significant benefits from either a combination of vitamin C and vitamin E or a combination of vitamin C and β-carotene. Methyl donors, including folate and methionine, are critical for maintaining normal cellular functions such as nucleotide synthesis and gene regulation. Data from most prospective observational studies support an inverse association and several folate intervention trials are under way.

Lifestyle Factors

Alcohol induces cellular proliferation, blocks methyl group donation, and inhibits DNA repair. In a meta-analysis of 27 case-control and cohort studies, consumption of two alcoholic beverages per day was associated with a modest, but statistically significant, 10% elevation in CRC risk. Cigarette smoking appears to increase CRC risk after a prolonged latency period of perhaps 3–4 decades. CRC incidence rates among tobacco users who began smoking in the distant past have usually been two to three times higher than those observed among never smokers. Physical activity appears to have a protective influence on CRC risk, which may be related to lower serum insulin concentrations among nonsedentary individuals. Overall, regular physical activity seems to reduce CRC risk by approximately 40–50%.

HERITABLE SYNDROMES

Although familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC) contribute to only 5–10% of all CRC cases, these syndromes are of substantial clinical and research importance. The germ-line mutations responsible for FAP (*APC* gene) and HNPCC (*MLH1*, *MSH2*, *MSH6*, *PMS1*, and *PMS2*, and perhaps other as yet unrecognized genes) are inherited in an autosomal dominant fashion, have a high degree of penetrance, and lead to malignant disease at a relatively early age. The major features of these syndromes are discussed below. Further details regarding FAP, HNPCC, and other heritable cancer syndromes are available in the American Gastroenterological Association's technical review on hereditary CRC and genetic testing.

Familial Adenomatous Polyposis

Germ-line mutations in the *APC* gene occur in approximately 1 in 5000 persons in the United States. Up to 20% of FAP probands represent new onset *APC* mutations with an unremarkable family history. The hallmark lesion of FAP is diffuse colorectal polyposis, with hundreds to thousands of adenomas typically developing at puberty or during adolescence. Extracolonic manifestations of FAP include the following: duodenal adenomas, gastric fundic gland hyperplasia, mandibular osteomas, supernumerary teeth, congenital hypertrophy of the retinal pigmented epithelium, desmoid tumors, epidermoid cysts, fibromas, lipomas, and thyroid, adrenal, or hepatobiliary cancers. Rarely, patients with germ-line *APC* mutations also develop medulloblastomas. Without prophylactic colectomy, FAP patients are almost universally diagnosed with CRC at a relatively young age (mean = 45 years). A subset of patients with attenuated FAP have relatively fewer adenomas (<100) and develop CRC at a slightly older age (mean = 56 years). Interestingly, colonic neoplasms are frequently found in the proximal colon in patients with attenuated FAP, for reasons that remain unclear.

Hereditary Nonpolyposis Colorectal Cancer

HNPCC is characterized by early onset CRC (mean = 44 years), which tends to occur in the proximal colon and exhibits the microsatellite instability phenotype. Up to 25% of HNPCC patients are found to have more than one CRC at the time of their initial diagnosis. Endometrial, ovarian, gastric, small bowel, hepatobiliary, and genitourinary cancer risks are also increased

within HNPCC kindreds. Clinical data can be used to identify families with HNPCC. According to the revised Amsterdam Criteria II, at least three relatives must have an HNPCC-associated cancer and (1) one person is a first-degree relative of the other two; (2) two or more successive generations are affected; and (3) one or more cancers were diagnosed before the patient was 50 years of age. Muir-Torre syndrome represents a subset of HNPCC patients with sebaceous neoplasms in addition to the above-mentioned clinical manifestations (ratio of men : women = 2 : 1).

EARLY DETECTION

The several-year dwell time of benign and presymptomatic malignant colorectal neoplasms seemingly offers an ample window of opportunity for effective early detection. To succeed, CRC screening and surveillance interventions must be accurate, efficient, safe, and affordable when periodically applied to appropriate at-risk populations. A general early detection algorithm for colorectal neoplasia is provided in Fig. 2. Salient features of the current screening and surveillance guidelines are presented below, followed by discussions of the available tools and emerging technologies.

Current Screening and Surveillance Guidelines

Average-Risk Screening and Surveillance

Screening should begin at age 50 years for asymptomatic adults who have no identifiable CRC risk factors. The American Cancer Society and other organizations endorse several options and testing intervals including the following: (1) fecal occult blood test every year; (2) flexible sigmoidoscopy every 5 years; (3) fecal occult blood test every year and flexible sigmoidoscopy every 5 years; (4) barium enema every 5 years; and (5) colonoscopy every 10 years. Selection of a particular option is ideally based on an informed discussion with patients that considers estimated effectiveness, risks, and costs of each approach.

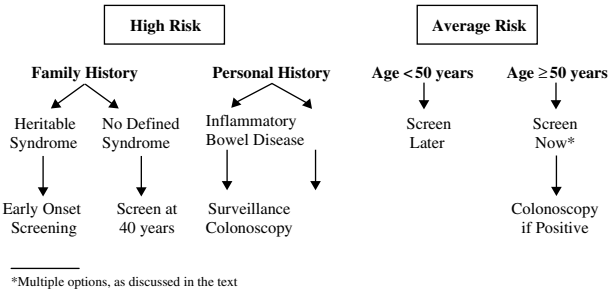


FIGURE 2 Basic algorithm for the early detection of colorectal neoplasia. Copyright Mayo Foundation.

After a complete clearing colonoscopy has been performed, surveillance colonoscopy is indicated at 3 years for patients with advanced or multiple adenomas at baseline or with a family history of CRC. Surveillance procedures can be delayed for 5 years among low-risk patients (i.e., those with only one or two diminutive or small adenomas and no family history of CRC). Once a negative surveillance colonoscopy has been documented, subsequent surveillance examinations may be deferred for 5 years. Patients with previously resected sessile adenomas >2 cm in diameter should be reendoscoped in 3–6 months. If residual adenomatous tissue is still present after two or three attempts at therapeutic colonoscopy, surgical consultation should be considered.

High-Risk Screening and Surveillance

For high-risk patients, the following early detection recommendations have been adopted:

- FAP-flexible sigmoidoscopy every year, beginning at puberty;
- HNPCC-colonoscopy every other year, beginning at age 20 and increasing to every year after age 40;
- Family history of colorectal neoplasia (not FAP) or HNPCC-colonoscopy every 5 years, beginning at age 40 or 5 years, before the youngest case diagnosis, whichever is earlier;
- Inflammatory bowel disease-colonoscopy every year after 8 years of pancolitis or after 15 years of distal colitis only.

Malignant Polyps and Post-CRC Resection

Malignant polyps are defined as neoplasms with dysplastic cells invading through the muscularis mucosae and can be treated endoscopically in some cases. The following criteria have been proposed as determinants of acceptable risk for endoscopic management: (1) the lesion has been completely excised and submitted to the pathologist in its entirety; (2) the depth of invasion, grade of differentiation, and completeness of excision can be accurately determined from the endoscopic resection specimen; (3) no signs of poor differentiation, vascular invasion, or lymphatic involvement are identified; and (4) the margin of excision is free of cancer cells.

For patients with a history of curatively resected CRC, clearing colonoscopy should be performed within 1 year of the initial operation. The first surveillance colonoscopy should be performed at 3 years and if

this examination is negative, the second surveillance examination can be delayed for 5 years.

Fecal Occult Blood Tests

Fecal occult blood testing has been widely practiced for more than 3 decades as an approach to colorectal cancer screening. Though a variety of fecal occult blood tests (FOBTs) are available, the guaiac-impregnated Hemoccult card has been most commonly used. Three large controlled trials have demonstrated that Hemoccult screening over more than 10 years significantly reduces colorectal cancer mortality, although reductions were modest at 15–33%. Furthermore, the impact of FOBT screening on colorectal cancer incidence was either negligible or slight. Such outcomes are consistent with a tool that misses most premalignant lesions and many early stage cancers. Indeed, direct comparisons of Hemoccult testing against colonoscopy in a screening population have revealed Hemoccult point sensitivities for cancer of 11–50% and for advanced adenomas of 5–25% with Hemoccult specificity ranging from 88 to 98%. The likelihood of finding a colorectal cancer or advanced adenoma with a positive Hemoccult test result (positive predictive value) has ranged from 8 to 20% depending on the test technique and population studied. Immunochemical FOBTs may offer some advantages in specificity, as they do not react with dietary or other fecal constituents that confound guaiac testing.

FOBT screening is noninvasive, relatively inexpensive, requires no cathartic preparation, and can be performed without a formal healthcare visit. These advantages must be balanced against its low sensitivity, relatively modest impact on colorectal incidence and mortality, and high false-positive rate.

Flexible Sigmoidoscopy

Case control studies have estimated that sigmoidoscopy screening reduces both incidence and mortality rates from colorectal cancer and that the benefit is preserved with screening frequencies as low as every 10 years. However, failure to inspect the proximal colon is an inherent shortcoming of sigmoidoscopy and case control studies suggest that sigmoidoscopy has no effect on mortality from proximal colon cancer. Although no rigorous sensitivity studies have been carried out, current length flexible sigmoidoscopes may detect only 30–50% of colorectal cancers due to incomplete insertion and other technical limitations in common practice. Moreover, the incidence trend toward more proximal cancers will continue to compromise sigmoidoscopic detection rates over time. Most practitioners advocate that the proximal colon should be evaluated if polyps are found on sigmoidoscopy and this approach has the potential to increase overall cancer detection rates slightly.

The advantages of sigmoidoscopy include its general availability, lack of sedation requirement, and lower cost compared to colonoscopy. Disadvantages include its incomplete inspection of the colon, modest effectiveness in overall cancer mortality reduction, and associated discomfort. Some surveys have indicated that fewer than half of those patients undergoing sigmoidoscopy are willing to return for repeat screening by this modality.

Barium Enema

Few data are available on the effectiveness of colorectal screening by barium enema radiography. The sensitivity of colon X rays for colorectal neoplasms >1 cm may be only approximately 50%, based on comparisons against colonoscopy in asymptomatic screening populations. Although barium X rays image the entire colorectum, viewing is limited to a two-dimensional plane with confounding due to superimposed radiodensities and other technical factors. Specificity may be no higher than 80–85%, meaning that 15–20% of normal patients will have false-positive test results.

The advantages of barium enema radiography include avoidance of sedation and imaging of the full length of the colorectum. This approach may have special attractiveness in individuals who are on anticoagulants or who have structural impediments to safe colonoscopy (e.g., fixed and narrowed sigmoid colons due to dense diverticulosis). Offsetting disadvantages include the uncertain benefit of this approach, the requirement for full cathartic preparation (and repreparation if a lesion is found), and a high false-positive rate.

Colonoscopy

Most clinicians consider colonoscopy to be the diagnostic gold standard for colorectal evaluation. However, owing to its expense, invasiveness, and small risk of fatal complications, colonoscopy had not been considered for average-risk screening until recently. Lacking controlled trials on the effectiveness of colonoscopy, various models suggest that colonoscopy is the most effective tool available to reduce colorectal cancer incidence and mortality. At a frequency of every 10 years beginning at age 50, screening colonoscopy is estimated to reduce colorectal cancer by 75–90%. Medicare now covers colonoscopy for average-risk screening.

The advantages of screening colonoscopy are its unparalleled point sensitivity and specificity, high

level of effectiveness in reducing colorectal cancer risk, and capacity to remove premalignant lesions without additional procedures. Disadvantages include the inconvenient and unpleasant bowel preparation, time away from daily activities, sedation, limited or delayed access in many areas, and small risk of serious complications. Furthermore, at a frequency of every 10 years, the most aggressive neoplasms (i.e., those with the shortest premalignant dwell time) may be missed.

Emerging Technologies

New approaches to colorectal cancer screening may offer patient-friendly improvements and encompass more accurate stool tests, especially those that target DNA markers, and minimally invasive structural evaluation by CT colonography (virtual colonoscopy).

DNA-Based Stool Testing

DNA-based stool testing represents a novel early detection method with considerable promise. Because DNA is released into the fecal stream continuously via exfoliation, rather than intermittently via bleeding, targeting this class of markers should lead to enhanced sensitivity over fecal occult blood testing. Moreover, since DNA originates directly from the neoplasm, rather than from the circulation, a higher level of specificity is possible. Known genetic alterations associated with colorectal carcinogenesis can be targeted as markers. Sensitive laboratory techniques allow for the detection of minute amounts of fecal DNA and these analytes also appear to be stable during transit and storage. Finally, dietary constituents and pharmaceutical agents are unlikely to interfere with fecal DNA testing.

Early investigations assessing single DNA markers (usually K-*ras*) demonstrated that the mutations present in tumor tissue can be detected in stool obtained from the same patient. However, because colorectal neoplasms are known to be genetically heterogeneous, no single universally expressed mutation has yet been identified. For example, mutant K-*ras* appears to be present in fewer than half of all large bowel tumors, which would restrict its maximum sensitivity for CRC detection to less than 50% if used as the sole analyte in a stool screening assay.

Early data suggest that the diagnostic yield is increased using multitarget stool assays directed at a spectrum of DNA alterations associated with CRC. Findings to date further suggest that fecal DNA screening may also detect supracolonic aerodigestive cancers at sensitivities comparable to those observed with colorectal neoplasia. As such, stool screening with DNA markers could have cancer control benefits beyond just the colorectum alone. However, it will be important to evaluate the implications of these preliminary results on other relevant program parameters, such as screening cost-effectiveness, and corroborate performance characteristics in large representative populations.

Computed Tomography Colonography

CT colonography represents a new minimally invasive alternative to conventional colorectal imaging that incorporates virtual reality technology and rapid CT scanners to yield detailed two and three-dimensional views of the entire colorectal surface. After a CT scan in the supine and prone positions, each obtained with a single breath-hold, diagnostic interrogation is performed on the virtual image rather than on the patient. Early studies by some indicate that cathartic bowel cleansing may be obviated, as stool can be subtracted digitally from the image, and such virtual preparation may represent an important incentive for compliance. CT colonography appears to detect colorectal neoplasms >1 cm at sensitivity rates of 70–95% with specificities of 85–95%, based on early observations in selected patient populations. Performance characteristics in large screening populations have not yet been determined.

Based on modeling using performance assumptions to date, CT colonography may not be cost-effective relative to screening colonoscopy unless it is substantially less expensive than colonoscopy and compliance is much higher. Technical refinements with CT colonography will certainly continue.

CHEMOPREVENTION

Chemoprevention can be defined as the use of chemical compounds to prevent, inhibit, or reverse carcinogenesis prior to the point when dysplastic epithelial cells invade across the basement membrane. In its broadest sense, chemoprevention includes the use of either nutritional or pharmaceutical agents. A variety of nutritional interventions have been described above (see Environmental Factors). The following section highlights three pharmaceutical agents with potential application to CRC chemoprevention.

Nonsteroidal Anti-Inflammatory Drugs

Extensive epidemiologic data suggest that regular use of nonsteroidal anti-inflammatory drugs (NSAIDs) may reduce CRC risk by approximately 40–60%. The chemopreventive effects of NSAIDs are thought to result from cyclooxygenase-2 (COX-2) inhibition. Recently, agents that selectively block the COX-2 enzyme isoform (for example, celecoxib and rofecoxib) have been developed. These drugs are being actively investigated among

high-risk CRC patients to determine whether they afford an improved risk : benefit profile.

Exogenous Estrogens

Estrogen compounds decrease hepatic bile acid synthesis, with a consequent lowering of secondary bile acid concentrations within the colorectal lumen. The estrogen receptor also appears to be functionally important in growth regulation in the large bowel mucosa, although the specific cellular pathways involved have yet to be defined. Based on a recent meta-analysis of 28 observational studies, ever use of estrogen compounds appears to decrease CRC risk by approximately 20%.

Ursodeoxycholic Acid

Ursodeoxycholic acid (UDCA) is a hydrophilic epimer of chenodeoxycholate. This compound is noncytotoxic, stimulates the expression of major histocompatibility complex antigens, and may serve to beneficially modulate cellular growth and differentiation through a protein kinase C-mediated pathway. When used to treat patients with liver disease, UDCA has been found to have an excellent safety profile. Impressive chemopreventive potential has been demonstrated in animal models of CRC. Two randomized, clinical trials among patients with a history of colorectal adenomas are currently under way. These data should provide strong evidence regarding the chemopreventive efficacy of UDCA.

TREATMENT

Surgery

Surgical excision is the mainstay of CRC treatment, especially when a curative outcome may be potentially achievable. The type of operation and extent of resection are dependent on multiple factors, including tumor location, size, and preoperative stage. In general, right hemicolectomy is performed for CRCs arising from the cecum, ascending colon, or hepatic flexure. The transverse colon and both flexures are typically removed when cancers originate from anywhere between the ascending colon and the descending colon. For distal colon cancers, left hemicolectomy is usually the procedure of choice. For any cancers above the rectum, at least a 5 cm margin of grossly uninvolved tissue should be obtained and regional lymph nodes should be aggressively sampled. Most colonic lesions can be resected with a primary anastomosis. Adenocarcinomas in the middle and upper rectum are usually removed by anterior resection. Cancers in the lower rectum (0–5 cm above the anal verge) may require preoperative chemoradiation therapy and/or abdominoperineal resection with a permanent colostomy.

Adjuvant Therapy

Although specific dosages and delivery methods may differ slightly across institutions, adjuvant chemotherapy is usually given to stage III colon cancer patients and typically includes 5-fluorouracil (5-FU) and leucovorin. A subset of stage II colon cancer patients may benefit from this regimen as well. Chemotherapy is often begun 3–5 weeks after surgery and is administered for a duration of 6 months. Common toxicities include stomatitis, diarrhea, and neutropenia. For rectal cancer patients, adjuvant therapy may be recommended for stage II or stage III disease. A combination of 5-FU and radiation therapy (preoperatively, postoperatively, or both) is used at most centers in the United States. Investigational agents such as irinotecan (CPT-11) and oxaliplatin have shown considerable promise in early clinical trials and may become part of standard chemotherapy regimens in the near future.

See Also the Following Articles

Cancer, Overview • Colectomy • Colitis, Ulcerative • Colonoscopy • Colorectal Adenomas • Colorectal Cancer Screening • Colostomy • Crohn's Disease • Diabetes Mellitus • Diet and Environment, Role in Colon Cancer • Familial Adenomatous Polyposis (FAP) • Lynch Syndrome/ Hereditary Non-Polyposis Colorectal Cancer (HNPCC) • Sigmoidoscopy

Further Reading

Ahlquist, D. A., and Shuber, A. P. (2002). Stool screening for colorectal cancer: Evolution from occult blood to molecular markers. *Clin. Chim. Acta* **315**, 157–168.

Bond, J. H. (2000). Polyp guideline: Diagnosis, treatment, and surveillance for patients with colorectal polyps. Practice Parameters Committee of the American College of Gastroenterology. *Am. J. Gastroenterol.* **95**, 3053–3063.

Calvert, P. M., and Frucht, H. (2002). The genetics of colorectal cancer. *Ann. Intern. Med.* **137**, 603–612.

Colditz, G. A., Cannuscio, C. C., and Frazier, A. L. (1997). Physical activity and reduced risk of colon cancer: Implications for prevention. *Cancer Causes Control* **8**, 649–667.

Giardiello, F. M., Brensinger, J. D., and Petersen, G. M. (2001). AGA technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* **121**, 198–213.

Gwyn, K., and Sinicrope, F. A. (2002). Chemoprevention of colorectal cancer. *Am. J. Gastroenterol.* **97**, 13–21.

Hawk, E., Lubet, R., and Limburg, P. (1999). Chemoprevention in hereditary colorectal cancer syndromes. *Cancer* **86**, 2551–2563.

Hobday, T. J., and Erlichman, C. (2002). Adjuvant therapy of colon cancer: A review. *Clin. Colorectal Cancer* **1**, 230–236.

Minsky, B. D. (2002). Adjuvant therapy of resectable rectal cancer. *Cancer Treat. Rev.* **28**, 181–188.

Murthy, S., Flanigan, A., and Clearfield, H. (2002). Colorectal cancer in inflammatory bowel disease: Molecular and clinical features. *Gastroenterol. Clin. North Am.* **31**, 551–564.

Potter, J. D. (1999). Colorectal cancer: Molecules and populations. *J. Natl. Cancer Inst.* **91**, 916–932.

Smith, R. A., Cokkinides, V., von Eschenbach, A. C., Levin, B., Cohen, C., Runowicz, C. D., Sener, S., Saslow, D., and Eyre, H. J. (2002). American Cancer Society guidelines for the early detection of cancer. *CA Cancer J. Clin.* **52**, 8–22.

Surveillance, Epidemiology, and End Results (SEER) Program Website. Available at http://seer.cancer.gov/.

Winawer, S. J., Zauber, A. G., O'Brien, M. J., Ho, M. N., Gottlieb, L., Sternberg, S. S., Waye, J. D., Bond, J., Schapiro, M., Stewart, E. T., *et al.* (1993). Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. The National Polyp Study Workgroup. *N. Engl. J. Med.* **328**, 901–906.

Colorectal Adenomas

AYAAZ ISMAIL AND PETER LANCE
University of Arizona, Tucson

colorectal adenocarcinoma Malignant neoplasm of the colon or rectum (colorectal cancer).

colorectal adenoma Benign neoplasm of the colon or rectum with malignant potential.

colorectal polyp Growth protruding from the mucosal surface into the bowel lumen.

ureterosigmoidostomy Surgical implantation of the ureters into the sigmoid colon.

Colorectal cancer is the second most common cause of cancer-related death in males and females in the United States. In 2003, it is estimated there will be 147,500 new cases of colorectal cancer and 57,100 deaths from the disease. Almost all colorectal cancers arise from adenomas, which are benign neoplasms. Adenomas that are removed cannot progress to become cancers. Screening for adenomas and their removal are, therefore, now a major part of colorectal cancer screening programs.

INTRODUCTION

Colorectal polyps are raised growths that protrude into the bowel lumen. They are described by their shape (sessile or pedunculated), size, and histological type. The common histological types are inflammatory, hyperplastic, hamartomatous, and adenomatous polyps. It is only adenomatous polyps (adenomas) that are neoplastic and therefore have malignant potential. Colorectal adenomas are classified histologically as tubular, tubulovillous, or villous. The epithelium of all adenomas is dysplastic, which is categorized as low or high grade. Advanced adenomas have the greatest potential for malignant transformation and are defined as adenomas of diameter ≥ 1 cm, or those of any size with villous histology or high-grade dysplasia. Advanced colorectal neoplasms are composed of colorectal adenocarcinomas and advanced adenomas. Adenomas are readily detected endoscopically using a sigmoidoscope or colonoscope and most can also be completely removed during the same examination by the procedure of polypectomy.

EPIDEMIOLOGY

Prevalence

Adenomas are uncommon before the age of 50 years in individuals at average risk for colorectal cancer. Prevalence rates from autopsy studies are higher than those in comparable groups determined colonoscopically. Adenoma prevalence is over 40% in subjects undergoing first colonoscopy in the seventh decade and approaches 60% in autopsy studies of subjects over the age of 65.

 Copyright 2004, Elsevier (USA). All rights reserved.

Murthy, S., Flanigan, A., and Clearfield, H. (2002). Colorectal cancer in inflammatory bowel disease: Molecular and clinical features. *Gastroenterol. Clin. North Am.* **31**, 551–564.

Potter, J. D. (1999). Colorectal cancer: Molecules and populations. *J. Natl. Cancer Inst.* **91**, 916–932.

Smith, R. A., Cokkinides, V., von Eschenbach, A. C., Levin, B., Cohen, C., Runowicz, C. D., Sener, S., Saslow, D., and Eyre, H. J. (2002). American Cancer Society guidelines for the early detection of cancer. *CA Cancer J. Clin.* **52**, 8–22.

Surveillance, Epidemiology, and End Results (SEER) Program Website. Available at http://seer.cancer.gov/.

Winawer, S. J., Zauber, A. G., O'Brien, M. J., Ho, M. N., Gottlieb, L., Sternberg, S. S., Waye, J. D., Bond, J., Schapiro, M., Stewart, E. T., *et al.* (1993). Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. The National Polyp Study Workgroup. *N. Engl. J. Med.* **328**, 901–906.

Colorectal Adenomas

AYAAZ ISMAIL AND PETER LANCE
University of Arizona, Tucson

colorectal adenocarcinoma Malignant neoplasm of the colon or rectum (colorectal cancer).

colorectal adenoma Benign neoplasm of the colon or rectum with malignant potential.

colorectal polyp Growth protruding from the mucosal surface into the bowel lumen.

ureterosigmoidostomy Surgical implantation of the ureters into the sigmoid colon.

Colorectal cancer is the second most common cause of cancer-related death in males and females in the United States. In 2003, it is estimated there will be 147,500 new cases of colorectal cancer and 57,100 deaths from the disease. Almost all colorectal cancers arise from adenomas, which are benign neoplasms. Adenomas that are removed cannot progress to become cancers. Screening for adenomas and their removal are, therefore, now a major part of colorectal cancer screening programs.

INTRODUCTION

Colorectal polyps are raised growths that protrude into the bowel lumen. They are described by their shape (sessile or pedunculated), size, and histological type. The common histological types are inflammatory, hyperplastic, hamartomatous, and adenomatous polyps. It is only adenomatous polyps (adenomas) that are neoplastic and therefore have malignant potential. Colorectal adenomas are classified histologically as tubular, tubulovillous, or villous. The epithelium of all adenomas is dysplastic, which is categorized as low or high grade. Advanced adenomas have the greatest potential for malignant transformation and are defined as adenomas of diameter ≥ 1 cm, or those of any size with villous histology or high-grade dysplasia. Advanced colorectal neoplasms are composed of colorectal adenocarcinomas and advanced adenomas. Adenomas are readily detected endoscopically using a sigmoidoscope or colonoscope and most can also be completely removed during the same examination by the procedure of polypectomy.

EPIDEMIOLOGY

Prevalence

Adenomas are uncommon before the age of 50 years in individuals at average risk for colorectal cancer. Prevalence rates from autopsy studies are higher than those in comparable groups determined colonoscopically. Adenoma prevalence is over 40% in subjects undergoing first colonoscopy in the seventh decade and approaches 60% in autopsy studies of subjects over the age of 65.

 Copyright 2004, Elsevier (USA). All rights reserved.

Distribution

Approximately 50% of adenomas diagnosed at colonoscopy are located distal to the splenic flexure (distal adenomas). The proportion of adenomas on the right side of the colon (proximal adenomas) increases with age and the proportion of all adenomas located on the right side of the colon has increased in recent decades. Individuals with a distal adenoma are more likely than those without a distal adenoma to have proximal adenomas, but approximately 50% of patients with advanced proximal adenomas do not have distal adenomas.

Conditions Associated with Adenomas

There are three conditions typically associated with adenomas:

1. Ureterosigmoidostomy. Up to 29% of patients with a ureterosigmoidostomy develop adenomas or even CRC at the ureterosigmoidostomy site. The latency period is long, with an average of 20 years, but shorter periods have been reported. Nitrosamines generated from the diverted urine are thought to be important etiologic factors.
2. Acromegaly. The risk for developing colorectal adenomas and CRC is sixfold higher in acromegalics than in control subjects. A family history of CRC further compounds this excess risk.
3. *Streptococcus* infection. *Streptococcus bovis* bacteremia and endocarditis have been associated with simple adenomas, adenomatous polyposis coli, and colon cancer. Colonoscopy has therefore been recommended for such patients. A similar association has also been reported with *Streptococcus agalactiae* infection.

ETIOLOGY

Hereditary

Individuals with the autosomally dominant hereditary conditions of familial adenomatosis polyposis (FAP) or hereditary nonpolyposis colorectal carcinoma (HNPCC) are at greatly increased risk for CRC. However, a large majority of colorectal neoplasms, termed sporadic, arise outside the context of FAP or HNPCC. Individuals with first-degree relatives (parent, sibling, or child) who have had colorectal adenomas or CRC are at increased risk for colorectal adenomas and CRC. The risk for individuals in this category is greatest for those with a first-degree relative diagnosed with CRC or adenomas at age <60 years. Individuals with no first-degree relatives diagnosed with colorectal neoplasms are deemed at "average risk" for colorectal adenomas and cancer.

Diet

Dietary factors are considered the most important determinants of risk for developing colorectal adenomas in average-risk individuals. Wide geographic variations in the prevalence of colorectal adenomas and cancer are attributed to differences in diet. Although the "Western" diet that is high in animal fat and low in fruit, vegetables, and fiber has been widely incriminated for the high prevalence of colorectal neoplasms in developed countries, definitive proof for a causal association between a Western diet and adenoma or CRC development is largely lacking. Recent randomized prospective clinical trials in individuals who had recently undergone colonoscopic polypectomy (removal of adenomas) failed to show any reduction of adenoma recurrence rates following consumption of diets low in animal fat or high in fruit and vegetable servings and fiber. There is reliable evidence that high levels of folate consumption and calcium supplements can reduce the risk for developing colorectal neoplasms.

Other Factors

Aspirin is a nonsteroidal antiinflammatory drug (NSAID) that acts by inhibiting the cyclooxygenase (COX) enzymes, COX-1 and -2. Daily aspirin ingestion over a period of years can reduce mortality from CRC by up to 50% and aspirin has recently been shown to reduce the rate of adenoma recurrence following polypectomy. The antineoplastic actions of aspirin are attributed to inhibition of COX-2 rather than COX-1, and COX-2 levels are increased in colorectal neoplasms. The serious and potentially life-threatening complications of aspirin are due largely to inhibition of COX-1. Selective NSAIDs that inhibit only COX-2 are now available. A number of large trials are currently in progress to determine whether the colorectal antineoplastic actions of the selective COX-2 inhibitor NSAIDs are equivalent to those of aspirin.

THE ADENOMA–CARCINOMA SEQUENCE

Abundant evidence indicates that almost all CRCs develop from benign adenomas. The neoplastic changes in adenomas constitute histological dysplasia and are confined to the epithelial cell layer. An adenoma progresses to become a CRC when dysplastic cells

transgress the muscularis mucosae to invade the bowel wall. Several lines of evidence support the concept of a colorectal adenoma–carcinoma sequence.

Epidemiology

The prevalence of both adenomas and CRC increases with age, but the prevalence of adenomas peaks 5 to 10 years earlier than the peak prevalence for CRC. This applies to sporadic adenomas and the adenomas of subjects with FAP.

Clinical

Subsequent incidence of CRC is profoundly reduced in subjects who have undergone adenoma polypectomy.

Morphologic

Large adenomas sometimes harbor adenocarcinomatous foci and adenomatous remnants are sometimes seen immediately adjacent to invasive cancer, implying a histological continuum from benign to malignant colorectal neoplasms. However, small foci of adenocarcinoma surrounded by normal mucosa are encountered very rarely.

Genetic

The progression from formation of the earliest adenoma to advanced adenoma and invasive cancer is driven by accumulation of genetic mutations in the neoplastic tissue (these are somatic mutations as opposed to the germ-line mutations of FAP and HNPCC). In the commonest mutational pathway, formation of the initial adenoma is heralded by mutations of both copies of the adenomatous polyposis coli (APC) gene in the colorectal epithelial cells that form the adenoma. In those adenomas that grow and have an aggressive phenotype, transition from a benign adenoma to an invasive cancer is marked by sequential mutations to both copies of the p53 gene. Mutations or silencing of a number of other genes are required for the adenoma–carcinoma sequence to be completed. The accumulated molecular genetic events that are the hallmark of colorectal carcinogenesis provide strong direct evidence linking benign and malignant colorectal neoplasms through the adenoma–carcinoma sequence.

NATURAL HISTORY OF COLORECTAL ADENOMAS

With adenoma prevalence rates as high as 50% or more in older populations, it is clear that only a small minority of adenomas progress to CRC. The most common adenoma is a diminutive (diameter $\leq$5 mm) tubular adenoma with low-grade dysplasia. Fewer than 10% of adenomas progress to become advanced adenomas and CRC. Advanced adenomas are those most likely to progress. When it occurs, progression from adenoma, most of which are of diameter $<$10 mm, to CRC is estimated to take at least 10 years on average. However, there have been very few studies in which adenomas have been identified and observed over a period of years rather than being removed when first diagnosed. In one radiological study using barium enema examinations, the time period for adenomas $>$10 mm in diameter to progress to become CRCs was from 2 to 5 years.

CLINICAL MANIFESTATIONS

A majority of adenomas are asymptomatic. A minority of large polyps may bleed sufficiently for an occult blood test to be positive without frank blood being visible in the stool. Frank bleeding from even large polyps is rare. Altered bowel habit and pain from colorectal neoplasia almost always signify CRC rather than an adenoma.

DIAGNOSIS

Because most adenomas are asymptomatic, they are usually detected by screening tests or as incidental findings in the course of investigations for unrelated symptoms. Screening of all subjects for colorectal neoplasms is now recommended, starting from the age of 50. Several modalities are available.

Fecal Occult Blood Testing

The surface epithelium of adenomas is intact so that bleeding is infrequent and intermittent. Therefore, the sensitivity of fecal occult blood testing (FOBT) for detection of adenomas is extremely poor and a majority remain undiagnosed if FOBT is the only screening tool. To improve sensitivity, it is recommended that flexible sigmoidoscopy be combined with FOBT.

Flexible Sigmoidoscopy

The 60-cm flexible sigmoidoscope has replaced the 25-cm rigid sigmoidoscope. Among asymptomatic individuals undergoing screening flexible sigmoidoscopy, 10–15% are found to have adenomas. The major limitation of flexible sigmoidoscopy as a screening tool is the inaccessibility of proximal colonic lesions to the flexible sigmoidoscope. The prevalence

of distal colorectal adenomas is higher in subjects who have a proximal advanced neoplasm than in those who do not. Nonetheless, approximately one-half of subjects with an asymptomatic proximal advanced neoplasm will have a negative flexible sigmoidoscopy. The proximal advanced neoplasms will, therefore, go undiagnosed if FOBT and flexible sigmoidoscopy are negative.

Double-Contrast Barium Enema

The sensitivity of double-contrast barium enema (DCBE) for detecting adenomas $\geq$6 mm in diameter is approximately 50%. It is an unreliable modality for detecting smaller adenomas. Flexible sigmoidoscopy should always be combined with DCBE because lesions in the distal rectum are often not visualized by DCBE.

Colonoscopy

Colonoscopy is indicated for all subjects with a positive FOBT or adenoma(s) diagnosed by flexible sigmoidoscopy. It is also increasingly being advocated as the primary screening tool for colorectal neoplasms. The sensitivity of colonoscopy for detection of advanced adenomas is >90% and somewhat lower for small adenomas. Colonoscopy, however, has several drawbacks as the primary screening tool for colorectal neoplasms in average-risk subjects. Colonic perforations of 1 per 1000 to 2000 have been reported, although recent studies suggest this may overestimate their frequency. The costs of colonoscopy are considerable. It is highly doubtful whether sufficient skilled colonoscopists could be trained in the foreseeable future to be able to provide screening colonoscopy for the entire United States population aged 50 years and older. Nonetheless, colonoscopy is regarded as the "gold standard" screening technique for colorectal adenomas and CRC.

Other Techniques

Virtual colonoscopy is another diagnostic tool, although the diagnostic accuracy of this procedure is not well established. After a short bowel preparation, patients undergo computed tomography with colonic insufflation. Mucosal images are reconstructed to identify adenomas and other mass lesions. Colonoscopy is indicated for all subjects with a mass lesion on virtual colonoscopy. Virtual colonoscopy has not yet been adequately tested outside the research setting and should not, therefore, be used for routine screening purposes.

In another procedure, fecal DNA from exfoliated cells can be extracted from stool and amplified by polymerase chain reaction. In this way, it is possible to detect somatic mutations and other abnormalities in DNA deriving from the exfoliated cells of asymptomatic colorectal neoplasms. Studies are in progress to determine the sensitivity and specificity of this approach to screening.

TREATMENT

Colonoscopy

All patients with a positive FOBT, flexible sigmoidoscopy, or DCBE should undergo a colonoscopy. Because most colon cancers arise from adenomas and visual examination of a polyp does not reliably differentiate adenomatous from nonadenomatous polyps, all polyps should be removed or at least biopsied and submitted for histological examination. The histological type determines subsequent management. Small adenomas are removed using cold or hot biopsies whereas larger polyps are removed using a snare with or without cautery. Very large or broad-based sessile polyps are removed piecemeal and any remnant polyp tissue can be ablated by a variety of thermal devices.

Endoscopic polypectomy is safe and carries a less than 1% risk of perforation or bleeding. This risk is increased when removing multiple polyps, very large polyps, and lesions in the cecum or ascending colon, which have a thinner wall.

Surgery

Surgical resection is needed for lesions that cannot be adequately removed endoscopically. Polyps found to harbor foci of carcinoma warrant special consideration. If the cancer cells are intramucosal and do not traverse the muscularis mucosae, they are considered to be noninvasive malignant polyps and polypectomy alone is curative. When the cancer cells have invaded beyond the muscularis mucosa, the polyp is considered to be a malignant polyp. Polypectomy alone of a malignant polyp carries a 10% risk of having cancer cells remaining in the bowel wall or surrounding lymph nodes. Several unfavorable factors have been identified, the presence of which increase the risk for an adverse outcome. These include poorly differentiated carcinomas, evidence of vascular or lymphatic invasion, involvement of resection margins, and involvement of the submucosa of the

bowel wall, which is likely in sessile malignant polyps. Surgical resection is indicated if any one of these unfavorable features is present.

Postpolypectomy Followup

Up to 50% of patients will develop recurrent adenomas following initial polypectomy. The annual rate of adenoma recurrence following polypectomy is from 10 to 15%. The likelihood of recurrence is greatest with multiple adenomas, large polyps, villous histology, and severe dysplasia. The optimal time period between examinations depends on the initial findings. Following an adequate examination and complete resection of all adenomas, surveillance colonoscopy should be performed in 3 to 5 years, depending on characteristics of the initial adenoma(s). If recurrent adenoma(s) are found at the first surveillance exam, colonoscopy should be repeated in 3 years. In the absence of recurrent adenomas, the surveillance interval can be extended to 5 years.

See Also the Following Articles

Cancer, Overview • Colonoscopy • Colorectal Adenocarcinoma • Colorectal Cancer Screening • Diet and Environment, Role in Colon Cancer • Familial Adenomatous Polyposis • Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) • Sigmoidoscopy • Virtual Colonoscopy

Further Reading

Janne, P. A., and Mayer, R. J. (2000). Chemoprevention of colorectal cancer. *N. Engl. J. Med.* **342**(26), 1960–1968.

Leslie, A., Carey, F. A., Pratt, N. R., *et al.* (2002). The colorectal adenoma-carcinoma sequence. *Br. J. Surg.* **89**, 845–860.

Sleisenger, M. H. (2002). Colonic polyps and polyposis syndromes. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease. Pathophysiology/Diagnosis/Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., Vol. 2, pp. 2175–2214. W. B. Saunders, Philadelphia.

Winawer, S., Fletcher, R., Rex, D., *et al.* (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale—Update based on new evidence. *Gastroenterology* **124**(2), 544–560.

Winawer, S. J., Zauber, A. G., Gerdes, H., *et al.* (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N. Engl. J. Med.* **329**, 1977–1981.

Colorectal Cancer Screening

DAVID JAGER AND MARIE L. BORUM
George Washington University

adenoma Benign epithelial tumor in which the cells form recognizable glandular structures or in which the cells are clearly derived from glandular epithelium.

colonoscope Fiber-optic flexible endoscope that permits visual examination of the entire colon.

colonoscopy Examination of the colon using th colonoscope.

familial adenomatous polyposis Autosomal dominant disease characterized by the development of hundreds to thousands of adenomas in the colon.

hereditary nonpolyposis colorectal cancer Autosomal dominant condition in which there is an increased risk of developing colorectal cancer, as well as ovarian, renal, pancreatic, and endometrial cancers.

Colorectal cancer is the third leading cause of cancer deaths in the United States. Approximately 130,000 new cases of colorectal cancer are diagnosed every year. It is found in all ethnic groups, with the highest prevalence occurring in African Americans. Individuals at the greatest risk for the development of colorectal cancer include those with a history of adenomatous polyps, previous colorectal cancer, inflammatory bowel disease, or an inherited syndrome that predisposes to colorectal cancer development. A family history of adenomatous polyps or colorectal cancer can also increase the risk for the development of colorectal cancer. In

Copyright 2004, Elsevier (USA). All rights reserved.

bowel wall, which is likely in sessile malignant polyps. Surgical resection is indicated if any one of these unfavorable features is present.

Postpolypectomy Followup

Up to 50% of patients will develop recurrent adenomas following initial polypectomy. The annual rate of adenoma recurrence following polypectomy is from 10 to 15%. The likelihood of recurrence is greatest with multiple adenomas, large polyps, villous histology, and severe dysplasia. The optimal time period between examinations depends on the initial findings. Following an adequate examination and complete resection of all adenomas, surveillance colonoscopy should be performed in 3 to 5 years, depending on characteristics of the initial adenoma(s). If recurrent adenoma(s) are found at the first surveillance exam, colonoscopy should be repeated in 3 years. In the absence of recurrent adenomas, the surveillance interval can be extended to 5 years.

See Also the Following Articles

Cancer, Overview • Colonoscopy • Colorectal Adenocarcinoma • Colorectal Cancer Screening • Diet and Environment, Role in Colon Cancer • Familial Adenomatous Polyposis • Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) • Sigmoidoscopy • Virtual Colonoscopy

Further Reading

Janne, P. A., and Mayer, R. J. (2000). Chemoprevention of colorectal cancer. *N. Engl. J. Med.* **342**(26), 1960–1968.

Leslie, A., Carey, F. A., Pratt, N. R., *et al.* (2002). The colorectal adenoma-carcinoma sequence. *Br. J. Surg.* **89**, 845–860.

Sleisenger, M. H. (2002). Colonic polyps and polyposis syndromes. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease. Pathophysiology/Diagnosis/Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., Vol. 2, pp. 2175–2214. W. B. Saunders, Philadelphia.

Winawer, S., Fletcher, R., Rex, D., *et al.* (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale—Update based on new evidence. *Gastroenterology* **124**(2), 544–560.

Winawer, S. J., Zauber, A. G., Gerdes, H., *et al.* (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N. Engl. J. Med.* **329**, 1977–1981.

Colorectal Cancer Screening

DAVID JAGER AND MARIE L. BORUM
George Washington University

adenoma Benign epithelial tumor in which the cells form recognizable glandular structures or in which the cells are clearly derived from glandular epithelium.

colonoscope Fiber-optic flexible endoscope that permits visual examination of the entire colon.

colonoscopy Examination of the colon using th colonoscope.

familial adenomatous polyposis Autosomal dominant disease characterized by the development of hundreds to thousands of adenomas in the colon.

hereditary nonpolyposis colorectal cancer Autosomal dominant condition in which there is an increased risk of developing colorectal cancer, as well as ovarian, renal, pancreatic, and endometrial cancers.

Colorectal cancer is the third leading cause of cancer deaths in the United States. Approximately 130,000 new cases of colorectal cancer are diagnosed every year. It is found in all ethnic groups, with the highest prevalence occurring in African Americans. Individuals at the greatest risk for the development of colorectal cancer include those with a history of adenomatous polyps, previous colorectal cancer, inflammatory bowel disease, or an inherited syndrome that predisposes to colorectal cancer development. A family history of adenomatous polyps or colorectal cancer can also increase the risk for the development of colorectal cancer. In

Copyright 2004, Elsevier (USA). All rights reserved.

addition, risk factors that have been noted to be associated with colorectal cancer development include obesity, physical inactivity, excess caloric intake, and centrally deposited adipose tissue.

INTRODUCTION

The stage at which colorectal cancer is diagnosed relates to patient survival. Screening for the malignancy has the potential to detect premalignant lesions and early cancers such that curative intervention can be initiated.

RISK STRATIFICATION

The initial step in colorectal cancer screening is the determination of an individual's risk status. Risk stratification determines when screening should be initiated and what tests should be performed. Guidelines by the United States Multisociety Task Force on Colorectal Cancer note that risk stratification can be accomplished by inquiring about several issues. It should be determined if the patient has had colorectal cancer or adenomatous polyps, if the patient has an illness that predisposes to colorectal cancer, or if the patient has a family member who has had an adenomatous polyp or colorectal cancer. The relationship of the affected family member to the individual and the age at which polyps or colorectal cancer developed in the relative should be determined.

If a person is asymptomatic and without a personal or family history that predisposes to colorectal cancer, the individual is considered to be of average risk and should begin colorectal cancer screening at 50 years of age. If a person has a first-degree relative or two or more second-degree relatives who have had adenomatous polyps or colorectal cancer at ≥ 60 years of age, they should be considered to be at average risk, but can begin screening at 40 years of age. If a person has a first-degree relative with colorectal cancer or adenomatous polyps at <60 years of age or two or more second-degree relatives with colorectal cancer at any age, they are considered to be at increased risk for colorectal cancer and screening should begin at age 40 years, or 10 years younger than the age at which the earliest diagnosis was made. If a person has a family history of familial adenomatous polyposis or hereditary nonpolyposis colorectal cancer, the patient is considered to be at increased risk and screening is initiated based on the particular condition.

RECOMMENDATIONS FOR AVERAGE-RISK INDIVIDUALS

Guidelines for colorectal cancer screening include fecal occult blood testing and endoscopic (flexible sigmoidoscopy or colonoscopy) or radiologic evaluation of the colon.

Fecal Occult Blood Testing

Fecal occult blood testing (FOBT), using a guaiac-based test, should be started at age 50 years and should be performed annually. FOBT is performed on three consecutive stools without hydration. There have been three randomized controlled studies that reported that FOBT reduces the risk of death from colorectal cancer. Screening is most effective when performed annually, as opposed to every 2 years. Dietary and medication restrictions are commonly recommended to reduce the false positive rate for the sensitive guaiac-based tests. It is recommended that individuals avoid red meat for 3 days, as well as avoiding radishes, broccoli, turnips, nonsteroidal antiinflammatory drugs, and vitamin C for 2 days prior to performing the test. It is recommended not to rehydrate the samples prior to FOBT despite the increased sensitivity, because it decreases the specificity of the testing.

Flexible Sigmoidoscopy

A flexible sigmoidoscopy allows for visualization of the distal portion of the colon. The current recommendation is that flexible sigmoidoscopy should be offered every 5 years. The performance of a flexible sigmoidoscopy can reduce colorectal cancer mortality by up to two-thirds for lesions that are within the view of the instrument. Additionally, a case-control study has demonstrated that the use of a rigid sigmoidoscopy for screening reduces mortality from distal colorectal cancer by 59%. If a polyp detected by sigmoidoscopy is ≥ 1 cm in size, it should be assumed to be adenomatous. A polyp <1 cm in size should be evaluated to determine if it is hyperplastic or adenomatous. Sigmoidoscopies detecting adenomatous polyps, when followed by a colonoscopy with removal of all polyps, have been associated with an 80% reduction in the incidence of colorectal cancer.

Barium Enema

A barium enema, performed every 5 years, is an alternative to the use of endoscopic examinations for colorectal cancer screening. This radiologic examination results in the visualization of the entire colon and can detect the presence of most large polyps. However, there have been no randomized controlled studies that demonstrate a reduction of mortality from colorectal cancer in individuals at average risk for the disease. Additionally, the barium enema does not allow for biopsies of suspicious lesions or removal of polyps.

This procedure is an option for those individuals who have a contraindication to or are unable to undergo endoscopic evaluation.

Colonoscopy

The colonoscopy is the most effective tool for the early detection of precancerous and cancerous lesions and can be performed every 10 years for colorectal cancer screening. The colonoscopy allows for the visualization of the entire colon and the biopsy or removal of colonic lesions. Two cohort studies have demonstrated that the colonoscopy can reduce the incidence of colorectal cancer in individuals with adenomatous polyps. Disadvantages for the colonoscopy include the increased cost, risk of perforation, and greater degree of invasiveness. The recommended 10-year interval between screenings is based on the rate at which adenomatous polyps can convert to cancer. It has been estimated that it may take >10 years for an adenomatous polyp to develop and convert to cancer. In a study of asymptomatic average-risk individuals with negative screening colonoscopies, a second colonoscopy 5 years later had a <1% incidence of advanced neoplasm.

RECOMMENDATIONS FOR HIGH-RISK PATIENTS

Family History of Adenomatous Polyp or Cancer

Any person with a first-degree relative with adenomatous polyps or colon cancer diagnosed at age <60 years or two first-degree relatives with colorectal cancer should have a colonoscopy at age 40 years, or 10 years before the earliest diagnosis in their family. The colonoscopy should be repeated every 5 years. In an individual who has a first-degree relative with colon cancer or adenomatous polyps at age ≥60 years or two second-degree relatives with colorectal cancer, screening should begin at age 40 years and should occur at intervals similar to those for individuals at average risk for cancer. Individuals with one second-degree or third-degree relative with colorectal cancer should be screened as an average-risk individual.

Family History of Familial Adenomatous Polyposis

Individuals with familial adenomatous polyposis (FAP) or those who are at risk for having FAP should have an annual endoscopic screening starting at age 10–12 years. FAP is an autosomal dominant disorder with adenomas appearing at an average age of 16 years. The cause of FAP is a mutation in the adenomatous polyposis coli gene (*APC*). The current recommendations suggest that in addition to endoscopic screening, genetic testing should be considered in individuals with FAP and relatives who are at risk for the disorder.

Hereditary Nonpolyposis Colorectal Cancer

Individuals with hereditary nonpolyposis colorectal cancer (HNPCC) or those at increased risk for HNPCC should have a colonoscopy every 1–2 years beginning at the age of 20–25 years, or 10 years earlier than the earliest cancer diagnosis in the family. The genetic testing for HNPCC should be offered to first-degree relatives of those patients with a known mismatch repair gene mutation. If the individual meets clinical criteria for HNPCC, genetic testing should also be performed.

Inflammatory Bowel Disease

It has been suggested that individuals with inflammatory bowel disease for >10 years are at increased risk for colon cancer development and should have a colonoscopy with surveillance biopsies every 1–2 years.

Personal History of a Polyp or Colon Cancer

Patients who have had a malignant adenoma, multiple adenomas, or a large sessile adenoma should have a followup colonoscopy within a short period of time. The interval of time for repeat screening is typically 1 year, but is based upon the physician's clinical judgment. If an individual has had less than three adenomas, the repeat colonoscopy can be performed in 3 years. If the individuals has one to two small adenomas, the followup colonoscopy can be done in 5 years. An individual's specific followup regimen is based on the number of polyps and the pathology of the polyps.

See Also the Following Articles

Barium Radiography • Cancer, Overview • Colonoscopy • Colorectal Adenocarcinoma • Diet and Environment, Role in Colon Cancer • Familial Adenomatous Polyposis (FAP) • Familial Risk of Gastrointestinal Cancers • Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) • Sigmoidoscopy

Further Reading

Giovannucci, E. (2002). Modifiable risk factors for colon cancer. *Gastroenterol. Clin. North Am.* **31**, 925–943.

Hofstad, B., and Vatn, M. (1997). Growth rate of colon polyps and cancer. *Gastrointest. Endosc. Clin. North Am.* 7, 345–362.

Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A. (1999). Cancer statistics, 1999. *CA Cancer J. Clin.* **49**, 8–31.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Lliadou, A., Kaprio, J., Koskenvuo, M., *et al.* (2000). Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85.

Mandel, J. S., Bond, J. H., Church, T. R., Snover, D. C., Bradley, G. M., Schuman, L. M., *et al.* (1993). Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N. Engl. J. Med.* **328**, 1365–1371.

Rex, D. K., Cummings, O. W., Helper, D. J., Nowak, T. V., McGill, J. M., Chiao, G. Z., *et al.* (1996). Five-year incidence of adenomas after negative colonoscopy in asymptomatic average-risk persons. *Gastroenterology* **111**, 1178–1181.

Rozen, P., Knaani, J., and Samuel, Z. (1999). Eliminating the need for dietary restrictions when using a sensitive guaiac fecal occult blood test. *Dig. Dis. Sci.* **44**, 756–760.

Selby, J. V., Friedman, G. D., Quesenberry, C. P., Jr., and Weiss, N. S. (1992). A case control study of screening sigmoidoscopy and mortality form colorectal cancer. *N. Engl. J. Med.* **326**, 653–657.

Towbridge, B., and Burt, R. (2002). Colorectal cancer screening. *Surg. Clin. North Am.* **82**, 943–957.

Winawer, S., *et al.* (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale—Update based upon new evidence. *Gastroenterology* **124**, 544–560.

Winawer, S. J., Zauber, A. G., Ho, M. N., O'Brien, M. J., Gottlieb, L. S. A., Sternberg, S. S., *et al.* (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N. Engl. J. Med.* **329**, 1977–1981.

Colostomy

DOUGLAS W. WILMORE

Harvard Medical School and Brigham and Women's Hospital

ostomate Individual undergoing an ostomy, usually an ileostomy or colostomy.

ostomy Operation to create an artificial opening through which body wastes may exit.

A colostomy is a surgically created passage between the colon and the skin of the abdominal wall, culminating in an opening, or stoma. The purpose of a colostomy is to divert the fecal stream from the colon distal to the site of the colostomy or to prevent or avoid use of the rectum for fecal elimination. An individual with a colostomy wears a bag or appliance that collects the fecal material expelled from the stoma. A person with a colostomy is referred to as an ostomate, and it is estimated that approximately 1 million individuals in the United States have undergone ostomies. To help patients deal with difficulties, ostomy organizations share information and collectively purchase supplies in bulk. Highly trained nurses in this field of gastroenterology teach patients about dealing with their colostomies and aid in the care of the stoma. Patients with colostomies are active, productive individuals and are generally unaffected by the colostomy, but may have life changes because of the disease that prompted this operation.

INDICATIONS

There are numerous indications for a colostomy. Diversion of the fecal stream may be necessary because of a distal bowel resection that has resulted in an anastomosis. A proximal colostomy to divert feces from the operative site may be required if a surgeon feel that a suture line in the colon is too fragile to receive intestinal contents. Often, colon obstruction and distension with fluid and gas is resolved by a colostomy proximal to the obstruction so that all the contents can be expelled, which allows the colon to decompress. At a later time, the patient undergoes an operation to allow more definitive repair of the obstructing site. A proximal colostomy is often formed when the colon perforates, either from diverticulitis, inflammatory bowel disease, cancer, or secondary to pelvic or abdominal trauma, to allow inflammation to subside; final resection and repair can be performed under more ideal operative

Copyright 2004, Elsevier (USA). All rights reserved.

Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A. (1999). Cancer statistics, 1999. *CA Cancer J. Clin.* **49**, 8–31.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Lliadou, A., Kaprio, J., Koskenvuo, M., *et al.* (2000). Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85.

Mandel, J. S., Bond, J. H., Church, T. R., Snover, D. C., Bradley, G. M., Schuman, L. M., *et al.* (1993). Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N. Engl. J. Med.* **328**, 1365–1371.

Rex, D. K., Cummings, O. W., Helper, D. J., Nowak, T. V., McGill, J. M., Chiao, G. Z., *et al.* (1996). Five-year incidence of adenomas after negative colonoscopy in asymptomatic average-risk persons. *Gastroenterology* **111**, 1178–1181.

Rozen, P., Knaani, J., and Samuel, Z. (1999). Eliminating the need for dietary restrictions when using a sensitive guaiac fecal occult blood test. *Dig. Dis. Sci.* **44**, 756–760.

Selby, J. V., Friedman, G. D., Quesenberry, C. P., Jr., and Weiss, N. S. (1992). A case control study of screening sigmoidoscopy and mortality form colorectal cancer. *N. Engl. J. Med.* **326**, 653–657.

Towbridge, B., and Burt, R. (2002). Colorectal cancer screening. *Surg. Clin. North Am.* **82**, 943–957.

Winawer, S., *et al.* (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale—Update based upon new evidence. *Gastroenterology* **124**, 544–560.

Winawer, S. J., Zauber, A. G., Ho, M. N., O'Brien, M. J., Gottlieb, L. S. A., Sternberg, S. S., *et al.* (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N. Engl. J. Med.* **329**, 1977–1981.

Colostomy

DOUGLAS W. WILMORE

Harvard Medical School and Brigham and Women's Hospital

ostomate Individual undergoing an ostomy, usually an ileostomy or colostomy.

ostomy Operation to create an artificial opening through which body wastes may exit.

A colostomy is a surgically created passage between the colon and the skin of the abdominal wall, culminating in an opening, or stoma. The purpose of a colostomy is to divert the fecal stream from the colon distal to the site of the colostomy or to prevent or avoid use of the rectum for fecal elimination. An individual with a colostomy wears a bag or appliance that collects the fecal material expelled from the stoma. A person with a colostomy is referred to as an ostomate, and it is estimated that approximately 1 million individuals in the United States have undergone ostomies. To help patients deal with difficulties, ostomy organizations share information and collectively purchase supplies in bulk. Highly trained nurses in this field of gastroenterology teach patients about dealing with their colostomies and aid in the care of the stoma. Patients with colostomies are active, productive individuals and are generally unaffected by the colostomy, but may have life changes because of the disease that prompted this operation.

INDICATIONS

There are numerous indications for a colostomy. Diversion of the fecal stream may be necessary because of a distal bowel resection that has resulted in an anastomosis. A proximal colostomy to divert feces from the operative site may be required if a surgeon feel that a suture line in the colon is too fragile to receive intestinal contents. Often, colon obstruction and distension with fluid and gas is resolved by a colostomy proximal to the obstruction so that all the contents can be expelled, which allows the colon to decompress. At a later time, the patient undergoes an operation to allow more definitive repair of the obstructing site. A proximal colostomy is often formed when the colon perforates, either from diverticulitis, inflammatory bowel disease, cancer, or secondary to pelvic or abdominal trauma, to allow inflammation to subside; final resection and repair can be performed under more ideal operative

Copyright 2004, Elsevier (USA). All rights reserved.

conditions. Finally, when a patient has complete resection of the rectum, or on some occasions the most distal colon, a colostomy is performed because the sphincter mechanisms of the rectum have been removed and controlled evacuation through this distal portal is not possible. This situation usually occurs because the patient has cancer in the distal bowel and the entire area is resected.

TYPES OF COLOSTOMIES

There are three general types of colostomies: temporary, permanent, and reservoir. Temporary colostomies are the most common and are usually performed because of an acute need, such as colonic obstruction, perforation, and/or abdominal sepsis. This type of colostomy is created by making a small puncture incision in the abdominal wall and pulling a loop of bowel through the incision. The bowel is not transected or opened at this time. A plastic rod is passed through the mesentery and under the loop to keep the colon above the skin surface and the bowel is covered with a nonadhesive dressing. Over the next several days this loop of bowel undergoes an inflammatory reaction and becomes adherent to the abdominal wall. When this seal is formed, the colon is opened and the gas and fluid contents are evacuated. In several days, a colostomy bag is placed over the stoma, and with time the bowel contracts and resides just above the surface of the abdominal skin. This series of events may be quite frightening to the patient, who has just undergone a major emergency operation. However, the one positive point is that the ostomy is only temporary, and in 3–6 months, depending on the patient's condition, the colostomy will be closed and the patient will be able to evacuate in a normal manner.

The second type of colostomy is permanent. It is formed in patients with cancer of the most distal bowel or with a rectum that has been resected. Because the sphincter muscles have been removed, fecal countenance is not possible, requiring a colostomy. In this type of colostomy (an end colostomy), the end of the bowel is pulled through a small incision in the abdominal wall, the lumen is opened, and the outer and inner lining of the colon is everted and sewn to the skin. The bowel is also attached to the inner layers of the abdominal wall. When this procedure is finished, the colostomy looks much like a rosebud, with the luminal mucosa visible, and is generally at the level or only slightly above the level of the abdominal skin. Patients frequently leave the operating room with a bag placed over this colostomy. Alternatively, the colostomy may have a temporary clamp placed transversely across the distal colon and in several days the bowel seals with the abdominal wall. The clamp is then removed, the bowel is opened, and a bag is applied.

The third type of colostomy is rarely utilized, and was conceived by surgeons who wanted to create a sphincterlike mechanism for the distal colon. A pouch, or reservoir, is created with two loops of bowel and a nipple valve is fashioned in the distal colon, just at the level of the abdominal wall. The patient can then insert a tube through the valve and evacuate the contents of the reservoir. This bypasses the need to wear a bag. Although the initial concept was appropriate, the reservoir colostomy valves rarely worked correctly and many of the pouches had to be revised and the more standard type of colostomy formed. Few if any surgeons perform the procedure at this time.

COLOSTOMY FUNCTION

One of the main functions of the colon is to dehydrate the fecal contents by absorbing water and sodium. This occurs gradually along the length of the colon, so that stool in the right (proximal) side of the colon is more liquid than that in the left (distal) side of the colon. This is important, because colostomies may be created in the transverse (middle) colon or in the descending (left) colon. Stool from the midcolon may be more liquid than that produced from a more distal site, such as the left lower colon.

A colostomy does not have a sphincter mechanism, and hence peristaltic movements of the bowel, which are not under voluntary control, cause stool evacuation. Thus, a patient with a colostomy must wear a bag at all times to collect the extruded feces. Two schools of thought exist in terms of regulating a colostomy. One group of physicians, primarily found in Europe, believes that the colon should function on its own, and hence a small amount of stool is evacuated from the stoma three to seven times a day. The patient empties the colostomy bag when convenient or when indicated. In the United States, patients are usually instructed to irrigate their colostomy once a day. This is a process similar to having an enema; fluid is instilled via a tube inserted into the colostomy and this stimulates peristalsis, which results in evacuation. Following this procedure, there is usually no additional evacuation from the stoma for the next 24 hours.

Patients with colostomies wear a bag, and the proper attachment of the bag to the abdominal skin is a process usually learned from an ostomy (stoma) nurse. The more chronic attachment devices use a soft rubber doughnut-shaped disk or wafer that fits closely around the stoma and is fixed to the skin. A flange is attached to

the disk or wafer and the flange accepts a plastic ring that is part of a specially constructed ostomy bag. This generally provides a watertight seal. The bag can thus be removed without disturbing the apparatus attachment to the skin. Only occasionally is the disk changed, which protects the skin surrounding the stoma.

A temporary approach uses a light plastic bag that has adhesive fixed to a portion of the outer surface around an appropriately sized hole.. This allows direct application of the bag to the skin around the stoma. The attached bag must be removed from the underlying skin each time the bag is changed, which causes undesirable irritation to the skin. Hence, this approach is only a temporary short-term solution to ostomy collection.

The ostomate should generally be able to live a near-normal life. Bathing and showering will not harm the stoma. Physical activity should not be limited and diet should be tailored to food preferences. Sexual activity is not prohibited. Most health insurance policies, including Medicare and Medicaid, cover the cost of equipment and supplies.

COMPLICATIONS

Early complications include wound infections, intestinal obstruction, prolapse of the bowel through the incision in the abdominal wall or through the stoma, bleeding from the exposed bowel, and excoriation of the skin around the stoma.

Long-term complications are similar to those that occur in the short term but, most often relate to bag attachment problems. Bag leakage and spillage result in soilage of clothing, creating social embarrassments. If the disk around the ostomy is not well fitted, skin excoriation can occur, and this is an annoying and painful problem. Ostomy nurses are specialists in dealing with these problems and have a variety of solutions for these mechanical complications. The colon ferments fiber and produces gas, which is often entrapped in the bag and on occasion causes odors. Dietary modification and some pharmacological approaches are often useful to solve these problems. Adequate hydration is necessary to ensure that a semisoft stool is present and can be evacuated without difficulty. Injury to the exposed mucosa occasionally occurs, resulting in bleeding. This is usually minor and self-limiting but may necessitate surgical care if the bleeding is prolonged. Stenosis and herniation of the colostomy are additional complications that require consultation with a physician.

COLOSTOMY CLOSURE

Individuals with temporary colostomies will be admitted to the hospital 3–6 months later for colostomy closure. Although this may seem like a minor procedure when compared to the colon resection, this operation involves opening the abdomen, resecting a small portion of bowel, and connecting the portions of colon distal and proximal to the colostomy. Some time in the hospital may be required and complications may occur. Wound infections are frequent, and ileus and poor food tolerance are initially the most common complications observed. Rehabilitation associated with normal bowel action usually progresses without incident.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Diverticulosis

Further Reading

Celestin, L. R. (1987). "Color Atlas of the Surgery and Management of Intestinal Stomas." Year Book Medical Publ., Chicago.

Rosenthal, M. J., and Rosenthal, D. M. (2002). Stomal care. *In* "ACS Surgery: Principles and Practice" (D. W. Wilmore, ed.), pp. 1675–1690. WebMD, New York.

United Ostomy Association, 19772 MacArthur Boulevard, Suite 200, Irvine CA 92612-2405 (phone, 1-800-826-0826; website, www.uoa.org).

Complementary and Alternative Medicine

JUDITH K. SHABERT
Harvard Medical School

holistic Describes therapies based on information about the "whole person," including spiritual, physical, mental, emotional, environmental, and social health.

placebo An inactive substance given to a participant in a research study as part of a test of the effects of another substance or treatment.

Over the past decade, there has been a growing interest in complementary and alternative medicine (CAM), a field that encompasses a wide variety of approaches ranging from herbal medicines to massage and from macrobiotics to magnet therapy. The therapies that are proven safe and effective gradually become incorporated into mainstream practices and thus the list of practices is continually changing. The National Center for Complementary and Alternative Medicine, a component of the National Institutes of Health, defines CAM as "those healthcare and medical practices that are not currently an integral part of conventional medicine." Using these criteria, CAM practices can be grouped in five major domains: alternative medical systems; mind–body interventions; biologically based treatments; manipulative and body-based methods; and energy therapies. Common to most of these practices is the concept of vitalism or the transfer of energy to the individual to support the healing process. This point will be emphasized in the following descriptions of this wide range of practices.

MAJOR DOMAINS OF CAM

Alternative Medical Systems

Alternative medical systems involve complete systems of theory and practice that have evolved independent of and often prior to the conventional biomedical approaches. These systems include shamanism and traditional healing in cultures throughout the world. Some systems originated 5000–6000 years ago includingAyurveda from India and traditional Oriental medicine originating in China and practiced throughout Asia.

Traditional Chinese Medicine

In Asia, the full complement of traditional Chinese medicine (TCM) treatment incorporates diet, acupuncture, and Chinese herbal preparations for each individual patient for optimal results. In Western cultures, acupuncture is often given in isolation from diet and herbal preparations.

Acupuncture is based on the concept of vital energy channels or qi (pronounced chee), which flows through the body and is essential for health. Vital energy channels are known as meridians; disruption or imbalance of qi results in illness. Two other concepts are central to TCM: Yin–Yang describes the polarity of nature, and the concept of Five Elements (earth, water, air, fire, and metal) describes the interrelatedness of the body's organs and energy flow.

There are 12 meridians in the body that flow close to the skin's surface with more than 450 acupuncture sites. Acupuncture sites are stimulated, depending on the specific ailment, by needles that penetrate the skin. The duration and frequency of the treatments are determined by the practitioner and by the response of the patient to treatment. Recommended conditions appropriate for acupuncture are found in Table I.

Ayurveda

Ayurveda is a 6000-year-old system that utilizes energetic forces (doshas) to guide healing and maintain health. This system believes in a strong connection between the mind and the body and utilizes the Laws of Nature to create health and maintain balance of the body, senses, mind, and soul. Ayurveda, like TCM, views a person as being composed of five primary elements including air, fire, water, and earth. Ether is the fifth element in Ayurveda versus metal in TCM.

The Ayurveda medical system incorporates three energetic forces or Tridoshas. Vata dosha is the most important of the three doshas and drives the other two doshas, Pitta dosha and Kapha dosha. Vata dosha controls body movement (mental and physical), eliminates waste, transmits sensory input to the brain, and assists in aspects of metabolism. When Vata dosha is the dominant dosha, its imbalance is powerful enough to upset the Pitta and Kapha doshas.

Pitta dosha represents fire or heat and is located at sites of transformation such as digestion and

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Conditions Appropriate for Acupuncture Therapy

Digestive	Emotional	Eye–Ear–Nose–Throat	Gynecological
Abdominal panic	Anxiety	Cataracts	Infertility
Constipation	Depression	Gingivitis	Menopausal symptoms
Diarrhea	Insomnia	Poor vision	Premenstrual syndrome
Hyperacidity	Nervousness	Tinnitis	
Indigestion	Neurosis	Toothache	
Miscellaneous	**Musculoskeletal**	**Neurological**	**Respiratory**
Addiction control	Arthritis	Headaches	Asthma
Athletic performance	Back pain	Migraines	Bronchitis
Blood pressure regulation	Muscle cramping	Neurogenic bladder dysfunction	Common cold
Chronic fatigue	Muscle pain/weakness	Parkinson's disease	Sinusitis
Immune system tonification	Neck pain	Postoperative pain	Smoking cessation
Stress reduction	Sciatica	Stroke	Tonsillitis

Source: World Health Organization (1979). "Viewpoint on Acupuncture." World Health Organization, Geneva, Switzerland.

metabolism, appetite, vision, regulation of body temperature, skin, comprehension, and bravery. Kappa dosha is the heaviest of the doshas and it provides structure and lubrication to the body. It is associated with bodily strength, fertility, and stability of mind and body.

An Ayurvedic practitioner assesses a person's body type and determines their dosha by feeling the pulse. When it is determined that one of the doshas is underactive or, more significantly, overactive, a diet that is specific to one of the doshas is prescribed to rebalance the doshas and remedy the disease. Herbs and aromatherapy are also utilized.

Homeopathy

Homeopathy is a medical system based on the concept of "Laws of Similars" and it incorporates tinctures produced from natural sources such as plants, animals, and minerals. Dr. Samuel Hahnemann developed homeopathy in the early 1800s.

Dr. Hahnemann discovered that when the bark of the cinchona tree, which was used to treat swamp fever (malaria), was given to healthy individuals, symptoms of swamp fever were elicited. This experiment led Hahnemann to test thousands of substances on healthy individuals. When a specific symptom occurred in a healthy person given a certain substance, this substance was then used to treat people who had this specific symptom.

The process of preparing homeopathic remedies incorporates potentization and succussion. The substance being prepared as a remedy is successively diluted with water (potentization) and shaken (succussion). When the final product is produced, it is unlikely that even one molecule of the original substance remains and that is the goal of homeopathic preparations. This phenomenon is called the Principle of Infinitesimal Dose.

Drops of a preparation are usually given every 3–4 h for each symptom of an illness. For example, if a person has insomnia due to fright, one type of preparation is given. If insomnia is due to exhaustion, another preparation is suggested.

Naturopathic Medicine

Naturopathic medicine is a system of medical care based on a number of different modalities. A basic tenet of naturopathy is the belief that the body has an innate ability to heal itself utilizing treatments that include clinical nutrition, homeopathy, herbal medicine, light therapy, therapeutic counseling, TCM, mechanical manipulation, and hydrotherapy. The interdependent parts of an individual including mind, body, spirit, and emotion are considered in formulating a treatment plan and in providing care to an individual.

Naturopathic medicine is not considered a solitary system of care and naturopaths work with other practitioners, including allopathic caregivers, to provide optimal care for the patient. Although naturopaths do treat patients, their emphasis is on patient education and preventive care.

Mind–Body Interventions

Herbert Benson, a cardiologist at Harvard Medical School, and John Kabat-Zinn, from the University of Massachusetts, were the first to research and prove the benefit of mind–body medicine. Relaxation exercises, visual imaging, and particularly biofeedback are the basis of this approach that is used for the treatment of many chronic health conditions. Most mind–body interventions developed or proven effective by Benson and Kabat-Zinn are now an integral part of mainstream medicine. Other forms of mind–body interventions

include art, dance, and music therapy, intuitive healing, prayer, and meditation.

Biofeedback

Biofeedback is a technique that trains the individual to alter brain activity and therefore modify blood pressure, muscle tension, heart rate, and other bodily functions that are ordinarily under autonomic (nonvoluntary) control. The training initially utilizes a device that triggers a flashing light or activates a beeper when an appropriate response is elicited. Thus, one can think about increasing blood flow to the hands, and a small temperature probe attached to a finger will indicate when this is accomplished. Such signals "teach" people to improve their health and performance by controlling bodily functions. Other signal systems respond to changes in blood pressure, heart rate, muscle tension, sweat activity, and activity of the brain waves.

Some of the conditions that are treated by this technique include migraine headaches, chronic pain, high blood pressure, cardiac arrhythmias, Raynaud's phenomena, and some seizure disorders.

Transcendental Meditation

Many forms of meditation are used to promote healing. In North America, the most institutionalized form is transcendental meditation (TM), brought to the Western world by Maharishi Mahesh Yogi. Followers of TM meditate twice a day on a mantra provided to them by a TM instructor following two weekends of instruction on its techniques.

Art, Dance, and Sound Therapy

All of these therapies have a psychological directive and art and dance therapies are also useful for individuals with physical disabilities. It is believed that the expression of self though body movement, the creation of art work, and listening to sound and music are therapeutic for improving self-esteem, recovery from abuse and trauma, and enhancement of social integration, to name a few benefits.

Biological-Based Therapies

This grouping of CAM practices is based on interventions using natural or biological substances. Examples include the practice of orthomolecular medicine, the provision of nutritional supplements (vitamins, minerals, herbs, hormones, and other products), special diets (macrobiotic, Atkins, The Zone), and chelation therapy.

Orthomolecular Medicine

Linus Pauling first used the term "orthomolecular" in an article in *Science* in the 1960s. The term describes the practice of preventing and treating disease by providing natural nutrients to the body in optimal amounts specific for each person. It is based on the concept of genetic individuality. Just as each person has a different skin color that determines tolerance to the sun, each person has unique biochemical needs based upon their genetic make-up that determines nutrient requirements. Orthomolecular medicine supports the use of megadose vitamins for physical and mental conditions (for example, the use of large doses of vitamin C to treat the common cold).

Herbal Medicine

Herbs are the oldest form of medical therapy and they are used throughout the world to treat a variety of conditions. In Europe, herbs are prescribed by allopathic physicians as frequently as are pharmaceutical drugs. Herbs can be provided as teas, tinctures, and capsules both internally and externally. Extensive research is being conducted on the effectiveness of herbs to treat a variety of conditions. A list of commonly used herbs is presented in Table II.

Macrobiotics

The use of the macrobiotic diet and the philosophy surrounding it were developed and promoted by Michio Kushi.

The macrobiotic diet is based loosely on the traditional Japanese diet and it also supports a healthy lifestyle. Brown rice, vegetable soup, vegetables, beans, and sea vegetables are the staples of the diet. Occasionally, fruit, nuts, and seeds are allowed. The diet eliminates most animal products but fish is occasionally permitted. The most restrictive form of the diet can result in vitamin B12 deficiency.

Chelation Therapy

The intravenous infusion of a chelation substance was initiated in the 1940s to treat individuals with heavy metal exposure by chelating the metal and facilitating its elimination from the body.

Today ethylenediaminetetraacetic acid (EDTA) chelation therapy is being utilized by some CAM physicians to treat patients with atherosclerotic and peripheral vascular disease.

Proponents believe that chelation therapy binds toxic metals such as lead, mercury, cadmium, copper, calcium, iron, arsenic, nickel, and aluminum. By chelating calcium from cholesterol-laden plaques in

TABLE II Commonly Used Herbs

Name	Use	Pharmacologal Effects	Potential side effects
Echinacea, purple cone flower root	Treatment of infectious diseases, particularly those of the respiratory tract	Immunostimulation, immunosuppression	Allergic reactions, immunosuppression with long-term use
Ephedra, ma huang	Promote weight loss, increase energy, treat asthma	Contains alkaloids including ephedrine; increases heart rate and blood pressure	Myocardial infarction, stroke, interaction with monoamine, oxidase inhibitors
Garlic, ajo	Decrease risk of atherosclerosis, decrease serum lipids	Inhibit platelet aggregation	Potential risk of bleeding
Ginkgo, duck foot tree	Use for cognitive disorders and peripheral vascular disease	Inhibits platelet-activating factor	Potential risk of bleeding
Ginseng	Protects body against stress	Lowers blood glucose, inhibits platelet aggregation	Hypoglycemia, increased risk of bleeding, interaction with warafin
Kava, awa	Reduces anxiety; used as a sedative	Sedation	Potential to increase sedative effects of other drugs
St. John's wort, amber, goat weed	Treatment of mild to moderate depression	Inhibit neurotransmitter reuptake	Affects a variety of drugs by altering liver metabolism
Valerian, all-heal	Treatment of insomnia	Sedation	Increases potency of other sedative drugs

vessels throughout the body, atherosclerotic tissue damage is resolved and vessel integrity is improved. Usually a patient is treated one to three times per week over several months. The intravenous solution containing EDTA is also supplemented with vitamins and minerals that replace nutrients eliminated through the urine during therapy and megadoses of antioxidant nutrients are also provided.

The Food and Drug Administration and numerous medical groups such as the American Medical Association are opposed to this therapy as a means of treating atherosclerosis because the double-blind studies testing chelation therapy have not shown benefit.

Light Therapy

Some individuals experience depression only in the winter months when the hours of sunlight are reduced. This is referred to as seasonal affective disorder. Light therapy utilizes lamps with the full spectrum of ultraviolet light. Exposure to these lamps for 1−2 h/day is effective is resolving depression is some of these individuals.

Hydrotherapy

Hydrotherapy is a general term that refers to treatments that use water for the relief of symptoms arising from various diseases or injuries. The therapy may vary by the temperature of the water (hot versus cold) or it may use water under pressure in the form of jets, whirlpools, or aerated bubbles. This form of therapy is often used for complaints related to the musculoskeletal system or the joints.

Internal hydrotherapy includes colonic irrigations and enemas to eliminate toxins and steam baths or steam inhalation to relieve respiratory congestion.

Environmental Medicine

Practitioners of environmental medicine support a holistic approach to the patient and believe that the thousands of artificial chemicals in the environment are related to ill health and allergies. Chronic exposure to environmental toxics leads to chronic health conditions that affect not only physical health but mental and emotional health as well. Treatment approaches in environmental medicine focus on decreasing exposure to environmental toxins and allergies and detoxification programs that may include chelation therapy and megavitamin therapy.

Manipulative and Body-Based Methods

All forms of massage and body manipulation performed by chiropractors and osteopaths are included in this grouping.

Massage Therapy

Massage therapy has been practiced for at least 4000 years. The goal of massage is to manipulate the soft tissues of the body to normalize those tissues. It is estimated that 7% of the population has at least one massage annually. Massage is being used to provide

relief from injury, stress, pain, and other acute and chronic conditions. The types of massage therapy are too numerous to list but include Swedish, deep tissue, hot stone, neonatal, and sports massage.

Oriental massage techniques such as Shiatsu, Amma, acupressure, and Tibetan point holding are based on the concept of qi and focus on the 12 meridians of the body.

Osteopathic Medicine

Osteopathic medicine is a system of health care based on the premise that disease is the result of structural and anatomical abnormalities. Treatment aims to rebalance the interrelationship of structure and function. Osteopaths undergo an educational program similar to that for medical doctors including residency training in all of the major medical specialties including surgery.

Osteopathic physicians differ from conventionally trained physicians because, in addition to the use of all allopathic practices, they use manipulative techniques to diagnose and treat illness. An extensive body of work supports the use of osteopathic techniques for musculoskeletal and nonmusculoskeletal problems.

Chiropractic

Chiropractic care is based on spinal manipulation and includes adjustment and manipulation of the articulations and adjacent tissues of the body. The practice is drug- and surgery-free. There are more than 50,000 chiropractic practitioners in the United States. In addition to spinal manipulation, many practitioners provide advice on and dispense herbs, books and videos, healthcare products, and nutritional supplements.

The National Guideline Clearinghouse provides information on chiropractic quality assurance and practice parameters.

Energy Therapies

Energy therapies are concerned with energy originating from within the body or energy applied to the body from outside sources such as magnet therapy. Many practices and systems of health care discussed previously incorporate the concept of energy therapy, such as acupuncture, Ayurveda, Shiatsu, and qi gong.

Therapeutic Touch

This method claims to assist the natural healing process by redirecting and rebalancing the energy fields of the body. The practitioner places his or her hands on or over certain parts of the body to redistribute the patient's energy or transfer his or her own energy to the patient.

This therapy is said to reduce pain and anxiety, promote relaxation, and stimulate the body's natural healing process.

VALIDATING THE EFFECTIVENESS OF CAM

Much of both allopathic and alternative medicine is based on anecdotal evidence or experiences; that is, a particular approach appeared to work for one patient so the practitioner institutes the same or a similar therapy on the next patient. Providers and payers are now asking for the "evidence" that therapies are effective, not only for allopathic treatments but also for CAM approaches. Within the past 15 years, the field of "evidence-based" medicine has grown starting with a group of investigators at Oxford University. Levels of evidence are used to evaluate information concerning various therapies. In the opinion of the Oxford group, the strongest evidence for a particular therapy is based on prospective, randomized, blinded, clinical trials and the weakest recommendations are based on case series or expert opinion without explicit clinical appraisal.

Some CAM treatments, particularly herbal therapies and nutritional supplements, can be evaluated by using conventional clinical trial techniques like those that are used to test new drugs. Difficulty arises in designing an appropriate study when a specific procedure involves physical manipulation such as with acupuncture, therapeutic touch, and chiropractic adjustments.

Another approach to evaluating the effectiveness of CAM is to gather a group of experts and review the evidence published on a specific topic. The experts then offer a final opinion or consensus on a topic and this information is made available in periodicals or on the Internet. An individual seeking allopathic or CAM treatment should be made aware of the available evidence concerning a particular therapy before embarking on a course of treatment.

Other research on CAM treatments is being supported by the National Institutes of Health and/or other funding sources such as industry.

RISKS ASSOCIATED WITH CAM

In addition to understanding the benefits of a particular therapy, an individual should appreciate the risks involved. This is particularly true of side effects associated with the ingestion of herbs or supplements (Table III). Because patients rarely inform their allopathic providers

TABLE III Some of the CAM Substances Associated with Specific Side Effects

Product	Common name	Toxin(s)	Side effects
Aconite	Monkshood, wolfbane	Aconite, hypaconitine	Nausea, vomiting, neuromuscular weakness, seizures, coma, cardiac effects (bradycardia, arrhythmia, and hypotension)
Chaparral tea			Liver failure
Chomper			May be contaminated with digitalis
Comfrey	Bruisewort, knitbone	Pyrrolizidine	Veno-occlusive disease, vomiting, hepatomegaly, hepatic necrosis
Jin bu huan			Sedative, hepatitis, bradycardia
Lobella			Vomiting, coma, tachycardia, respiratory distress
Ma huang	Ephedra	Ephedrine, pseudoephedrine	Hypertension, heart attack stroke, arrhythmias, headaches, seizures, tremors, anxiety, hallucinations
Pennyroyal		Pulegone	Liver and renal failure, nausea, vomiting, abdominal pain, shock, alterations in mental status (delirium, confusion, agitation, seizures)
Plantain leaf			May be contaminated with digitalis
Yohimbe			Seizures, renal failure

Adapted from Cassileth, B. R. (1998). "The Alternative Medicine Handbook: The Complete Reference Guide to Alternative and Complementary Therapies." W. W. Norton & Co., New York, and Mack, R. B. (1998). "Something wicked this way comes"—Herbs even witches should avoid. *Contemp. Pediatr.* **15**(6), 49–64, with permission.

of their use of CAM therapies—or alternatively allopathic providers rarely ask a patient or are knowledgeable about such therapies—there is concern that deleterious herb–supplement–drug interactions may occur. This concern has prompted anesthesiologists to advise patients to discontinue all herbal medicines for 1–7 days before an operation in order to prevent relevant side effects from interactions of herbal products and anesthetic agents.

There is also concern that a patient with a specific disease will avoid a conventional therapy that could potentially cure the disease and choose a CAM treatment that may have variable or negligible effects on the disease. High-risk patients, such as those with cancer or autoimmune deficiency syndrome, may be more likely to pursue such a course. A variety of information databases are available to help individuals in this decision-making process.

TRENDS IN THE USE OF CAM

Over the past decade, the use of CAM has grown rapidly in the United States, Japan, Australia, and Europe. In 1990, approximately 33% of Americans used an alternative therapy and that percentage increased to approximately 42% by 1997. Individuals spent more than $27 billion on these therapies and such expenditures were in the most part not covered by insurance and were paid for out-of-pocket.

Allopathic physicians are increasingly participating in CAM therapies, although the inclusion of CAM instruction as required curriculum in medical schools is quite variable. One survey reported that 60% of physicians from a wide range of specialties recommended alternative therapy to their patients at least once and almost one-half of the physicians surveyed had used alternative therapies themselves.

A variety of reasons are offered for the recent success of CAM. Some suggest that the popularity reflects a failing in our present medical system and relates to problems in care delivery (the growth of health maintenance organizations and the 10-min patient visit). Others suggest that this trend reflects an issue related to patient control and choice, which usually are not associated with present methods of allopathic care. Finally, most of the CAM approaches are holistic and are seen by the patient as being more "natural" than allopathic treatments. Whatever the reasons, it is quite clear that CAM will play a greater role in medical care in the future.

See Also the Following Article

Functional (Non-Ulcer) Dyspepsia

Further Reading

About chelation. Available at http://holisticonline.com/chelation. Accessed October 22, 2001.

About Shiatsu. Available at http://holisticonline.com/shiatsu/hol. Accessed October 22, 2001.
American Chiropractic Association. Available at http://www.amerchiro.org. Accessed October 22, 2001.
American Dance Therapy Association. Available at http://www.adta.org. Accessed October 22, 2001.
American Massage Therapy Association. Available at http://www.amtamassage.org. Accessed October 22, 2001.
American Music Therapy Association. Available at http://www.musictherapy.org. Accessed October 22, 2001.
American Osteopathic Association. Available at http://www. aoanet.org. Accessed October 22, 2001.
Astin, J. A. (1998). Why patients use alternative medicine: Results of a national study. *J. Am. Med. Assoc.* **279**, 1548–1553.
Borkan, J., Neher, J. O., Anson, O., and Smoker, B. (1994). Referrals for alternative therapies. *J. Family Practice* **39**(6), 545–550.
Eisenberg, D. M., Davis, R. B., Ettner, S. L., Appel, S., Wilkey, S., Van Rompay, M., and Kessler, R. C. (1998). Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. *J. Am. Med. Assoc.* **280**, 1569–1575.
Guidelines for chiropractic assurance and practice parameters. Available at http://www.guideline.gov/FRAMESETS/guideline_fs.asp?guideline=1048&sSearch_string=.
Longwood Herbal Task Force, authoritative site on herbs. Available at http://www.mcp.edu/herbal. Accessed October 22, 2001.
Mind–Body Medical Institute. Available at www.stress.org/benson or www.mbmi.org. Accessed October 22, 2001.
National Center for Complementary and Alternative Medicine. Available at http://nccam.nih.gov. Accessed October 22, 2001.
Reiki Association. Available at http://www.reikiassociation.org. Accessed October 22, 2001.
Rolf Institute. Available at http://www.rolf.org. Accessed October 22, 2001.
Sifton, D. W. (ed.) (1999). "The PDR Family Guide to Natural Medicines and Healing Therapies." Three Rivers Press, New York.
Wetzel, M. S., Eisenberg, D. M., and Kaptchuk, T. J. (1998). Courses involving complementary and alternative medicine at U. S. medical schools. *J. Am. Med. Assoc.* **280**(9), 784–787.

Computed Tomography (CT)

RICHARD M. GORE, VAHID YAGHMAI, GERALDINE M. NEWMARK, JONATHAN W. BERLIN, AND FRANK H. MILLER

Northwestern University and Evanston Northwestern Healthcare

computed tomography (CT)-colonography Multidetector CT acquires a volume data set of the abdomen. With computer software, a virtual, volume-rendered display of the colon is created, permitting detection of polyps and cancer.

contrast-enhanced computed tomography images obtained with intravascular and intralumenal contrast material. This is the preferred scanning protocol for most intra-abdominal pathology.

helical computed tomography (CT) CT technology that acquires a volume of data rather than acquiring data slice by slice as in older technology.

intralumenal contrast A dilute solution (2%) of barium or gatrografin is ingested to distend and opacify the hollow abdominal viscera. This is helpful in differentiating the gut from masses, lymph nodes, and abscesses.

intravascular contrast Iodinated contrast material is injected into an antecubital fossa vein to assess bowel wall enhancement, to identify and characterize focal lesions in the liver, spleen, pancreas, adrenal glands, kidneys, and bladder, and to perform computed tomography–angiography. Usually, 150 cc of contrast is injected at a rate of 2–5 cc/s.

multidetector computed tomography (CT) An advancement in helical CT in which multiple detectors, rather than a single detector, are employed to acquire the data. This technology allows for extremely rapid scans, optimizing intravenous contrast levels.

mural stratification Computed tomography visualization of all three layers of the gut: mucosa, submucosa, and muscularis propria. Most benign causes of gastrointestinal pathology result in mural thickening of the gut with preservation of mural stratification. Malignant disorders generally cause wall thickening with loss of mural stratification.

noncontrast computed tomography (CT) scans Images obtained on CT without intravascular or intralumenal contrast material. It is a good technique for evaluating kidney and ureter stones but limits the sensitivity in detecting gastrointestinal pathology.

Copyright 2004, Elsevier (USA). All rights reserved.

About Shiatsu. Available at http://holisticonline.com/shiatsu/hol. Accessed October 22, 2001.
American Chiropractic Association. Available at http://www.amerchiro.org. Accessed October 22, 2001.
American Dance Therapy Association. Available at http://www.adta.org. Accessed October 22, 2001.
American Massage Therapy Association. Available at http://www.amtamassage.org. Accessed October 22, 2001.
American Music Therapy Association. Available at http://www.musictherapy.org. Accessed October 22, 2001.
American Osteopathic Association. Available at http://www. aoanet.org. Accessed October 22, 2001.
Astin, J. A. (1998). Why patients use alternative medicine: Results of a national study. *J. Am. Med. Assoc.* **279**, 1548–1553.
Borkan, J., Neher, J. O., Anson, O., and Smoker, B. (1994). Referrals for alternative therapies. *J. Family Practice* **39**(6), 545–550.
Eisenberg, D. M., Davis, R. B., Ettner, S. L., Appel, S., Wilkey, S., Van Rompay, M., and Kessler, R. C. (1998). Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. *J. Am. Med. Assoc.* **280**, 1569–1575.
Guidelines for chiropractic assurance and practice parameters. Available at http://www.guideline.gov/FRAMESETS/guideline_fs.asp?guideline=1048&sSearch_string=.
Longwood Herbal Task Force, authoritative site on herbs. Available at http://www.mcp.edu/herbal. Accessed October 22, 2001.
Mind–Body Medical Institute. Available at www.stress.org/benson or www.mbmi.org. Accessed October 22, 2001.
National Center for Complementary and Alternative Medicine. Available at http://nccam.nih.gov. Accessed October 22, 2001.
Reiki Association. Available at http://www.reikiassociation.org. Accessed October 22, 2001.
Rolf Institute. Available at http://www.rolf.org. Accessed October 22, 2001.
Sifton, D. W. (ed.) (1999). "The PDR Family Guide to Natural Medicines and Healing Therapies." Three Rivers Press, New York.
Wetzel, M. S., Eisenberg, D. M., and Kaptchuk, T. J. (1998). Courses involving complementary and alternative medicine at U. S. medical schools. *J. Am. Med. Assoc.* **280**(9), 784–787.

Computed Tomography (CT)

RICHARD M. GORE, VAHID YAGHMAI, GERALDINE M. NEWMARK, JONATHAN W. BERLIN, AND FRANK H. MILLER

Northwestern University and Evanston Northwestern Healthcare

computed tomography (CT)-colonography Multidetector CT acquires a volume data set of the abdomen. With computer software, a virtual, volume-rendered display of the colon is created, permitting detection of polyps and cancer.

contrast-enhanced computed tomography images obtained with intravascular and intralumenal contrast material. This is the preferred scanning protocol for most intra-abdominal pathology.

helical computed tomography (CT) CT technology that acquires a volume of data rather than acquiring data slice by slice as in older technology.

intralumenal contrast A dilute solution (2%) of barium or gatrografin is ingested to distend and opacify the hollow abdominal viscera. This is helpful in differentiating the gut from masses, lymph nodes, and abscesses.

intravascular contrast Iodinated contrast material is injected into an antecubital fossa vein to assess bowel wall enhancement, to identify and characterize focal lesions in the liver, spleen, pancreas, adrenal glands, kidneys, and bladder, and to perform computed tomography–angiography. Usually, 150 cc of contrast is injected at a rate of 2–5 cc/s.

multidetector computed tomography (CT) An advancement in helical CT in which multiple detectors, rather than a single detector, are employed to acquire the data. This technology allows for extremely rapid scans, optimizing intravenous contrast levels.

mural stratification Computed tomography visualization of all three layers of the gut: mucosa, submucosa, and muscularis propria. Most benign causes of gastrointestinal pathology result in mural thickening of the gut with preservation of mural stratification. Malignant disorders generally cause wall thickening with loss of mural stratification.

noncontrast computed tomography (CT) scans Images obtained on CT without intravascular or intralumenal contrast material. It is a good technique for evaluating kidney and ureter stones but limits the sensitivity in detecting gastrointestinal pathology.

Copyright 2004, Elsevier (USA). All rights reserved.

The noninvasive evaluation of gastrointestinal disorders has been revolutionized over the past 20 years by technical advances in cross-sectional imaging techniques: ultrasound, computed tomography (CT), and magnetic resonance imaging. CT has evolved as the premier imaging examination in most clinical situations. It has earned this role by virtue of its ability to provide a global perspective of the gut, solid abdominal and pelvic organs, mesenteries, omenta, peritoneum, retroperitoneum, extraperitoneum, and subperitoneum, uninhibited by the presence of bowel gas, fat, or surgical wounds. The advent of multidetector CT has further advanced the clinical utility of CT by permitting thinner contiguous images to be obtained without respiratory misregistration and allowing multiple data acquisitions to be made during different phases of a single intravenous contrast bolus injection. This advance has also shortened examination times and the acquired data set can be displayed as angiographic renderings, multiplanar reformatted images, and surface- or volume-rendered images of the abdominal viscera or lumen without exposing the patient to additional radiation. In this article, the CT features of the most common gastrointestinal disorders are discussed and the utility of CT is placed in a clinical perspective.

THE SOLID ABDOMINAL ORGANS

Liver

Diffuse Disease

Cirrhosis The computed tomography (CT) findings in cirrhosis are varied and depend on the cause and duration of the hepatic injury. Typically, the liver has an irregular nodular appearance with inhomogenous contrast enhancement. There is widening of the fissures and gallbladder fossa, a notch in the posterior aspect of the right hepatic lobe, atrophy of the right lobe with hypertrophy of the caudate lobe (Fig. 1) and lateral segment of the left lobe. In the later stages of cirrhosis, the overall liver volume is diminished. Whereas all patients pathologically demonstrate regenerating nodules, these are infrequently demonstrated on CT. Siderotic nodules have higher attenuation than liver and other soft tissue organs and are readily apparent on unenhanced CT. Because of the insensitivity of CT to depicting regenerating nodules, it cannot reliably identify the transformation process of regenerative nodules to dysplastic nodules.

The gallbladder, hepatic flexure of the colon, and greater omentum often become more superficial in location due to atrophy of the right hepatic lobe and thus are at risk for injury during liver biopsy. Patients with immune-mediated liver disease, such as chronic active hepatitis and sclerosing cholangitis, often have

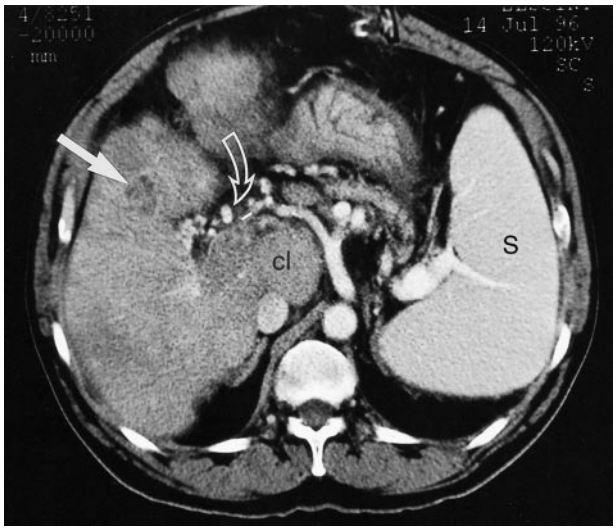


FIGURE 1 Cirrhosis of the liver with a hypovascular hepatoma (straight arrow). The liver has a nodular contour with enlargement of the caudate lobe (cl) and there is cavernous transformation of the main portal vein (curved arrow). Note the splenomegaly (s).

enlarged lymph nodes in the porta hepatis. CT is also helpful in showing signs of portal hypertension: varices, ascites, splenomegaly, and increased density of mesenteric and omental fat.

Steatosis There is an excellent correlation between hepatic parenchymal CT attenuation and the amount of hepatic triglycerides found on liver biopsy specimens. Intracellular fat lowers the CT attenuation of the liver, so that the liver is less dense than the spleen (Fig. 2). Normally the CT density of the liver is equal to or greater than that of the spleen.

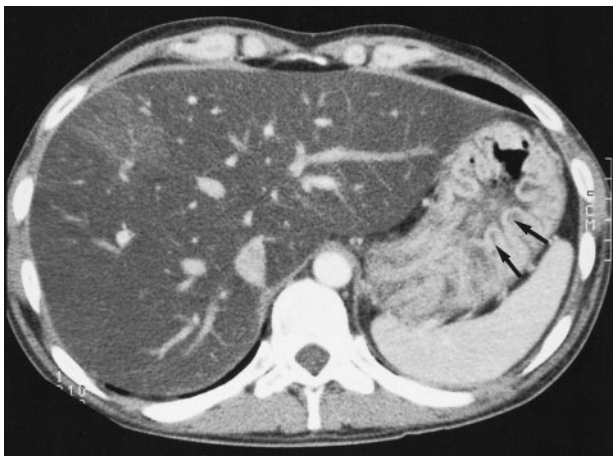


FIGURE 2 Diffuse hepatic steatosis in a patient with alcoholic gastritis. The density of the liver is diffuse low, contrasting with the hepatic vessels. The thickened rugal folds of the stomach (arrows) attest to the gastritis.

When fat deposition is focal, it can be confused with a hepatic mass. Focal fat, however, does not displace adjacent blood vessels or cause focal contour abnormalities and may change its appearance rapidly. In questionable cases, the diagnosis of focal fat can be confirmed by magnetic resonance imaging (MRI).

Iron overload Hemosiderosis and hemochromatosis increase the density of the liver on CT. Indeed, there is a linear relationship between CT attenuation values and hepatic iron content. CT can therefore establish the diagnosis, quantitate the amount of iron in the liver, and follow the efficacy of phlebotomy therapy. High hepatic parenchymal attenuation can also be seen in patients receiving amiodarone or gold therapy, glycogen storage disease, calcification from shock liver, and previous exposure to thorotrast.

Hepatitis The primary role of CT in patients with hepatitis is to exclude focal masses or a hepatocellular carcinoma. CT findings in hepatitis are usually nonspecific. Hepatomegaly, gallbladder wall thickening, and hepatic periportal lucency are the major CT findings in patients with acute viral hepatitis. This lucency is due to fluid and lymphedema surrounding the portal veins and has also been reported in patients with acquired immunodeficiency syndrome (AIDS), trauma, neoplasm, liver transplants, liver transplant rejection, and congestive heart failure. Early in the disease, multiple regenerating nodules may be seen in an atrophied liver. CT has shown that hepatic necrosis is repaired by hypertrophy of regenerating nodules rather than by an increase in the number of nodules. These nodules give rise to the macronodular pattern of postnecrotic cirrhosis. Rarely, the nodules may be hypodense and simulate metastases on CT.

In patients with chronic active hepatitis, lymphadenopathy can be found in the porta hepatis, gastrohepatic ligament, and retroperitoneum in 65% of cases and may be the only CT abnormality in a patient with severe hepatic dysfunction. CT can also follow the course of immunosuppression by observing the reduction of lymph node size with therapy.

Focal Hepatic Lesions

Malignant neoplasms

Metastases Liver metastases are the most common malignant tumor of the liver and CT is preferred over MRI as the initial screening test because of its shorter examination time, greater ability to detect extrahepatic disease, lower cost, and greater availability. On noncontrast scans, metastases most commonly appear hypodense relative to normal liver. Metastases can occasionally appear hyperdense due to hemorrhage, calcification, or fatty infiltration of the liver.

The appearance of metastatic lesions following the intravenous administration of contrast is dependent on the intrinsic vascularity of the lesion (hypovascular versus hypervascular) and the timing of scanning following contrast administration. During the hepatic arterial phase of contrast enhancement, hypervascular metastases either appear uniformly hyperdense or contain a hyperdense ring relative to the surrounding liver. Hypovascular lesions appear hypodense when compared to adjacent hepatic parenchyma during the arterial phase. During the portal venous phase of contrast enhancement, most metastatic lesions will appear hypodense compared to normal hepatic parenchyma (Fig. 3). Hypervascular metastases include renal cell carcinoma, islet cell tumors, leiomyosarcoma, and thyroid carcinoma. Metastases may also show a complete peripheral ring of contrast enhancement. The CT findings of metastases are not pathognomonic and can be simulated by hepatomas, abscesses, hematomas, and complicated cysts.

Hepatocellular carcinoma Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and the most common type of cancer worldwide. Since nearly 90% of these tumors occur in patients with cirrhosis or chronic viral infection, detection can be obscured by the hepatic morphologic changes that attend these disorders. On noncontrast scans, HCC presents as a single hypodense or isodense hepatic mass or as multiple masses. In patients with fatty infiltration, however, the lesions can appear hyperdense. Calcifications are present in nearly 7% of lesions. HCC is most commonly hyperdense (Fig. 4) during arterial-phase imaging and becomes hypodense during the portal

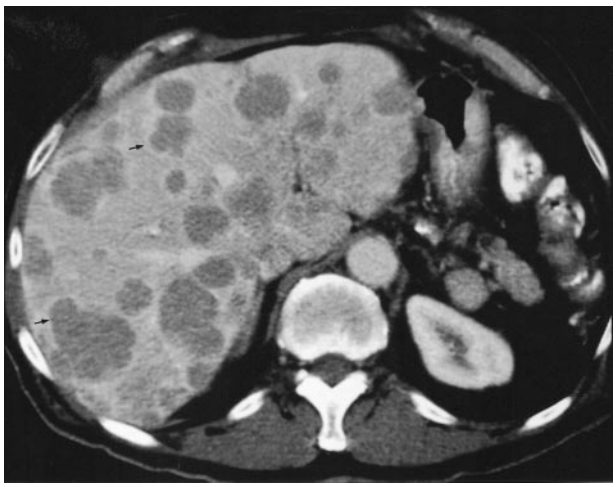


FIGURE 3 Liver metastases from colon cancer. Multiple low-attenuation masses are identified, showing a thin peripheral rim of enhancement (thin small arrows).

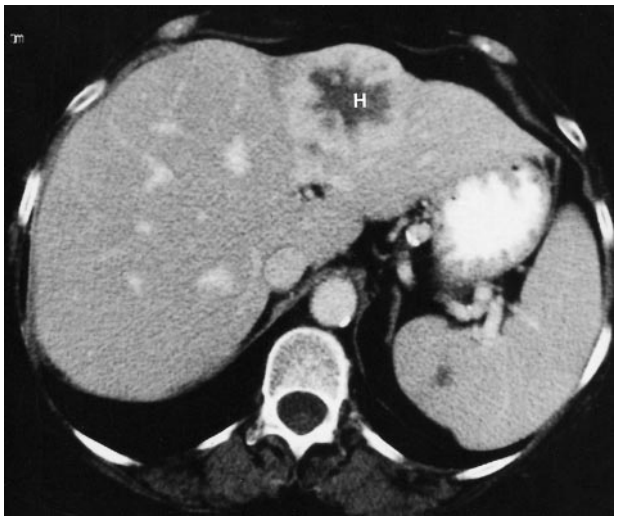


FIGURE 4 Hepatocellular carcinoma. There is a large peripherally hypervascular mass in the lateral segment of the left lobe of the liver (H). Necrosis is present centrally.

venous phase. On CT, HCC can appear as a focal, solitary mass, as a multifocal process, or as a diffuse, infiltrating process. Encapsulated HCC will demonstrate a hypodense rim on contrast-enhanced CT and demonstrate delayed contrast fill-in.

Cholangiocarcinoma Cholangiocarcinoma is the second most common primary malignant liver tumor. On noncontrast CT scans, cholangiocarcinomas appear as an irregular round or oval hypodense mass. Initial contrast enhancement is limited and may be peripheral or septated. The majority of these lesions demonstrate homogenous delayed contrast enhancement on scans performed 6 to 30 min following contrast administration.

Benign neoplasms

Hemangiomas Hemangiomas are the most common benign liver tumors and are usually less than 5 cm in size. Helical CT can definitively establish the diagnosis in more than 90% of cases by using the following criteria: peripheral globular or nodular enhancement, a lack of continuity of enhancing tissue, isodense or hyperdense enhancement relative to the aorta (Fig. 5), and centripetal "fill-in" of the lesion. When globular enhancement is demonstrated in CT, no further work-up is needed.

Focal nodular hyperplasia Focal nodular hyperplasia (FNH) is the second most common benign hepatic neoplasm. This lesion contains a central scar and is characteristically subcapsular in location. On noncontrast CT scans, FNH is typically a well-defined lesion that is isodense or hypodense compared to normal hepatic parenchyma. After contrast administration, FNH characteristically shows robust enhancement that washes out quickly. The central stellate scar is initially hypodense, but will accrete the contrast material on delayed images.

Adenomas Adenomas typically occur in young women taking oral contraceptives, athletes abusing anabolic steroids, and patients with glycogen storage disease. Adenomas have a propensity to bleed, which significantly affects their imaging appearance. These lesions may appear hypodense or hyperdense on noncontrast CT scans depending on the presence (see Fig. 5) or absence of hemorrhage. These lesions enhance rapidly on CT due to the predominantly hepatic arterial supply of the tumor. Hepatic adenomas tend to appear well defined due to the presence of a tumor capsule.

Vascular Disorders

Passive hepatic congestion Passive hepatic congestion is caused by stasis of blood within the liver parenchyma due to compromise of hepatic venous drainage. It is a common complication of congestive heart failure and constrictive pericarditis, wherein elevated central venous pressure is directly transmitted from the right atrium to the hepatic veins because of their close anatomic relationship. The liver becomes tensely swollen as the hepatic sinusoids engorge and dilate with blood. The CT findings of passive hepatic congestion include dilation of the inferior vena cava and the hepatic veins. The elevated central venous pressure permits retrograde opacification and enhancement of the inferior vena cava and hepatic veins following contrast injection (Fig. 6). Contrast-enhanced CT scans may show an inhomogenous,

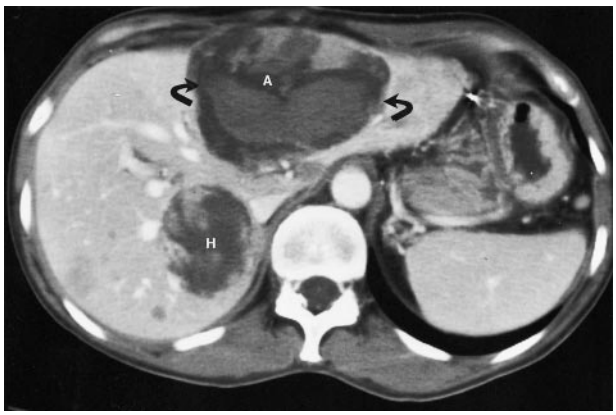


FIGURE 5 Hepatic adenoma and hemangioma. Globular peripheral enhancement, isodense with the aorta, is diagnostic of hepatic hemangioma (H). The adenoma (A) shows central hemorrhage (curved arrows) and inhomogenous enhancement.

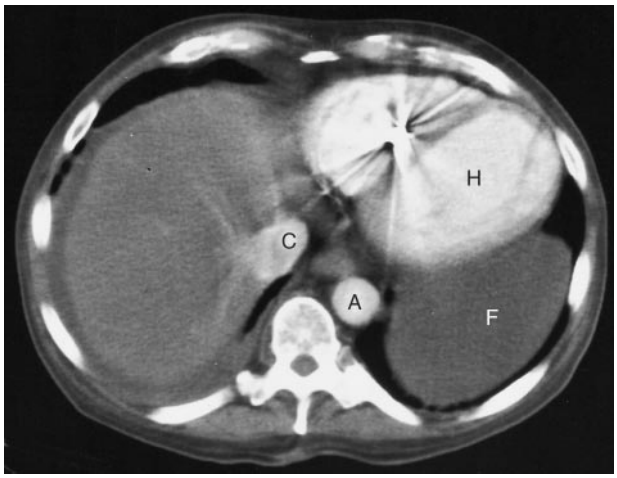


FIGURE 6 Passive hepatic congestion. The heart (H), aorta (A), and inferior vena cava (C) should not enhance to this degree so early after contrast administration, indicating reflux of contrast through the failing right heart. Note ascites (F).

mottled, reticulated-mosaic pattern of parenchymal contrast enhancement. Ancillary findings include cardiomegaly, hepatomegaly, intrahepatic periportal lucency due to perivascular lymphedema, pleural effusions, ascites, and pericardial effusions.

Budd–Chiari syndrome Budd–Chiari syndrome comprises a diverse group of conditions associated with hepatic venous outflow obstruction, at the level of either the large hepatic veins or the suprahepatic segment of the inferior vena cava. Diagnosis of this disorder is frequently difficult because of its protean causes and clinical manifestations. On noncontrast CT scans, diffuse hepatic hypodensity is associated with global liver enlargement and ascites. Hyperdense thrombus may be seen in the inferior vena cava and hepatic veins. Rarely, a calcified caval membrane is identified. After the injection of contrast material, the liver typically shows patchy enhancement. This is due to the hepatic congestion, which causes portal and sinusoidal stasis. In most patients, however, the central regions of the liver, including the caudate and part of the left lobe, may enhance normally and appear hyperdense compared with the more peripheral parts of the liver, which show decreased enhancement (Fig. 7). Later, a classic "flip-flop" pattern may develop, in which the contrast material from the normally enhanced central liver washes out so that this region becomes relatively hypodense compared with the peripheral zones, which are slowly accreting contrast material. The subcapsular portions of the liver may enhance normally, as they have independent venous drainage through systemic subcapsular veins. Intravascular thrombus is best seen in the acute phase of the Budd-Chiari syndrome, as a low-density mass in the lumen of the hepatic veins and inferior vena cava. Portal vein thrombus may be seen in up to 20% of cases. Compression of the inferior vena cava is also demonstrated.

Portal vein thrombosis Portal vein thrombosis is the leading cause of presinusoidal hypertension in the United States. Portal vein thrombosis appears as a low-density central zone surrounded by an intensely enhanced periphery on contrast-enhanced scans (Fig. 8). There is also transient inhomogenous enhancement of the periportal hepatic parenchyma. Enlargement of the occluded vein increases the likelihood of the presence of tumor thrombus. Streaky enhancement of the clot can be seen in tumor thrombus. On precontrast scans, the portal vein contents may be hyperdense due to the high protein content of concentrated red blood cells.

Biliary Tract

Inflammatory Diseases

The CT findings of sclerosing cholangitis (Fig. 9) include the following: intrahepatic duct stenoses, dilated peripheral ducts with no apparent connection to the central ducts, irregular intrahepatic duct dilation with a beaded appearance, and mural thickening, irregularity, and abnormal enhancement of the common bile duct.

Acute cholangitis is usually seen in the clinical setting of biliary obstruction and CT shows dilation of the intrahepatic bile ducts, high-attenuation debris in the duct lumen, and thickening and abnormal enhancement of the bile duct wall. Frank liver abscesses that manifest

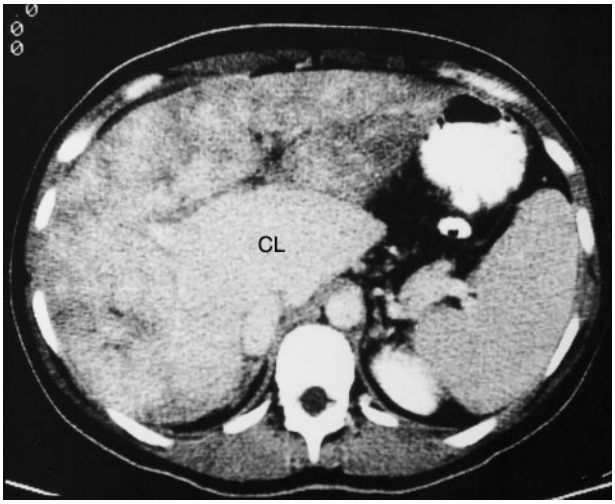


FIGURE 7 Budd-Chiari syndrome. Note the enlarged, enhancing caudate lobe (CL) and hypodense hepatic periphery with mottled enhancement.

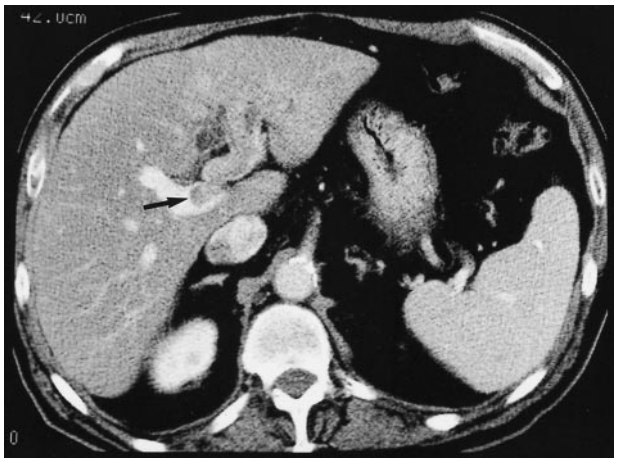


FIGURE 8 Portal vein thrombosis. Low-density tumor thrombus (arrow) is present within the left portal vein in this patient with hepatoma.

as low-attenuation areas in contiguity with the biliary tract may result from acute suppurative cholangitis.

Cholangiocarcinoma

Cholangiocarcinoma of the extrahepatic bile ducts is usually infiltrating, causing stenosis of the duct, but may appear exophytic or intralumenal. Characteristically, the tumor is hypodense early after contrast administration and later shows retention of contrast on delayed images.

Gallbladder

Gallstones

CT is less sensitive in the detection of gallstones than ultrasound: 75% compared to 98%. The CT appearance of gallstones is variable, depending on their composition, pattern of calcification, and the presence of lamination, fissuring, or gas. Stones with a high cholesterol content are difficult to see because they are isodense with the surrounding bile. Well-calcified stones are easily detected on CT. Stones that are denser than bile may be seen due to a rim or nidus of calcification. The CT attenuation of gallstones correlates more closely with the cholesterol content of the stones than with the calcium content. On CT, gallstones can be simulated by the enhancing mucosa of a contracted gallbladder wall or neck, which often fold upon themselves.

Acute Cholecystitis

Although ultrasound is the preferred method for diagnosing acute cholecystitis, CT is frequently the initial examination ordered because the diagnosis is unclear. Mural thickening greater than 3 mm in the setting of a distended gallbladder and enhancement of the inflamed wall are the most sensitive helical CT findings of acute cholecystitis (Fig. 10). Transient focal increased attenuation of the liver may develop adjacent to the inflamed gallbladder as a result of hepatic arterial hyperemia and early venous drainage. Less specific signs include pericholecystic fluid, haziness of the pericholecystic fat, and increased attenuation of the gallbladder bile. Helical CT can also depict the complications of acute cholecystitis, such as perforation and gangrene. Intramural or intralumenal gas is present in emphysematous cholecystitis.

Choledocholithiasis

Patients with choledocholithiasis typically present with acute right upper quadrant pain, fever, jaundice, and pancreatitis. Thin collimation (3 mm) scans are

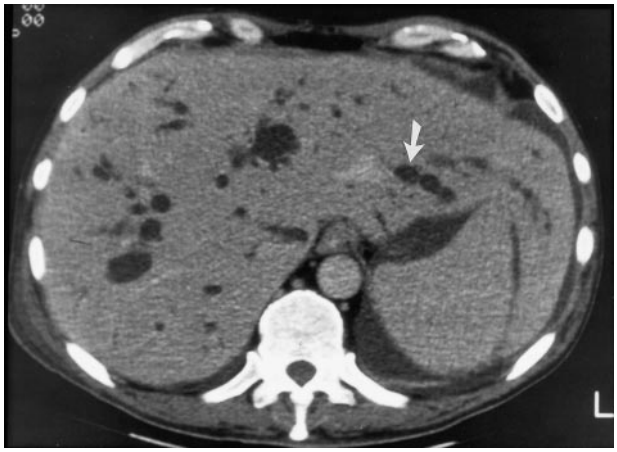


FIGURE 9 Sclerosing cholangitis. Multiple areas of heading of focally ectatic intrahepatic (arrow) are evident.

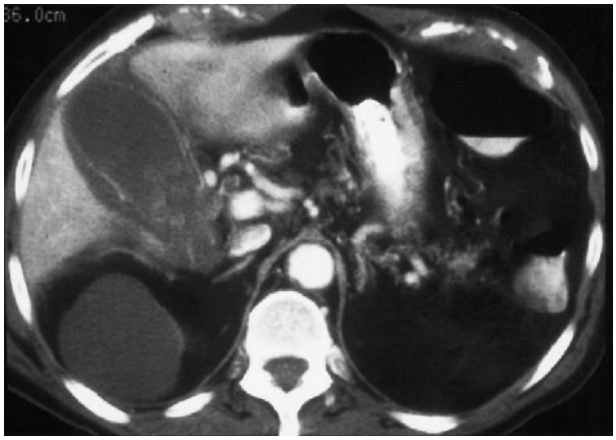


FIGURE 10 Acute cholecystitis. The gallbladder wall is thickened and shows mural edema. Multiple stones are also evident.

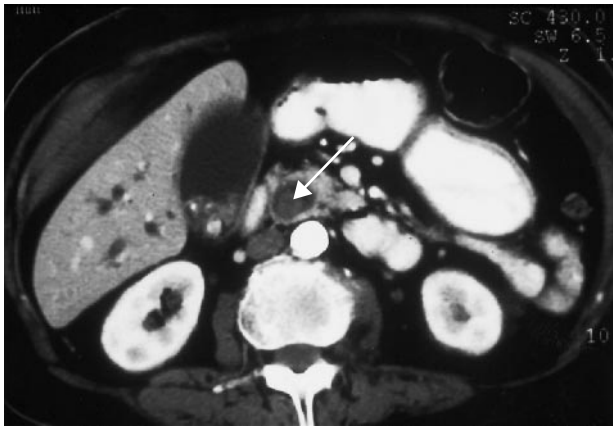


FIGURE 11 Choledocholithiasis. A high-density stone is present within the lumen of a dilated common bile duct (arrow).

needed to optimize the detection of stones on CT. A high-density nidus may be visualized within the duct or alternating low- and high-density rings of mixed cholesterol–calcium stones may be seen (Fig. 11). Biliary dilation may be evident proximally. Helical CT has a sensitivity of 88%, a specificity of 97%, and an accuracy of 94% in the detection of choledocholithiasis. Positive intralumenal and intravascular contrast material can obscure the detection of peripherally calcified stone, however.

Neoplasms

Carcinoma of the gallbladder is the fifth most common malignancy of the gastrointestinal tract, responsible for nearly 7000 deaths annually in the United States. This neoplasm has three major patterns of presentation pathologically and on CT: (1) focal or diffuse mural thickening; (2) an intralumenal polypoid mass (Fig. 12), usually larger than 2 cm, originating in the gallbladder wall; and (3) most commonly (45–65%), a subhepatic mass replacing or obscuring the gallbladder, often invading adjacent liver.

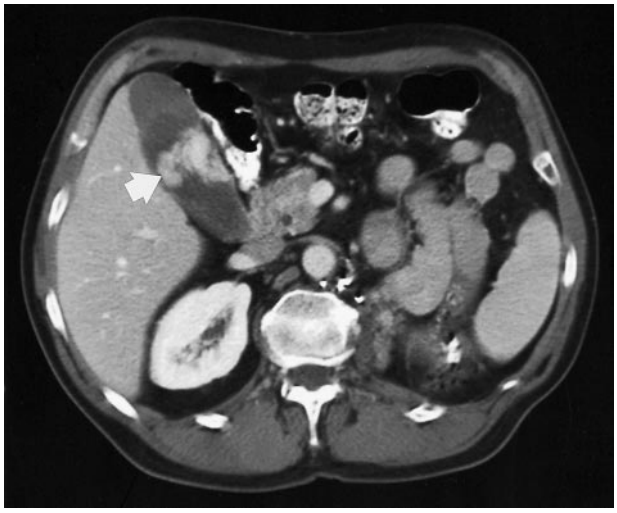


FIGURE 12 Gallbladder carcinoma. This neoplasm manifests as an enhancing mass (arrow) in the gallbladder lumen.

Pancreas

Neoplasms

Adenocarcinoma On CT, most adenocarcinomas present as a focal hypodense (Fig. 13) mass on contrast-enhanced scans. Atrophy of the pancreas and dilation of the pancreatic duct are often seen upstream of the neoplasm. Helical CT is also quite useful in assessing resectability of the tumor. Liver metastases, vascular encasement, and involvement of adjacent lymph nodes suggest unresectability.

Islet cell tumors Most islet cell tumors are strikingly hypervascular lesions and consequently enhance robustly following contrast administration. Most are less than 2 cm in size.

Cystic pancreatic neoplasm Serous, microcystic cystadenomas have small ($\leq$1 cm), often innumerable cysts interspersed within a dense fibrous honeycomb. Central calcifications occur in 30% of these usually benign tumors. Mucinous cystadenomas and cystadenocarcinomas present as macrocystic ($\geq$2 cm) unilocular cysts or as intraductal neoplasms. Calcification is rare and mural nodules may be present. On the basis of CT findings, it is difficult to differentiate benign from malignant mucinous tumors.

Pancreatitis

Acute pancreatitis Helical CT plays a vital role in the clinical management and staging of patients with acute pancreatitis. CT can establish the presence of hemorrhage or necrosis within the gland itself and identify the extension of the inflammatory process into adjacent organs. The CT findings of acute pancreatitis reflect edema of the gland and surrounding fat and may be normal in up to 28% of mild cases. The entire gland may become diffusely enlarged with an irregular contour. In mild cases, the peripancreatic fat contains wisps of high attenuation, the vascular margins are cuffed, and the fascial planes are thickened. Mild peripancreatic inflammation may be present around an otherwise normal appearing gland. Segmental pancreatitis occurs in 10–18% of patients and is usually associated with stone disease. Typically, the gland shows uniform enhancement.

In more advanced cases, intraglandular intravasation of pancreatic fluid leads to the formation of multiple, small, intrapancreatic fluid collections. In

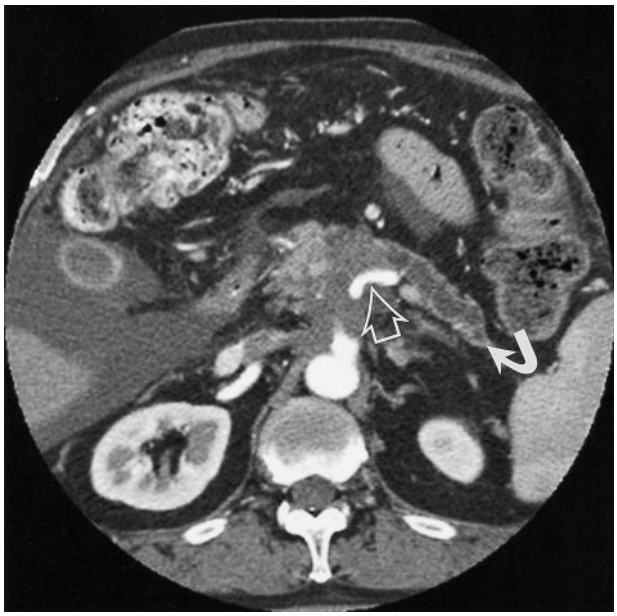


FIGURE 13 Pancreatic adenocarcinoma. CT shows a hypodense mass that encases the splenic artery (open arrow). Note the upstream dilation of the pancreatic duct and parenchymal atrophy (curved arrow).

necrotizing pancreatitis, the gland becomes enlarged and is often enveloped by high-attenuation exudates. Necrotic parenchyma shows decreased or no enhancement that is sharply demarcated from normally enhancing viable tissue (Fig. 14). The body and tail are usually involved; the head is spared because of its rich collateral vascular network. Enhancing islands of viable tissue may be scattered throughout the gland. The poorly

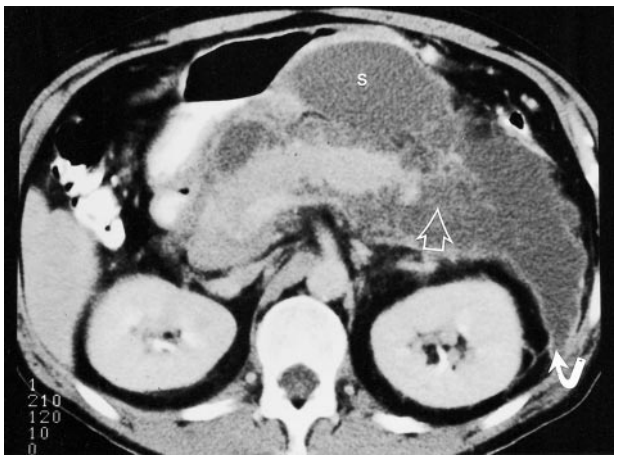


FIGURE 14 Necrotizing pancreatitis. There is poor enhancement of the pancreatic tail (open arrow), suggesting necrosis. Note the large fluid collections in the lesser sac (S) and the anterior interfascial plane (curved arrow).

defined peripancreatic exudates obliterate the peripancreatic fat, dissect fascial planes, and penetrate through fascial and peritoneal boundaries and ligaments. These collections typically accumulate in the lesser sac, anterior pararenal space, and anterior interfascial space. Helical CT is also useful in demonstrating vascular complications, such as pseudoaneurysms and splenic and portal vein thrombosis. Delineation of necrosis by helical CT can help predict patient outcome.

Chronic pancreatitis The CT features of chronic pancreatitis include the following: dilation of the main pancreatic duct, parenchymal atrophy, pancreatic calcifications, intra- and peripancreatic fluid collections, focal pancreatic enlargement, and biliary tract dilation. It is often difficult to differentiate focal mass-like chronic pancreatitis from adenocarcinoma on the basis of CT features alone.

HOLLOW ABDOMINAL VISCERA

Esophagus

Neoplasms

On CT, esophageal carcinoma produces either concentric mural thickening or a focal, asymmetric soft tissue mass arising from the wall. CT is useful for tumor staging, assessing response to therapy, and diagnosing complications of surgery. Tumor invasion of the aorta is unlikely if the tumor involves the aortic circumference by $\leq 45°$; contact $\geq 90°$ is highly suggestive of invasion and contact in the range of 45° to 90° is indeterminate. Aortic invasion is also predicted when the small triangular deposit of fat between the esophagus, aorta, and spine is obliterated. The preoperative identification of tracheobronchial invasion is important because it may make tumor resection, particularly by the transhiatal route, more difficult or even impossible. The demonstration of a tracheobronchial fistula or extension of the tumor into the airway lumen is a specific sign of invasion. Tracheobronchial invasion should be suspected if an esophageal tumor either causes inward bowing of the posterior tracheal or bronchial wall or displaces the trachea and/or bronchi away from the spine.

Varices

The complete extent of esophageal and abdominal varices is well depicted on CT. Varices produce thickening of the esophageal wall, a scalloped contour, and enhancing intralumenal protrusions following intravenous contrast injection.

Stomach

Neoplasms

Malignant tumors

Adenocarcinoma Gastric carcinomas on CT appear as a region of focal, inhomogenous mural thickening (Fig. 15). Scirrhous and polypoid forms, with or without ulceration, are typically seen. Gastric carcinoma typically spreads to the liver, lymph nodes, pancreas, duodenum, transverse mesocolon, and peritoneal cavity. Direct extension of proximal tumors can occur via the gastrohepatic ligament to the porta hepatis and left lobe of the liver. The transverse colon may be involved by tumor spread through the gastrocolic ligament and splenic involvement may be seen from gastrosplenic ligament invasion. Direct extension may be seen into the pancreas and duodenum. The use of helical CT techniques that optimize intravenous contrast levels and minimize respiratory misregistration helps to identify lymph node involvement and vascular encasement. Carcinomatosis with diffuse peritoneal seeding, malignant ascites, and adnexal and pelvic metastases are well depicted by CT.

Lymphoma On CT, gastric lymphomas tend to produce significantly greater wall thickening than adenocarcinoma. Lymphomas also tend to be more homogenous and more diffusely involve the stomach than adenocarcinomas.

Leiomyosarcomas On CT, these lesions tend to be large and inhomogenous, to ulcerate, and to have a greater degree of extragastric growth than leiomyomas.

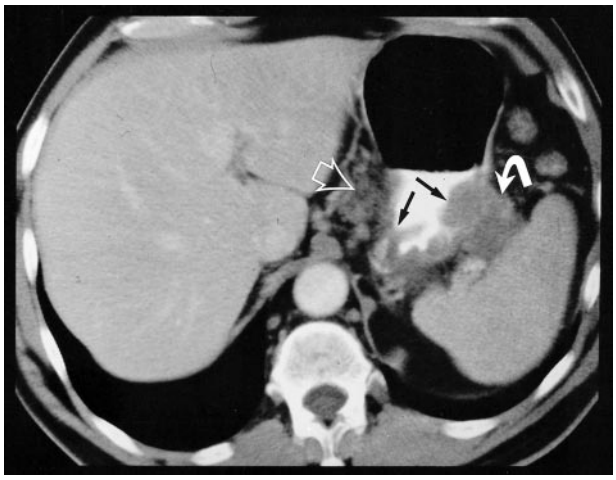


FIGURE 15 Gastric cancer. There is mural thickening with polypoid folds of the proximal gastric body (black arrows). There is tumor spread into the gastrohepatic ligament (open arrow) and gastrosplenic ligament (curved arrow).

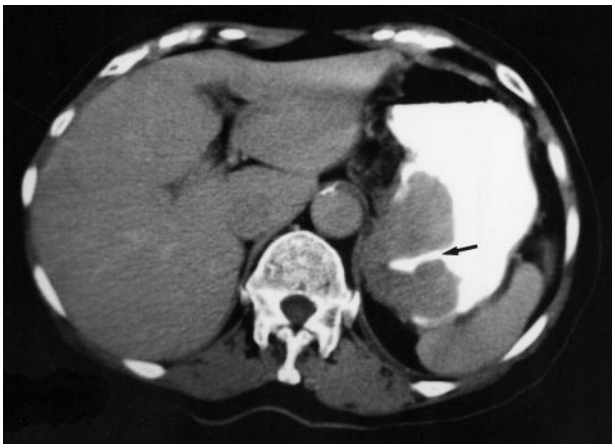


FIGURE 16 Gastric leiomyoma. A large, ulcerating (arrow), endophytic mass is present near the cardia of the stomach.

Benign tumors

Leiomyomas These lesions tend to be less than 4 cm in size, are homogenous, and show endolumenal growth as on CT (Fig. 16). Differentiation from leiomyosarcomas may be impossible.

Lipomas These well-circumscribed lesions have fat (low) attenuation on CT.

Gastritis

On CT, gastritis manifests as mural thickening, often with submucosal edema. CT has no role in the primary evaluation of gastritis but can be useful in showing complications of peptic ulcer disease, e.g., perforation (Fig. 17) or fistula.

Small Bowel

Crohn's Disease

Crohn's disease is manifested on CT by bowel wall thickening ranging between 1 and 2 cm. During the acute, noncicatrizing phase of Crohn's disease, the small bowel and colon maintain mural stratification and often have a target or double-halo appearance. There is a soft-tissue-density ring (corresponding to mucosa), which is surrounded by a low-density ring with an attenuation near water or fat (corresponding to submucosal edema or fat infiltration, respectively), which in turn is surrounded by a higher-density ring (muscularis propria). Inflamed mucosa and serosa may show significant contrast enhancement following bolus intravenous contrast administration and the intensity of enhancement correlates with the clinical activity of disease (Fig. 18).

The CT demonstration of mural stratification, i.e., the ability to visualize distinct mucosal, submucosal,

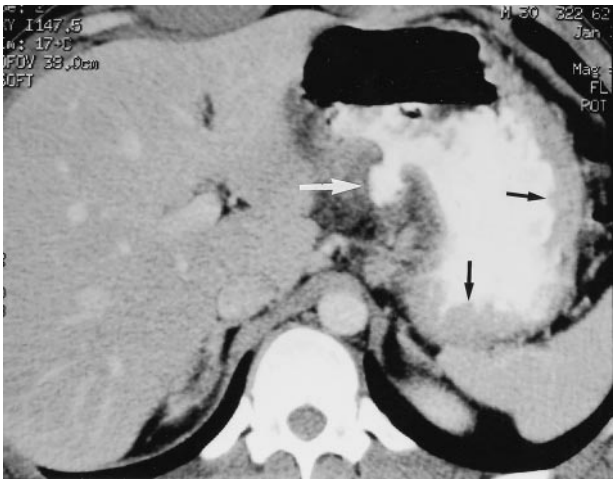


FIGURE 17 Penetrating benign gastric ulcer. A large benign gastric ulcer (white arrow) is seen penetrating into the gastrohepatic ligament. The thick gastric folds (black arrows) attest to the gastritis.

and muscularis propria layers, indicates that transmural fibrosis has not occurred and that medical therapy may be successful in ameliorating lumen compromise. The edema and inflammation of the bowel wall that cause mural thickening and lumen obstruction are reversible to some extent. A modest decrease in wall thickness often produces a dramatic increase in lumen cross-sectional area and resolution of the patient's obstructive symptoms.

In patients with long-standing Crohn's disease and transmural fibrosis, mural stratification is lost so that the affected bowel wall typically has homogenous attenuation on CT. Homogenous attenuation of the thickened bowel wall in the presence of good intravascular contrast-medium levels and thin-section scanning suggests irreversible fibrosis so that anti-inflammatory agents may not provide a significant reduction in bowel wall thickness. If these segments become sufficiently narrow, surgery or stricturoplasty may be necessary to relieve the patient's obstruction.

CT can readily differentiate the extralumenal manifestations of Crohn's disease. Fibrofatty proliferation, also known as "creeping fat" of the mesentery, is the most common cause of separation of bowel loops seen on small bowel series in patients with Crohn's disease. On CT, the sharp interface between bowel and mesentery is lost and the attenuation value of the fat is elevated by 20 to 60 housefield units (HU) due to the influx of inflammatory cells and fluid. Mesenteric adenopathy with lymph nodes ranging in size between 3 and 8 mm may also be present. Contrast-enhanced CT scans often show hypervascularity of the involved mesentery manifesting as vascular dilation, tortuousity, prominence, and wide spacing of the vasa recta. These distinctive vascular changes have been called "vascular jejunization of the ileum" or the "comb sign." Identification of this hypervascularity should suggest active disease and may be useful in differentiating Crohn's disease from lymphoma or metastases, which tend to be hypovascular lesions. On CT, a phlegmon produces loss of definition of the surrounding organs and a "smudgy" or "streaky" appearance of the adjacent mesenteric or omental fat. Initially, abscesses appear as a mass of soft tissue attenuation due to the influx of inflammatory cells. With maturation, the abscess undergoes liquefactive necrosis centrally and highly vascularized connective tissue develops peripherally. As a result, this lesion has a low-attenuation center with an enhancing rim (Fig. 19).

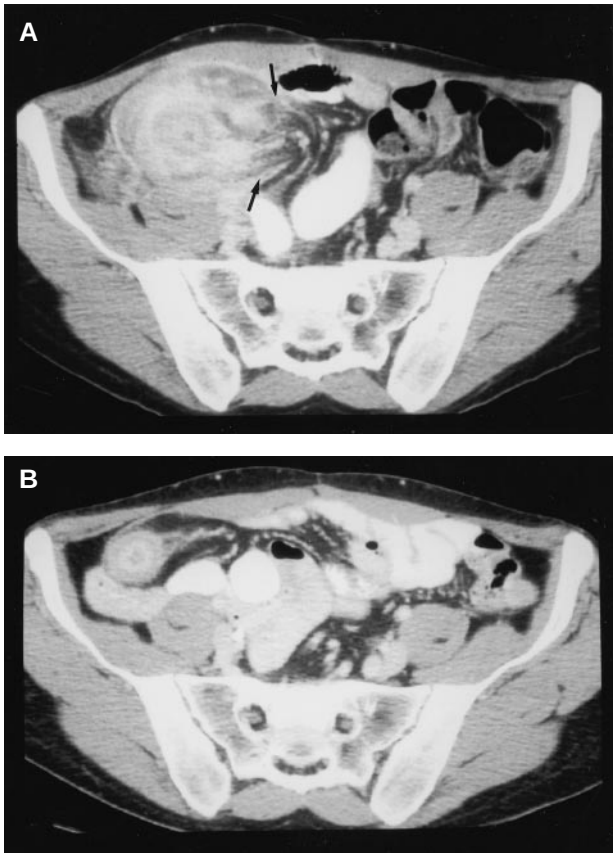


FIGURE 18 Crohn's disease. (A) There is marked mural thickening and abnormal enhancement of the distal ileum. Note the inflammation of the adjacent mesentery (arrows). (B) Note the dramatic mural and mesenteric improvement following medical therapy.

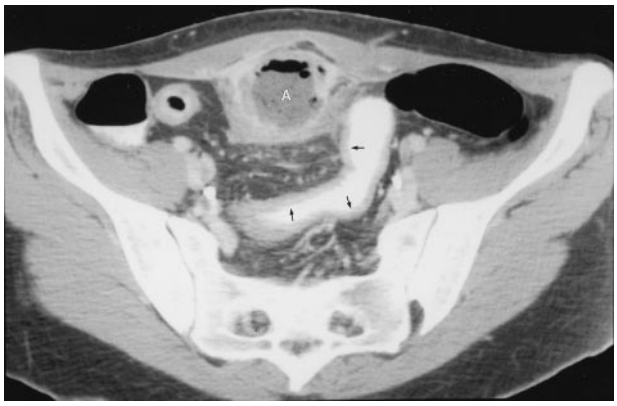


FIGURE 19 Crohn's disease with mesenteric abscess. There is mural thickening of the distal ileum (black arrows), which has an adjacent mesentery that contains engorged vasa recta. Note the gas-containing abscess (A) that has a thick, enhancing wall.

Bowel Obstruction

Small and large bowel obstruction accounts for approximately 20% of acute abdominal surgical conditions. Helical CT has replaced conventional contrast studies in these patients because it can more reliably answer the following questions:

1. Is an obstruction present?
2. What is the level of obstruction?
3. What is the cause of the obstruction?
4. What is the severity of the obstruction?
5. Is the obstruction simple or is it a closed loop?
6. Is strangulation or ischemia present?

It is important to differentiate between simple versus closed-loop obstruction since the former can be managed conservatively, whereas the latter requires prompt surgical intervention. Scans in these patients are best obtained without oral contrast material because intralumenal fluid and gas serve as a natural contrast agent. Intravenous contrast is important in assessing intestinal perfusion and ischemia as well as delineating the size, configuration, and patency of the mesenteric vessels.

Demonstration of a transition zone between dilated and decompressed bowel is the CT hallmark of bowel obstruction. Careful inspection of the transition zone and lumenal contents will often demonstrate the underlying etiology of the obstruction. CT is most helpful in patients with internal and external hernias, neoplasms, gallstone ileus, various forms of enteroenteric intussusception and afferent loop obstruction following a Billroth II gastrectomy. If no mass, hernia, intussusception, stone (Fig. 20), abscess, or inflammatory thickenings are present, then adhesion is the most likely diagnosis. The typical adhesion has a beak-like narrowing and the affected gut may be difficult to appreciate depending on the orientation of the loop relative to the axial plane.

An incarcerated or closed-loop obstruction manifests as a loop-shaped, fluid-filled structure, causing the proximal segments to dilate with gas and fluid. The mesenteric vessels have a radial distribution as they become stretched and converge toward the U- or C-shaped loop. There are typically two adjacent collapsed round/oval/triangular segments that represent the afferent and efferent entry points or the torsion site. The mesenteric vasculature may have an unusual course. When ischemia develops, the bowel wall may be thickened and have a target appearance due to submucosal edema. Additionally, there may be poor or delayed enhancement of the involved bowel wall. Fluid and hemorrhage may collect in the mesentery, bowel wall, and/or lumen of the involved segment. The mesentery becomes hazy in appearance and ascites may develop.

In patients with high-grade small bowel obstruction, CT has a reported sensitivity of 90–96%, a specificity of 91–96%, and an accuracy of 90–95%. CT is less accurate in patients with low-grade obstruction.

Neoplasms

Small bowel lymphoma may demonstrate a large homogenous mass or a large ulcerating mass, often with associated enlargement of mesenteric and retroperitoneal lymph nodes (Fig. 21). Carcinoid tumors classically manifest as a variably calcified mass with radially oriented linear densities in the perienteric fat. Resulting desmoplasia causes bowel loops to be drawn in toward the mass.

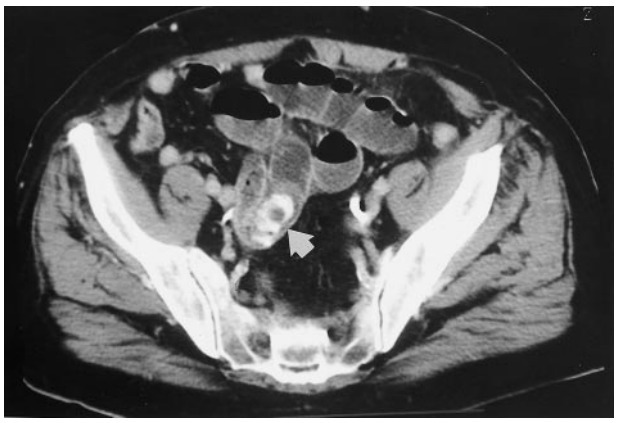


FIGURE 20 Gallstone ileus. A calcified gallstone (arrow) is causing small bowel obstruction

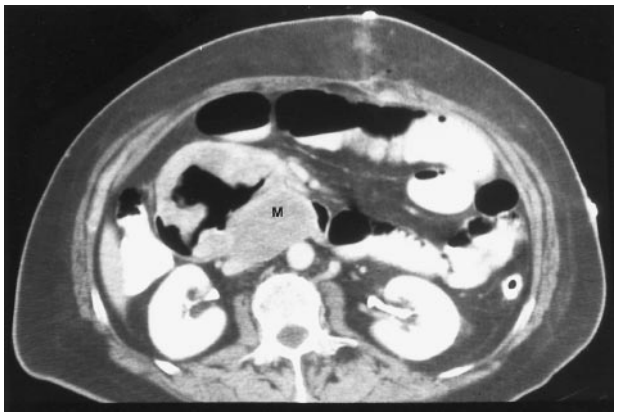


FIGURE 21 Small bowel lymphoma. There is a large cavitating mass (M) in the ileum.

Colon

Ulcerative Colitis

In acute, fulminant ulcerative colitis (Fig. 22), deep ulcerations may be identified associated with increased intralumenal fluid. Mural thickening and lumen narrowing are common CT features of subacute and chronic ulcerative colitis. The mucosa becomes thickened due to hypertrophy of the muscularis mucosae in chronic ulcerative colitis. Additionally, the lamina propria is thickened due to round cell infiltration in both acute and chronic ulcerative colitis. The submucosa becomes thickened due to the deposition of fat (Fig. 23) or, in acute and subacute cases, edema. Submucosal thickening further contributes to lumenal narrowing.

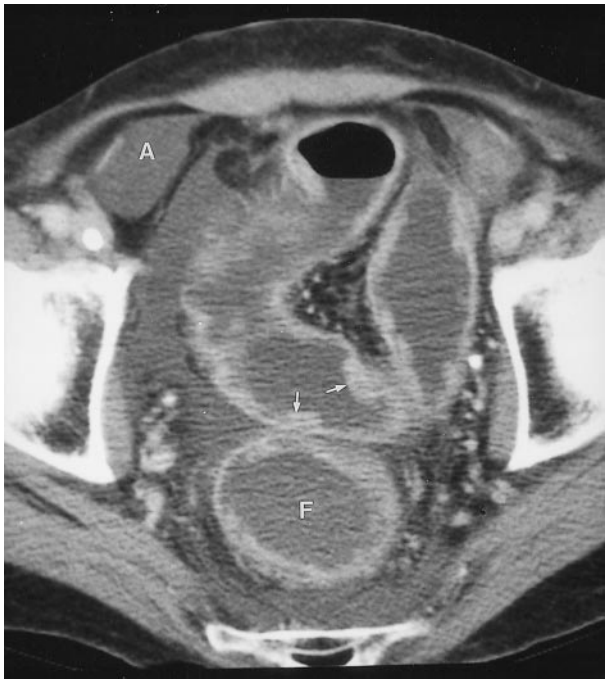


FIGURE 22 Acute ulcerative colitis. Large areas of ulceration leave inflammatory pseudopolyps (arrows) in their wake. The increased lumenal fluid (F) attests to the copious diarrhea. Note the ascites (A).

On CT, these mural changes produce a target or halo appearance when axially imaged: the lumen is surrounded by a ring of soft tissue density (mucosa, lamina propria, hypertrophied muscularis mucosae), which is surrounded by a low-density ring (edema or fatty infiltration of the submucosa), which in turn is surrounded by a ring of soft tissue density (muscularis propria). This mural stratification is not specific and can also be seen in Crohn's disease, infectious enterocolitis, pseudomembranous colitis, ischemic and radiation enterocolitis, mesenteric venous thrombosis, bowel edema, and graft-versus-host disease.

Diverticulitis

Diverticulitis occurs in 10–25% of patients with known diverticulosis. Clinical misdiagnosis rates ranging from 34 to 67% have been reported. The role of CT in these patients is to confirm the diagnosis, establish the presence of complications (e.g., abscess), provide a road map for percutaneous or surgical therapy, and suggest alternative diagnoses in patients in whom diverticulitis has been excluded.

Inflammatory change in the pericolic fat is the CT hallmark of diverticulitis (Fig. 24), seen in 98% of patients. In mild cases, there is only minimal haziness of the adjacent fat. Small fluid collections, fine linear strands, and extralumenal gas bubbles may be visualized. In more severe cases, phlegmon or frank abscess formation can occur. Diverticula are evident in >80% of cases and symmetric mural thickening in excess of 4 mm is seen in approximately 70% of patients. In some cases, contrast material collects in an arrowhead shape adjacent to the inflamed colonic diverticulum—the arrowhead sign. The offending inflamed diverticulum may appear as a rounded, paracolic out-pouching centered within the paracolic inflammation with soft tissue, calcium, barium, or air attenuation.

Appendicitis

Acute appendicitis is the most common abdominal surgical emergency, affecting approximately 250,000 people annually in the United States. Although the correct diagnosis can be made in most patients on the basis of history, physical examination, and laboratory tests, diagnosis is uncertain in the 20–33% of patients who present with atypical features. The diagnosis is most

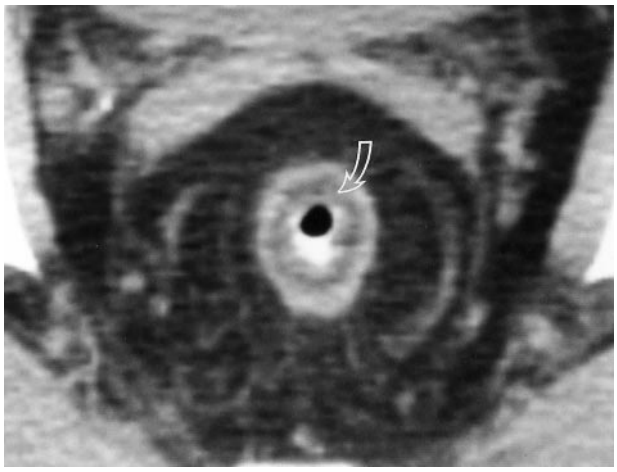


FIGURE 23 Chronic ulcerative colitis. CT scan of the rectum shows mural thickening and submucosal fat deposition (arrow).

difficult to establish in infants and young children, the elderly, and women of reproductive age. In the past, an average negative laparotomy rate of 20% was acceptable. The widespread use of helical CT in patients with suspected appendicitis has been shown to positively affect patient outcome and increase the number of positive laparotomies. Studies evaluating the efficacy of helical CT show sensitivities of 90–100%, specificities of 83–97%, and accuracies of 93–98% for the diagnosis of acute appendicitis.

The CT findings in acute appendicitis reflect the extent and severity of inflammation. In mild disease, the appendix appears as a distended (6–15 mm diameter) fluid-filled structure that shows circumferential, symmetric mural thickening (Fig. 25). Homogenous

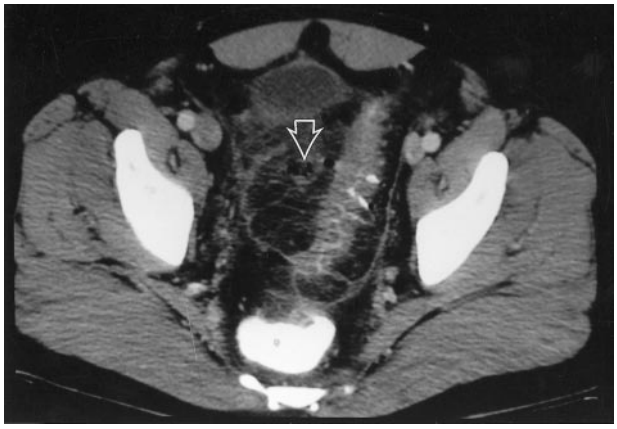


FIGURE 24 Diverticulitis. There is mural thickening of the sigmoid colon associated with inflammatory change and gas bubbles (arrow) in the sigmoid mesocolon.

dense contrast enhancement of the wall is typical but a target sign may be seen when the appendix is axially viewed. Periappendiceal inflammation manifests as slight haziness of the fat of the mesoappendix. A calcified appendicolith is more reliably demonstrated on CT than on plain films.

With disease progression and perforation, the appendix becomes fragmented, destroyed, and replaced by a phlegmon or abscess. There may be associated mural thickening of the adjacent distal ileum and cecum. In these cases, the specific diagnosis of appendicitis can be made if an appendicolith is seen within the abscess or phlegmon.

Epiploic Appendagitis

This unusual condition occurs when an epiploic appendage of the colon develops inflammation, torsion, or ischemia. Epiploic appendagitis can simulate appendicitis and right- and left-sided diverticulitis both clinically and on CT scans. The inflamed appendage presents as a small fat attenuation mass with a hyperattenuating rim that abuts the serosal surface of the colon (Fig. 26). At the center of the lesion, a small round or linear hyperdense focus may be seen and is thought to represent vascular thrombosis. Epiploic appendagitis also produces mass effect and focal thickening of the adjacent bowel, infiltration of the mesenteric fat, and focal thickening of the surrounding peritoneum.

Other Colitides

Infectious enterocolitides Gastroenteritis and the infectious enterocolitides are responsible for nearly 70% of emergency room visits prompted by abdominal pain. Most cases are self-limited and do not require imaging. In atypical cases, colicky abdominal pain rather than diarrhea may be the predominant symptom. The CT scan may be normal or show nonspecific mural thickening in more severe cases of *Shigella*, invasive *Escherichia coli*, *Salmonella*, *Yersinia*, and *Entamoeba*.

In pseudomembranous colitis, potent antibiotics disrupt the normal bacterial flora of the colon, leading to overgrowth of *Clostridium difficile*. The release of its enterotoxins causes mucosal inflammation and the development of pseudomembranes consisting of mucous and inflammatory debris. On CT, there is mural thickening averaging 15–20 mm, with a target or halo pattern due to submucosal edema (Fig. 27). Contrast material caught between thick haustra may simulate deep ulcerations and produce an accordion-like appearance. The lumen may be completely effaced as well. Ascites and pericolic inflammatory changes accompany these features.

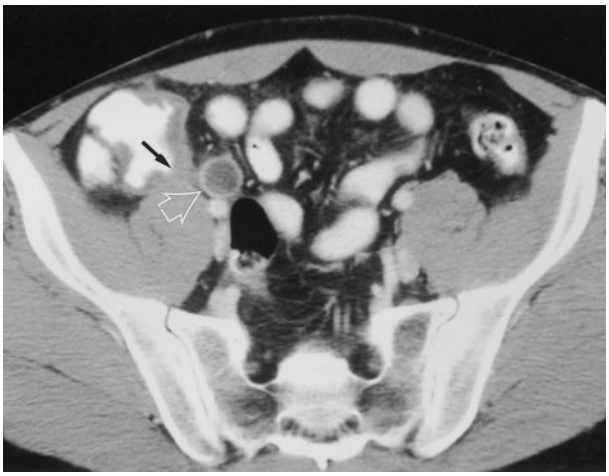


FIGURE 25 Appendicitis. There is dilation of the appendix with increased mural enhancement (open arrow). There is reactive thickening of the adjacent cecum (black arrow).

Helical CT is most useful in differentiating the panoply of inflammatory, infectious, and neoplastic disorders that can cause the acute abdomen in AIDS patients. Infections such as cryptosporidiosis (Fig. 28) and cytomegalovirus produce thickening of the gut wall with edema of the submucosa and increased enhancement of the mucosa.

Typhlitis Neutropenic colitis is an acute inflammatory and necrotizing process that affects the cecum of immunocompromised patients with profound neutropenia. In this disorder, there is ulceration of the mucosa followed by bacterial and fungal invasion. The CT features are nonspecific and include segmental mural thickening of the cecum (Fig. 29), intramural regions of edema or necrosis, pericolic fluid, and perienteric stranding. In advanced cases, pneumatosis intestinalis and frank perforation may develop.

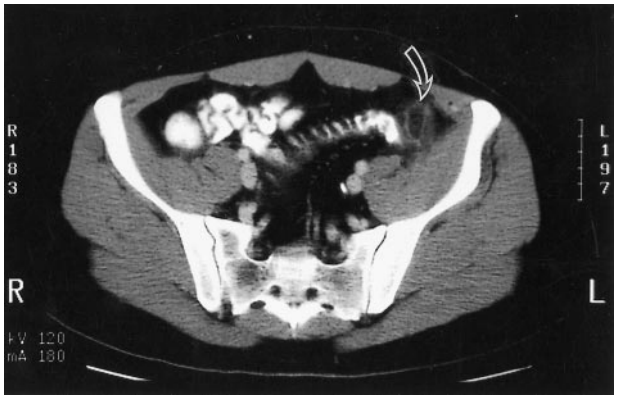


FIGURE 26 Epiploic appendagitis. The fat-density epiploic appendage is surrounded by high-density inflammatory changes (curved arrow).

Neoplasms

Colorectal cancer (CRC) is usually diagnosed by barium studies, colonoscopy, or sigmoidoscopy. Although these techniques provide superb visualization of the mucosa, they cannot determine the depth of mural invasion by the tumor or the extent of metastatic involvement. Preoperative staging aims to accurately assess tumor extent in order to individualize patient therapy, facilitate evaluation of treatment results, assess risk of disease recurrence, and determine prognosis. By virtue of their cross-sectional imaging format and ability to demonstrate the extent of wall invasion, extralumenal tumor spread, and lymph node and distant metastases, CT has become the primary means for radiologic staging of CRC (Fig. 30).

The preoperative accuracy of CT in staging CRC is limited, with reported sensitivity rates of only 48–77%. This is due to the inability of CT to accurately assess the degree of mural invasion and the difficulty in differentiating reactive from malignant lymph nodes. Despite these limitations, preoperative CT is commonly used to provide a baseline exam for comparison with postoperative imaging studies and to help detect metastatic lesions prior to surgery. Additional indications for preoperative CT include suspected contiguous organ invasion by tumor and unusual tumor histology, such as lymphoma.

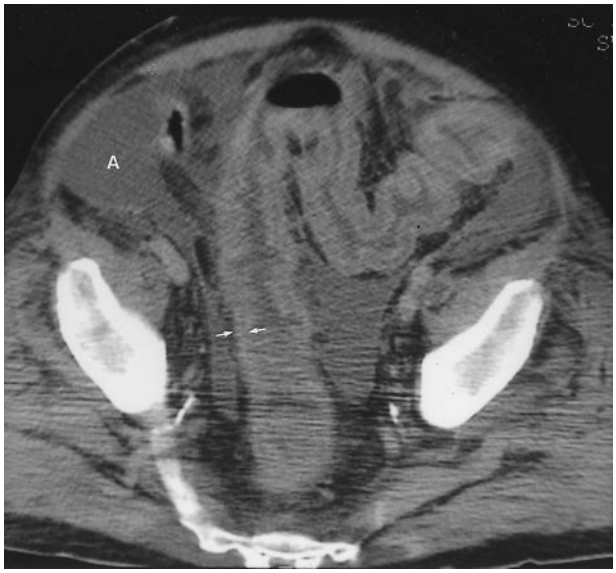


FIGURE 27 Pseudomembranous colitis. There is a proctosigmoiditis associated with diffuse mural thickening (arrows) and submucosa edema. A, ascites.

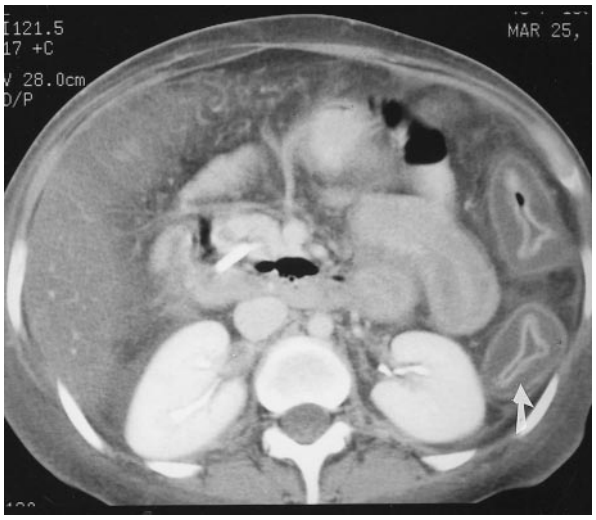


FIGURE 28 Cryptosporidial colitis. There is marked submucosal edema (arrow) evident in the region of the distal transverse and proximal descending colon in this AIDS patient with a CD4 lymphocyte count <50. Note the enhancing mucosal and muscularis propria layers.

The sensitivity of conventional CT approaches 75% in the detection of the initial colorectal tumor. Lesions less than 3 to 5 mm are usually not detectable. This is probably not clinically significant in most cases, since the risk of malignancy in a polyp less than 1 cm in diameter is less than 1%. Most often, the inability to detect small lesions of the colon and rectum on CT is due to incomplete bowel distension and difficulty in the differentiation of adherent fecal material from true colonic polyps.

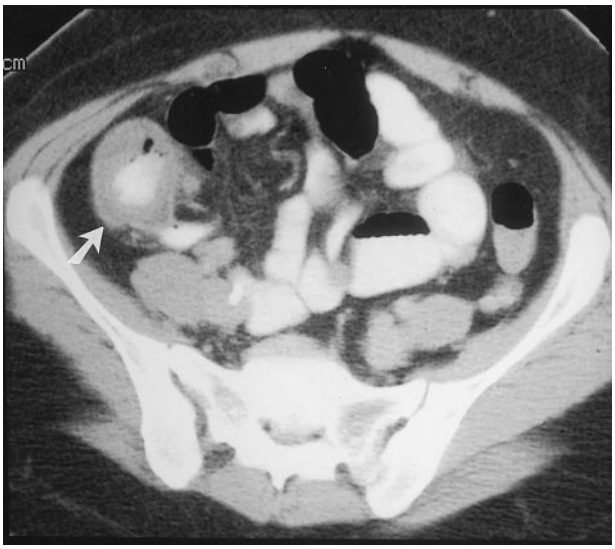


FIGURE 29 Typhlitis. There is focal mural thickening of the cecum (arrow) in this patient with leukemia and profound neutropenia.

Detection of colonic tumors on CT can be optimized by using a scanning protocol that acquires thin-section images (3–5 mm) obtained during peak enhancement of a rapidly delivered contrast bolus ($\geq$3 ml/s). Some authors advocate scanning from the symphysis pubis cephalad to the iliac crests to maximize visualization of mural enhancement. Certain centers also perform pelvic CT in both supine and prone positions to minimize adherent fecal material. Other techniques used to maximize tumor detection include instilling air into the rectum with a Foley catheter to increase distension of the colon and rectum. Helical CT scanning with breath-holding should be used if possible to minimize respiratory motion artifacts and spatial misregistration. Images should be displayed in soft tissue, liver, and bone windows to maximize lesion conspicuity.

CRC on CT most commonly manifests as a soft tissue mass that narrows the colonic lumen. The mass, if large enough, may undergo central necrosis, causing low attenuation. Mucinous primary adenocarcinomas may rarely contain calcium, although this more commonly occurs with hepatic metastasis than in the primary tumor. Occasionally, necrotic tumor masses may contain gas and perforate into adjacent structures. In such cases, differentiation from an abscess may be difficult. Another relatively common appearance of primary tumors of the colon is focal wall thickening. The normal, distended colonic wall should not measure more than 3 mm in thickness on CT. Wall thickness between 3 and 6 mm is considered indeterminate but concerning for neoplasia and measurements greater than 6 mm are abnormal. Malignant wall thickening most often has a nodular appearance and is associated with lumenal narrowing. However, the thickening can occasionally appear symmetric and involve the entire circumference of the bowel wall. In such cases, malignancy can be difficult to differentiate from benign causes of wall thickening including inflammation and mural edema.

The pericolonic and perirectal fat should appear uniformly low in density without soft tissue stranding on CT. Invasion beyond the bowel wall should be suspected if a focal mass deforms the contour of the outer colonic wall, with or without soft tissue stranding into the adjacent fat. Extramural invasion may be present when there is loss of fat planes between the colon and adjacent muscles such as the levator ani, piriform, obturator, and gluteal muscles. Unfortunately, the loss of fat planes is a nonspecific sign in cross-sectional imaging and can also be seen with cachexia and lymphatic or vascular obstruction. Therefore, the diagnosis of

extramural tumor extension should be made with caution. Additionally, patients with microinvasion through the bowel often demonstrate normal-appearing pericolonic and perirectal fat on CT, leading to understaging of the tumor.

CT detection of malignant lymphadenopathy has traditionally been based on size criteria. Lymph nodes greater than 1 cm in size are considered abnormal, except in the perirectal fat, where any lymphadenopathy is considered pathologic. Unfortunately, size criteria are based only on statistical probability. In reality, many nodes smaller than 1 cm are malignant and nodes larger than 1 cm are caused by reaction to a number of benign inflammatory conditions. As a result, true differentiation of benign from malignant lymphadenopathy can be achieved only with biopsy.

Gross tumor invasion into the adjacent pelvic musculature is manifested by muscle enlargement and occasionally an enhancing intramuscular mass. Bladder invasion with rectovesical fistula should be considered in patients with gas in the bladder and no history of Foley catheterization.

CT virtual colonography (Fig. 31) takes the detection of CRC and polyps one step further. In this technique, CT data are displayed with a two- and three-dimensional format. Viewing these images at a rate of 15–30 per second creates the illusion of viewing the colon endolumenally. The bowel preparation for cross-sectional colonography is similar to that for colonoscopy and barium enema. Colonic distension prior to virtual colonography is accomplished with CO_2 or room air. To minimize motion artifacts, data are obtained during a single helical CT acquisition. The sensitivity for detection of polyps greater than 1 cm in size has been reported to be as high as 75% with a specificity greater than 90%. Virtual colonography also offers the advantage of allowing visualization of the colon proximal to a distal obstructing tumor, an area that is inaccessible endoscopically. This is particularly valuable, because the rate of a second synchronous CRC is between 1.5 and 9.0% and the rate of multiple coexistent adenomatous polyps is between 27 and 55%. This advantage suggests that virtual colonography will likely play a larger role in the radiographic staging of colon carcinoma.

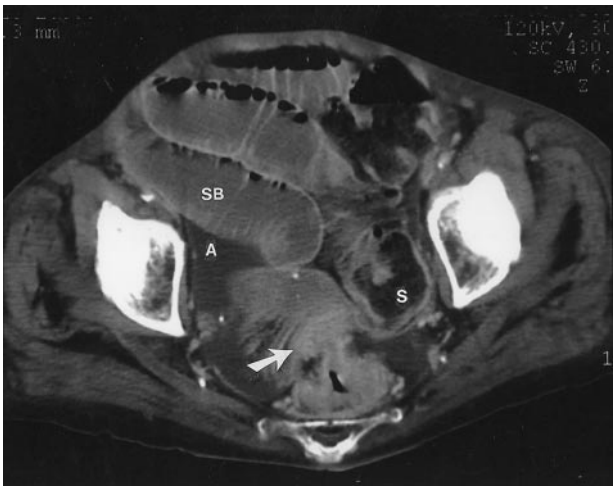


FIGURE 30 Annular constricting rectal cancer (arrow) causing large and small bowel obstruction. SB, dilated fluid-filled small bowel; S, dilated sigmoid colon; A, ascites.

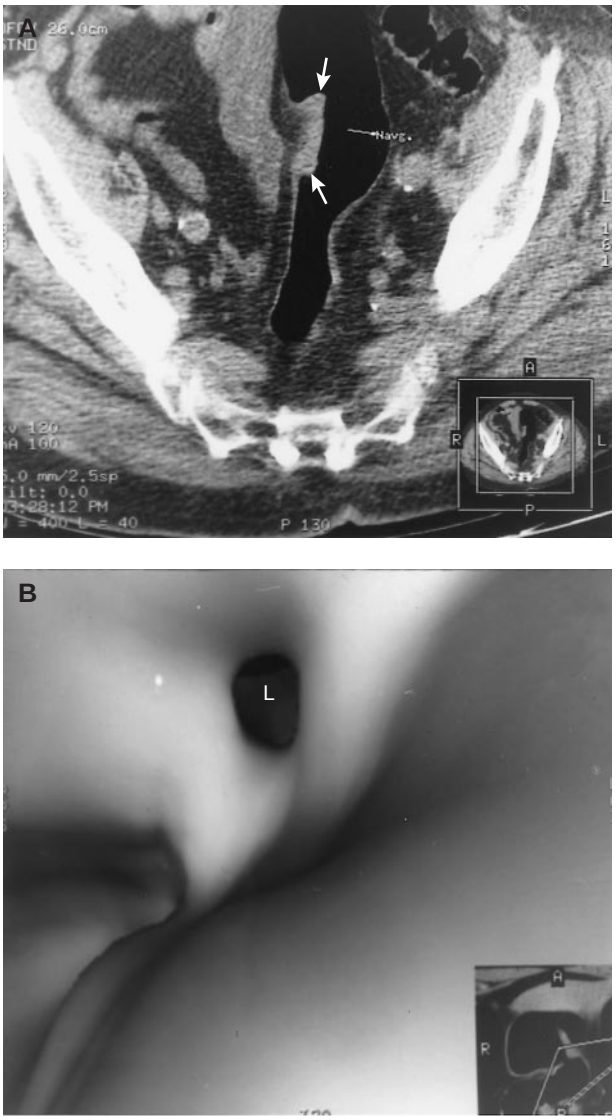


FIGURE 31 Flat distal sigmoid cancer. (A) Tumor (arrows) visualized on conventional axial image. (B) Lumen (L) narrowing as depicted on 3D virtual colonographic view.

Hepatic metastases from CRC on CT most commonly appear as low-attenuation lesions on noncontrast images. Surgical resection of hepatic metastasis is attempted if only one lobe of the liver is involved with less than four lesions, the remaining hepatic lobe has a normal appearance and function, and no focal adenopathy or distant spread of tumor is present. Patients undergoing hepatic lobectomy for metastatic disease should also have intraoperative ultrasound, as this is the most sensitive imaging technique for detection of hepatic metastases.

Peritoneal carcinomatosis, caused by spread of tumor over peritoneal surfaces, is usually well demonstrated by CT and appears as linear or nodular peritoneal thickening that enhances with contrast. Adrenal metastases on CT appear as low-density masses in the adrenal gland and osseous metastases on CT are most commonly lytic and associated with cortical disruption.

Overall sensitivity rates in the detection of recurrent colon cancer are similar with CT and MRI. For detecting the recurrence of rectal tumors, TRUS and MRI appear to have better detection rates. Despite this, however, the radiographic evaluation for recurrent tumor is most commonly performed with CT. With easy access to CT scanners, overall sensitivity rates of 69–95%, and relative speed in completing the procedure, CT will likely remain the imaging modality of choice for detecting tumor recurrence in the near future. Limitations of CT for detection of postoperative recurrence are similar to those of preoperative studies and primarily consist of difficulty in distinguishing malignant adenopathy and microinvasive spread of tumor into pericolonic and perirectal fat. Detection of tumor recurrence is further complicated by several benign entities that can simulate recurrence on cross-sectional imaging studies. The most important of these are postradiation changes, including radiation colitis and radiation fibrosis. Additionally, patients receiving chemotherapy are often more susceptible to inflammatory conditions, which can mimic tumor recurrence as well.

See Also the Following Articles

Barium Radiography • Magnetic Resonance Imaging (MRI) • Picture Archiving and Communication Systems (PACS) • Radiology, Interventional • Ultrasonography • Virtual Colonoscopy

Further Reading

Gore, R. M., and Levine, M. S. (2002). "Textbook of Gastrointestinal Radiology," 2nd Ed. W. B. Saunders, Philadelphia, PA.

Jacobs, J. (2001). Acute conditions of the abdomen and pelvis. *Semin. Roentgenol.* **36**, 81–175.

Levine, M. S. (2000). Colorectal cancer: Diagnosis, screening and staging. *Semin. Roentgenol.* **35**, 325–408.

Martinez-Noguera, A. (2002). Cirrhosis and chronic hepatitis. *Semin. Ultrasound CT MR* **23**, 3–140.

Mindelzun, R. E. (2000). Unenhanced CT in the evaluation of acute abdominal pain. *Semin. Ultrasound CT MR* **20**, 61–147.

Raymond, H. W., Zwiebel, W. J., and Swartz, J. D. (2000). Acute abdominal imaging. *Semin. Ultrasound CT MR* **21**, 2–93.

Constipation

ARNOLD WALD
University of Pittsburgh Medical Center

colonic transit Measurement of the time it takes for inert markers to pass through the colon and anorectum, traditionally measured while the subject consumes a high-fiber diet and uses no laxatives.

dietary fiber That portion of plant food that escapes digestion and which is composed of either insoluble or soluble components. Insoluble fiber retains water within the cellular structures whereas soluble fiber stimulates growth of colonic bacteria; both actions increase fecal mass.

laxative Substance that acts to promote the passage of stools modestly and without violent action.

pelvic floor dyssynergia Condition in which normal defecation is impeded by the unconscious contraction of the puborectalis and external anal sphincter muscles.

In the United States, there are at least 2.5 million visits to medical facilities per year for constipation, mostly to physicians in general and family practice. Eighty-five percent of patients receive at least one prescription, with laxatives being the most frequently prescribed. For both sexes, visits increase with increasing age but there is a higher frequency of visits for women. Over $800 million was spent in the United States for laxatives in 1994. Additional costs for diagnostic testing and surgical procedures for constipation are unknown. For a small but clinically important subset of individuals, constipation appears to be refractory to conventional treatment.

INTRODUCTION

The definition of constipation has often been based on a stool frequency of less than three per week. Yet only 6% of individuals who self-report constipation have infrequent bowel movements. Constipated individuals more frequently complain of defecatory straining or a sense of incomplete defecation. An international working committee has recommended that the definition of constipation should include either two or more of the following complaints that are present for at least 3 months (not necessarily consecutive), in the previous 12 months:

1. Straining with at least 25% of bowel movements.
2. Sensation of incomplete evacuation after at least 25% of bowel movements.
3. Hard or lumpy stools with at least 25% of bowel movements following laxatives.
4. Stools less frequent than three per week.
5. Sensation of anorectal obstruction on at least 25% of defecations.
6. Manual maneuvers to facilitate defecation on at least 25% of attempts.

There are two important implications of this definition: infrequent defecation alone is insufficient for the diagnosis of chronic constipation and defecatory difficulty in the absence of infrequent defecation is sufficient for and comprises the majority of patients who complain of chronic constipation.

CAUSES

Constipation is a heterogeneous disorder that often is hypothesized to be associated with disordered movement through the large intestine, the anorectum, or both. Constipation has been associated with a large number of diseases or as a side effect of many drugs (Tables I and II). Diseases may produce constipation by having a direct effect on gastrointestinal function or may result in physical or mental impairments that produce or exacerbate constipation. An example of the latter is physical immobility, which can lead to fecal retention. The most devastating disorders are neurologic diseases, which may have direct effects on the large intestine, produce generalized weakness or voluntary muscle dysfunction, and lead to inanition or physical and emotional impairments.

TABLE I Some Diseases Associated with Chronic Constipation

Neurogenic disorders	Nonneurogenic disorders
Autonomic neuropathy	Hypothyroidism
Diabetes mellitus	Hypercalcemia
Hirschsprung's disease	Systemic sclerosis
Intestinal pseudo-obstruction	
Multiple sclerosis	
Parkinson's disease	
Spinal cord injury	

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Some Drugs Associated with Constipation

Anticholinergics	Diuretics
Anticonvulsants	5-HT3 antagonists
Antihypertensives	Granisetron
Anti-Parkinson drugs	Ondansetron
Calcium channel blockers	Opiates
Cation-containing agents	Tricyclic antidepressants
Aluminum (antacids, sucralfate)	
Bismuth	
Iron supplements	

PATHOPHYSIOLOGY

Most constipated patients have no obvious cause to explain their symptoms. Contrary to a commonly held belief, there is no established relationship to decreased dietary fiber or fluid intake, although fiber supplements are often effective when treating some individuals with constipation. Several studies do suggest that colonic transit is affected by decreased calorie intake.

One approach to assessing idiopathic constipation is to classify patients according to patterns of colonic transit.

1. Normal colonic transit: Patients with infrequent defecation and normal transit constipation may misperceive bowel frequency and often exhibit increased psychosocial distress. Alternatively, they complain about difficult or unsatisfactory defecation, which is not identified using colonic transit assessment techniques. The patients experiencing difficult or unsatisfactory defecation may have disorders of anorectal function or outlet dysfunction, but identification of such abnormalities is not universal in these patients.
2. Slow colonic transit: Slow colonic transit can result from decreased propulsion (hypomotility) or may be associated with increased distal motility (hypermotility) and retropulsion of markers. The term "colonic inertia" should be reserved for cases in which transit in the proximal colon is delayed without evidence of outlet delay.

In most patients with colonic inertia, there is little or no increase of colonic motor activity after meals or following administration of bisacodyl and cholinergic agents. Such findings suggest abnormalities of the enteric nerve plexus; this is supported by the recent demonstration of decreased interstitial cells of Cajal, which serve as intestinal pacemakers in the colon of patients with colonic inertia. Other studies have shown decreased numbers of serotonin-containing cells. In some patients, colonic slowing may be secondary to anorectal dysfunction, and colonic motor activity is normal.

Another form of slow colonic transit is thought to occur in some patients with irritable bowel syndrome, in which nonpropulsive segmenting contractions are more pronounced and result in retarded movement of colon contents though the left colon.

Outlet Dysfunction

Several anorectal disorders may contribute to disordered defecation. These include megarectum, failure of internal anal sphincter relaxation associated with congenital aganglionosis (Hirschsprung's disease), some large rectoceles in which there is misdirection of stool into a pouch (associated with weakness of the rectovaginal septum), and pelvic floor dyssynergia, in which ineffective defecation is associated with failure to relax, or inappropriate contraction of, the puborectalis and external anal sphincter muscles.

A diagnosis of outlet dysfunction should be considered when there is impaired expulsion of a water-filled balloon from the rectum (patients are asked to attempt this in privacy). The diagnosis may then be confirmed by at least two of the following studies: anorectal manometry, anal sphincter electromyelogram (EMG), and defecography. If studies are not diagnostic, other causes for ineffective expulsion should be considered.

TREATMENT OF UNCOMPLICATED CONSTIPATION

The initial management of functional chronic constipation includes patient education, behavior modification, dietary changes, and laxatives or enemas. Patient education should include reassurance and an explanation as to normal bowel habits. Efforts are made to reduce excessive use of laxatives and cathartics, increase fluid and fiber intake, and utilize postprandial increases in colonic motility by instructing patients to attempt to defecate after certain meals. Dietary fiber and bulk laxatives such as psyllium or methylcellulose are the most physiologic approaches. Cereal fiber cell walls generally resist digestion and retain water within their cellular structures, whereas citrus fruit and legume fibers stimulate the growth of colonic flora, which increases fecal mass. Wheat bran is one of the most effective fiber laxatives, and there is a clear dose response to fiber with respect to fecal output. Particle size appears to be crucially important and the large particle size of cereal products appears to enhance fecal bulking effects.

TABLE III Types of Laxatives and Their Effects

Softening of formed stools (1–3 days)	Soft–semifluid stools (6–12 hours)
Bulk agent	**Saline laxatives (low dose)**
Dietary fiber	Milk of magnesia
Psyllium	Magnesium sulfate
Methylcellulose	**Diphenylmethanes**
Calcium polycarbophil	Bisacodyl (oral)
Nonabsorbed substances	**Anthraquinones**
Polyethylene glycol[a]	Senna (oral)
Lactulose[a]	Cascara sagrada
Sorbitol[a]	Aloe

[a] Available by prescription only.

Those patients who respond poorly or who do not tolerate fiber may require other laxatives. In general, bulk agents, nonabsorbed substances, and mineral oil are used daily in varying doses, whereas saline laxatives, diphenylmethanes, and anthraquinones are used no more than two to three times weekly (Table III).

OTHER TREATMENTS OF CONSTIPATION

Behavioral Approaches

Habit training has been successfully employed in children with severe constipation. A modified program may also be effective in many adults with neurogenic constipation, dementia, or physical impairment. Initially, patients should be disimpacted if necessary and the colon evacuated. This can be accomplished with enemas followed by drinking a solution containing polyethylene glycol until cleansing is complete. After bowel cleansing, osmotic laxatives are given to produce one stool at least every other day. The patient is instructed to use the bathroom after eating breakfast to take advantage of meal-stimulated increases in colonic motility. An enema or bisacodyl suppository is administered if there is no defecation after 2 days to prevent recurrence of fecal impaction.

A modified program may be used in demented or bedridden patients with fecal impactions. After disimpaction and bowel cleansing with enemas or polyethylene glycol-containing solutions, a fiber-restricted diet together with cleansing enemas once or twice per week will assist nursing management by decreasing buildup of stool and recurrence of fecal impaction.

Another behavioral approach is biofeedback. This is used to correct inappropriate contraction of the pelvic floor muscles and the external anal sphincter during defecation in patients with pelvic floor dyssynergia.

Pharmacologic Approaches

Misoprostol, a prostaglandin analogue, has been used with some success in patients with severe constipation but is not a good choice in young women because of its abortifacient properties. Oral colchicine has been reported to be effective in several small studies. Prokinetics such as metoclopramide and erythromycin have little effect on colonic motility. There is enthusiasm for serotonin (or 5-hydroxytryptamine) receptor agonists, and these are being tested.

5-Hydroxytryptamine (5-HT) is found throughout the body, but a majority of the receptors are located along the gastrointestinal tract. The 5-HT subtype 4 (5-HT4) agonists selectively stimulate colonic propulsive activity and increase stool frequency. A selective 5-HT4 agonist is currently being tested after having been shown to accelerate large bowel transit and improve symptoms associated with constipation-predominant irritable bowel syndrome (IBS).

Surgical Approaches

Surgery to treat constipation remains controversial. Subtotal colectomy with ileorectal anastomosis can dramatically ameliorate incapacitating constipation in carefully selected patients. If surgery is undertaken, at least three criteria should be met: (1) the patient has chronic, severe, and disabling symptoms from constipation that are unresponsive to medical therapy; (2) the patient has slow colonic transit of the inertia pattern; and (3) the patient does not have intestinal or anorectal dysmotility, as demonstrated by radiologic or manometric studies. Many studies have documented that greater than 90% of carefully studied patients have satisfactory long-term results. However, this operation is associated with significant morbidity and some mortality. A common cause of morbidity is small bowel obstruction, and many patients require laparotomy at some stage. Some patients appear to develop intestinal pseudo-obstruction despite normal preoperative studies. Finally, abdominal pain and bloating are a problem in many patients who have been followed for long periods of time. The selection of that subset of constipated patients who benefit greatly from this surgery remains a key issue. The presence of abdominal pain that is not relieved with bowel cleansing should lead to extreme caution in operating on such patients.

See Also the Following Articles

Colonic Motility • Defecation • Dietary Fiber • Laxatives

Further Reading

Bharucha, A. E., and Phillips, S. F. (2001). Slow transit constipation *Gastroenterol. Clin. North Am.* **30**, 77–95.

Knowles, C. H., Scott, M., and Lumiss, P. J. (1999). Outcome of colectomy for slow transit constipation. *Ann. Surg.* **230**, 627–638.

Prather, C. M., Camilleri, M., and Zinsmeister, A. R. (2000). Tegaserod accelerates orocecal transit in patients with constipation predominant irritable bowel syndrome. *Gastroenterology* **118**, 463–468.

Sonnenberg, A., and Koch, T. R. (1989). Physician visits in the United States for constipation. *Dig. Dis. Sci.* **34**, 606–611.

Tramonte, S. M., Brand, M. B., and Mulrow, C. D. (1997). The treatment of chronic constipation in adults: A systematic review. *J. Gen. Int. Med.* **12**, 15–24.

Wald, A. (1999). Approach to the patient with constipation. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., pp. 910–925. Lippincott Williams & Wilkins, Philadelphia.

Wald, A. (2000). Constipation. *Med. Clin. North Am.* **84**, 1231–1246.

Wald, A. (2001). Outlet dysfunction constipation. *Curr. Treat. Options Gastroenterol* **4**, 293–297.

Wofford, S. A., and Verne, G. N. (2000). Approach to patients with refractory constipation. *Curr. Gastroenterol. Rep.* **2**, 389–394.

Constipation, Pediatric

MARC A. BENNINGA
Emma Children's Hospital, Amsterdam

constipation In infants and children, constipation can be defined by a stool frequency of less than 3 per week, but the passage of painful, hard or lumpy stools and stool retention are symptoms of constipation even when the defecation frequency is 3 or more per week.

encopresis The voluntary or involuntary passage of a normal bowel movement in the underwear (or other unorthodox locations), after the age of 4 years, occurring on a regular basis without any organic cause.

soiling The involuntary seepage of feces that is frequently associated with fecal impaction and which reflects staining of the underwear.

Many caretakers regard constipation and encopresis as trivial symptoms, which will eventually disappear. However, constipation in childhood is not always simple to treat and often requires prolonged follow-up after a comprehensive explanation and demystification of its pathophysiology and appropriate medical treatment. Long-term follow-up studies show that more than 40% of the children have persistent complaints even after 5 years of intensive medical and behavioral treatment. Moreover, in approximately 30% of the patients, childhood constipation proceeds into adulthood constipation. Importantly, apart from shame and fear of discovery, it often leads to social withdrawal, low self-esteem, and depression. The etiology of chronic constipation remains poorly understood although our understanding has increased in all areas relevant to this condition such as epidemiology, clinical classification, psychological disturbance, changes in motility, and abnormalities in enteric neurochemistry.

DEFINITION

One of the problems in correctly defining constipation is that constipation is a symptom rather than a disease. Furthermore, constipation is often differently interpreted by patients and their caretakers and physicians. Physicians often define constipation on the basis of defecation frequency, whereas healthy subjects and patients define constipation in terms of function (straining) and consistency (hard stools). Recently, a group of experts in the field of pediatric gastroenterology categorized defecation disorders using symptom-based diagnostic criteria. These criteria have provided clinicians with a method for standardizing their manner of defining clinical disorders and have allowed researchers from various fields to study the physiology and treatment of the same disorders. Disorders of defecation in children include infant dyschezia, functional constipation, and functional retention (Table I).

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Bharucha, A. E., and Phillips, S. F. (2001). Slow transit constipation *Gastroenterol. Clin. North Am.* **30**, 77–95.

Knowles, C. H., Scott, M., and Lumiss, P. J. (1999). Outcome of colectomy for slow transit constipation. *Ann. Surg.* **230**, 627–638.

Prather, C. M., Camilleri, M., and Zinsmeister, A. R. (2000). Tegaserod accelerates orocecal transit in patients with constipation predominant irritable bowel syndrome. *Gastroenterology* **118**, 463–468.

Sonnenberg, A., and Koch, T. R. (1989). Physician visits in the United States for constipation. *Dig. Dis. Sci.* **34**, 606–611.

Tramonte, S. M., Brand, M. B., and Mulrow, C. D. (1997). The treatment of chronic constipation in adults: A systematic review. *J. Gen. Int. Med.* **12**, 15–24.

Wald, A. (1999). Approach to the patient with constipation. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., pp. 910–925. Lippincott Williams & Wilkins, Philadelphia.

Wald, A. (2000). Constipation. *Med. Clin. North Am.* **84**, 1231–1246.

Wald, A. (2001). Outlet dysfunction constipation. *Curr. Treat. Options Gastroenterol* **4**, 293–297.

Wofford, S. A., and Verne, G. N. (2000). Approach to patients with refractory constipation. *Curr. Gastroenterol. Rep.* **2**, 389–394.

Constipation, Pediatric

MARC A. BENNINGA
Emma Children's Hospital, Amsterdam

constipation In infants and children, constipation can be defined by a stool frequency of less than 3 per week, but the passage of painful, hard or lumpy stools and stool retention are symptoms of constipation even when the defecation frequency is 3 or more per week.

encopresis The voluntary or involuntary passage of a normal bowel movement in the underwear (or other unorthodox locations), after the age of 4 years, occurring on a regular basis without any organic cause.

soiling The involuntary seepage of feces that is frequently associated with fecal impaction and which reflects staining of the underwear.

Many caretakers regard constipation and encopresis as trivial symptoms, which will eventually disappear. However, constipation in childhood is not always simple to treat and often requires prolonged follow-up after a comprehensive explanation and demystification of its pathophysiology and appropriate medical treatment. Long-term follow-up studies show that more than 40% of the children have persistent complaints even after 5 years of intensive medical and behavioral treatment. Moreover, in approximately 30% of the patients, childhood constipation proceeds into adulthood constipation. Importantly, apart from shame and fear of discovery, it often leads to social withdrawal, low self-esteem, and depression. The etiology of chronic constipation remains poorly understood although our understanding has increased in all areas relevant to this condition such as epidemiology, clinical classification, psychological disturbance, changes in motility, and abnormalities in enteric neurochemistry.

DEFINITION

One of the problems in correctly defining constipation is that constipation is a symptom rather than a disease. Furthermore, constipation is often differently interpreted by patients and their caretakers and physicians. Physicians often define constipation on the basis of defecation frequency, whereas healthy subjects and patients define constipation in terms of function (straining) and consistency (hard stools). Recently, a group of experts in the field of pediatric gastroenterology categorized defecation disorders using symptom-based diagnostic criteria. These criteria have provided clinicians with a method for standardizing their manner of defining clinical disorders and have allowed researchers from various fields to study the physiology and treatment of the same disorders. Disorders of defecation in children include infant dyschezia, functional constipation, and functional retention (Table I).

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Childhood Functional Defecation Disorders: ROME II—Criteria

	Diagnostic criteria
Infant dyschezia	At least 10 min of straining and crying before successful passage of soft stools in an otherwise healthy child
Functional constipation	In infants and preschool children at least 2 weeks of: 1. Scybalous, pebble-like, hard stools for a majority of stools; or 2. Firm stools $\leq$2 times/week; and 3. No evidence of structural, endocrine, or metabolic disease
Functional fecal retention	From infancy to 16 years old, a history of at least 12 weeks of: 1. Passage of large-diameter stools at intervals $<$2 times/week 2. Retentive poisturing, avoiding defecation by contracting pelvic floor and gluteal muscles

Adapted from Rasquin-Weber, *et al.* (1999).

EPIDEMIOLOGY

Constipation represents the chief complaint in 3% of pediatric outpatient visits and 10 to 25% of pediatric gastroenterology visits in the United States. Constipation is reported in 26 to 74% of children with cerebral palsy. In the Netherlands, 1% of children ages 0–4 years, but hardly any in the range 4–15 years, visit the general practitioner for complaints of constipation. American and British parents reported constipation in 16 and 34% of toddlers, respectively. In contrast to constipation in adults, constipation in childhood seems to be more common in boys than in girls (2 : 1), although a 1 : 1 ratio is also described.

PHYSIOLOGY

Defecation is elicited by the presence of fecal material in the rectum due to peristaltic propagation of colonic motility. Consequently, sensory stimuli in the anal canal provoke a sudden drop in the tone of the internal anal sphincter. By voluntary control, defecation starts with relaxation of the puborectalis and the levator. The distension of the rectum evokes a wave of contractions of the rectum and defecation can be completed by a voluntary increase in intra-abdominal pressure. The act of defecation depends on maturational control, a process that can be trained when the child slowly discovers the ability to control the pelvic muscles. First, the child feels the satisfaction and gains approval from parents by using the potty chair. Later, toileting changes to something that is done for self-approval and comfort. Before the age of 4 years, most children acquire autonomy, and defecation becomes a private, hardly thought about activity.

Normally, a decline in stool frequency from more than four stools per day during the first week of life to one to two per day at 4 years of age is observed, with a corresponding increase in stool size. Approximately 97% of 1- to 4-year-old children pass stool three times daily to once every other day.

CLINICAL PRESENTATION

The most important complaint of constipated children is a combination of a low defecation frequency and the involuntary loss of feces (Table II). Soiling often happens several times a day, and in the case of severe constipation and fecal retention, it occurs even at night. Often, once a week to once a month, a very large amount of stool is produced, which may clog the toilet. Preceding this often painful and difficult evacuation of feces, soiling frequency increases and the children complain of

TABLE II Common Clinical Presentation of Constipation

Feature	Percentage (%)
Soiling/encopresis	75–90
Defecation frequency $<$3/week	75
Large stools	75
Straining during defecation	35
Pain during defecation	50–80
Retentive posturing	35–45
Abdominal pain	10–70
Abdominal distension	20–40
Anorexia	10–25
Vomiting	10
Poor appetite	25
Enuresis/urinary tract infection	30
"Psychological problems"	20
Physical examination	
Abdominal mass	30–50
Anal prolapse	3
Fissures/hemorrhoids	5–25
Fecal impaction	40–100

abdominal pain and poor appetite. These symptoms disappear immediately after the production of this large amount of stool. In approximately 30% of constipated children, there are complaints of urinary tract infections and enuresis. Most constipated children do have abdominal and/or rectal fecal impaction upon physical examination.

PATHOPHYSIOLOGY

In 90 to 95% of constipated children, no obvious cause can be identified. In some babies, an acute episode of constipation may occur associated with a change in diet (i.e., changing from human milk to cow's milk). This results in the passage of dry hard stools, sometimes with traces of fresh blood, often the consequence of anal fissures. The pain caused by the anal fissures leads to understandable motivation of the child to avoid defecation and induces the persistence of this cycle of problems. In "infant dyschezia," it is speculated that neonates fail to coordinate increased intra-abdominal pressure with relaxation of the pelvic floor. Parents need to be reassured that this condition is part of the child's learning process for which no intervention is indicated.

"Retentive posturing" is a hallmark of the development and/or persistence of constipation in children. When the child experiences an urge to defecate, he or she assumes an erect posture and holds the legs stiffly together to forcefully contract the pelvic and gluteal muscles. Consequently, the rectum accommodates to its content and the urge to defecate disappears. The retained stools become progressively more difficult to evacuate, leading to a vicious cycle in which the rectum is increasingly distended by abnormally firm and large fecal contents. Finally, chronic rectal distention may cause overflow soiling, loss of rectal sensitivity, and eventually loss of the urge to defecate. This aberrant behavior might lead to the unconscious contraction of the external sphincter during defecation, better known as anal sphincter dyssynergy. More than 50% of children with fecal retention show this abnormal defecation pattern. For a long time, this paradoxical contraction of the anal sphincter complex was thought to be the major pathophysiological mechanism in childhood constipation. However, normalization of this pattern with biofeedback training did not correlate with successful treatment.

In a minority of patients, a deficiency in dietary fiber and/or insufficient fluid intake may also cause constipation.

There are rare organic causes of constipation such as anatomical, neurological, endocrine, and metabolic disorders (Table III). Hirschsprung's disease is characterized by the absence of ganglion cells from the distal rectum to a variable length up to the duodenum and must be considered in any child of any age with severe constipation.

DIAGNOSTIC TOOLS

A careful medical history together with a thorough physical examination is the cornerstone of the diagnostic work-up of children with constipation. Adjunctive procedures are available, such as an abdominal X ray, barium enema, anorectal and colonic manometry, rectal biopsies, and magnetic resonance imaging of the spine. However, the value of these tests is limited and often not necessary to make a correct diagnosis. They should be reserved for appropriate clinical indications.

The medical history of children with constipation should include questions about the age of onset of bowel problems, stool frequency, consistency and size of stools, painful defecation, and retentive posturing. Furthermore, information about the encopresis frequency, the amount of feces lost in the underwear, the time of occurrence (day and/or night), and the situation in which encopresis occurs (while sitting at the computer, during play outside) is of major importance. Other questions should focus on abdominal pain, loss of appetite, urinary tract problems, neuromuscular development, and psychological or behavioral problems. Finally, medical treatment and previous treatment strategies in relationship to the defecation problems should be discussed. In addition, it is essential to ask for possible contributing events, such as a death in the family, the birth of a sibling, school problems, and sexual abuse.

After the history is obtained, a complete physical (and neurological) examination should be performed. Abdominal examination gives information concerning the accumulation of gas or feces. Perianal inspection provides information about the position of the anus, perianal feces, erythema, fissures, hemorrhoids, and scars (sexual abuse). Anorectal digital examination is essential to define sphincter pressure, the amount and consistency of stool in the rectum, and the voluntary contraction and relaxation of the anal sphincter. Neurological exam may suggest problems that were previously unappreciated (such as tethered cord).

TREATMENT

Despite the high prevalence of childhood constipation, large randomized controlled therapeutic trials are lacking. The treatment of constipation is therefore mainly based on experience and a small number of clinical trials. The treatment of constipation in childhood is

TABLE III Conditions Associated with Childhood Constipation

Anal stenosis
Anorectal malformation (imperforate anus, anteriorly located anus)
Aganglionosis and/or abnormal myenteric plexus
1. Hirschsprung's disease
2. Acquired Chagas' disease
3. Chronic intestinal pseudo-obstruction
Spina bifida
Meningomyelocele
Spinal cord tumor
Hypothyroidism
Hypercalcemia
Diabetes mellitus? diabetes insipidus
Infantile renal acidosis
Anesthetics
Antidepressant agents
Hemantinics (iron)
Opiates
Vincristine

based on four important phases: (1) education, (2) disimpaction, (3) prevention of reaccummulation of feces, and (4) follow-up.

It is important to treat constipation early in childhood. Large studies in the United States and Great Britain showed that constipated children <2 years of age responded better to treatment than children >2 years of age. Early treatment is important to prevent development of severe constipation, fecal soiling, or both.

The first essential in treatment is to gain the family's confidence and explain the mechanism of the symptoms. It is crucial for both the child and the parent that the physician alleviate guilt, make the child and parents comfortable talking about the loss of feces, explain that soiling is a consequence of the full rectum, and decrease feelings of shame. A positive nonaccusatory approach is almost invariably accepted with relief by the child and parents. Furthermore, it should be explained that the treatment is often of long duration and marked by periods of improvement alternating with periods of deterioration.

Treatment of severe fecal impaction in the rectum should begin by administering enemas for 3 consecutive days. Once disimpaction has been achieved, it is essential to begin an oral daily laxative immediately and continue this treatment for months, or longer if necessary, to prevent reaccumulation of retained stools and overcome stool withholding if present. The correct dose is that which produces daily soft stool without side effects. In the United States, milk of magnesia is often used as oral laxative, whereas in Europe, the most common laxative drug prescribed is lactulose. Recently, good results using polyethylene glycol have been obtained in children with constipation. However, the success of this compound in these children was 50 to 60%, which was not significantly better than that obtained with older laxatives.

A combination of oral laxatives and regular toilet training (sitting on the toilet for 5 to 10 min within 30 min after each meal) combined with parental praise and reward is more successful than oral laxatives alone. Moreover, long-term success is higher in children who are able to maintain regular toileting.

The role of biofeedback training in constipated children with paradoxical sphincter contraction and/or diminished rectal sensation is limited. Two large randomized trials in such children showed that biofeedback training could change the abnormal sphincter pattern but was not more effective than laxative treatment.

Psychological referral is indicated in children who fail intensive medical treatment and in those with severe emotional problems or serious family problems. In these patients, a combination of psychiatric and pediatric treatment strategies is preferred.

Frequent follow-up in these patients is of major importance since relapse of symptoms occurs in 50% of the children. After 3 months, the use of laxatives can be reduced, provided that a normal bowel habit is maintained. Approximately 50% of children with chronic constipation require treatment and close follow-up for at least 6 to 12 months.

See Also the Following Articles

Colonic Motility • Defecation • Dietary Fiber • Laxatives

Further Reading

Di Lorenzo, C. (1994). Constipation. *In* "Pediatric Gastrointestinal Motility Disorders" (P. E. Hyman, ed.), 1st Ed., pp. 129–143. Academy Professional Information Services, New York.

Loening-Baucke, V. (1993). Chronic constipation in children. *Gastroenterology* **105**, 1557–1564.

Loening-Baucke, V. (1993). Constipation in early childhood: Patient characteristics, treatment, and long-term follow up. *Gut* **34**, 1400–1404.

Loening-Baucke, V. (1995). Biofeedback treatment for chronic constipation and encopresis in childhood: Long-term outcome. *Pediatrics* **96**, 105–111.

Loening-Baucke, V. (2002). Polyethylene glycol without electrolytes for children with constipation and encopresis. *J. Pediatr. Gastroenterol. Nutr.* **34**, 372–377.

Nolan, T., Debelle, G., Oberklaid, F., and Coffey, C. (1991). Randomised trial of laxatives in treatment of childhood encopresis. *Lancet* **338**, 523–527.

Rasquin-Weber, A., Hyman, P. E., Cucchiara, S., Fleisher, D. R., Hyams, J. S., Milla, P. J., and Staiano, A. (1999). Childhood functional gastrointestinal disorders. *Gut* **45**(Suppl. 2), 60–68.

Staiano, A., Andreotti, M. R., Greco, L., Basile, P., and Auricchio, S. (1994). Long-term follow-up of children with chronic idiopathic constipation. *Digest. Dis. Sci.* **39**, 561–564.

van der Plas, R. N., Benninga, M. A., Büller, H. A., Bossuyt, P. M., Akkermans, L. M., Redekop, W. K., and Taminiau, J. A. (1996). Biofeedback training in treatment of childhood constipation: A randomised controlled study. *Lancet* **348**, 776–780.

van Ginkel, R., Büller, H. A., Boeckxstaens, G. E., van der Plas, R. N., Taminiau, J. A., and Benninga, M. A. (2001). The effect of anorectal manometry on the outcome of treatment in severe childhood constipation: A randomized, controlled trial. *Pediatrics* **108**, E9–E16.

Cow Milk Protein Allergy

GLENN T. FURUTA

Children's Hospital Boston and Harvard Medical University

Cow milk, egg, peanut, soy, wheat, and fish antigens account for 90% of food allergic reactions. This article will focus on clinical syndromes associated with allergic reactions to cow milk proteins. Cow milk is not only a very common food product but also a potent allergen that has the capacity to induce a wide variety of different clinical reactions that affect both children and adults. In addition, the severity of the reactions ranges from systemic anaphylaxis to benign milk protein allergy of infancy; this diverse spectrum of clinical syndromes suggests different pathogenetic pathways for a common antigen.

INTRODUCTION

A wide spectrum of immunoglobulin E (IgE)-mediated and non-IgE-mediated food responses exist. For instance, the classical food allergic reaction to peanuts offers the best example of an IgE-mediated hypersensitivity reaction. Upon ingestion of or contact with peanuts, patients can develop the rapid onset of anaphylactic symptoms. Further evaluation shows elevated radioimmunosorbent test (RAST) for peanut IgE and abnormal skin prick testing. In contrast, many patients with food allergic responses, characterized by delayed reactions and intestinal mucosal pathology, often do not have abnormal RAST or positive skin testing for specific antigens. For instance, patients with cow milk protein colitis of infancy develop rectal bleeding and often no elevation of RAST for milk; further evaluation reveals eosinophilia of the rectosigmoid mucosa. A new era is dawning in the field of food allergic reactions as clinicians and scientists begin to define the clinical and pathophysiological features of the ever-broadening spectrum of food allergic reactions.

DEFINITIONS

Food Intolerance and Food Allergic/ Hypersensitivity Responses

Adverse food reactions refer to undesirable physical responses to an ingested food product. These reactions can be divided into two subcategories, food intolerances and food hypersensitivity/allergic reactions. Food intolerances are reflective of digestive deficiencies (i.e., lactose intolerance), toxins (i.e., bacterial toxins derived from infected foods such as from *Shigella* or *Campylobacter*), or biochemical properties of the food (i.e., monosodium glutamate, caffeine). Three features characterize food allergic/hypersensitivity reactions. First, they are reproducible reactions to a specific food product. Second, and often the more difficult part of the immunological response should be characterized in the diagnosis, clinic setting or with laboratory analysis. For instance, currently the availability of accurate tests is limited to RAST, skin prick testing, and double-blind placebo control food challenge. In

 Copyright 2004, Elsevier (USA). All rights reserved.

Rasquin-Weber, A., Hyman, P. E., Cucchiara, S., Fleisher, D. R., Hyams, J. S., Milla, P. J., and Staiano, A. (1999). Childhood functional gastrointestinal disorders. *Gut* **45**(Suppl. 2), 60–68.

Staiano, A., Andreotti, M. R., Greco, L., Basile, P., and Auricchio, S. (1994). Long-term follow-up of children with chronic idiopathic constipation. *Digest. Dis. Sci.* **39**, 561–564.

van der Plas, R. N., Benninga, M. A., Büller, H. A., Bossuyt, P. M., Akkermans, L. M., Redekop, W. K., and Taminiau, J. A. (1996). Biofeedback training in treatment of childhood constipation: A randomised controlled study. *Lancet* **348**, 776–780.

van Ginkel, R., Büller, H. A., Boeckxstaens, G. E., van der Plas, R. N., Taminiau, J. A., and Benninga, M. A. (2001). The effect of anorectal manometry on the outcome of treatment in severe childhood constipation: A randomized, controlled trial. *Pediatrics* **108**, E9–E16.

Cow Milk Protein Allergy

GLENN T. FURUTA

Children's Hospital Boston and Harvard Medical University

Cow milk, egg, peanut, soy, wheat, and fish antigens account for 90% of food allergic reactions. This article will focus on clinical syndromes associated with allergic reactions to cow milk proteins. Cow milk is not only a very common food product but also a potent allergen that has the capacity to induce a wide variety of different clinical reactions that affect both children and adults. In addition, the severity of the reactions ranges from systemic anaphylaxis to benign milk protein allergy of infancy; this diverse spectrum of clinical syndromes suggests different pathogenetic pathways for a common antigen.

INTRODUCTION

A wide spectrum of immunoglobulin E (IgE)-mediated and non-IgE-mediated food responses exist. For instance, the classical food allergic reaction to peanuts offers the best example of an IgE-mediated hypersensitivity reaction. Upon ingestion of or contact with peanuts, patients can develop the rapid onset of anaphylactic symptoms. Further evaluation shows elevated radioimmunosorbent test (RAST) for peanut IgE and abnormal skin prick testing. In contrast, many patients with food allergic responses, characterized by delayed reactions and intestinal mucosal pathology, often do not have abnormal RAST or positive skin testing for specific antigens. For instance, patients with cow milk protein colitis of infancy develop rectal bleeding and often no elevation of RAST for milk; further evaluation reveals eosinophilia of the rectosigmoid mucosa. A new era is dawning in the field of food allergic reactions as clinicians and scientists begin to define the clinical and pathophysiological features of the ever-broadening spectrum of food allergic reactions.

DEFINITIONS

Food Intolerance and Food Allergic/ Hypersensitivity Responses

Adverse food reactions refer to undesirable physical responses to an ingested food product. These reactions can be divided into two subcategories, food intolerances and food hypersensitivity/allergic reactions. Food intolerances are reflective of digestive deficiencies (i.e., lactose intolerance), toxins (i.e., bacterial toxins derived from infected foods such as from *Shigella* or *Campylobacter*), or biochemical properties of the food (i.e., monosodium glutamate, caffeine). Three features characterize food allergic/hypersensitivity reactions. First, they are reproducible reactions to a specific food product. Second, and often the more difficult part of the immunological response should be characterized in the diagnosis, clinic setting or with laboratory analysis. For instance, currently the availability of accurate tests is limited to RAST, skin prick testing, and double-blind placebo control food challenge. In

 Copyright 2004, Elsevier (USA). All rights reserved.

certain circumstances, intestinal endoscopy may offer clues to the presence of an allergic reaction but offer no benefit in identifying the etiology. Last, symptoms and intestinal pathology normalize with removal of the food antigen.

SYNDROMES ASSOCIATED WITH COW MILK PROTEINS

Esophagitis

A variety of upper intestinal symptoms, including vomiting, dysphagia, food impaction, heartburn, chest pain, irritability, and failure to thrive, can be seen in milk protein allergic responses. Two different allergic diseases leading to esophagitis have been recently defined. First, eosinophilic esophagitis (EE) is characterized by upper gastrointestinal symptoms (as noted above) that do not respond to acid-blocking medications, specific histological features with large numbers of eosinophils in the squamous epithelium, and normal pH probe monitoring of the distal esophagus. EE occurs in children over 1 year of age and throughout adulthood. Studies to date indicate that as many as three-quarters of patients have a personal or family history of allergic diseases. Physical examination usually reveals evidence of atopic diseases, including eczema, asthma, or allergic rhinitis/conjunctivitis. As many as 60% of reported patients with EE have abnormal RAST or skin prick testing for cow milk or soy proteins, suggesting evidence of an IgE-mediated reaction, but removal of these proteins does not always bring about remission. Diagnostic endoscopy reveals esophageal eosinophilic inflammation affecting both the proximal and distal esophagus, but not the stomach or duodenum. Some studies have demonstrated therapeutic success with the elimination of cow milk protein from the diet. Others have induced remission with an elemental formula or systemic or topical corticosteroids. The pathogenesis for this EE is not certain but murine studies demonstrate the importance of interleukin-5 (IL-5), an eosinotropic cytokine. The incidence of long-term complications from EE is unknown but esophageal strictures have been identified in both adults and children.

Another group of patients with upper gastrointestinal symptoms related to allergies are infants who have gastroesophageal reflux disease and cow milk protein allergy. These babies present with symptoms of vomiting and irritability. Mucosal biopsy of the esophagus demonstrates mild eosinophilic inflammation of the esophagus and gastric or duodenal abnormalities. pH probe monitoring of the distal esophagus is often abnormal. RAST and skin testing demonstrate sensitization to cow milk protein. Removal of the protein from the diet leads to clinical improvement. Children outgrow this condition and do not have further complications.

Allergic Gastritis

Milk proteins can induce symptoms associated with gastritis including abdominal pain, vomiting, nausea, early satiety, failure to thrive, gastrointestinal bleeding, and obstruction. Patients with milk protein allergy-induced gastritis can present at any age from the neonatal period to adolescence. As many as half of patients who develop the physical reactions to milk proteins demonstrate normal RAST and lack peripheral eosinophilia. When endoscopy is performed, mucosal biopsies show dense eosinophilic infiltration primarily affecting the gastric antrum. Removal of cow milk will relieve symptoms within a few days to weeks. It has been a longstanding belief that the majority of infants affected with cow milk protein are also affected with soy protein allergy. This finding has led to the practice of switching to a protein hydrolysate-based formula as the initial treatment. In fact, recent evidence suggests that less than 15% of infants develop soy allergy and therefore soy formulas are a reasonable, less costly initial therapeutic alternative.

Allergic Eosinophilic Gastroenteritis

Eosinophilic infiltration of the mucosa, muscular, or serosal layer of the stomach, small intestine or colon characterizes this disease. This anatomic classification system assists in explaining the wide variety of upper and lower intestinal symptoms occurring in this condition. For instance, mucosal disease can present with diarrhea, failure to thrive, bleeding and abdominal pain; muscular disease can present with symptoms of obstruction and serosal disease can manifest with eosinophilic ascites. While this classification system offers a convenient way to identify patients, clinical overlap certainly exists. Laboratory analysis often reveals anemia, peripheral eosinophilia, hypoalbuminemia, and, in a minority of patients, increased IgE antibodies to specific foods including milk proteins. Treatment with a diet restricted in the identified food allergen can lead to remission in 50% of patients. In those not responding, the use of protein hydrolysate formulas, elemental formulas, or systemic corticosteroids is indicated. Typically, patients have a prolonged course with intermittent exacerbations and remissions.

Food Protein-Induced Enterocolitis Syndrome

Food allergic responses sometimes involve the small and large intestine, leading to symptoms including severe abdominal pain, vomiting, diarrhea with bleeding, and hypotension. This constellation of symptoms presents in the first year of life and is associated with a non-IgE-mediated reaction to cow milk or soy proteins. At challenge, children develop vomiting and diarrhea within several hours. Villous atrophy and colonic inflammation characterize the histological features. Over 80% of patients respond quickly to a protein hydrolysate formula and the remainder require amino acid-based preparations. In rare circumstances, intravenous nutrition is required. The majority of patients outgrow this illness by 3 years of age; if the allergen is soy protein, a longer course can be expected.

Allergic Proctitis

The best characterized clinical syndrome associated with the ingestion of cow milk proteins is allergic proctitis of infancy. Well-appearing, thriving infants develop loose, mucousy, bloody stools upon the ingestion of cow milk or breast milk of mothers consuming milk and milk products. Interestingly, over half of affected infants are reported to be allergic to breast milk. Bloody stools begin within the first 3 months of life but can occur as early as the first week. Stool cultures are negative for pathogens and stool analysis may reveal eosinophils or evidence of degranulation with Charcot Leyden crystal proteins. Flexible sigmoidoscopic examination shows focal eosinophilic colitis and affected infants respond to a formula free of the offending protein. If the mother desires to continue breast-feeding, cow milk and milk products should be eliminated from her diet. The pathogenesis of allergic colitis is unknown. Recent evidence has identified increased numbers of T lymphocytes within the affected mucosa, increased production of IL-5, and decreased production of interferon-γ, a cytokine associated with the development of oral tolerance. Long-term prognosis is good and most infants outgrow the allergy by 1 year of age.

Constipation Associated with Cow Milk Allergy

In contrast to the bloody diarrhea associated with allergic proctitis of infancy or food protein-induced enterocolitis syndrome, several studies suggest that cow milk protein allergy may be associated with constipation through an immunologically mediated response. Children typically present during the first few years of life with symptoms of constipation or partial obstruction soon after beginning cow milk. In the initial report of this condition, over three-quarters of the subjects showed evidence of IgE sensitization to cow milk. Endoscopic evaluation of the distal rectal mucosa revealed mild eosinophilic inflammation without crypt invasion. Removal of cow milk increased the number of stools and reintroduction led to a decline in the number of stools per day.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Constipation • Eosinophilic Gastroenteritis • Food Allergy • Food Intolerance • Gastritis • Proctitis and Proctopathy

Further Reading

Butt, A. M., Murch, S. H., Ng, C.-L., *et al.* (2002). Upregulated eotaxin expression and T cell infiltration in the basal and papillary epithelium in cow's milk associated reflux oesophagitis. *Arch. Dis. Child.* **87**, 124–130.

Cavataio, F., Carroccio, A., and Iacono, G. (2000). Milk-induced reflux in infants less than one year of age. *J. Pediatr. Gastroenterol. Nutr.* **30**, S36–S44.

De Bouissieu, D., Matrarazzo, P., and DuPont, C. (1997). Allergy to extensively hydrolyzed cow milk proteins in infants: Identification and treatment with an amino acid formula. *J. Pediatr.* **131**, 744–747.

Fox, V., Nurko, S., and Furuta, G. T. (2002). Eosinophilic esophagitis—It's not just kids stuff. *Gastrointest. Endosc.* **56**, 260–270.

Iacono, G., Cavataio, R., Montalto, G., *et al.* (1998). Intolerance of cow's milk and chronic constipation. *N. Engl. J. Med.* **339**, 1100–1104.

Klemola, T., Vanto, T., Juntenen-Backman, K., *et al.* (2002). Allergy to soy formula and to extensively hydrolyzed whey formula in infants with cow's milk allergy: A prospective, randomized study with a follow up to the age of 2 years. *J. Pediatr.* **140**, 219–224.

Odze, R. D., Wershil, B. K., Leichtner, A. M., *et al.* (1995). Medical progress, allergic colitis in infants. *J. Pediatr.* **126**, 163–170.

Sicherer, S. H., Eigenmann, P. A., and Sampson, H. A. (1998). Clinical features of food-protein enterocolitis syndrome. *J. Pediatr.* **133**, 214–219.

Sicherer, S. H. (2000). Determinants of systemic manifestations of food allergy. *J. Allergy Clin. Immunol.* **106**, S251–257.

Zeiger, R. S., Sampson, H. A., Bock, S. A., *et al.* (1999). Soy allergy in infants and children with IgE-associated cow's milk allergy. *J. Pediatr.* **134**, 614–622.

Crohn's Disease

WILLIAM S. MOW AND MARIA T. ABREU
University of California, Los Angeles and Cedars-Sinai Medical Center

anti-*Saccharomyces cerevesiae* An antibody that recognizes *Saccharomyces cerevesiae* (baker's yeast). It is thought that this antibody represents an altered immune response to bacteria that share cross-reacting epitopes with oligomannosidic cell wall epitopes on baker's yeast. This antibody is seen in 60% of CD patients.

azathioprine (AZA) Pro-drug of 6-mercaptopurine (6-MP). AZA is metabolized to make 6-MP. AZA is commonly used in the treatment of Crohn's disease and works by interfering with DNA replication and cell proliferation.

Crohn's disease A disorder of chronic intestinal inflammation of unclear etiology and no known cure.

cyclosporin A calcineurin inhibitor that suppresses T helper cell responses to interleukin-1 (IL-1), thus blocking the immune response via its effect on calcineurin-mediated IL-2 mRNA transcription and production.

extraintestinal manifestations Areas with disease involvement outside of the intestinal tract in patients with Crohn's disease. They can affect any organ, mucosal, or epithelial surface.

fissure A tear in the anal canal associated with Crohn's disease.

fistula A connection between the gastrointestinal tract and other organs resulting from ongoing inflammation or obstructive stricturing. Fistulas can be entero-entero, entero-vesicular, entero-renal, entero-cutaneous, recto-vaginal, peri-anal, etc.

granulomas An uncommon finding in Crohn's disease (CD), but when found they are highly specific for this diagnosis. The granulomas seen in CD are of the noncaseating epitheliod form.

IBD1 The first gene locus identified in inflammatory bowel disease. It is located in the peri-centromeric region of chromosome 16 and contains the NOD2 gene.

infliximab A humanized monoclonal antibody directed toward tumor necrosis factor α that has shown great efficacy in the therapy of Crohn's disease.

6-mercaptopurine Anti-metabolic drug used commonly in treatment of Crohn's disease; it works by interfering with in DNA replication and cell proliferation.

methotrexate (MTX) A dihydrofolate reductase inhibitor. MTX and its metabolites inhibit the enzymes responsible for folate metabolism that result in anti-proliferative and cytotoxic effects on hematopoietic cells. This drug has been used with some efficacy in Crohn's disease.

NOD2 A gene within the IBD1 locus that has been associated with Crohn's disease. It codes for a protein involved in the innate immune response to bacteria products.

perinuclear anti-neutrophil cytoplasmic antibody A distinct form of anti-neutrophil cytoplasmic antibody seen in inflammatory bowel disease. It is characterized by a perinuclear staining pattern and is seen in up to 65% of ulcerative colitis patients.

probiotics Therapies involving ingestion of bacteria that are thought to have anti-inflammatory properties.

stricture or stenosis A narrowed area of the gut lumen commonly seen in Crohn's disease, usually secondary to scarring or inflammation.

tacrolimus A potent drug that works similarly to cyclosporin by blocking T-cell reactivity to interleukin-1 (IL-1) and therefore T-cell production of IL-2.

tumor necrosis factor α A cytokine that is important in the inflammation seen in Crohn's disease and rheumatoid arthritis secondary to its effect on T helper 1 immune responses.

Crohn's disease (CD) is a chronic disorder characterized by patchy transmural inflammation that may affect any portion of the gastrointestinal tract, but most commonly involves the ileum and colon. In addition to transmural inflammation, 10–28% of biopsy specimens demonstrate noncaseating epitheliod granulomas that are pathognomonic of this disorder. Patients with this disease suffer from chronic abdominal pain, malnutrition, and diarrhea. Therefore, CD may result in great morbidity and increased health care costs. Although the exact etiology is still under investigation, the current paradigm is that genetically susceptible individuals develop aberrant T-cell-mediated inflammation in response to lumenal bacteria. The incidence of CD appears to be on the rise in Westernized countries, with an incidence of between 3.1 and 14.6 per 100,000 person–years and an overall prevalence of between 0.1 and 0.5%. Because CD is a chronic disorder, it must be assessed and treated appropriately to minimize morbidity and health care costs in these patients. New and established therapies for CD have resulted in dramatic improvements in the quality of life for patients with CD.

PATHOGENESIS

There is strong evidence that Crohn's disease (CD) is a genetically based disorder in which individuals with a predisposition for this disease develop an abnormal

 Copyright 2004, Elsevier (USA). All rights reserved.

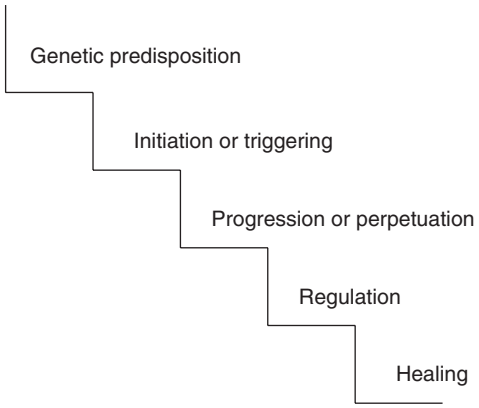


FIGURE 1 Steps in the development of Crohn's disease.

mucosal immune response to environmental triggers (Figs. 1 and 2). Although there is a strong genetic role in this disease, environmental factors also play a role in the pathogenesis of CD.

Environmental Triggers

Environmental triggers in CD may consist of bacteria, viruses, or food antigens. In humans, the exact relationship between CD and bacteria is less clear than in animal models. There are over 10^{10} bacteria per milliliter of stool residing in the human colon. Although this disease can affect any part of the gastrointestinal tract, a majority of patients have disease in the areas of the bowel with the highest bacterial load, the ileum, cecum, and colon. Studies have shown that diversion of the fecal stream can ameliorate disease. Additionally, there is literature supporting the use of therapies that can alter the bacterial flora, such as probiotics or antibiotics, for treatment of this disease. Animal models of inflammatory bowel disease (IBD) have provided much insight

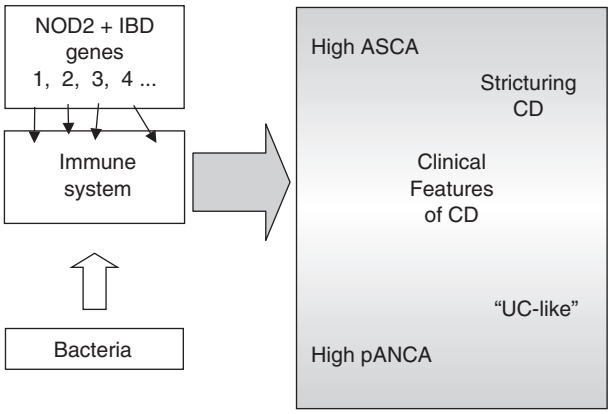


FIGURE 2 Influence of genes and bacteria in the clinical manifestations of Crohn's disease.

into the pathogenesis of CD in humans. These models have supported the knowledge that numerous genes can create the same clinical disease, that bacterial flora are key for the expression of the disease, that disease-specific effector lymphocytes can transfer disease to immune-deficient mice, that T helper 1 (T_H1) cytokines such as tumor necrosis factor-α (TNF-α), interleukin-12 (IL-12), and interferon-γ (IFN-γ) mediate disease, and that regulation of mucosal inflammation is controlled by T cells.

Many papers that both support and refute the role of a specific microbial pathogen as the etiologic cause of CD have been published. Most of this research has been directed toward *Mycobacterium paratuberculosis* and the measles virus but has been inconclusive. Evidence in animal models of IBD supports the concept that lumenal bacteria play a key role in the initiation of inflammation in the mucosa. More recently, a unique bacterial DNA sequence has been identified within lamina propria mononuclear cells in CD patients. This DNA sequence is present in the affected colonic mucosa, but not in uninvolved CD mucosa, in ulcerative colitis (UC) patients, or in normal controls. This DNA codes for a peptide termed I2. This peptide can function as a T-cell superantigen, a protein that can induce a T-cell response in a non-major histocompatibility complex (MHC)-restricted fashion. This strong immunologic response is credited to specific, direct reactivity of I2 and the MHC class II protein. Further research has identified the organism expressing the I2-containing gene as *Pseudomonas fluorescens*. *P. fluorescens* is part of the normal human intestinal flora and is usually found within the terminal ileum and colon. It is possible that this organism may play a role in initiating mucosal inflammation in some CD patients or may colonize the intestine of CD patients as a result of treatment.

Although the exact role of bacteria in the pathogenesis of Crohn's disease is unclear, considerable research exists to support the role of microorganisms in IBD. Much of this research has been performed on animal models of CD in which researchers can control the bacteria in the environment, creating the ability to associate specific bacteria with the onset and progression of disease. Current evidence supports the concept that bacteria are important for the development of the mucosal immune system and colonization with bacteria is important to the development of disease in many animal models. For example, two mouse models, the IL-$10^{-/-}$ and the Gαi$2^{-/-}$ mice, have several similarities to human CD, such as a predominant T_H1 response of $CD4^+$ T cells and transmural inflammation. In these models, disease does not develop if the mice are raised in germ-free or specific microorganism-free environments, supporting

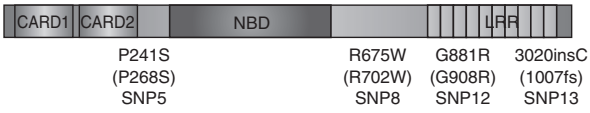
the role of bacteria in these animal models. In the C3H/HeJBir murine colitis model, animals develop a spontaneous colitis characterized by $CD4^+$ T cells that have specific reactivity to bacterial antigens. In addition, these bacteria-reactive $CD4^+$ T cells can transfer disease to recipient mice with severe combined immunodeficiency. These studies demonstrate that genetically susceptible hosts require intestinal flora and the subsequent development of bacterial-reactive T cells for intestinal inflammation.

An increasing incidence of CD in Westernized nations has led to theories that environmental changes that occur with industrialization may contain the environmental triggers for CD. A delicate balance between the gut microflora and the mucosal immune response may be disrupted with alteration in the bacterial flora secondary to refrigeration and improved sanitation. Literature supports the role of helminthic parasites in the suppression of CD. A decline in the incidence of helminthic infections is seen in industrialized societies. Helminth infections are associated with a T helper 2-type response that may counteract a robust T helper 1-type response that is associated with CD.

Genetics

CD is most prevalent in people of Northern European and Jewish descent. Genetic disorders such as Turner's syndrome and Hermansky-Pudlak syndrome are associated with CD. Those suffering from other disorders, e.g., glycogen storage disease type 1b patients, have a predisposition to "CD-like" disease characterized by chronic colonic inflammation, which at times can involve circumferential ulcerations. First-degree relatives of patients with CD have 15 times the normal risk of developing IBD. Approximately 44% of monozygotic twins share the diagnosis of CD, compared with only 3.8% of dizygotic twins, indicating that genetics plays an important role in this disease. Heredity, however, is not the only factor that dictates whether or not a genetically predisposed person develops disease.

It is likely that many genes are involved in the development of Crohn's disease. Extensive research on families with IBD has identified linkage regions on five chromosomes for these disorders. The identified regions are IBD1 on chromosome 16 (NOD2 region), IBD2 on chromosome 12, IBD3 on chromosome 6 (the human leukocyte antigen region), IBD4, which is another linkage area at chromosome 14q11−q12, and IBD5 at chromosome 5q31, which is a dense area of cytokine genes. In the pericentromeric region of IBD1, a gene called NOD2, which confers susceptibility to development of Crohn's disease, was recently described.



- Carriage of NOD2 allelic variants:
  - Crohn's disease 27– 39% (49% of all DCMs)
  - Control population 14–16% (22.5% of all DCMs)
  - UC population 12–14% (12%)
- 27 rare mutations account for 19% of disease-causing mutations

FIGURE 3 Allelic variants of NOD2 are associated with Crohn's disease. Data from Hugot, J. P. *et al.* (2001). *Nature* **411**, 599; Ogura, Y., *et al.* (2001). *Nature* **411**, 603; Hampe, J., *et al.* (2001). *Lancet* **357**, 1925; Lesage, S. *et al.* (2002). *Am. J. Hum. Genet.* **70**, 845; Cuthbert, A. P., *et al.* (2002). *Gastroenterology* **122**, 867; Ahmad, T., *et al.* (2002). *Gastroenterology* **122**, 854.

This NOD2 gene is specific for CD and is not associated with UC (Fig. 3). Its protein product is expressed by monocytes and activates nuclear factor κB (NF-κB), an important factor that is involved in the mucosal immune response to bacterial products such as lipopolysaccharides. The exact mechanism by which mutations in NOD2 confer susceptibility to CD is not yet understood. The mutations that occur in NOD2 result in diminished activation of NF-κB in response to bacterial products. CD patients with NOD2 mutations are characterized by having a fibrostenosing disease phenotype predominantly of the small bowel.

Mucosal Immunology

The mucosal immune system of the gut must achieve a delicate balance between tolerance to the lumenal flora and defense against pathogenic bacteria that may gain access through the oral route. Below the epithelial layer, naive, undifferentiated T cells await stimulation from antigen-presenting cells to activate them and provide guidance with respect to which type of T cell they will become. Interaction between co-stimulatory molecules on the surface of T cells, e.g., CD40 ligand (CD40L or CD154), and their cognate receptors on antigen-presenting cells, e.g. CD40, is required for full T-cell activation. Because T cells are the principal directors of the immune response, the characterization of the immune response is based on the pattern of cytokines or the surface markers expressed by mature T cells, e.g., T_H1, T_H2, T_H3, or T regulatory 1(Tr1). T_H1 and T_H2 cytokines are effector cytokines that may have important roles in determining the clinical manifestations associated with CD or ulcerative colitis. Antigen-presenting cells, such as dendritic cells and macrophages, also play a critical role in shaping the immune response through secretion of cytokines that

guide the differentiation of T cells. IL-12, a macrophage-derived cytokine, and IL-18, a macrophage- and epithelium-derived cytokine, shift the immune response toward a T_H1-type response and are elevated in the mucosa of patients with CD (Fig. 4). As a result, T_H1 cytokines such as IFN-γ, IL-2, and TNF-α are frequently elevated in the mucosa of CD patients. In animal models, mice lacking the IL-10 gene (IL-10 knockout mice) develop a Crohn's disease-like inflammatory bowel disease. In these mice, IFN-γ and IL-12 levels peak coincident with the onset of the inflammation at week 10.

Transforming growth factor-β (TGF-β) and IL-10 are secreted by Th3 and Tr1 cells, respectively. In several animal models, delivery of TGF-β or IL-10 can ameliorate colitis. Because these cytokines appear to have beneficial effects, attempts have been made to harness the power of Tr1 and Th3 cells. Studies have been hampered by an inability to isolate these cell populations. In addition, TGF-β is associated with increased fibrosis and is unlikely to be tolerated systemically. Likewise, studies using recombinant human IL-10 to treat CD have been associated with flu-like symptoms and little clinical benefit. In contrast, the TNF-α antagonist infliximab has been dramatically effective in the treatment of CD. Studies have demonstrated that infliximab results in apoptosis of TNF-α-expressing T cells in the mucosa. However, in clinical practice, cytokine measurements are not currently used to make decisions about patient care as they are cumbersome and are not sufficiently specific because of the broad range of normal and abnormal cytokine levels.

DIAGNOSIS

The diagnosis of CD is based on a combination of clinical, endoscopic, and radiologic findings. Serologic and pathologic findings and genetic testing are available and may be used as adjuncts to diagnosis and management. Most patients will present with abdominal pain, weight loss, fever, diarrhea, bloody stools, and poor growth (in pediatric patients). The differential diagnosis includes infectious diarrhea (viral, bacterial, mycobacterial, or parasitic), ulcerative colitis, collagenous or microscopic colitis, colorectal carcinoma, polyposis, nonsteroidal anti-inflammatory drug- or medication-related gastrointestinal inflammation, diverticulitis, human immunodeficiency virus, immunodeficiency, and celiac disease.

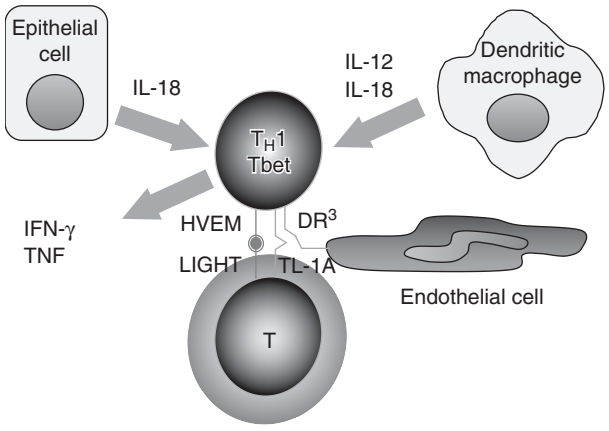


FIGURE 4 T_H1 development in Crohn's disease.

The work-up for a patient suspected to have CD includes basic laboratory tests including electrolytes, liver function tests (including albumin), complete blood count with differential and platelet count, C-reactive protein, and erythrocyte sedimentation rate. These tests point to problems associated with CD, such as anemia, malnutrition, and active inflammatory disease. An upper gastrointestinal series with small bowel follow-through (UGI-SBFT) should be performed to assess for areas of disease involvement. Small-bowel enteroclysis provides similar information and in general is not thought to be more sensitive than UGI-SBFT. In select cases, however, CD may be missed with UGI-SBFT and seen with enteroclysis and vice versa.

Colonoscopy with ileal intubation and biopsies is generally necessary to confirm the diagnosis of CD. Endoscopic lesions seen in CD include linear ulcers, stellate ulcers, aphthous ulcers, inflammation, polyps, erosions, strictures, fistulae, and bleeding/friability. In contrast to UC, the inflammation may be both microscopically and macroscopically patchy. It is important that the endoscopist define areas of disease involvement during the colonoscopy. Ideally, endoscopy should include intubation of the ileum for evaluation and biopsies, as this area is frequently affected in CD. Video capsule enteroscopy is a new method of assessing the small bowel of patients with CD and offers the ability to visually assess the small bowel in its entirety. At present, capsule studies are contraindicated in patients with CD-related strictures or stenosis.

Since CD can affect any part of the gastrointestinal tract, the disease has diverse manifestations. The most common area of disease is the ileum and ileocecal area. Up to 40% of CD patients have involvement of this area and present with abdominal pain and diarrhea, but they may also have fever, abdominal mass, and weight loss (Fig. 5). Some patients presenting with right lower quadrant pain are diagnosed with appendicitis and the diagnosis of CD is made at the time of surgery. Approximately 30% of CD patients have disease limited to the small bowel. These patients also present with

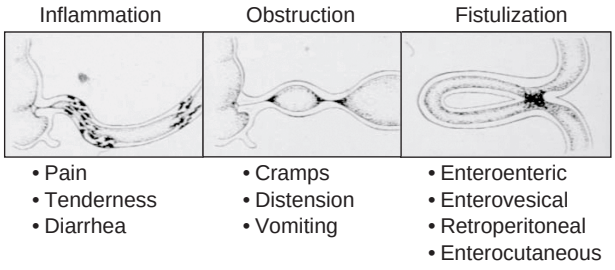


FIGURE 5 Clinical presentations of Crohn's disease.

diarrhea, abdominal pain, fever, and abdominal mass, but they may have more severe problems with malabsorption, steatorrhea, weight loss, and malnutrition, depending on the extent of small-bowel involvement. In 25–30% of CD patients, inflammation is limited to the colon. These patients can be difficult to distinguish from UC because of similarity of symptoms, including bloody diarrhea, weight loss, abdominal pain, urgency, and tenesmus. Endoscopic features, such as skip lesions, rectal sparing, and right-sided colonic disease, are supportive of the diagnosis of CD colitis.

Approximately 18% of patients will have perianal disease at the time of presentation. The spectrum of perianal disease ranges from hemorrhoids and skin tags to perianal fistula and abscesses. In addition, CD patients may have extraintestinal symptoms including, but not limited to, arthralgias, episcleritis, nodular scleritis, eythema nodosum, pyoderma gangrenosum, and psoriasis.

SEROLOGIC TESTS

In recent years, serologic tests have been used with increased frequency to help diagnose and define disease. These blood tests are moderately specific and sensitive for detecting IBD, but are generally insufficient as the sole diagnostic or screening tool. If there is a high level of suspicion for CD, diagnosis should be made by endoscopy with biopsies of the diseased areas. The main antibody tests used in IBD are anti-neutrophil cytoplasmic antibody (ANCA) and anti-*Saccharomyces cerevisiae* antibody (ASCA). These antibodies are directed toward self-antigens and microbial antigens, supporting the hypothesis that these disorders represent a loss of tolerance to self or to commensal organisms.

Distinct from ANCA in Wegener's granulomatosis, ANCA in IBD is characterized by a perinuclear staining pattern, hence the term pANCA. Of the serologic markers, pANCA is most prevalent in IBD. pANCA is positive in 15% of CD patients, 65% of UC patients, and less than 5% of non-IBD controls. This autoantibody is strongly associated with UC. In CD, patients expressing pANCA have "UC-like" features, such as left-sided colitis and pan-colitis. CD patients expressing pANCA tend to have isolated colonic disease and it has been shown that these patients have a blunted clinical response to therapy with anti-tumor necrosis factor therapies, such as infliximab.

ASCA is an antibody that recognizes *S. cerevesiae* (baker's yeast). It is thought that this antibody represents an altered immune response to bacteria that share cross-reacting epitopes with oligomannosidic cell wall epitopes on baker's yeast. This antibody is seen in 60% of CD patients, 5% of UC patients, and less than 5% of non-IBD patients. High levels of ASCA expression have been associated with younger age of onset, fibrostenosis, internal-penetrating phenotype, and an increased risk for multiple surgeries.

In addition to pANCA and ASCA, antibodies to I2 and *Escherichia coli* outer membrane porin C (OmpC) are also used to assist in diagnosis. Serum reactivity to I2 is seen in up to 54% of CD patients. The exact role of antibodies to OmpC has not been defined, but serum reactivity to this bacterial antigen is seen in up to 55% of CD patients. Further research on serologic responses to ANCA, ASCA, I2, and OmpC has indicated that CD patients can be grouped into four categories based on these serologic responses. ASCA and pANCA have been shown to play a role in the work-up of chronic abdominal pain in pediatric patients when used as a tool to determine whether invasive tests are necessary. Pancreatic autoantibodies have also been found in up to 27% of CD patients, but they are absent in healthy controls and ulcerative colitis patients. In addition to these serologies, there are many other tests being developed that show promise in helping to diagnose IBD and predict disease prognosis and course in attempts to better direct therapy and minimize morbidity.

PATHOLOGY

CD is characterized by acute and chronic transmural inflammation of any part of the gastrointestinal tract. The typical pathologic features include acute and chronic inflammation (esophagitis, ileitis, colitis), ulcerations, and crypt abscesses. Noncaseating granulomas are typical of CD, but can also be seen in gastrointestinal infections such as *Yersinia enterorolitica*, *Chlamydia trachomatis*, and fungal infections. The presence of caseating granulomas should always raise suspicion for tuberculosis. When present, noncaseating granulomas are useful in helping to diagnose CD. However, they are rarely seen on biopsy specimens and the absence of granulomas does not rule out the diagnosis of CD. The

location from which biopsy specimens are taken is essential to help differentiate CD from UC. Many colonic biopsy specimens from UC and CD patients are indistinguishable unless there are granulomas present and thus the endoscopic description of disease extent is essential for the diagnosis. Patchy microscopic inflammation is suspicious for CD.

TREATMENT

Treatments for CD can be broadly defined as those directed toward protection of the epithelium, those directed at the mucosal immune response, and those directed at modifying the lumenal contents. These therapies are generally delivered in a tiered approach depending on symptoms and the phase of disease (Fig. 6).

Epithelial Protection

Numerous studies on aminosalicylate-containing products, including sulfasalazine and mesalamine (5-ASA) products, have been performed. These drugs are known to inhibit cyclooxygenase and 5-lipoxygenase pathways and may have other functions that help in the treatment of IBD including suppression of antibody secretion and action as oxygen radical scavengers. Recent studies have found that mesalamine inhibits activation of NF-κB and thus may exert its anti-inflammatory effect through this mechanism. Sulfasalazine at a dose of 4–6 g/day has been evaluated in four double-blind studies, of which three studies demonstrated the drug to be better than placebo in achieving remission. This drug seems to have its best effect in mild to moderately active colonic CD and not in small-bowel CD. This drug has not been demonstrated to maintain remission; however, these studies were not performed using high doses and it is likely that patients who achieve remission on sulfasalazine can be maintained in remission on this drug as well.

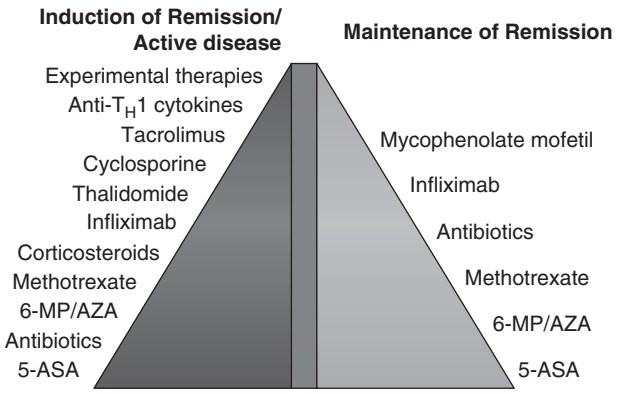


FIGURE 6 Therapeutic pyramid for Crohn's disease.

Non-sulfa-containing 5-aminosalicylate formulations are widely used for treatment of UC and CD. Many studies have been performed to evaluate the ability of these drugs to achieve remission in CD patients. The largest of these studies examined a dose range of ethyl-cellulose-coated granules of mesalamine (1, 2, and 4 g/day) and found that only the high-dose treatment group had a statistically significant reduction in Crohn's disease activity index (CDAI). The conclusions that can be made from studies on mesalamine are that it is effective in mildly to moderately active CD at doses greater than 3.2 g/day. Studies to evaluate maintenance of remission suggest a role for higher doses (>3 g/day). However, meta-analysis of these studies failed to show significant efficacy in the prevention of recurrence following medical remission. Studies on postoperative prophylaxis for CD are more promising than those on treatment of active disease. High-dose mesalamine demonstrated a trend toward reduction of recurrence at ileocolic anastomoses. It is likely there are subgroups of CD patients who are responsive to 5-ASA products, but it is not yet possible to identify these patients.

IL-11, a cytokine produced by mesenchymal cells, has a strong effect on thrombopoiesis and also acts to enhance the intestinal mucosal barrier. Recombinant human IL-11 has been evaluated in trials of CD patients. Phase IIa dose–response studies were performed but showed no overall benefit. However, there was a trend toward improvement in the 16 μg/kg per dose group. In phase IIb/III trials, 148 patients with active CD were enrolled and received subcutaneous doses of 15 μg/kg weekly or 7.5 μg/kg twice weekly. The rate of remission in the 15 μg/kg weekly group at 6 weeks of therapy was significantly higher than that observed in the placebo group, but was not very high (40% versus 16%). Patients in this trial also had dose-related increases in platelet counts. Although no thrombotic events occurred in this trial, it is possible that with continued therapy patients may experience these complications.

Growth hormone (GH) has been investigated for use in CD with the rationale that GH may reverse the catabolic effects of chronic intestinal inflammation. A randomized, placebo-controlled trial using human recombinant growth hormone in 37 CD patients with active disease was conducted. GH-treated patients had a statistically significant decrease in their CDAI compared with placebo-treated patients. However, this study did not report remission rates and therefore additional studies are needed before recommendations on GH use can be made. The dose of GH used in this study

was very high [5 mg administered subcutaneously (sc) the first day followed by 1.5 mg sc daily] and in the range used for catabolism cited with acquired immune deficiency syndrome. There were two patients who were found to have a tumor during the study. The long-term risks may therefore be unacceptable.

Immunomodulators

Inhibition of Trafficking as a Strategy to Treat CD

Glucocorticoids have multiple effects on inflammation including diminished production of cytokines, apoptosis of lymphocytes, and reduced expression of adhesion molecules that regulate leukocyte trafficking to sites of inflammation. Glucocorticoids have a definite role in inducing remission in CD patients with moderate to severe symptoms. Prednisone and prednisolone have been studied extensively, but the use of these drugs is complicated by side effects, such as adrenal suppression, Cushingoid features, hypertension, diabetes, and infection. Although these drugs have potent effects on moderate to severe CD, they have little long-term efficacy. Two large trials, the National Cooperative Crohn's Disease Study in 1979 and the European Cooperative Crohn's Disease Study in 1984, demonstrated that prednisone at doses greater than 40 mg was better than placebo in achieving remission in acute disease. Greater than 80% of CD patients apparently respond to glucocorticoids and patients with high disease activity as measured by CDAI, previous surgeries, and perianal disease were less likely to respond. Another corticosteroid, budesonide CIR (controlled ileal release), is formulated to allow for delivery to the ileal and cecal areas because it is coated with a methacylic acid copolymer (Eudragit), which dissolves at a pH of 5.5 or higher. It is a more potent steroid than prednisone and has fewer systemic side effects secondary to a high first-pass metabolism in the liver. Placebo-controlled trials have supported the role of this drug in mild to moderate ileal–colonic CD. In one of the first studies, a dose of 9 mg daily for 8 weeks was superior to other doses and to placebo in achieving remission. In other studies, budesonide at 9 mg once daily has been compared to 40 mg of prednisone daily. After 8 weeks of treatment, remission rates for daily budesonide and prednisone were similar, 53–66%. In addition to providing good results in induction of remission in mild to moderate CD, budesonide has been effective in maintaining remission. In a placebo-controlled 1-year study, it was demonstrated that time to relapse in the group given 6 mg daily of budesonide was longer (>250 days) than in the placebo group (<100 days). The prolonged use of these medications has indicated that their efficacy may wane after a year of continuous treatment. Long-term follow-up studies on CD patients treated with budesonide, prednisone, and non-steroid-based therapies for over 2 years indicated that budesonide does not spare patients the osteoporotic effects of other systemic glucocorticoids. In fact, budesonide-treated patients in some trials had a greater decrease in bone mineral density than patients treated with prednisone. Results of a prospective study comparing budesonide to prednisone for active CD demonstrate, however, that budesonide is not as harmful to bone as prednisone.

A new therapy, natalizumab, currently in phase III trials, is a humanized monoclonal antibody to $\alpha 4$ integrin. It inhibits lymphocyte binding to the mucosal addressin cell adhesion molecule expressed on high endothelial venules in the intestine by blocking the $\alpha 4$ receptor on lymphocytes. Studies on humans and animals indicate that these $\alpha 4$ integrins play a key role in the migration of lymphocytes to the gut. $\alpha 4$ integrin blockade for the therapy of IBD was first used on cotton-topped tamarinds with spontaneous colitis. These animals had complete resolution of their symptoms with treatment. In a human trial of natalizumab, patients with mildly to moderately active CD had a statistically significant decrease in CDAI and many achieved remission by week 2. Further studies are being performed using higher doses of natalizumab and examining rates of response and remission. Studies on this drug show that it holds promise in the treatment of CD, but it has not yet been approved by the Food and Drug Administration and more studies are needed.

Antimetabolites

Since the initial trials of 6-mercaptopurine (6-MP) and azathioprine (AZA), these drugs have become the mainstays of maintenance therapy for CD. Studies have been both supportive and contradictory, but 6-MP and AZA remain the primary medications used for maintenance of remission. The benchmark study on this drug was a double-blind placebo-controlled crossover study, which showed that 6-MP at a dose of up to 1.5 mg/kg/day was effective in producing a clinical response in active CD in up to 72% of patients compared to 14% for placebo. In addition, 75% of the 6-MP-treated patients were able to wean or discontinue steroids compared to $<40\%$ of placebo-treated patients. Fistulas closed in 36% of 6-MP-treated patients compared to 6% of placebo-treated patients. Finally, it was observed that the mean time to clinical response was approximately 3 months. In many cases, however, 3 months was the earliest time point at which patients were seen. This study solidified the role of 6-MP in treatment

of active CD with the goals of minimizing steroids, inducing remission, and treating active fistulas.

AZA has been also evaluated for maintenance of remission. Meta-analysis of numerous controlled studies evaluating the role of AZA in maintenance of remission supports its use. Overall, 67% of patients had continued clinical response compared with an odds ratio of 3.09 (95% confidence interval CI, 2.45 to 3.91), favoring response in patients treated with 6-MP/AZA. Although this study did not reach statistical significance, there is a wealth of long-term clinical experience supporting the use of AZA. However, studies have failed to establish the length of time required to achieve and to maintain remission. Studies have addressed whether AZA withdrawal after several years is safe in CD patients in long-term remission. Although initial studies suggested that >4 years of AZA use was associated with maintenance of remission before drug withdrawal, more recent studies have found that AZA withdrawal is associated with a greater number of relapses than continued use of the drug. Studies have failed to conclusively determine the duration of therapy that is adequate for permanent remission of disease. Most studies seem to indicate that those patients that are taken off of 6-MP or AZA tend to relapse, whereas those that are maintained on one of these medicines continue to do well. In addition, the long-term use of these drugs has been associated with an increased risk of malignancy. Treatment of fistulas has also been evaluated in controlled trials and has been summarized in a meta-analysis that showed that approximately half of patients with fistulas had a favorable response to 6-MP/AZA, whereas only approximately 20% of controls showed improvement. Finally, its role in postoperative prophylaxis has also been evaluated and results from a few studies are very promising.

More recently, research has focused on the metabolism of 6-MP by an enzyme called thiopurine methyltransferase (TPMT) (Fig. 7). It appears that this enzyme is responsible for methylation of the drug and production of 6-methylmercaptopurine (6-MMP), the blood level of which appears to correlate with the hepatotoxic effects of the drug. There has also been much research directed at 6-thioguanine (6-TG), which has been correlated with clinical response to 6-MP/AZA (Fig. 8). Methods for measurement of TPMT activity and determination of genotype are now commercially available and may help to minimize toxicity. Less than 1% of patients have an absence of TPMT activity, 10% of patients have intermediate activity, and the remainder have normal activity. Many patients with normal TPMT activity will have preferential

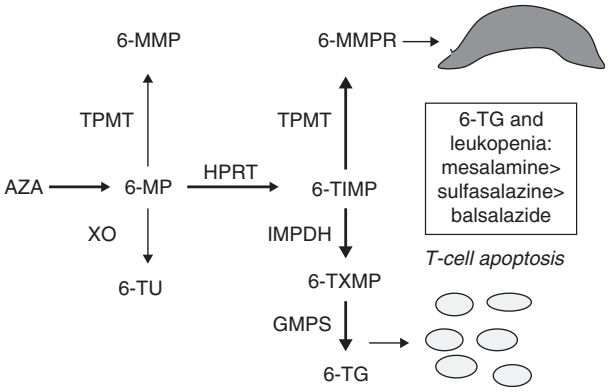


FIGURE 7 6-MP/AZA mechanism of action. From Lowry *et al*. (2001). *Gut*, with permission.

metabolism of 6-MP toward 6-methylmercaptopurine, which has a stronger association with hepatotoxicity than 6-TG.

Methotrexate (MTX) is a dihydrofolate reductase inhibitor. MTX and its metabolites inhibit the enzymes responsible for folate metabolism that result in antiproliferative and cytotoxic effects on hematopoietic cells. This drug has been studied in CD patients with chronic active disease and steroid dependence for over 3 months. Patients were randomized to receive 25 mg intramuscularly of methotrexate weekly or placebo injections. After 16 weeks of therapy, more patients treated with methotrexate were in remission and off steroids than those given the placebo. Although this drug is effective for maintenance of remission, it does carry significant risks for hepatotoxicity, teratogenicity, bone marrow toxicity, and other side effects.

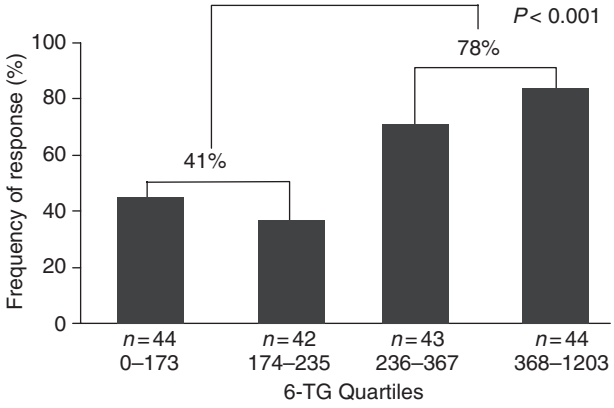


FIGURE 8 6-TG level correlates with clinical response. Reprinted from Dubinsky, M. C., *et al*. (2002). *Gastroenterology* **118**, 705–713, with permission from Elsevier.

Calcineurin Inhibitors

Cyclosporine (CSA) is potent drug that has revolutionized organ transplantation. It is a cyclic protein derived from the fungus *Tolypocladium inflatum*. Cyclosporine is a calcineurin inhibitor that suppresses T helper cell responses to IL-1, thus blocking the immune response via its effect on calcineurin-mediated IL-2 mRNA transcription and production. There have been several trials on the use of cyclosporine in Crohn's disease. Many of these reports have yielded promising results, with the main risks being the side effects of the drug. The first demonstration of efficacy of oral CSA (5–7 mg/kg/day) was in active steroid-refractory CD, where it was superior to placebo. Several uncontrolled series have suggested that CSA can be used in active CD as a bridge to anti-metabolite therapy. One of the major advantages of this medication is its prompt onset of action. Response rates to cyclosporine are high (>60%) and the benefits of cyclosporine are generally noticeable within 2 weeks of starting therapy. However, few patients who achieve a response to this medication maintain the benefit for prolonged periods of time. Thus, although a high proportion of inflammatory and fistulizing CD patients remained in remission while on oral cyclosporine, almost all of the patients relapsed when cyclosporine was discontinued. Moreover, this medication has several side effects and toxicities that make continuous intravenous infusion risky if done incorrectly. Significant side effects include neurotoxicity and renal toxicity; thus, patients taking this medication should be followed closely with laboratory tests and cyclosporine levels should be monitored carefully. The standard oral dose of cyclosporine is 5–10 mg/kg/day divided twice daily. All patients on cyclosporine have an increased risk of opportunistic infections and should receive antibiotic prophylaxis for *Pneumocystis carinii* pneumonia and prompt evaluation for infection-related symptoms.

Tacrolimus is a newer agent with actions similar to those of cyclosporine, but with improved oral absorption compared to its older relative. There have been a handful of reports on the use of this drug in active inflammatory and fistulizing CD. A high percentage of pediatric patients (>60%) with steroid-refractory UC or CD colitis responded to therapy and were discharged. The drug was discontinued, however, after 2–3 months and a majority of patients underwent colectomy within 1 year of receiving tacrolimus. This drug has also been used for short-term treatment of CD patients as well as CD patients with fistulizing disease with some beneficial effects. Despite these promising reports, there remains a need for a good prospective study to evaluate the use of this drug in CD. Moreover, tacrolimus-treated patients reportedly have a high frequency of side effects, such as paresthesias, insomnia, tremor, and headache. As with cyclosporine, patients treated with tacrolimus are at risk for opportunistic infections and similar side effects, such as renal, neurologic, and bone marrow toxicity, can occur. It is essential to monitor drug levels and follow patients with laboratory tests.

T-Cell Regulation

Tumor Necrosis Factor-α-Modulating Drugs

T-cell activation is a central event in the pathogenesis of Crohn's disease. Because T-cell activation results in cytokine secretion and because certain cytokines may be detrimental when overexpressed, biologists and clinicians have examined the effect of cytokine blockade in patients with CD.

TNF-α is a potent pro-inflammatory cytokine that plays an important role in the mucosal inflammation seen in CD. There are many different drugs aimed at blocking the function of this pro-inflammatory protein (Fig. 9). Infliximab is a chimeric monoclonal antibody with human constant regions and murine variable regions. It is approximately 25% murine-derived protein. CDP 571 is a humanized version of a murine antibody toward TNF-α. This drug has human IgG4 constant regions and κ light chains in which specific murine protein residues were inserted to confer reactivity to TNF-α. This drug is reported to be 90% less immunogenic than murine antibodies. Finally, etanercept is a soluble TNF-α receptor fusion protein that has been used extensively for rheumatoid arthritis.

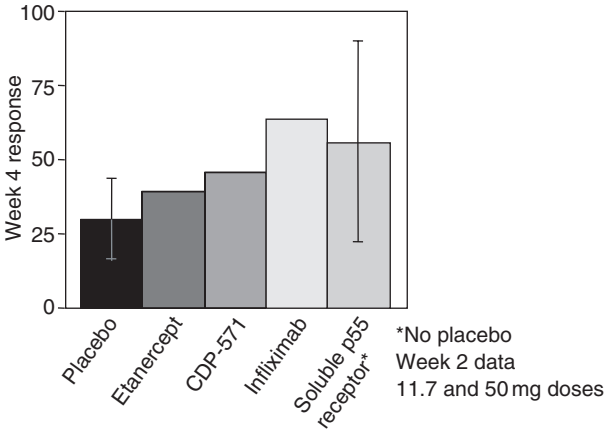


FIGURE 9 Comparison of anti-TNF-α strategies for Crohn's disease.

The initial trials on infliximab showed that it had great efficacy in patients with severe refractory CD. More than 70% of patients will respond within 2 weeks of the first infusion of this drug at a dose of 5 mg/kg given intravenously. Further studies have confirmed that infliximab has a role in the maintenance of remission in CD that is refractory to traditional therapies. A recent study demonstrated that CD patients can achieve long-term remission with repeated dosing. In addition, infliximab has shown efficacy in perianal fistulas, with fistula closures in greater than 50% of patients.

The genetically engineered human monoclonal antibody to TNF-α, CDP571, has also been studied. For reasons that have yet to be determined, clinical response to this medication is less than that for infliximab. Infusion reactions and other side effects that were associated with infliximab were also seen with CDP571. This drug may have a role in the therapy of CD, but it must be compared with the superior efficacy of infliximab. Differences may be accounted for by the fact that infliximab induces apoptosis of activated lymphocytes and binds more avidly to TNF-α. Etanercept has also been studied in a double-blinded placebo-controlled trial at a dose similar to that used for rheumatoid arthritis. Rates of remission for etanercept and placebo were similar. Thus, this drug seems to have the least anti-TNF-α activity of the medications discussed thus far.

Therapy with murine-derived anti-TNF-α agents is not without risks. Patients may develop acute IgE-mediated reactions to infusions of these agents as a result of the immunogenic murine portions of these drugs. Acute hypersensitivity reactions are characterized by symptoms of anaphylaxis, hypotension, urticaria, angioedema, and respiratory difficulty. Other patients may have a T-cell-mediated delayed-type hypersensitivity reaction characterized by gradual onset of arthralgias, myalgias, fever, and acute respiratory distress. Because TNF-α is required to contain infections by intracellular pathogens, therapy with anti-TNF-α agents has also resulted in opportunistic infections such as tuberculosis, histoplasmosis, aspergillosis, coccidiomycosis, and listeriosis. Due to increased infections by tuberculosis, all patients must receive a purified protein derivative skin test prior to their first infusion. Physicians treating patients with anti-TNF-α medications should have a low threshold to thoroughly evaluate symptoms of fever, night sweats, productive cough, etc. Other side effects include exacerbation of multiple sclerosis symptoms, lupus-like reactions, and autoimmune antibody formation. The effects of these drugs on a patient's risk for malignancy are not known. Regardless of the risks, infliximab has changed the face of treatment of CD and is the first effective biological therapy for this chronic disorder.

Thalidomide is another agent used in the therapy of CD. Thalidomide reduces TNF-α mRNA stability and has been found to reduce TNF-α and IL-12 levels in CD patients. Two open-label trials have addressed the use of this medication in CD. A high rate of response, comparable to that obtained with infliximab, was observed. Patients who completed this trial were able to reduce their steroid dosage by at least 50%. In these open-label studies, thalidomide also seems to provide therapeutic benefit for perianal fistulas. Since this medication has effects on TNF-α production, it has also been studied in patients who have achieved remission following infusions of infliximab. In this study, over 80% of patients were able to maintain remission for almost a year. Despite the promising effects of thalidomide in these trials, the teratogenic effects of this drug continue to haunt users and prescribing doctors. It is recommended that all patients (both male and female) receiving this medication be made aware of its teratogenic effects and use effective contraception. In addition, this medication has frequent side effects such as drowsiness, peripheral neuropathy, edema, and dermatitis. Because of the efficacy of thalidomide, newer agents called selective cytokine inhibitory drugs are being developed with the hopes of harnessing the TNF-α suppressive effects while reducing the side effects and teratogenicity. CC-1088 is the first of these drugs and is currently in clinical trials for use in CD.

SURGERY AND POSTOPERATIVE RECURRENCES

Despite advances in knowledge about the pathogenesis of CD and the development of new potent biologic therapies for this disorder, many CD patients will fail the above outlined medical therapies or have advanced disease at the time of presentation. In these cases, it is important for these patients to be evaluated by an experienced surgeon with expertise in IBD. Many of these patients have had surgeries previously and are at risk for intra- and postoperative complications. Surgery is not curative in CD (unlike ulcerative colitis) and therefore it is usually considered predominantly in patients with complications such as abscesses, perforation, obstruction, hemorrhage, medically refractory disease, cancer, fulminant colitis, toxic megacolon, sepsis, and certain fistulas. It is estimated that up to 80% of CD patients with ileal disease will need surgical intervention for

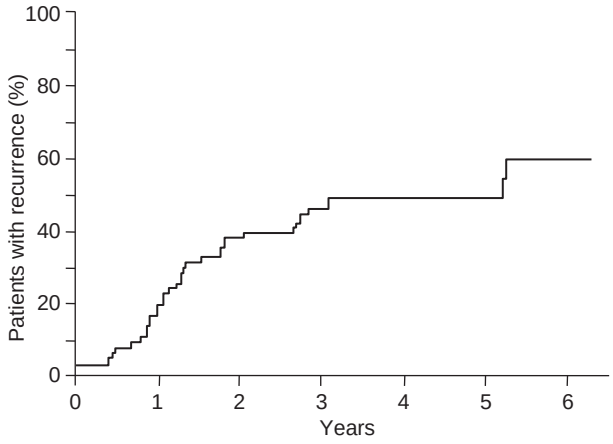


FIGURE 10 Recurrence of Crohn's disease symptoms after surgery. Reprinted from McLeod, R. S., *et al.* (1997). *Gastroenterology* **113**, 1823, with permission from Elsevier.

their disease at some time in their life. The overall goal of surgery for CD is to relieve the symptoms while sparing the bowel. Twenty-two percent of operations on patients with CD are secondary to obstruction. Obstruction in CD is secondary to strictures of the small intestines necessitating either small-bowel resection or stricturoplasties in cases in which bowel length must be preserved.

Because surgery is not curative, select patients should be placed on therapy to prevent a postoperative recurrence. As many as 50% of CD patients will have recurrence following surgical resection and need further surgical intervention (Fig. 10). There have been limited studies on 6-MP/AZA and mesalamine examining postoperative prophylaxis. It is generally believed that patients with severe recurrent surgical disease should be placed on stronger postoperative medications such as 6-MP/AZA, whereas those patients with their first surgery may be placed on mesalamine products alone. Postoperative recurrence rates in patients who were placed on mesalamine were somewhat lower than those for patients placed on placebo. Similar results have been found with the anaerobic antibiotic metronidazole.

EXTRAINTESTINAL MANIFESTATIONS

It is not uncommon for patients with Crohn's disease to have manifestations of this disorder in organs distant to the intestinal tract. These manifestations are not always associated with clinical disease activity, but often the aim of therapy revolves around increased medical therapy for the underlying chronic inflammatory bowel disease. These manifestations are thought to be a consequence of ongoing inflammation in the intestines and can affect any organ, mucosal, or serosal surface. Some of the more common extraintestinal manifestations include pyoderma gangrenosum, erythema nodosum, arthritis, primary sclerosing cholangitis, uveitis, episcleritis, anklylosing spondylitis, thrombosis, pancreatitis, thrombosis, autoimmune hepatitis, diabetes, and osteoporosis.

SUMMARY

The exact etiopathogenesis of CD is unknown but knowledge of the genes and immune effectors playing a role in this disease has dramatically increased. The next few years will witness the identification of more disease susceptibility genes and bacterial factors resulting in CD. This should permit improved treatment for CD that targets specific steps in the inflammatory cascade.

Acknowledgments

This work was supported by National Institutes of Health GI Training Grant T32 DK07180–27.

See Also the Following Articles

Colitis, Ulcerative • Diarrhea • Growth Hormone • TH1, TH2 Responses • Transforming Growth Factor-β (TGF-β) • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Abreu, M. T., Taylor, K. D., Lin, Y.-C., Hang, T., Gaiennie, J., Landers, C. J., *et al.* (2002). Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* **123**, 679–688.

Abreu, M. T., Vasiliauskas, E. A., Kam, L., and Dubinsky, M. (2001). Use of serologic tests in Crohn's disease. *Clin. Perspect. Gastroenterol.*, 155–164.

Binder, V. (1998). Genetic epidemiology in inflammatory bowel disease *Digest. Dis.* **16**, 351–355.

Camma, C., Giunta, M., Rosselli, M., and Cottone, M. (1997). Mesalamine in the maintenance treatment of Crohn's disease: A meta-analysis adjusted for confounding variables. *Gastroenterology* **113**, 1465–1473.

Dubinsky, M. C. and Lamothe, S., Yang, H. Y., Targan, S. R., Sinnett, D., Theoret, Y., *et al.* (2000). Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. *Gastroenterology* **118**, 705–713.

Feagan, B. G., Rochon, J., Fedorak, R. N., Irvine, E. J., Wild, G., Sutherland, L., *et al.* (1995). Methotrexate for the treatment of Crohn's disease. The North American Crohn's Study Group Investigators. *N. Engl. J. Med.* **332**, 292–297.

Greenberg, G. R., Feagan, B. G., Martin, F., Sutherland, L. R., Thomson, A. B., Williams, C. N., *et al.* (1994). Oral budesonide for active Crohn's disease. Canadian Inflammatory Bowel Disease Study Group. *N. Engl. J. Med.* **331**, 836–841.

Hanauer, S. B., Feagan, B. G., Lichtenstein, G. R., Mayer, L. F., Schreiber, S., Colombel, J. F., *et al.* (2002). Maintenance infliximab for Crohn's disease: The ACCENT I randomised trial. *Lancet* **359**, 1541–1549.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., *et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., *et al.* (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606.

Pearson, D. C., May, G. R., Fick, G. H., and Sutherland, L. R. (1995). Azathioprine and 6-mercaptopurine in Crohn disease. A meta-analysis. *Ann. Intern. Med.* **123**, 132–142.

Present, D. H., Korelitz, B. I., Wisch, N., Glass, J. L., Sachar, D. B., and Pasternack, B. S. (1980). Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double-blind study. *N. Engl. J. Med.* **302**, 981–987.

Present, D. H., Rutgeerts, P., Targan, S., Hanauer, S. B., Mayer, L., van Hogezand, R. A., *et al.* (1999). Infliximab for the treatment of fistulas in patients with Crohn's disease. *N. Engl. J. Med.* **340**, 1398–1405.

Quinton, J. F., Sendid, B., Reumaux, D., Duthilleul, P., Cortot, A., Grandbastien, B., *et al.* (1998). Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: Prevalence and diagnostic role. *Gut* **42**, 788–791.

Crohn's Disease, Pediatric

JOHANNA ESCHER AND ATHOS BOUSVAROS
Children's Hospital Boston

Crohn's disease Inflammatory disease of the intestine; may affect any portion of the intestinal tract from the mouth to the anus.

inflammatory bowel disease Refers to the two chronic inflammatory disorders of the gastrointestinal tract, ulcerative colitis and Crohn's disease.

ulcerative colitis Inflammatory disease of the colon and rectum.

Crohn's disease is an immune-mediated inflammatory disease of the intestine that may affect any portion of the intestinal tract from the mouth to the anus. The hallmarks of Crohn's disease that distinguish it from ulcerative colitis are transmural intestinal inflammation, fistulizing disease, and noncaseating granulomas in the bowel. The disease is usually localized and focal, most commonly affecting the ileum alone, the ileum and the cecum, or the ileum and the entire colon. Gastritis and upper intestinal tract inflammation are present in 30% of cases. Although 20% of Crohn's disease cases are diagnosed in patients under age 18 years, the disease is rare in children under age 5 years. Diagnosis is established by barium radiography, colonoscopy, and upper endoscopy with multiple biopsies. Treatment of the child and adolescent with the disease involves a combination of medical therapy, nutritional supplementation, properly timed surgery, and psychological support.

PATHOPHYSIOLOGY AND EPIDEMIOLOGY

Although the exact pathophysiology of Crohn's disease (CD, regional enteritis) has not been fully defined, there is both a genetic and an environmental component. Evidence supporting a genetic component includes the increased risk of developing CD among first-degree relatives of an index case (approximately 5–10%) and the high concordance for CD among monozygotic twins (approximately 50%). Mutations in the human macrophage *NOD2* gene, which regulates the immune response to bacterial lipopolysaccharide, are present in 25% of patients with CD. An individual inheriting one mutant *NOD2* allele has a 2.6-fold increase in risk of developing CD, and an individual inheriting two mutant *NOD2* alleles has a 40-fold increase in risk. Mutations in the *NOD2* gene are most commonly identified in patients who develop ileal disease and

 Copyright 2004, Elsevier (USA). All rights reserved.

Hanauer, S. B., Feagan, B. G., Lichtenstein, G. R., Mayer, L. F., Schreiber, S., Colombel, J. F., *et al.* (2002). Maintenance infliximab for Crohn's disease: The ACCENT I randomised trial. *Lancet* **359**, 1541–1549.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., *et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., *et al.* (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606.

Pearson, D. C., May, G. R., Fick, G. H., and Sutherland, L. R. (1995). Azathioprine and 6-mercaptopurine in Crohn disease. A meta-analysis. *Ann. Intern. Med.* **123**, 132–142.

Present, D. H., Korelitz, B. I., Wisch, N., Glass, J. L., Sachar, D. B., and Pasternack, B. S. (1980). Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double-blind study. *N. Engl. J. Med.* **302**, 981–987.

Present, D. H., Rutgeerts, P., Targan, S., Hanauer, S. B., Mayer, L., van Hogezand, R. A., *et al.* (1999). Infliximab for the treatment of fistulas in patients with Crohn's disease. *N. Engl. J. Med.* **340**, 1398–1405.

Quinton, J. F., Sendid, B., Reumaux, D., Duthilleul, P., Cortot, A., Grandbastien, B., *et al.* (1998). Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: Prevalence and diagnostic role. *Gut* **42**, 788–791.

Crohn's Disease, Pediatric

JOHANNA ESCHER AND ATHOS BOUSVAROS
Children's Hospital Boston

Crohn's disease Inflammatory disease of the intestine; may affect any portion of the intestinal tract from the mouth to the anus.

inflammatory bowel disease Refers to the two chronic inflammatory disorders of the gastrointestinal tract, ulcerative colitis and Crohn's disease.

ulcerative colitis Inflammatory disease of the colon and rectum.

Crohn's disease is an immune-mediated inflammatory disease of the intestine that may affect any portion of the intestinal tract from the mouth to the anus. The hallmarks of Crohn's disease that distinguish it from ulcerative colitis are transmural intestinal inflammation, fistulizing disease, and noncaseating granulomas in the bowel. The disease is usually localized and focal, most commonly affecting the ileum alone, the ileum and the cecum, or the ileum and the entire colon. Gastritis and upper intestinal tract inflammation are present in 30% of cases. Although 20% of Crohn's disease cases are diagnosed in patients under age 18 years, the disease is rare in children under age 5 years. Diagnosis is established by barium radiography, colonoscopy, and upper endoscopy with multiple biopsies. Treatment of the child and adolescent with the disease involves a combination of medical therapy, nutritional supplementation, properly timed surgery, and psychological support.

PATHOPHYSIOLOGY AND EPIDEMIOLOGY

Although the exact pathophysiology of Crohn's disease (CD, regional enteritis) has not been fully defined, there is both a genetic and an environmental component. Evidence supporting a genetic component includes the increased risk of developing CD among first-degree relatives of an index case (approximately 5–10%) and the high concordance for CD among monozygotic twins (approximately 50%). Mutations in the human macrophage *NOD2* gene, which regulates the immune response to bacterial lipopolysaccharide, are present in 25% of patients with CD. An individual inheriting one mutant *NOD2* allele has a 2.6-fold increase in risk of developing CD, and an individual inheriting two mutant *NOD2* alleles has a 40-fold increase in risk. Mutations in the *NOD2* gene are most commonly identified in patients who develop ileal disease and

 Copyright 2004, Elsevier (USA). All rights reserved.

fibrostenosing disease. Genetic linkage studies have identified several other candidate genes, including a locus on chromosome 5 (5q31–q33) associated with early-onset CD. Otherwise, genotype–phenotype correlations have not yet been demonstrated in the other candidate loci.

Although genetic loci transmit risk for the disease, environmental factors appear to influence the development of disease. Environmental triggers influencing the onset of disease include dietary factors, exogenous microbial pathogens such as atypical mycobacteria, the host's endogenous bacterial flora, medications (nonsteroidal antiinflammatory drugs and oral contraceptives), and other environmental toxins (tobacco smoke). Studies have failed to identify any consistent environmental triggers that promote the onset of CD, although use of nonsteroidal antiinflammatory drugs may be a risk factor for relapse.

Interaction of the predisposed individual with the environment leads to an activated mucosal immune system characterized by a predominance of CD4+ T helper (Th) cells of the Th1 phenotype, and by activated macrophages. Proinflammatory cytokines such as tumor necrosis factor α, interleukin-1, interleukin-6, and interleukin-12 are present in the intestinal mucosa. The activated immune cells then recruit additional leukocytes, with the end result being the production of a wide variety of inflammatory mediators of granulomatous inflammation and tissue ulceration.

The incidence of Crohn's disease is approximately 5–10 new cases/100,000 individuals per year, and is similar in children and adults. Epidemiologic studies from Sweden suggest a doubling of disease incidence in the past decade. The disease incidence is higher in northern latitudes (e.g., North America and Scandinavia), and in individuals of Jewish descent. Genetic syndromes associated with CD in children include Hermansky–Pudlak syndrome, glycogen storage disease type 1b, and Turner's syndrome. It is unclear whether the diseases seen in these syndromes are genetically similar to idiopathic CD.

DIAGNOSIS AND EVALUATION

The child or adolescent with Crohn's disease will typically present in one of two ways. Occasionally, children present acutely with bloody diarrhea or severe abdominal pain from severe ileitis, intestinal stricture, or abdominal abscess. Given that the most common area of abdominal tenderness in patients with CD is the right lower quadrant, patients with CD presenting in emergency room settings are often thought to have acute appendicitis. More commonly, children present with a chronic history of low-grade abdominal pain, fever, weight loss, recurrent perianal infections, or growth failure. Up to 50% of children with Crohn's disease will have a decrease in rate of growth (height) prior to the onset of any intestinal symptoms. The general physical examination may reveal an underweight child with poorly localized abdominal tenderness. Perianal skin tags, fissures, or fistulae will be identified in 30% of patients. Other physical findings identified may include eye inflammation, mouth ulcers, erythema nodosum, digital clubbing, and arthritis, but these are seen in fewer than 10% of patients at presentation. Laboratory studies of children with CD will commonly demonstrate a microcytic anemia, hypoalbuminemia, or an elevated erythrocyte sedimentation rate or C-reactive protein. Antibodies to neutrophilic and microbial antigens, including perinuclear antineutrophil cytoplasmic antibodies (pANCAs) and anti-*Saccharomyces cerevisiae* antibodies (ASCAs) can be utilized as a serologic "screen" for inflammatory bowel disease (IBD) in children, but use of these tests is limited by their 10% false positive rate. Although a positive result for antibody to *S. cerevisiae* is found in only 40–50% of children with Crohn's disease, this antibody is highly specific (95–98%) for CD.

When CD is suspected, other conditions causing similar symptoms should be ruled out. In some children, it may be difficult to distinguish CD from ulcerative colitis (UC), and these children are given the diagnosis of indeterminate colitis. Other illnesses that may present with rectal bleeding can be excluded by obtaining stool cultures for enteric pathogens and *Clostridium difficile*, and by considering other diagnoses, such as Henoch–Schonlein purpura, Behcet's disease, or vasculitis. Localized right lower quadrant or ileal disease can be caused by *Yersinia* infection, intestinal tuberculosis, appendicitis, or lymphoma. The differential diagnosis of an abdominal abscess includes a perforated appendix, vasculitic perforation, malignant typhilitis, or trauma. In adolescent females, gynecologic disease must be considered.

The definitive diagnosis of Crohn's disease and the localization of the disease are established with radiography and endoscopy. A barium upper gastrointestinal (GI) exam with small bowel follow through may identify mucosal nodularity, ileal and cecal narrowing, enteric fistula, or obstruction. In children, it is recommended that both colonoscopy and upper endoscopy with multiple biopsies be performed under general anesthesia as part of the initial evaluation. Crohn's disease in the colon is characterized by discontinuous colitis with normal intervening areas (skip areas), as well as superficial and deep ulcers. Endoscopy of the terminal

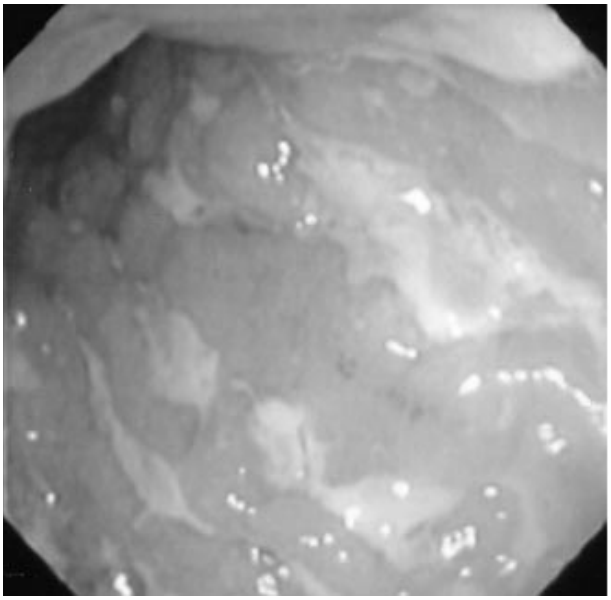


FIGURE 1 Severe ileitis with linear ulcerations and nodularity in a child with newly diagnosed Crohn's disease.

ileum may identify linear ulcerations with nodularity (cobblestoning) or ileal stenosis (Fig. 1). In patients with CD of the upper GI tract, esophagogastroduodenoscopy will identify duodenal erosions ("moth-eaten" folds) in children with duodenitis, or gastric erythema, nodularity, and aphthous lesions in those with gastritis. Histology of biopsy specimens can show nonspecific inflammation (neutrophilic colitis and gastritis), but will identify noncaseating granulomas in 25% of patients. The most reliable findings that differentiate CD from UC are the presence of small bowel involvement, perianal fistulizing disease, or granulomas on endoscopic biopsy. In 10% of cases, the patient may be given a diagnosis of "indeterminate colitis" if the clinician cannot differentiate CD from UC after complete evaluation.

Once diagnosis and disease location have been established, the child should be evaluated for other conditions associated with Crohn's disease. These may include dermatologic conditions (erythema nodosum and pyoderma gangrenosum), ocular involvement (uveitis and episcleritis), osteoporosis, and arthritis. Arthritis in CD is typically nonerosive, asymmetric, and affects large joints, including the hips, knees, and wrists. Abnormal liver function tests are most often caused by medications used to treat CD (e.g., 6-mercaptopurine) or hepatic steatosis. However, primary sclerosing cholangitis, an immune-mediated disease causing scarring and stenosis of the biliary tree, is present in 1–2% of patients with CD.

TREATMENT

Therapy for inflammatory bowel disease (ulcerative colitis and Crohn's disease) is indicated for induction of remission of disease activity, maintenance of remission, and prevention of relapse. In children and adolescents with IBD, the aim of therapy should include the eradication of abdominal symptoms, normalization of laboratory parameters such as hematocrit and sedimentation rate, and establishment of normal growth and pubertal development. Because the disease has a potential for long-term impact on physical growth and psychosocial development, psychosocial support is a necessary component of the treatment program. The unique therapeutic features of pediatric-onset inflammatory bowel disease are listed in Table I.

The three primary treatment modalities for pediatric Crohn's disease are medication, enteral nutrition, and surgery. In general, initial treatment consists of medication combined with nutritional treatment or supplementation. Nutritional therapy may be used as the primary initial treatment in active Crohn's disease in children or as an adjunct to medical therapy in the management of weight loss or growth failure. If the disease is medically refractory, surgery of a limited segment of bowel may eradicate all symptoms and facilitate catch-up growth during puberty.

Medication

Few clinical trials have focused on children with CD, and medical practice is mostly based on evidence from studies of adult CD patients. Dosages are extrapolated from adult dosages and adjusted according to body weight or body surface area. In addition, the duration of treatment is largely determined by empiric assessment of response and from adult studies.

There are five categories of current drug treatment options: aminosalicylates, corticosteroids, antibiotics, immunomodulators (azathioprine/6-mercaptopurine), and monoclonal antibodies (infliximab). The choice of treatment depends on the disease activity and behavior, and whether induction or maintenance of remission is the goal. Typically, induction of remission involves medications with greater efficacy and a more rapid onset of action, but a less favorable side-effect profile. In contrast, maintenance therapy involves long-term treatment with medications that have a slower onset, but fewer side effects. Table I lists indications, dosing information and efficacy rates for the medications that are commonly used.

TABLE I Medication Used in Pediatric Crohn's Disease

Treatment	Dose	Clinical response
Induction of remission		
Salicylates	Mesalazine, oral and/or rectal, 50–75 mg/kg/day in two to four divided doses, maximum 6 g/day	40–60% remission
	Sulfasalazine, oral and/or rectal, 50–100 mg/kg/day in three to four divided doses, maximum 4–6 g/day	40–60% remission
Steroids	Predniso(lo)ne, intravenously if needed, then oral, 1–2 mg/kg/day in one to two divided doses, maximum 40–60 mg/day, 2–4 weeks, then taper to zero slowly using oral agents	60–80% remission
	Budesonide, oral, 9 mg/day in one dose, 8 weeks, then taper to zero	50% remission
Antibiotics	Metronidazole, oral, 15–30 mg/kg/day in three to four divided doses, maximum 1500 mg/day	Not well studied
	Ciprofloxacin, oral, 20 mg/kg/day in two divided doses	Not well studied
Immunomodulator	Infliximab (anti-TNF), infusion, 5 mg/kg (specific protocol must be followed)	60–80% remission
Maintenance of remission		
Immunomodulator	Azathioprine, oral 1.5–2.5 mg/kg/day	60–80% maintenance of remission without steroids
	6-Mercaptopurine, oral, 1–2 mg/kg/day	60–80% maintenance of remission without steroids
	Methotrexate, subcutaneously, 10–15 mg/m^2/week, increase if necessary, maximum 25 mg/week	40% maintenance of remission without steroids
	Infliximab (anti-TNF), infusion, 5 mg/kg, every 8 weeks or on individual basis; fistulae, repeat weeks 2 and 6 (specific protocol must be followed)	Not well studied

Induction of Remission

In moderate to severely active disease and in hospitalized patients, medium-dose systemic corticosteroids will rapidly reduce pain and diarrhea and bring about weight gain. Prednisone or prednisolone treatment can generally be given orally, but in ill patients requiring hospitalization, methylprednisolone is preferred. Approximately 70–80% of patients will experience improvement of symptoms on prednisone therapy, usually within 2–4 weeks after therapy is started. However, corticosteroid side effects, including hypertension, hyperglycemia, infection, moon face, mood changes, osteopenia, and growth arrest, limit the duration of prednisone usage. In addition, many patients will experience a recurrence of symptoms once prednisone administration is stopped. For these reasons, corticosteroid use should ideally be limited to short-term induction of remission, and a medication useful for maintenance of remission should be started simultaneously. Budesonide is a locally acting corticosteroid that has minimal systemic bioavailability. Budesonide induces a remission in approximately 60% of children with mild to moderate Crohn's disease located in the ileum, cecum, and/or ascending colon, with fewer side effects than conventional corticosteroids. As with prednisone, maintenance budesonide treatment is not indicated because of the potential for impairment of linear growth and osteopenia.

Aminosalicylates and Antibiotics

Because these agents have a favorable safety profile with few side effects, physicians have utilized them for both induction and maintenance therapy. Aminosalicylates utilized in the treatment of ulcerative colitis include mesalamine, sulfasalazine, and balsalazide. Although all of these have been studied and have received United States Food and Drug Administration (FDA) approval for ulcerative colitis, they may also be effective in the treatment of Crohn's colitis. In addition, two formulations of mesalamine (Pentasa and Asacol) release drug into the distal small bowel and have been studied as induction therapies in Crohn's disease. In mild disease, the aminosalicylates may induce remission, but the results of meta-analyses in adults indicate that this occurs at a rate only slightly greater than that of placebo. Side effects of aminosalicylates, although rare, include proteinuria, skin rashes, headaches, and pancreatitis. Similarly, although open-label series and a few randomized

trials suggest a role for ciprofloxacin or metronidazole in treatment of small bowel or colonic Crohn's disease, the efficacy of antibiotic treatment is limited. One important role for antibiotic treatment is in the therapy of acute abdominal infections or chronic perianal fistulae. The principal side effects of metronidazole therapy include nausea, vomiting, a disulfiram-like reaction if the drug is taken with alcohol, and peripheral neuropathy with long-term use.

Enteral Nutrition

Enteral nutrition (as a 6-week course of exclusive elemental or polymeric formula, ingested orally or by nasogastric tube) can induce remission in 60–80% of children with CD. With enteral feeding, the patient gains weight, micronutrient deficiencies are corrected, and growth and pubertal development are promoted, while avoiding the systemic toxicity of corticosteroid therapy. Disadvantages with use of enteral feeds as primary induction therapy include the relapse of disease once patients start eating again, the poor palatability of these formulas (making oral ingestion difficult), and the discomfort and social stigma associated with nasogastric tube feeding. Elemental diet administration via nasogastric tube is commonly utilized as induction therapy for children with Crohn's disease in Europe; in contrast, corticosteroids are more commonly used as primary induction therapy in the United States.

Infliximab

Infliximab, a chimeric antibody to tumor necrosis factor, is very effective in inducing clinical remission within 1–2 weeks. Treatment is given in the hospital by intravenous infusion. This biologic agent has now been evaluated in several open-label noncontrolled studies of children and adolescents with severe, treatment-resistant Crohn's disease. Efficacy seems to be similar in children and adults, with remission rates of approximately 65–80% after a single infusion. In patients who respond, remission can be maintained by administration of repeat infusions, and a commonly utilized treatment protocol involves the administration of three initial infusions at 0, 2, and 6 weeks, followed by infusions every 2 months. In children with perianal fistulae refractory to antibiotics and 6-mercaptopurine, infliximab therapy may result in cessation of fistulous drainage with closure of the tract. Adverse effects include severe infusion reactions (e.g., chest pain, rashes, or anaphylaxis) and increased risk of infections (herpes zoster or tuberculosis). At the present time, infliximab treatment is primarily reserved for cases of refractory disease in patients who have persistently active disease despite therapy with aminosalicylates, corticosteroids, azathioprine, or 6-mercaptopurine.

MAINTENANCE OF REMISSION

The majority of patients with Crohn's disease located in the terminal ileum or colon, aminosalicylates may be effective in inducing and subsequently maintaining remission. Meta-analysis of a large series of controlled randomized trials suggests a slight superiority over placebo. High doses of aminosalicylates (up to 80 mg/kg/day) may be required to see this effect.

In patients who continue to have active disease or growth failure, early introduction of azathioprine (AZA) or its metabolite 6-mercaptopurine (6-MP) is advocated as maintenance treatment. These medications inhibit lymphocyte proliferation, may induce antiinflammatory cytokine production, and promote mucosal healing. Studies suggest that 6-MP and AZA are effective steroid sparing agents in approximately 65–75% of adults and children with steroid-refractory or steroid-dependent disease. In addition, a randomized placebo-controlled trial of 6-MP in children with newly diagnosed Crohn's disease demonstrated clear efficacy in relapse prevention and reduction of cumulative corticosteroid dose. However, the efficacy of AZA or 6-MP may not be seen until 3–6 months after the institution of therapy. For this reason, these drugs are not effective induction therapies, and must be combined with some other treatment in the acutely ill patient. Measurement of erythrocyte 6-thioguanine (6-TG) levels may help optimize treatment. In addition, *a priori* determination of thiopurine methyltransferase (the principal enzyme metabolizing 6-MP) activity prior to the institution of therapy may help guide dosing and minimize the risk of myelosuppression.

Adverse effects of 6-MP/AZA, although infrequent, are potentially serious. Mild gastrointestinal side effects may occur in 20–30% of pediatric patients, and hypersensitivity reactions (pancreatitis, high fevers, rash, and arthralgia) may develop in 5–10% in the first weeks of therapy. Aminotransferase elevation (more than twice normal) can be found in almost 15% of patients and is associated with high levels of 6-methylmercaptopurine, the principal inactive metabolite of 6-MP. Bone marrow suppression and hepatotoxicity are reversible with dose reduction, but require frequent blood monitoring, especially during the early phases of treatment. Opportunistic infections such as varicella or herpes zoster may occur. In adults, continuation of maintenance treatment has been shown to be useful even after 4 years. Whether AZA and 6-MP will be associated with an increase in malignancy in pediatric patients who undergo

long-term therapy is unknown, but most adult studies suggest little if any increase in risk.

If azathioprine/6-MP treatment fails, in case of steroid dependency or toxicity, repeat infliximab infusions are indicated. AZA or 6-MP treatment is generally continued as maintenance treatment. In children with refractory Crohn's disease of short (less than 2 years) duration, the clinical response to one infusion of infliximab (5 mg/kg) lasts longer than it does in patients with "late" Crohn's disease. An alternative to repeated infliximab infusions for refractory patients is conversion to therapy with methotrexate.

OTHER MEDICATIONS FOR REFRACTORY DISEASE

In Crohn's disease, methotrexate (MTX) may serve as an alternative to azathioprine and 6-MP, if these drugs are not tolerated or are ineffective. Experience and literature are limited for application to Crohn's disease in children, but controlled trials with adults demonstrate superiority of MTX over placebo both in induction and maintenance of remission. Parenteral administration is preferred, because gastrointestinal absorption is variable in IBD patients. MTX is a folate antagonist, thus concomitant administration of folic acid is advised. Side effects include oral ulcers, myelosuppression, and increased risk of infection. Two potential concerns related to chronic methotrexate use are pulmonary fibrosis and hepatoxicity with fibrosis (or even cirrhosis). However, experience with long-term MTX therapy for rheumatoid arthritis patients indicates that severe hepatotoxicity is rare.

Thalidomide is an inhibitor of tumor necrosis factor α (TNFα) production and of angiogenesis. Small open-label trials suggest efficacy in adults and children with steroid-dependent Crohn's disease, including children with refractory oral aphthous ulcers. Side effects include drowsiness, leukopenia, and peripheral neuropathy. Use of this drug is restricted because of the enormous risk of birth defects if the drug is ingested by pregnant women.

Cyclosporine may serve as a rescue treatment in children with treatment-resistant colitis, with rapid onset of action within 1–2 weeks during high-dose intravenous administration. A placebo-controlled trial has demonstrated efficacy of high-dose cyclosporine as an induction agent, but subsequent controlled trials demonstrated that low-dose cyclosporine does not maintain remission. Tacrolimus (FK-506), a drug with a mechanism of action similar to that of cyclosporine, is well absorbed after oral dosing. Topical tacrolimus has been utilized in the treatment of perianal disease, and oral tacrolimus has been utilized in the treatment of severe colitis and fistulizing perianal Crohn's disease. The use of cyclosporine or tacrolimus is limited by their side-effect profile, which includes nephrotoxicity, myalgia, headache, hyperglycemia, and risk of infection.

NUTRITIONAL MANAGEMENT AND SUPPORT

Growth failure represents a common, serious complication unique to the pediatric age group of IBD patients. Nutritional deficiencies, caused by inadequate dietary intake in relation to overall nutrient requirements, appear to be a major factor related to growth failure in children and adolescents with Crohn's disease. In addition to its positive effect on growth, nutritional therapy has been advocated as primary therapy for disease activity in these children. Continuation of enteral nutrition in a nocturnal regimen may promote maintenance of remission. There is no consensus on whether elemental or polymeric liquid formulas are preferred in induction of remission. Although the exact mechanism of the beneficial effect of nutrition on CD is unknown, proposed actions include restoration of anabolism, decreased gut motility, reduction of antigenic load, and changes in bowel flora. In addition, the liquid nature of the diet (and its ease of transport through the diseased and/or narrowed small bowel) may be responsible for the effect. In severely painful perianal disease, elemental feeding can minimize fecal output while maintaining good nutritional status.

When used as primary treatment of active disease, the elemental or polymeric formula is administered either orally or by nasogastric tube for 6 weeks. Calorie and protein intake should equal 150% or more of the required daily intake. If this is not tolerated, a semielemental (oligomeric) formula may provide relief. Except for clear liquids, no other food or drink is allowed. The child can receive nutritional treatment at home and is able to attend school. After 6 weeks, a normal diet is reintroduced while nocturnal formula administration is continued; the recommended regimen is 5 nights per week, because this may prolong the duration of remission.

NUTRITIONAL THERAPY OF NONMALNOURISHED OUTPATIENTS

Many adults and children with CD feel well on medical therapy, do not require nasogastric tube feeding, and are either normal weight or slightly underweight. Although

clinically asymptomatic, they remain at risk for nutritional complications of IBD and benefit from dietary counseling. There is no specific diet universally recommended for patients with CD and UC. Low-residue, low-fiber diets may be palliative for patients with narrowed regions of bowel (i.e., CD patients with inflammatory or fibrotic strictures). Fish oil capsules may have some efficacy in the prevention of relapse, because they may inhibit release of leukotrienes in the bowel. Patients should take a multivitamin because of reports of vitamin A, D, and E deficiencies. Patients on maintenance sulfasalazine should receive at least 1 mg of folic acid per day, because sulfasalazine interferes with luminal folate absorption. Adolescents should ingest 1200−1500 mg of total daily calcium, either through their diet or as supplements.

SURGERY

Surgery is primarily reserved for Crohn's disease patients who do not respond to medications or who develop complications (abscess, fistula, or stricture) that can only be treated surgically. In one study of 204 children with Crohn's disease from 1968 to 1994, 46% of patients with CD required surgery for complications within 3 years after diagnosis. The primary indications for surgery were medically refractory disease, intestinal abscess, or stricture. Surgery is most effective in patients with a limited region of disease (e.g., terminal ileal or ileocecal Crohn's disease) and usually involves resection of a diseased bowel segment and anastomosis. In some patients with multiple small bowel strictures, strictureplasty (widening of the strictured area) may eliminate the need for bowel resection. In patients with extensive Crohn's colitis, full colonic resection and permanent end-ileostomy may be necessary. After a bowel resection and anastomosis, Crohn's disease recurs in approximately 50% of patients within 5 years, with the most common site of disease resection being at the anastomosis. Postoperative prophylactic therapy, especially with 6-mercaptopurine, may delay the onset of disease recurrence; mesalamine and metronidazole may have some short-term efficiency as well.

PSYCHOSOCIAL SUPPORT

The chronic relapsing nature of CD and its profound effects on body appearance and image may lead to anxiety, depression, and psychosocial dysfunction in a subset of patients. Children with CD may be shorter than their classmates in high school because of growth failure and pubertal delay. The presence of surgical scars, nasogastric tubes, or ostomies may also lead to estrangement from their peers. Many children and teenagers develop "medication fatigue" from taking over 10 pills a day for several years. Family conflicts may arise over a child's poor appetite, especially if parents do not recognize that the disease is causing a child's anorexia. Although psychological counseling is often beneficial, many teenagers are either "too busy" or are reluctant to undergo counseling. Thus, the pediatrician and pediatric gastroenterologist need to address the above issues openly during outpatient visits. Discussion groups and summer camp programs provided by organizations such as the Crohn's and Colitis Foundation of America (**www.ccfa.org**) may help restore a child's self-esteem and provide support from adults and peers with the disease.

CANCER RISK

Patients with ulcerative and Crohn's colitis are at increased risk for the development of colon cancer. In general, the risk is minimal in patients who have had their disease for less than 10 years. However, rare cases of colon cancer have been reported in adolescents with inflammatory bowel disease. Therefore, children with Crohn's colitis involving the majority of the colon should probably be enrolled in colon cancer surveillance programs no later than 10 years after the onset of disease.

See Also the Following Articles

Colitis, Ulcerative (Pediatirc) • Diarrhea, Pediatric • Growth Hormone • TH1, TH2 Responses • Transforming Growth Factor-β (TGF-β) • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Dubinsky, M. C., Lamothe, S., Yang, H. Y., Targan, S. R., Sinnett, D., Theoret, Y., and Seidman, E. G. (2000). Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. Gastroenterology **118**(4), 705−713.

Faubion, W., Loftus, E., Harmsen, W., *et al.* (2001). The natural history of corticosteroid therapy for inflammatory bowel disease: A population-based study. *Gastroenterology* **121**, 255.

Gisbert, J. P., Gomollon, F., Mate, J., and Pajares, J. M. (2002). Role of 5-aminosalicylic acid (5-ASA) in treatment of inflammatory bowel disease: A systematic review. *Dig. Dis. Sci.* **47**, 471.

Griffiths, A., Ohlsson, A., Sherman, P., and Sutherland, L. (1995). Meta-analysis of enteral nutrition as a primary treatment of active Crohn's disease. *Gastroenterology* **108**, 1056.

Hyams, J. S., Markowitz, J., and Wyllie, R. (2000). Use of infliximab in the treatment of Crohn's disease in children and adolescents. *J. Pediatr.* **137**, 192−196.

Kundhal, P., Zachos, M., Holmes, J. L., *et al.* (2001). Controlled ileal release budesonide in pediatric Crohn disease: Efficacy and effect on growth. *J. Pediatr. Gastroenterol. Nutr.* **33**, 75–80.
Markowitz, J., Grancher, K., Kohn, N., *et al.* (2000). A multicenter trial of 6-mercaptopurine and prednisone in children with newly diagnosed Crohn's disease. *Gastroenterology* **119**, 895–902.
Motil, K. J., and Grand, R. J. (1996). Inflammatory bowel disease. *In* "Nutrition in Pediatrics" (W. A. Walker and J. B. Watkins, eds.), 2nd Ed., pp. 516–533. B. C. Decker Inc., Hamilton, Ontario.
Patel, H. I., Leichtner, A. M., Colodny, A. H., and Shamberger, R. C. (1997). Surgery for Crohn's disease in infants and children. *J. Pediatr. Surg.* **32**, 1063.

Cryptosporidium

DEREK A. MOSIER
Kansas State University

anthroponosis A cycle of infection characterized by transmission between humans.
monoxenous Requires only one host to complete the entire life cycle.
zoonosis A cycle of infection characterized by transmission between animals and humans.

Cryptosporidiosis is a diarrheal disease caused by the protozoan *Cryptosporidium*. The most common manifestation of infection is transient diarrhea in immunocompetent infants and adults. However, infection in immunosuppressed persons can cause persistent and life-threatening disease.

ETIOLOGY AND EPIDEMIOLOGY

Cryptosporidium parvum is the most common etiology of mammalian cryptosporidiosis. An anthroponotic genotype is responsible for most human cases (Fig. 1). However, zoonotic genotypes also cause disease in humans and various other species, particularly cattle. The *Cryptosporidium* life cycle is similar to that of other monoxenous coccidia, with the added features of recycling and amplification of asexual stages, production of autoinfective thin-walled oocysts, and the lack of a requirement for sporulation of oocysts in the environment (Fig. 1). Therefore, small numbers of oocysts (as few as 10–100) can induce severe infection and shedding of large numbers of oocysts. Sporadic disease occurs by contact with infected persons or animals, such as in daycare centers or animal facilities. Disease outbreaks are most often associated with drinking contaminated water. The incidence of infection in the general population is approximately 1–2%, with a seroprevalence rate of 25–35%. Approximately 3–5% of immunosuppressed populations, particularly persons with acquired immunodeficiency syndrome (AIDS), are infected. Cryptosporidiosis occurs with greatest prevalence in less developed countries. Incidence rates of 8–9% in infants and up to 50% in AIDS patients are common in these countries.

CLINICAL FEATURES

First described in 1976, human cryptosporidiosis is now recognized worldwide as an important cause of diarrhea. Infection is most common in 1- to 5-year-old infants, but can occur in persons of any age. Following an approximately 5- to 7-day incubation, there is watery diarrhea, often accompanied by fever, nausea, vomiting, abdominal cramps, weight loss, and anorexia. Symptoms subside spontaneously after 1–2 weeks; however, asymptomatic shedding of oocyts can continue for several more weeks.

In AIDS patients and other immunosuppressed individuals, diarrhea can be intractable and contribute to fatality due to dehydration and cachexia. The stool may contain over 10^{10} oocysts and fluid loss often exceeds 10 liters daily. Spontaneous remission in immunosuppressed persons is rare.

Copyright 2004, Elsevier (USA). All rights reserved.

Kundhal, P., Zachos, M., Holmes, J. L., *et al.* (2001). Controlled ileal release budesonide in pediatric Crohn disease: Efficacy and effect on growth. *J. Pediatr. Gastroenterol. Nutr.* **33**, 75–80.
Markowitz, J., Grancher, K., Kohn, N., *et al.* (2000). A multicenter trial of 6-mercaptopurine and prednisone in children with newly diagnosed Crohn's disease. *Gastroenterology* **119**, 895–902.
Motil, K. J., and Grand, R. J. (1996). Inflammatory bowel disease. *In* "Nutrition in Pediatrics" (W. A. Walker and J. B. Watkins, eds.), 2nd Ed., pp. 516–533. B. C. Decker Inc., Hamilton, Ontario.
Patel, H. I., Leichtner, A. M., Colodny, A. H., and Shamberger, R. C. (1997). Surgery for Crohn's disease in infants and children. *J. Pediatr. Surg.* **32**, 1063.

Cryptosporidium

DEREK A. MOSIER
Kansas State University

anthroponosis A cycle of infection characterized by transmission between humans.
monoxenous Requires only one host to complete the entire life cycle.
zoonosis A cycle of infection characterized by transmission between animals and humans.

Cryptosporidiosis is a diarrheal disease caused by the protozoan *Cryptosporidium*. The most common manifestation of infection is transient diarrhea in immunocompetent infants and adults. However, infection in immunosuppressed persons can cause persistent and life-threatening disease.

ETIOLOGY AND EPIDEMIOLOGY

Cryptosporidium parvum is the most common etiology of mammalian cryptosporidiosis. An anthroponotic genotype is responsible for most human cases (Fig. 1). However, zoonotic genotypes also cause disease in humans and various other species, particularly cattle. The *Cryptosporidium* life cycle is similar to that of other monoxenous coccidia, with the added features of recycling and amplification of asexual stages, production of autoinfective thin-walled oocysts, and the lack of a requirement for sporulation of oocysts in the environment (Fig. 1). Therefore, small numbers of oocysts (as few as 10–100) can induce severe infection and shedding of large numbers of oocysts. Sporadic disease occurs by contact with infected persons or animals, such as in daycare centers or animal facilities. Disease outbreaks are most often associated with drinking contaminated water. The incidence of infection in the general population is approximately 1–2%, with a seroprevalence rate of 25–35%. Approximately 3–5% of immunosuppressed populations, particularly persons with acquired immunodeficiency syndrome (AIDS), are infected. Cryptosporidiosis occurs with greatest prevalence in less developed countries. Incidence rates of 8–9% in infants and up to 50% in AIDS patients are common in these countries.

CLINICAL FEATURES

First described in 1976, human cryptosporidiosis is now recognized worldwide as an important cause of diarrhea. Infection is most common in 1- to 5-year-old infants, but can occur in persons of any age. Following an approximately 5- to 7-day incubation, there is watery diarrhea, often accompanied by fever, nausea, vomiting, abdominal cramps, weight loss, and anorexia. Symptoms subside spontaneously after 1–2 weeks; however, asymptomatic shedding of oocyts can continue for several more weeks.

In AIDS patients and other immunosuppressed individuals, diarrhea can be intractable and contribute to fatality due to dehydration and cachexia. The stool may contain over 10^{10} oocysts and fluid loss often exceeds 10 liters daily. Spontaneous remission in immunosuppressed persons is rare.

Copyright 2004, Elsevier (USA). All rights reserved.

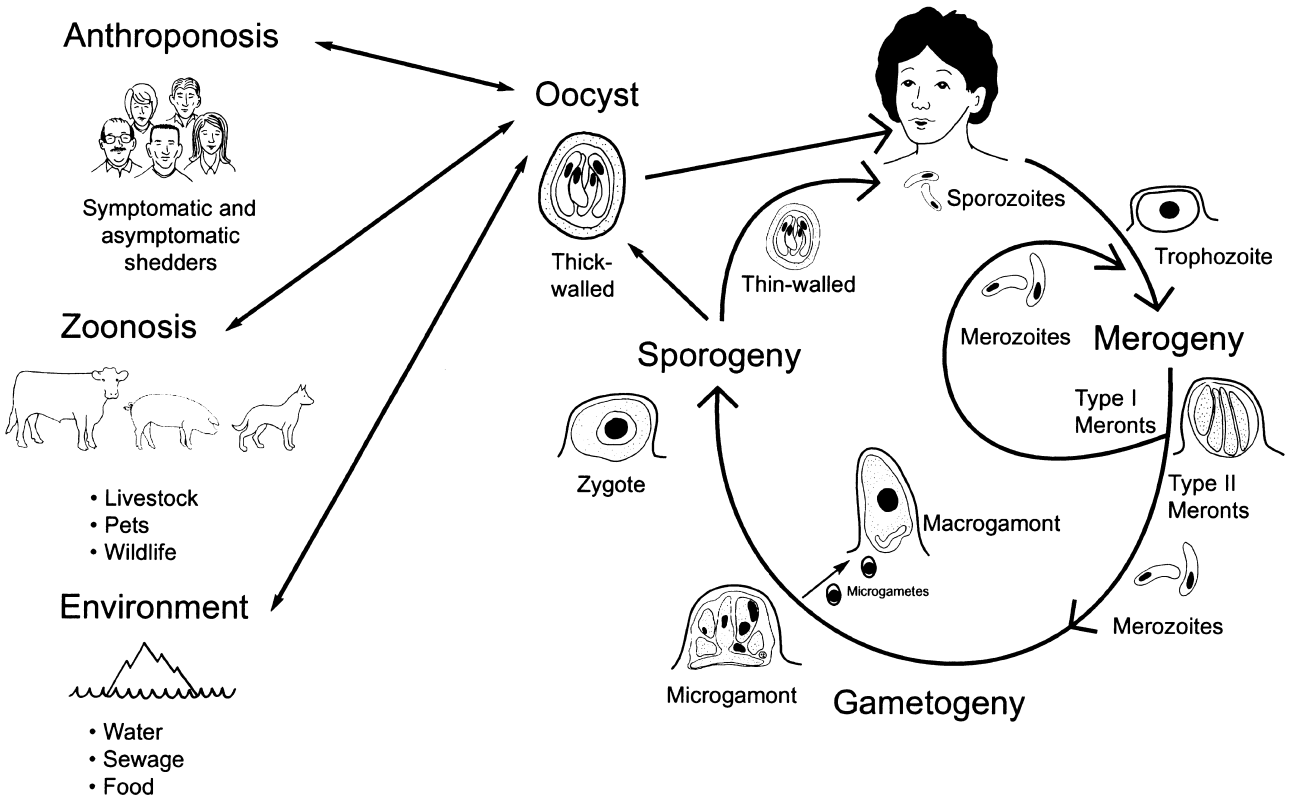


FIGURE 1 Sources of infection (left) and the life cycle of *Cryptosporidium* within an infected host (right). Important features of the life cycle include the amplification of infection by recycling of type I merozoites to produce additional type I meronts and the production of an autoinfective thin-walled oocyst in addition to the excreted thick-walled oocyst.

PATHOGENESIS AND IMMUNITY

Cryptosporidium are intracellular, extracytoplasmic epithelial cell parasites. Enteric infection causes villous atrophy with minimal inflammation in the lower jejunum and ileum. Diarrhea results from the decreased absorptive area and prostaglandin-induced inhibition of NaCl absorption in the villi accompanied by increased Cl^- secretion in the crypts. Gastrointestinal infection in immunosuppressed persons can affect the entire tract from the esophagus to the colon, including the biliary and pancreatic duct systems.

Cell-mediated responses are critical for preventing or resolving *Cryptosporidium* infection. Th-1 type $CD4^+$ T cells activate macrophages and cytotoxic $CD8^+$ T cells that help mediate protection. $CD4^+$ T cells may also interact with infected enterocytes to induce apoptosis and nitric oxide production. Interferon-γ prevents epithelial invasion and may inhibit intracellular development of *Cryptosporidium* stages. It also up-regulates the expression of various epithelial receptors that can interact with cytotoxic effector cells.

Humoral immune responses are less important, but neutralizing antibodies may inhibit autoinfection by intralumenal stages of the organism. Variable resistance to initial infection can be provided by preexisting maternal antibodies or from antibodies induced by prior exposure.

SIGNIFICANCE

Over 100 different therapeutic agents for cryptosporidiosis have been investigated *in vivo*, yet there is no consistently effective and safe, or approved, treatment. Oocysts are frequently present in lakes, rivers, and other surface waters due to contamination by human, domestic animal, or wildlife waste. Treated water supplies are also often contaminated. Organisms are notably resistant to many disinfectants, including chlorine, and its size (4–6 μ) allows it to pass through many filtration systems. The relative resistance of oocysts allows them to remain infectious in water or the environment for months. Therefore, waterborne infection is a considerable public health concern due to the potentially fatal consequences in immunosuppressed persons and the possibility of high morbidity in the general population. In less developed countries with inadequate sanitation

systems, this concern is even greater due to higher rates of exposure and the contributing effects of malnutrition and concurrent disease.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Diarrhea • Parasitic Diseases, Overview

Further Reading

Fayer, R. (ed.) (1997). "Cryptosporidium and Cryptosporidiosis." CRC Press, Boca Raton, FL.

Petry, F. (ed.) (2000). "Contributions to Microbiology, Vol. 6, Cryptosporidiosis and Microsporidiosis." Karger, New York.

Cystic Fibrosis

MARK E. LOWE AND ROBERT J. ROTHBAUM

Washington University School of Medicine and St. Louis Children's Hospital, Missouri

cirrhosis Chronic disease of the liver marked by progressive destruction and regeneration of liver cells, leading to diffuse fibrosis and nodules of liver cells.

exocrine pancreas Portion of the pancreas that synthesizes and secretes the components of pancreatic juice. The juice contains digestive enzymes, water, and bicarbonate. The pancreatic acinar and ductal cells comprise the cellular components of the exocrine pancreas.

ileus Failure of downward progress of the intestinal contents because of disordered propulsive motility of the bowel.

meconium Contents of the fetal intestine that are normally expelled after birth.

pancreatic insufficiency Decreased secretion of digestive enzymes or insulin to the extent that malabsorption or diabetes mellitus appears. Malabsorption typically manifests with steatorrhea, the presence of more than 7% of dietary fat in the stool.

Cystic fibrosis is the most common lethal genetic defect in the Caucasian population. This autosomal recessive disorder affects between 1 in 200 and 1 in 4500 newborns in different ethnic groups. A defect in the CF transmembrane conductance regulator causes the disease. The conductance regulator localizes to the apical membrane of epithelial cells in a variety of organs, including pancreas, lung, liver, and intestine, where it regulates chloride conductance and water flow and may function in the transport of other ions.

INTRODUCTION

Cystic fibrosis (CF) can be caused by over 900 different mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). The most common mutation, ΔF508, which is a 3-base-pair deletion, affects about 70% of CF genes. Approximately 50% of patients with CF are homozygous for this mutation and have classical CF with pulmonary and pancreatic dysfunction. CFTR mutations fall into five major classes. Mutations designated as classes I, II, and III result in complete loss of CFTR function because of defective protein production, abnormal protein processing, and abnormal regulation of chloride conductance, respectively. In general, these mutations cause more severe disease manifestations. Mutations designated as classes IV and V produce milder symptoms and produce mutant CFTRs with decreased conductance properties or cause decreased synthesis of normally active CFTRs. Individuals with CF may be homozygous for one genetic mutation or heterozygous for two different mutations.

The degree of organ dysfunction caused by a CFTR mutation depends on the properties of the CFTR and on the physiology of the involved organ. More severe mutations of the CFTR (classes I, II, and III) produce greater organ dysfunction and damage than do milder mutations (classes IV and V). Damage is generally

Copyright 2004, Elsevier (USA). All rights reserved.

systems, this concern is even greater due to higher rates of exposure and the contributing effects of malnutrition and concurrent disease.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Diarrhea • Parasitic Diseases, Overview

Further Reading

Fayer, R. (ed.) (1997). "Cryptosporidium and Cryptosporidiosis." CRC Press, Boca Raton, FL.

Petry, F. (ed.) (2000). "Contributions to Microbiology, Vol. 6, Cryptosporidiosis and Microsporidiosis." Karger, New York.

Cystic Fibrosis

MARK E. LOWE AND ROBERT J. ROTHBAUM

Washington University School of Medicine and St. Louis Children's Hospital, Missouri

cirrhosis Chronic disease of the liver marked by progressive destruction and regeneration of liver cells, leading to diffuse fibrosis and nodules of liver cells.

exocrine pancreas Portion of the pancreas that synthesizes and secretes the components of pancreatic juice. The juice contains digestive enzymes, water, and bicarbonate. The pancreatic acinar and ductal cells comprise the cellular components of the exocrine pancreas.

ileus Failure of downward progress of the intestinal contents because of disordered propulsive motility of the bowel.

meconium Contents of the fetal intestine that are normally expelled after birth.

pancreatic insufficiency Decreased secretion of digestive enzymes or insulin to the extent that malabsorption or diabetes mellitus appears. Malabsorption typically manifests with steatorrhea, the presence of more than 7% of dietary fat in the stool.

Cystic fibrosis is the most common lethal genetic defect in the Caucasian population. This autosomal recessive disorder affects between 1 in 200 and 1 in 4500 newborns in different ethnic groups. A defect in the CF transmembrane conductance regulator causes the disease. The conductance regulator localizes to the apical membrane of epithelial cells in a variety of organs, including pancreas, lung, liver, and intestine, where it regulates chloride conductance and water flow and may function in the transport of other ions.

INTRODUCTION

Cystic fibrosis (CF) can be caused by over 900 different mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). The most common mutation, ΔF508, which is a 3-base-pair deletion, affects about 70% of CF genes. Approximately 50% of patients with CF are homozygous for this mutation and have classical CF with pulmonary and pancreatic dysfunction. CFTR mutations fall into five major classes. Mutations designated as classes I, II, and III result in complete loss of CFTR function because of defective protein production, abnormal protein processing, and abnormal regulation of chloride conductance, respectively. In general, these mutations cause more severe disease manifestations. Mutations designated as classes IV and V produce milder symptoms and produce mutant CFTRs with decreased conductance properties or cause decreased synthesis of normally active CFTRs. Individuals with CF may be homozygous for one genetic mutation or heterozygous for two different mutations.

The degree of organ dysfunction caused by a CFTR mutation depends on the properties of the CFTR and on the physiology of the involved organ. More severe mutations of the CFTR (classes I, II, and III) produce greater organ dysfunction and damage than do milder mutations (classes IV and V). Damage is generally

Copyright 2004, Elsevier (USA). All rights reserved.

more extensive in organs with a high level of CFTR expression or in organs in which the CFTR regulates other ion channels or operates in parallel with other anion exchangers. In organs with ion transporters that have overlapping functions with the CFTR, damage may depend on the ability of these transporters to compensate for the lack of functioning CFTRs.

Decreased or absent CFTR function results in diminished secretion of electrolytes and water by the ductular epithelium. Consequently, the concentration of macromolecules in the lumen of the affected duct increases significantly, causing proteins to precipitate and form plugs that further slow duct flow and produce duct obstruction. Thus, organs that have levels of secretions with high protein concentration and slow flow through ducts are most susceptible to damage from mutant CFTRs. The epididymis and vas deferens, the pancreas, and, to a lesser extent, the bile ducts represent vulnerable organs. In the lung, macromolecules move through narrow passages at an air–fluid interface and any decrease in water content of pulmonary secretions, such as occurs with a defective CFTR, causes precipitation of macromolecules in small airways. The increased tenacity of pulmonary secretions and small airway plugging inhibit clearance of pathogens from the airways. Recurrent or chronic infection then contributes to the pulmonary damage. In contrast, the sweat gland has abnormal electrolyte and fluid flux, but does not sustain any histological damage. Presumably, the low protein concentration and high flow rates of their secretions protect sweat glands from damaging precipitates.

TABLE I Clinical Manifestations of Cystic Fibrosis

Respiratory
- Sinusitis
- Nasal polyps
- Atelectasis
- Emphysema
- Bronchiectasis
- Recurrent Infections

Gastrointestinal
- Meconium ileus
- Jejunoileal atresia
- Meconium peritonitis
- Distal intestinal obstruction syndrome
- Rectal prolapse

Pancreas
- Pancreatic insufficiency
  - Malnutrition
  - Vitamin deficiencies
  - Steatorrhea
- Diabetes
- Acute pancreatitis

Skeletal
- Delayed bone age
- Osteopenia
- Hypertrophic pulmonary osteoarthopathy

Hepatobiliary
- Gallstones
- Focal biliary cirrhosis
- Neonatal cholestasis
- Cirrhosis
- Portal hypertension

Reproductive System
- Females
  - Decreased fertility
- Males
  - Sterility
  - Absent vas deferens, epididymis, and seminal vesicles

Eye
- Venous engorgement
- Retinal hemorrhage

Other
- Salt depletion
- Heat stroke
- Hypertrophy of the apocrine glands

CLINICAL PICTURE

Because many cell types express mutant CFTRs and these cells have variable degrees of dysfunction, the clinical manifestations of CF are myriad (Table I). Most patients with CF are diagnosed within the first months of life and 80% are diagnosed before age 6 years (Fig. 1). Infants typically have protracted cough, wheezing, and tachypnea. Toddlers and school-aged children exhibit refractory wheezing or recurrent pneumonia and may have milder respiratory problems such as recurrent sinusitis or nasal polyps.

Most CF patients have exocrine pancreatic insufficiency, resulting in maldigestion and malabsorption. Stool complaints are common, ranging from bulky and frequent stools to diarrhea. At times, oil may separate from the stools. In infancy, weight gain is often poor or absent because of altered digestion and absorption. The severity and progression of CF vary considerably. Some patients deteriorate rapidly and die within the first year. Most patients have milder courses and with appropriate therapy can reach adulthood (Fig. 2). A few have mild disease or monosymptomatic disease and do well over a long period of time. Progressive pulmonary disease often dominates the

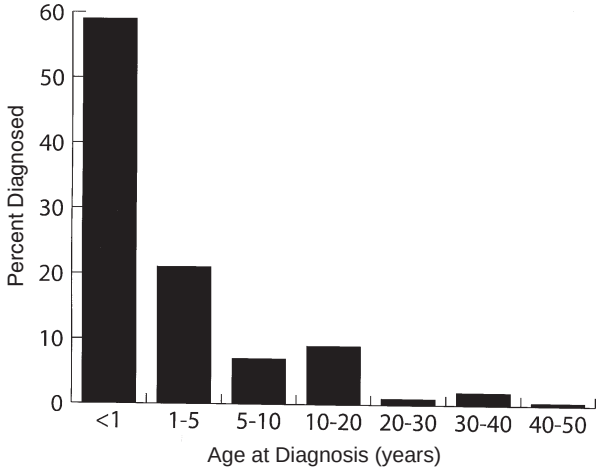


FIGURE 1 Age at diagnosis of cystic fibrosis. Data are from the National Cystic Fibrosis Registry.

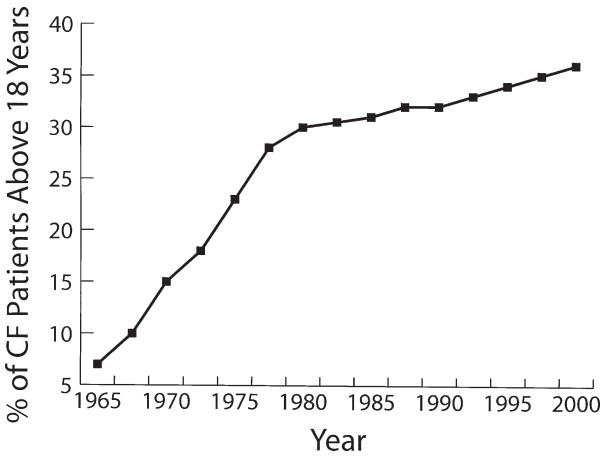


FIGURE 2 The percentage of cystic fibrosis patients older than 18 years, from 1965 to 2000. Data are from the National Cystic Fibrosis Registry.

clinical course, although a variety of gastrointestinal manifestations may predate pulmonary symptoms and may even be more problematic than the pulmonary disease. The only clinical manifestation that correlates with genotype is pancreatic function. Patients with pancreatic insufficiency generally carry two class I, II, or III mutations. In contrast, patients with pancreatic sufficiency generally carry at least one class IV or V mutation.

DIAGNOSIS

The diagnosis of CF depends primarily on measuring the concentration of electrolytes in sweat stimulated by pilocarpine iontophoresis. Patients with CF have abnormal sweat tests from birth. The sodium and chloride concentrations are elevated above 60 mEq/liter in at least 98% of patients with CF. Levels below 40 mEq/liter are normal. Levels between 40 and 60 mEq/liter are indeterminate and not diagnostic. Generally, the sweat chloride concentration exceeds the sodium concentration, particularly in infants. Several situations can produce misleading sweat test results. False negative sweat tests can occur in malnourished or edematous infants with CF. Other disorders besides CF can cause elevated sweat electrolyte concentrations (Table II). Steroids, diuretics, and the amount of sodium in the diet can affect the results. Accurate sweat collection and analysis require experience and expertise, and either negative or positive results can stem from improperly performed sweat tests. To avoid these pitfalls, an initial positive sweat test should be confirmed by testing at a CF center.

Genetic testing for CFTR mutations can aid in diagnosis, but current methods may not be adequate in many populations. The large number of disease-causing mutations in the CFTR gene and the marked variation in the distribution and frequency of mutations within geographic regions and ethnic groups make clinical testing for mutated alleles particularly difficult. In a handful of populations, 98–100% of the mutated alleles have been identified, and suspected patients can be screened for these mutations. Such certainty is impossible in the genetically heterogeneous populations residing in many geographic regions, and testing in these populations relies on commercially available panels that screen for the 70 most common mutations. Less common or unique mutant alleles will not be identified unless broader screening procedures become available. New techniques, such as denaturing ion-pair reverse-phase high-performance liquid chromatography, show promise for rapidly analyzing the complete coding sequence of the CFTR gene. In the near future, rapid and complete screening of the CFTR gene for mutations will be possible.

These new genetic tools create the potential for prenatal diagnosis and genetic counseling. In 1997, a consensus panel convened by the National Institutes of Health offered recommendations for genetic testing in families at risk for CF and for the general population. The panel concluded that genetic testing for CF should be offered to adults with a positive family history of CF, to partners of people with CF, to couples currently planning a pregnancy, and to couples seeking prenatal care, provided that adequate education and appropriate genetic testing and counseling services are available to all persons being tested. Effective counseling depends on extensive and thorough review of the implications, potential results, and possible outcomes of mutated alleles. For some CFTR mutations, a thorough discussion may not be possible because the effect of the mutation on CFTR function is unknown.

TABLE II Conditions Associated with Elevated Sweat Electrolytes

Cystic fibrosis
Glycogen storage disease, type I
Ectodermal dysplasia
Adrenal insufficiency
Fucosidosis
Mucopolysaccharidosis
Familial hypothyroidism
Mauriac's syndrome
Environmental deprivation syndrome

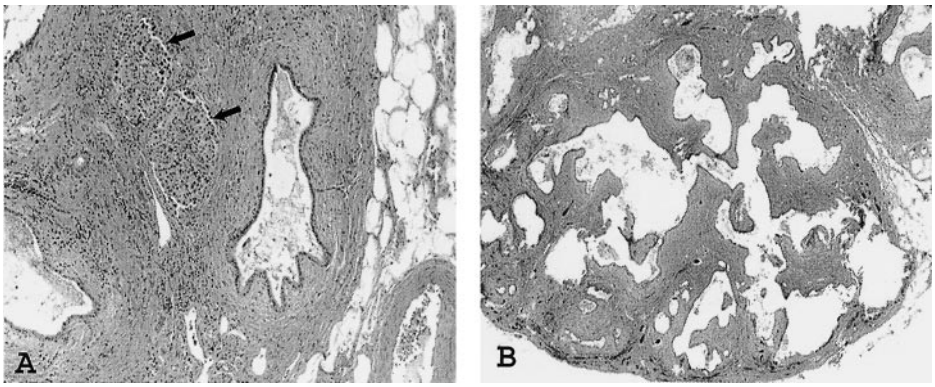


FIGURE 3 The histology of the pancreas in CF. (A) A section of pancreas with two preserved islets (black arrows) also includes a cystic region containing inspissated material. The cyst is surrounded by fibrous tissue. Fatty change is apparent on the left side of the section. (B) A lower power view of the same pancreas section illustrates the marked cystic changes typical of the CF pancreas. No normal acinar tissue is apparent.

PANCREATIC DYSFUNCTION

From 85 to 90% of CF patients have exocrine pancreatic insufficiency with malabsorption of fat and protein and decreased bicarbonate and fluid production. Pancreatic function is often deficient at or shortly after birth. In some individuals with CF, pancreatic function deteriorates over years before exocrine pancreatic insufficiency develops. A minority of patients who remain pancreatic sufficient still secrete lower than normal amounts of digestive enzymes, fluid, and bicarbonate.

Pathology of the Pancreas

Like the clinical course, the histopathology of the pancreas in CF varies considerably (Fig. 3). Patients with mild disease can have a normal pancreas. Severely affected patients have a shrunken, cystic, and fibrotic pancreas with fatty changes. In these patients, the gland may be difficult to identify at postmortem. Pancreatic damage results from obstruction of small ducts by precipitated proteins and cellular debris. Involvement of large ducts, usually stenosis, is much less common. Cystic spaces filled with eosinophilic, calcium-containing concretions develop secondary to duct blockage and acinar cell damage. Mild inflammatory changes and fibrosis develop around damaged acini. Even as fibrosis progresses, the islets of Langerhans are spared until later in life.

Radiographic imaging studies are usually not helpful in diagnosing or in predicting the degree of pancreatic insufficiency. Computer tomography (CT) and magnetic resonance imaging provide the most information about the extent of pancreatic damage. CT delineates fatty replacement and small pancreatic volume (Fig. 4). Magnetic resonance imaging reveals fat deposition as increased signal on T1-weighted images and identifies

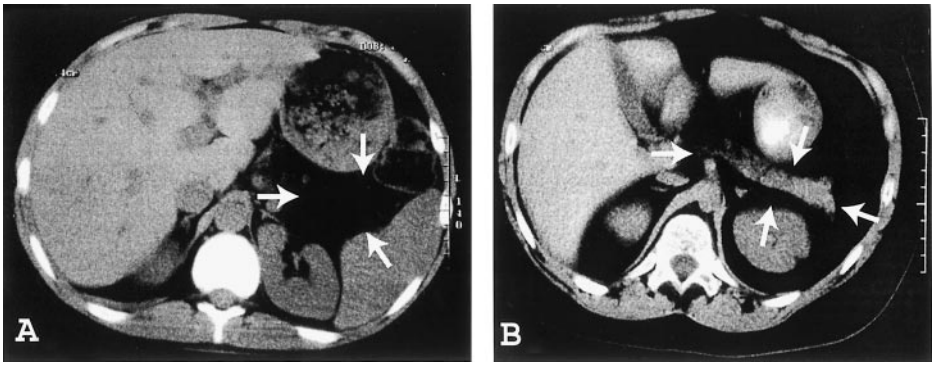


FIGURE 4 Computer tomography scan of the abdomen of a patient with CF. (A) The pancreas is replaced with fat. The arrows delineate the region of the pancreas. (B) The arrows identify an atrophic, fatty pancreas.

pancreatic fibrosis as low-signal intensity on both T1- and T2-weighted images, and may show cystic changes. Occasionally, both modalities reveal abnormalities of the main pancreatic duct.

Exocrine Dysfunction

When the secretion of lipase and trypsin falls below 2% of normal, pancreatic insufficiency results. At this level, 60% or more of ingested fat and around 50% of ingested protein are not digested or absorbed. A nonpancreatic lipase, gastric lipase, probably accounts for the residual lipolysis, and a compensatory increase in gastric lipase activity may occur. An estimated 85–90% of CF patients have pancreatic insufficiency.

Infants are most severely affected by the malabsorption that results from exocrine pancreatic insufficiency. Infants from 2 to 4 months old may present with hypoalbuminemia, edema, and severe normochromic, normocytic anemia. Fat-soluble vitamin deficiencies are most common in infancy prior to diagnosis. Vitamin A deficiency may produce pseudotumor cerebri with sixth nerve palsy and poor feeding, or xerophthalmia. Coagulopathy from vitamin K deficiency is common. Biochemical deficiencies of vitamin E and vitamin D can be detected, but clinical symptoms from either are infrequent.

Tests of Pancreatic Function

The direct measurement of enzyme activities in pancreatic juice collected from the duodenum after stimulation with cholecystokinin and secretin represents the standard for establishing pancreatic insufficiency. This testing provides quantitative results and can reliably detect patients with decreased pancreatic function who do not yet have pancreatic insufficiency. Because this test is invasive and is not widely available, other tests of pancreatic function have been proposed. Usually, protein absorption is measured by indirect tests such as the bentiromide test. In this test, *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid, a chymotrypsin substrate, is given orally and the liberated *p*-aminobenzoic acid is measured in the urine. A modification of this test using a radiolabeled substrate is not suitable for children. The pancreolauryl test is based on another oral substrate, fluorescein dilaurate, which is cleaved by esterases to release fluorescein. The fluorescein is then measured in the urine. Both methods have been tested in CF patients and have good sensitivity and specificity in patients with advanced pancreatic failure. Most other tests are measures of fat malabsorption. Qualitative Sudan stain on stools can be helpful in identifying patients with severe steatorrhea, but it performs poorly in identifying patients with moderate fat loss. The [^{14}C]triolein breath test has been advocated as a method for identifying fat malabsorption, but is not universally available. Even when available, interpretation of the results can be difficult because many factors affect the data. Altered metabolism of triolein, slow intestinal transit, and impaired secretion of CO_2 (which occurs in chronic lung disease) all affect the results. Many centers use the 72-hour fecal fat collection to determine fat malabsorption. Like all of the tests for pancreatic function, fecal fat collection has drawbacks. The fecal collection is cumbersome and is often distasteful for the families to perform. The results are affected by the level of dietary fat intake and by the kind of fat ingested, making an accurate dietary history mandatory to determine the percentage of fat absorbed accurately. Mineral oils and dietary medium-chain triglycerides may cause errors in fat measurement. Finally, fecal fat collections do not distinguish between pancreatic and nonpancreatic causes of fat malabsorption. Most often, infants and others who consume an *ad libitum* diet and are undernourished, and who also have a diagnostic sweat test result or mutant CFTR alleles by genetic testing, are presumed to have pancreatic insufficiency.

Treatment of Pancreatic Insufficiency

Standard treatment for pancreatic insufficiency begins with pancreatic enzyme replacement with extracts from porcine pancreas. Over the years, these preparations have evolved from powders to enteric-coated, encapsulated beads that dissolve at pH 5 to 6, allowing the release of the supplemental enzymes in the proximal small bowel where they are most effective. Importantly, the introduction of enteric-coated capsules has decreased the number of capsules taken per meal, improved the stool pattern, and decreased the abdominal complaints. Still, some treated patients continue to have steatorrhea and azotorrhea even with the enteric-coated preparations.

Several hypotheses attempt to explain the decreased effectiveness of enzyme replacement in some patients. Common to multiple theories is the effect of pH. An acidic pH dramatically slows the release of enzymes from the enteric-coated beads. Even the small difference between pH 5.5 and 5.2 can delay enzyme release up to threefold. A pH below 6.0 may decrease the activity of pancreatic lipase, the predominant lipase in pancreatic extracts, and a pH below 5.0 will irreversibly inactivate pancreatic lipase. Low pH will cause many bile salts to precipitate. The decreased bile acid concentration limits

the formation of mixed micelles of bile acids and fatty acids. Because mixed micelles are required for efficient absorption of fatty acids, steatorrhea results. Several studies document lower than normal pH in the duodenum of CF patients, presumably due to decreased secretion of pancreatic bicarbonate to neutralize gastric acid. Based on this information, agents that suppress gastric acid production, such as omeprazole or lansoprazole, are often prescribed in conjunction with pancreatic enzymes. Although there are no studies of the long-term effectiveness of this approach to increase duodenal pH, some patients benefit from acid suppression over the short term.

CF patients should consume age-appropriate diets. Dietary fat should not be restricted. To prevent deficiencies, all patients should receive multivitamins in addition to daily oral fat-soluble vitamin supplementation. Increased doses of multivitamins may provide adequate amounts of vitamins A and D, but vitamins E and K may require individual supplementation. The efficacy of treatment is judged by the presence of normal growth, by the absence of abdominal complaints, and by biochemical measurements of fat-soluble vitamins in serum. If the patient has slow weight gain or growth, the intake of energy-producing foods and protein should be determined because inadequate nutrient intake is the main cause of poor growth even in patients with undertreated steatorrhea. Low levels of food intake can result from complications of the disease, such as vomiting, anorexia, abdominal pain, and psychosomatic problems such as depression, fatigue, and altered body image.

Nutritional therapy should first aim to increase voluntary oral intake and to maximize fat absorption. The patient should be advised to eat meals and snacks with high caloric content. Some patients may accept fortified milkshakes or powders. Fat absorption may be improved by changing to a different enzyme preparation, by altering the timing of enzyme intake in relation to the meal, or by starting a histamine-2 (H2) receptor antagonist or a proton pump inhibitor. In patients who do still not grow appropriately, nocturnal nasogastric tube or gastrostomy feedings of a defined formula may improve growth. A minority of patients will require parenteral nutrition for acute support until adequate enteral nutrition can be tolerated.

Endocrine Dysfunction

As CF individuals survive into adulthood, progressive damage to the islets of Langerhans reduces insulin secretion, producing glucose intolerance and diabetes mellitus. Although diabetes can develop in young children, the mean age of onset is around age 20. The prevalence of diabetes mellitus increases at a rate of about 5% per year, and by 30 years of age 50% of patients are diabetic. The reasons that some patients develop diabetes are not fully understood. Generally, diabetes develops in patients with class I, II, or III mutations. These patients have impaired insulin secretion and increased insulin clearance rates, leading to insulinopenia. Because the development of diabetes in CF patients is often insidious, the diagnosis should be considered whenever there is a change in growth or pulmonary function, particularly in older patients. Adequate therapy can improve both growth and pulmonary function. Renal, ophthalmologic, and autonomic complications of diabetes can arise in CF patients.

Acute Pancreatitis

Approximately 1% of CF patients suffer recurrent episodes of acute pancreatitis and pancreatitis may be the presenting manifestation of CF. Acute pancreatitis occurs in patients who have pancreatic sufficiency. In such patients, the diagnosis of CF may be delayed because the patients lack other symptoms of CF. CF pancreatitis presents with typical symptoms of abdominal pain and vomiting. Elevations of serum lipase and amylase are present. In one series, patients averaged 3.5 attacks, with a range of one to seven. As with other types of acute pancreatitis, treatment is supportive. The usual course of the acute illness is 3 to 5 days. Complications occur infrequently.

HEPATOBILIARY DISEASE

Damage to the liver and biliary tree in patients with CF was originally described in 1938. Since then, various reports have identified an association of CF with biliary cirrhosis, portal hypertension, cholelithiasis, and bile duct abnormalities. Because the more severe fibrotic hepatic disorders develop insidiously over more than a decade, the impact of CF-related liver disease has become apparent only as the life expectancy of CF patients has improved, and the symptoms and complications of chronic liver disease now affect more CF patients. No marker reliably identifies patients with early liver disease and no specific therapy ameliorates the progression to severe disease.

Pathogenesis of Hepatobiliary Disease

In the human liver, the CFTR localizes to the apical domain of the epithelial cells of the intrahepatic and extrahepatic bile ducts and the gallbladder. Hepatocytes and other liver cells do not express detectable levels of

CFTRs. Presumably, the abnormal biliary secretions present in patients with CF result from the ability of CFTRs to secrete chloride through a low-conductance channel. Most likely, this chloride conductance creates a negative potential and osmotic gradient in the duct lumen. Together, these forces drive sodium and water into the duct lumen. Additionally, CFTRs may regulate anion exchanger activity in cholangiocytes and affect HCO_3^- flux across the cell and increase the secretion of chondroitin sulfate in CF biliary epithelium, which would further increase the viscosity of bile.

In this model, defective CFTR function alters the composition, viscosity, flow, and alkalinity of bile. The increased viscosity, decreased flow, and increased concentrations of macromolecules encourage precipitation of biliary components. The resultant obstruction of small biliary ductules and the accumulation of potentially cytotoxic bile acids damage hepatocytes and bile duct epithelial cells, causing the release of proinflammatory cytokines, growth factors, or free radicals. These factors activate hepatic stellate cells to synthesize and secrete collagen and attract neutrophils, macrophages, and lymphocytes to the damaged regions. Further cytokine release from the inflammatory cells recruits additional stellate cells into the developing lesion, and they add to the developing fibrosis. Over years to decades, the focal damage and fibrosis progress to bridging fibrosis and to multilobular, macronodular cirrhosis and portal hypertension.

Although decreased or absent CFTR function must contribute to the liver disease in CF, significant liver disease does not develop in most patients. Autopsy series show that the majority of older patients have focal biliary cirrhosis. Yet, few CF individuals manifest clinical liver disease. Other modifying factors, such as genetic predisposition or environmental agents, must influence the onset and severity of liver disease. Potentially, specific CFTR mutations could affect the progression to liver disease, but no obvious correlation exists between the CFTR genotype and the development of liver disease. Other possible explanations include differences in how the epithelial cells degrade mutant CFTRs, differences in how well other chloride conductance pathways compensate for the secretory defect, and differences in the immune response. A recent study did find a correlation between liver disease and the human leukocyte antigen (HLA) haplotype B7-DR15-DQ6 in male CF patients, suggesting that the immune response may indeed predispose to liver disease in some CF individuals. An additional study identified mutations in the glutathione transferase gene as a predisposing factor in CF-related liver disease.

Other types of liver disease, specifically neonatal cholestasis and hepatic steatosis, can affect CF patients. The abnormal biliary secretions and flow probably contribute to the pathophysiology of neonatal cholestasis because inspissated, eosinophilic secretions in bile canaliculi and intrahepatic bile ducts represent the predominant pathological changes in the liver. Factors apart from abnormal biliary secretions and flow probably underlie the pathogenesis of hepatic steatosis. Protein calorie malnutrition strongly associates with hepatic steatosis, and nutritional deficiencies such as essential fatty acid deficiency may produce fatty change in the liver of many CF patients. Additional mechanisms, such as increased levels of cytokines and ethanol ingestion, have been proposed as the basis for steatosis in well-nourished patients. It is unclear if steatosis is a precursor for permanent liver damage in CF patients.

Prevalence of Hepatobiliary Disease

The lack of diagnostic markers for CF liver disease makes the identification of all CF patients with liver disease virtually impossible, a problem that complicates any determination of prevalence and incidence of hepatobiliary disease in CF patients. Consequently, all studies probably underestimate the true risk for CF liver disease. An evaluation of voluntary reporting from CF centers in the United States to the CF Foundation National Cystic Fibrosis Registry indicated a prevalence rate of only 1% for hepatic cirrhosis, but found that liver disease accounted for 2.3% of deaths. Several large studies evaluated liver disease prospectively with biochemical markers and imaging criteria and found evidence for liver disease in 18–37% of CF patients.

Clinical Features of Liver Disease in CF

Neonatal Cholestasis

Fewer than 2% of infants with CF develop cholestasis. Half of those infants have meconium ileus compared to 15% of the total CF population. Biochemical studies, clinical findings, and liver histology are not specific and may suggest other disorders. Total and direct bilirubin levels are elevated and hepatomegaly is sometimes present. Acholic stools and a liver biopsy showing bile duct proliferation can occur and suggest biliary atresia. Giant cell transformation and intrahepatic paucity of bile ducts are also described. Most infants with neonatal liver disease do well. The cholestasis can continue for up to 8 months, but usually clears sooner. Hepatic fibrosis may develop and persist, but progression to chronic liver disease is unlikely.

Hepatic Steatosis

Diffuse hepatic steatosis occurs in up to 60% of CF patients. Most patients are identified by the detection of hepatomegaly on physical exam. Splenomegaly and other symptoms of liver disease are absent. Frequently, such patients are malnourished. Serum transaminases may be modestly elevated, but biochemical indicators of liver function are normal. The presence of steatosis does not correlate with the development of chronic liver disease or with worsening of pulmonary function.

Elevation of Serum Transaminases or of Alkaline Phosphatase

Asymptomatic elevations of serum transaminases or alkaline phosphatase are commonly found in CF patients. The frequency of elevated transaminases varies with age. Over 50% of patients below 1 year of age have elevated transaminases. By 2 years of age, the percentage drops to around 40% and continues to drop until it reaches about 10% by age 8 years. Similarly, 30–40% of patients less than 1 year of age have elevated serum alkaline phosphatase. The percentage decreases to approximately 5% by 8 years of age. Most patients have elevations of serum enzymes less than 1.5 times the upper limit of normal. Elevated serum transaminases or alkaline phosphatase or both are more common in patients with the ΔF508 mutation (class I). Of 267 patients who were homozygous for the ΔF508 mutation, 45% had enzyme elevations compared to only 20% in 25 pancreatic sufficient patients with class IV or V mutations. The prevalence rates for other genotypes ranged from 34 to 60%. The magnitude of the enzyme elevation does not predict the development of chronic liver disease.

Cirrhosis

Autopsy series estimate the prevalence of focal biliary cirrhosis at 10% in infants with CF within 3 months of birth, at 25% in children 1 year of age, and at 70% in adults. The available data are old and life expectancies have increased significantly since the publication of the last study in 1979. Consequently, the prevalence rate may be considerably different in CF patients at this time. Most patients with focal biliary cirrhosis do not progress to multilobular cirrhosis, which has a prevalence between 5 and 15%.

Because multilobular cirrhosis and portal hypertension develop insidiously, chronic liver disease is frequently first suspected on palpation of a hard-edged irregular liver and an enlarged spleen at a scheduled office visit. Other physical signs of chronic liver disease, such as dilated and prominent superficial abdominal veins or spider hemangiomas, may be present. Digital clubbing and oxygen desaturation can occur with cirrhosis even in patients without significant pulmonary disease. Impaired bile flow can decrease fat absorption, resulting in steatorrhea, weight loss, and clinical signs of fat-soluble vitamin deficiency. Serum transaminases and alkaline phosphatase are mildly elevated or normal. Initially, liver synthetic function is preserved and, although synthetic function can deteriorate, the complications of portal hypertension, variceal bleeding, and intractable ascites most often threaten the health of the CF patient. Recurrent hemorrhage from varices, intractable ascites, and hepatic synthetic failure most commonly provide the impetus for liver transplantation (Fig. 5).

Diagnostic Evaluation of Liver Disease

Even though a large proportion of CF patients have hepatomegaly or elevated serum transaminases, only a small fraction will develop hepatobiliary disease to an extent that alters their clinical course. To identify these individuals, there should be regular patient history reviews for bruising or bleeding, abdominal swelling, change in stool pattern, weight loss, and fatigue. Careful palpation and percussion of the liver and spleen and examination for cutaneous manifestations of chronic

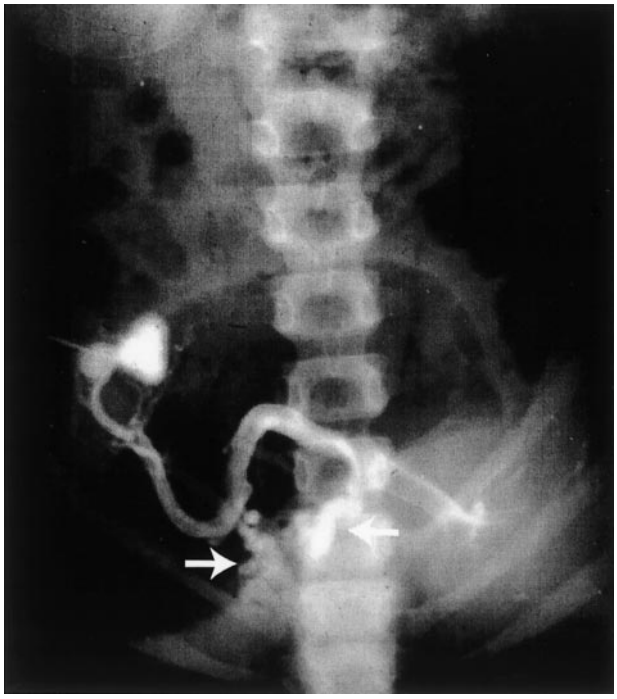


FIGURE 5 A venogram after a splenic injection shows varices (arrows) in a patient with both CF and cirrhosis.

liver disease and of nutritional deficiencies are required. Laboratory studies should include serum transaminase levels, total and direct bilirubin, alkaline phosphatase, γ-glutamyl transferase (GGT), total protein, serum albumin, prothrombin time, and a complete blood count. If significant liver involvement is suspected, then other potential causes of liver disease should be evaluated by specific testing. Multiple reports have documented the coexistence of CF-related liver disease with other causes of liver disease. α1-antitrypsin deficiency, chronic viral hepatitis, Wilson's disease, autoimmune hepatitis, hemochromatosis, sclerosing cholangitis, and drug-induced liver injury have all been reported in CF patients with liver disease.

Imaging studies and endoscopy have variable utility in CF-related liver disease. Ultrasonography of the liver, biliary tree, and hepatic vasculature can provide critical information in evaluating patients with newly suspected liver disease. Ultrasound coupled with Doppler ultrasound can assess the patient for gallstones, ascites, bile duct dilatation, and abnormal flow in the hepatic vessels. Computer tomography can exclude mass lesions or suggest steatosis. Magnetic resonance imaging is quite sensitive for fatty changes. Upper endoscopy is the most sensitive method for detecting esophageal or gastric varices and should always be employed in CF patients with upper gastrointestinal bleeding. Early detection of esophageal varices in patients with evidence of portal hypertension may allow prophylactic use of beta-blocker therapy to prevent variceal bleeding. This approach must be weighed against the potential effects of beta-blockers on airway function.

Liver biopsies are done infrequently in CF patients for several reasons. The heterogeneous nature of the hepatic lesion increases sampling error. Histology does not predict the likely clinical course nor does the type of lesion dictate therapy. Liver biopsy can differentiate between steatosis and focal biliary cirrhosis and may detect diffuse fibrosis or cirrhosis (Fig. 6). At times, this information may influence clinical decisions. If a liver biopsy is required, a prebiopsy ultrasound can direct the biopsy away from dilated hepatic veins and the often hyperexpanded right lower lobe of the lung.

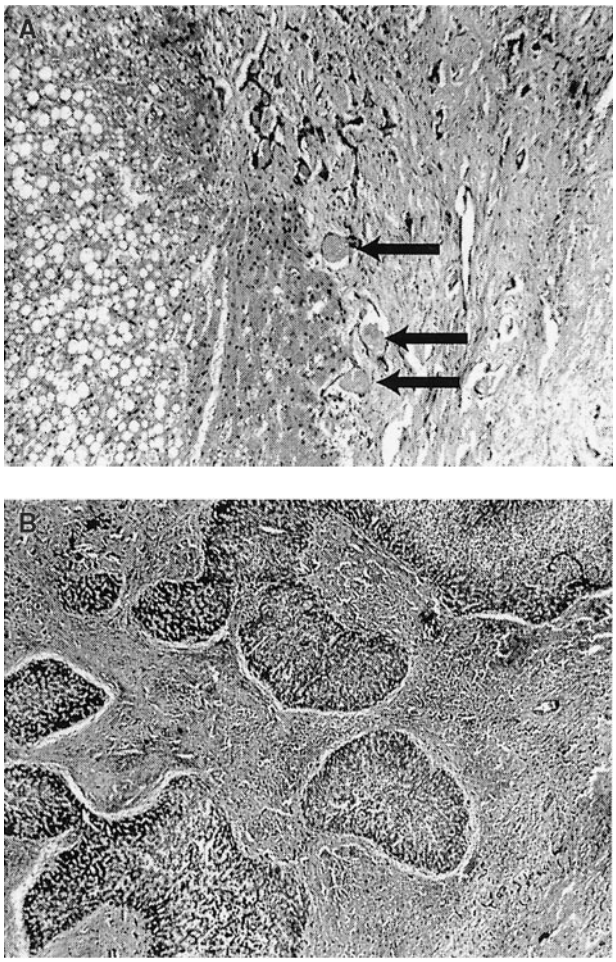


FIGURE 6 Histology of multilobular cirrhosis. (A) Fibrosis and fatty change are present. The arrows mark inspissated material in sinusoids. (B) Marked fibrosis and nodular regeneration are present.

Medical Therapy of CF-Related Liver Disease

The treatment of hepatic steatosis centers on improving the patient's nutritional status. The goal is to maximize the intake of energy-producing foods, protein, and fat-soluble vitamins. A thorough evaluation by a nutritionist can suggest ways to improve nutrition. Fat malabsorption can be quantitated and pancreatic enzyme supplements can be adjusted to optimize fat absorption. Deficiencies of fat-soluble vitamins and essential fatty acids can be determined by biochemical measurements and corrected as indicated. Older patients and well-nourished patients should be questioned about alcohol ingestion and drug use. New-onset diabetes mellitus should be considered.

The treatment of patients with disease characterized by fibrosis or cirrhosis centers on preventing and treating the complications of portal hypertension and cirrhosis. No therapy alters the progression to cirrhosis. Improvements in the biochemical indices of liver injury, in the symptom of pruritus, and in essential fatty acid status have been reported in uncontrolled trials of urosodeoxycholic acid (UDCA) therapy in CF patients with liver disease. UDCA is a potent choleretic and may

protect hepatocytes from injury by displacing toxic bile acids in the enterohepatic circulation and by increasing bile flow as determined by hepatobiliary scintigraphy. No study has shown that UDCA changes the course or outcome in CF-related liver disease or that UDCA prevents the development of liver disease in patients with little or no evidence of liver abnormalities. Despite this lack of evidence, many researchers still recommend UDCA treatment of all patients with fibrosis or cirrhosis.

For patients with life-threatening complications of portal hypertension, with severe synthetic dysfunction, or with growth failure on a background of severe liver disease, liver transplantation offers the only therapeutic option. Because wait times for a donor liver can be long, patients should be referred for transplant evaluation before they are desperately ill. The 1-year survival is about 80% and many patients show improvement in lung function. The outcome is much poorer in patients who require combined organ transplants.

Prevalence and Clinical Features of Biliary Disease in CF

Microgallbladder

Asymptomatic microgallbladder is the most common biliary tract abnormality in CF patients. It is found in about 30% of patients. The contents of the gallbladder are thick and colorless and the submucosa contains mucus-filled cysts. The cystic duct may be undersized or obstructed. Associated abnormalities in the hepatic or common bile ducts have not been described.

Cholelithiasis and Cholecystitis

The prevalence of gallstones and cholecystitis is 1–10% in patients with CF. The frequency may increase with age, but this correlation requires additional confirmation. Although the gallstones are radiolucent, they do not contain cholesterol. Instead, the stones in CF patients are composed of calcium bilirubinate and protein. Comparison of the biliary lipid concentrations between bile from CF patients with and without gallstones shows no differences that might explain why only some CF patients develop stones. When symptomatic, patients have right upper quadrant abdominal pain that may radiate to the right shoulder. Sometimes vague or diffuse abdominal pain is the only complaint in CF patients. Jaundice, nausea and vomiting, and pruritus are less commonly present. Laboratory studies may be normal or show elevated alkaline phosphatase or bilirubin.

Bile Duct Disorders

Abnormalities of the extrahepatic or intrahepatic bile ducts occur in less than 2% of patients. Structural changes in the larger intrahepatic bile ducts similar to the strictures and beading typical of sclerosing cholangitis can be present, particularly in adult patients. The interpretation of these duct abnormalities is complicated in CF patients because of the high probability that these patients have biliary cirrhosis and inspissated thickened biliary secretions. Still, these patients may have the same clinical manifestations found in sclerosing cholangitis, including hepatomegaly and pruritus associated with elevated alkaline phosphatase and GGT. Another abnormality, stenosis of the intrapancreatic common bile duct by compression from pancreatic fibrosis, has also been reported in a few CF patients.

Diagnosis of Biliary Disease

Various imaging studies aid in the identification of biliary disease. Ultrasonography and computer tomography reveal gallstones and duct dilatation. Endoscopic retrograde cholangiopancreatography (ERCP) discloses anatomic abnormalities of bile ducts and identifies gallstones. Because it is an invasive procedure, ERCP should be reserved for diagnosis in selected patients. Magnetic resonance cholangiopancreatography has revealed cholelithiasis and strictures and dilatation of the extrahepatic and intrahepatic bile ducts in a series of CF patients and may provide a noninvasive alternative to ERCP.

Treatment of Biliary Disease

ERCP provides a therapeutic option to laparotomy for treating biliary disease. Impacted gallstones can be removed, strictures can be dilated, and biliary stents can be placed through the endoscope. Because UDCA does not dissolve the gallstones in CF patients, patients with symptomatic gallbladder disease require laparoscopic or surgical cholecystectomy for treatment.

INTESTINAL DISORDERS IN CYSTIC FIBROSIS

The expression of mRNA encoding CFTRs occurs at high levels in the duodenal mucosa and decreases steadily throughout the length of the small intestine. A decreasing gradient of expression also occurs along the axis from the crypt to villus tip in both the small intestine and the colon. Other high-expressing cells reside in the Brunner's glands of the duodenum and are

scattered sparingly throughout the duodenum and jejunum. The consequence of abnormal CFTRs in the intestine is decreased chloride, sodium, and water secretion into the intestinal lumen by the crypt cells. The decreased water content results in thickened intestinal secretions that have abnormal mucus composition and concentrations. The CF intestine contains more mucus than normal intestine and the secreted CF mucus has abnormal biochemical properties with increased fucosylation and sulfation and decreased sialylation. Together, the decreased water content and abnormal mucus account for the increased viscosity of intestinal secretions in CF patients and predispose CF patients to intestinal obstruction.

Viscous secretions generate many of the histological and radiological findings in the intestine of CF patients. Variable quantities of inspissated secretions reside within the lumen of the mucosal glands of the small intestine. Brunner's glands are often dilated and have flattening of the epithelial cells lining the lumen. Thickened small bowel folds, nodular filling defects, and variable dilatation of small bowel loops are radiographic findings in upper gastrointestinal barium studies in CF patients. In the colon, collections of mucus frequently dilate the colonic crypts. Proximal colonic wall thickening without stricture, pericolonic fat proliferation, or mesenteric infiltration can be seen by computer tomography of the abdomen in many CF patients. The appendix also shows changes that suggest the diagnosis of CF. In the appendix, mucus-filled goblet cells line dilated crypts and eosinophilic casts of the crypts appear in the lumen of the appendix.

Meconium Ileus

About 10–15% of newborns with CF present with neonatal meconium ileus or its complications. In these infants, the meconium contains greatly decreased water and high calcium content. Other salts and minerals, such as sodium, potassium, copper, and zinc, are decreased. The development of meconium ileus does not correlate with the presence or absence of pancreatic insufficiency or with any genotype.

Pathology of Meconium Ileus

Viscous inspissated meconium produces distal small bowel obstruction. Typically, the distal ileum is narrow and the colon beyond the obstruction is small (a microcolon). Small bowel ischemia and perforation can occur antenatally. Leakage of meconium into the peritoneum causes an intense inflammatory reaction and can produce meconium cysts, meconium ascites, adhesions, or intraabdominal calcifications. In some infants, the intrauterine perforation seals spontaneously and restores intestinal continuity. Volvulus can occur in the fetus or in the newborn. Fetal volvulus can cause intestinal atresia.

Clinical Features of Meconium Ileus

Most infants with uncomplicated meconium ileus present with signs of intestinal obstruction, abdominal distension, and bilious vomiting within 48 hours of birth. These infants generally appear well at birth. In contrast, infants with complicated meconium ileus present earlier and often appear ill at birth or shortly after birth. Dilated, firm loops of bowel may be visible or palpable, frequently in the right lower quadrant.

Plain X-ray films of the abdomen characteristically show distended loops of bowel with few or no air–fluid levels. Gas trapped in the thick meconium may be found throughout the ileum. Calcifications indicate meconium peritonitis, and mass effect from a meconium pseudocyst may be appreciated. A contrast enema is diagnostic and often therapeutic (Fig. 7). During the enema, contrast fills the microcolon, refluxes into the small bowel, and outlines filling defects caused by the inspissated meconium, establishing the diagnosis. The hypertonicity and mild mucosal irritation of the contrast media may draw fluid into the

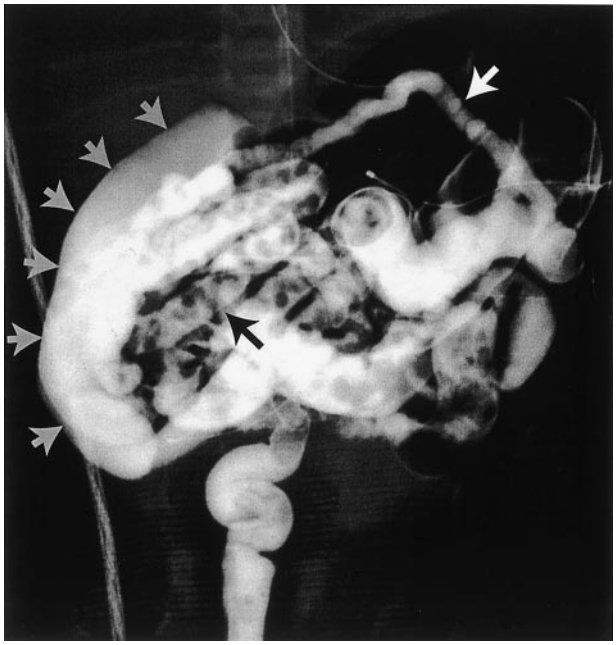


FIGURE 7 Barium enema in a patient with CF and meconium ileus. The white arrow identifies the microcolon. The black arrow shows filling defects in the distal small bowel. The gray arrows mark a dilated loop of more proximal small bowel.

meconium and, along with the mechanical action of the enema, mobilize the obstruction in 50–75% of affected infants.

Additional treatment options include other irrigating solutions and surgery. *N*-Acetylcysteine (Mucomyst) cleaves disulfide bonds in the mucoproteins and reduces the viscosity of the meconium. Surgery is required in infants who fail to respond to irrigation or who have complicated meconium ileus. In these cases, bowel resection with primary anastomosis or stoma formation is generally performed. Improvements in treatment have increased survival of infants with meconium ileus to over 95%.

All infants with meconium ileus, meconium peritonitis, jejunoileal atresia, or volvulus should have a sweat test done; 80% or more of infants with meconium ileus will have CF, as will 30–50% of infants with meconium peritonitis. The presence of jejunoileal atresia increases the risk that the patient has CF 210-fold. Another meconium obstruction syndrome, meconium plug syndrome, carries a much lower risk for CF. Meconium plug syndrome presents with signs and symptoms of bowel obstruction like meconium ileus, but the contrast enema reveals a normal caliber colon and obstruction of the colon with mucus plugs. As in meconium ileus, the enema frequently relieves the obstruction.

Distal Intestinal Obstruction Syndrome

Small bowel obstruction also presents problems in 15% of older CF patients. In the late 1960s, a published report described patients with distal small bowel obstruction secondary to inspissated intestinal contents in the terminal ileum. Originally named "meconium ileus equivalent," the term "distal intestinal obstruction syndrome" (DIOS) is preferred today.

Pathogenesis of DIOS

Multiple mechanisms probably contribute to DIOS. The presence of steatorrhea, either from noncompliance with enzyme therapy or ineffective enzyme therapy, is associated with DIOS. Slowing of intestinal motility, perhaps by the release of neurotensin or by narcotic pain relievers, may allow the accumulation of partially digested food residues in the ileum. Other potential factors include bowel dilatation, previous laparotomy, or inherent dysmotility. Occasionally, the syndrome appears in patients who have no obvious inciting factors.

Clinical Features of DIOS

DIOS presents with acute or chronic symptoms of partial or complete bowel obstruction. Rarely does the

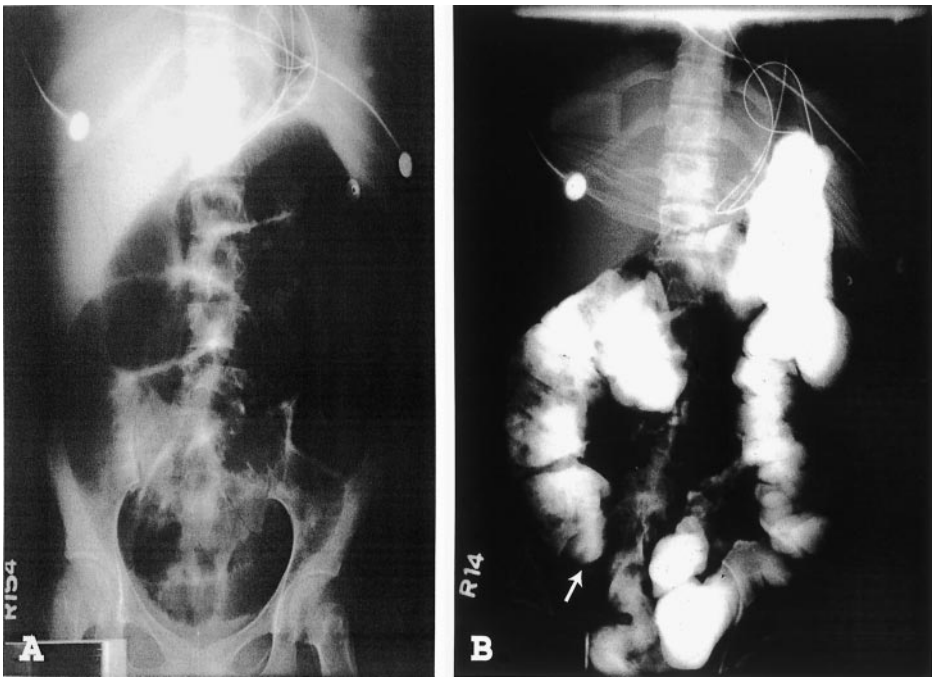


FIGURE 8 Radiography of distal intestinal obstruction syndrome. (A) An older patient with markedly dilated bowel loops. (B) A water-soluble contrast study identifies an obstruction (white arrow).

syndrome occur in patients younger than 5 years and most cases occur in patients older than 15 years. The cardinal features include crampy abdominal pain in the right lower quadrant or lower abdomen, a palpable abdominal mass, and decreased frequency of defecation. In about 10% of patients, abdominal distension and bilious vomiting predominate. Pain is the most common complaint in the chronic form of DIOS. Often, meals trigger the pain and anorexia develops as the patients try to avoid pain episodes. The pain may remit for weeks or months only to return at a later time. In both forms of DIOS, abdominal plain films identify bubbly fecal material in the right lower quadrant with or without bowel dilatation (Fig. 8). Other causes of partial or complete small bowel obstruction, such as intussusception, appendicitis, postoperative adhesions, Crohn's disease, or fibrosing colonopathy, may exist in patients who lack the typical history, physical findings, and radiological findings of DIOS.

The treatment of DIOS focuses on nonoperative relief of the distal small bowel obstruction. A hypertonic, water-soluble contrast enema confirms the diagnosis and may dislodge the impaction from the distal ileum. Patient distress may limit the ability to reflux contrast into the proximal small bowel and thus may thwart therapeutic efforts. In that case, patients without complete bowel obstruction can be treated by intestinal lavage with a balanced electrolyte solution such as Golytely. The lavage should be given orally or by nasogastric tube until the rectal effluent is almost clear. Lavage can also treat chronic DIOS. Some patients will have recurrent episodes of DIOS and may benefit from intermittent oral intake of intestinal lavage solution in addition to regular oral doses of pancreatic enzymes.

Rectal Prolapse

Rectal prolapse was seen in about 20% of patients with CF in early studies. More recent data from the Cystic Fibrosis Registry estimate that this complication now happens in fewer than 2% of CF individuals. Still, CF patients account for about 10% of all patients presenting with rectal prolapse, most presenting in the first year of life and experiencing recurrent episodes. Adequate pancreatic enzyme supplementation almost always stops the prolapse. A few patients require surgery.

Fibrosing Colonopathy

Fibrosing colonopathy, first reported in 1994, is a disorder associated with the administration of high doses of pancreatic enzymes. The colon characteristically has dense submucosal fibrosis and may be foreshortened or strictured. Histologically, eosinophilia, mild cryptitis, regenerative epithelium, and disruption of the muscularis mucosa are seen. Most commonly, patients present with a distended abdomen and abdominal pain. In addition to high-dose administration of pancreatic enzymes, predisposing factors may include young age, prior intestinal surgery, and DIOS. The withdrawal of all high-dose formulations of pancreatic enzymes from the market has been associated with a decreased prevalence of this complication.

See Also the Following Articles

Bile Flow • Cirrhosis • Exocrine Pancreatic Insufficiency • Gallstones, Pathophysiology of • Neonatal Cholestasis and Biliary Atresia • Pancreatic Function Tests

Further Reading

Henrion-Caude, A., Flamant, C., Roussey, M., Housset, C., Flahault, A., Fryer, A. A., Chadelat, K., Strange, R. C., and Clement, A. (2002). Liver disease in pediatric patients with cystic fibrosis is associated with glutathione S-transferase P1 polymorphism. *Hepatology* **36**, 913–917.

Lopez, J. M., and Grand, R. J. (1998). Hereditary and childhood disorders of the pancreas. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, M. H. Sleisenger, and S. Klein, eds.), 6th Ed., Vol. 1, pp. 782–808. W. B. Saunders Company, Philadelphia.

Office of Rare Diseases of the National Institutes of Health. (1997). Genetic testing for cystic fibrosis. Web site, retrieved at http://rarediseases.info.nih.gov/news-reports/workshops/cf000414.html.

Rothbaum, R. J. (1997). Cystic fibrosis: Pancreatic, hepatic, and intestinal manifestations. *In* "Hepatobiliary, Pancreatic and Splenic Disease in Children: Medical and Surgical Management" (W. F. Balistreri, R. Ohi, T. Todani, and Y. Tsuchida, eds.), pp. 459–478. Elsevier Science, New York.

Sokol, R. J., Durie, P. R., and Cystic Fibrosis Foundation Hepatobiliary Disease Consensus Group. (1999). Recommendations for management of liver and biliary tract disease in cystic fibrosis. *J. Pediatr. Gastroenterol. Nutr.* **28**, S1–S13.

Walters, S. (2001). Survival in cystic fibrosis—What do the statistics mean? Web site, retrieved at http://www.cfstudy.com/sarah/Survival%20in%20cystic%20fibrosis.pdf.

Cytochrome P450

CLARK R. GREGG
Veterans Affairs North Texas Health Care System and University of Texas Southwestern Medical School, Dallas

oxidative metabolizing enzymes Render drugs and other xenobiotics both less biologically active and more water soluble for renal or biliary excretion. The principal enzymes involved in this process are the microsomal mixed-function oxidases of the cytochrome P450 system, so named because of their absorbance maximum at 450 nm under laboratory conditions.

P450 induction Increased synthesis of P450 enzymes following extended exposure to certain drugs or substances, resulting in accelerated metabolism of other drugs.

P450 inhibition Specific inhibition of cytochrome P450 enzymes by certain drugs or foods; delays the metabolism and augments the pharmacologic effects of other drugs that are substrates of the same enzyme.

The cytochrome P450 enzymes are pivotally important in the biotransformation of drugs and are the most important mediators of drug–drug interactions. The clinical relevance of pharmacokinetic P450 drug interactions may not always be apparent, but the adverse effects may occasionally be severe. Inhibitory P450 interactions can in some situations be purposely used for pharmacologic benefit.

ENZYMATIC BIOTRANSFORMATION OF DRUGS

More than 30 distinguishable cytochrome P450 enzymes have been described. They are found in highest concentration in the liver and, in the case of certain subfamilies, in the small intestinal mucosal enterocytes and other organs. Many P450 enzymes have versatile substrate specificities and can metabolize drugs, toxins, and carcinogens as well as normal endogenous compounds. Cytochrome P450 enzymes are assigned a distinctive nomenclature and are grouped into families, with subgrouping (e.g., CYP 1A2, CYP 3A4) according to biochemical relatedness. The human P450 enzymes that are most important in drug metabolism are CYP 1A2, the CYP 2C family, CYP 2D6, and CYP 3A4. These isoenzymes differ in their pharmacogenetics, substrate specificities, inducibility, and susceptibility to inhibition by competing substrate drugs. In the intestine and liver, CYP 3A4 shares with the P-glycoprotein efflux transporter an important role in the presystemic metabolism and elimination of many commonly used drugs.

CYTOCHROME P450 AND DRUG INTERACTIONS

Exposure to certain drugs and other substances can induce the synthesis of some P450 enzymes, which accelerates the metabolism of other drugs that are substrates of these enzymes. The most important inducers of P450 enzymes are anticonvulsants, the antimicrobial drug rifampin, chronic alcohol consumption, and cigarette smoking. Following P450 induction, the enhanced metabolism of other drugs will diminish their pharmacologic activity. These effects may persist for weeks following cessation of the inducing drug.

Certain drugs can inhibit specific P450 isoenzyme-mediated metabolism of other (object) drugs in a dose-dependent fashion. This inhibition may result in an immediate decrease in the metabolism of the object drugs, resulting in their accumulation and possible toxicity. Because a range of variation in different P450 activities exists among individuals in a population, the clinical significance of this inhibition may be difficult to predict in a particular person. Commonly used drugs that, by inhibiting specific P450 enzymes, may cause clinically significant drug interactions include cimetidine, amiodarone, selective serotonin reuptake inhibitors (SSRIs), and a variety of antimicrobials (e.g., macrolides, sulfonamides, azole antifungals, and the HIV-1 protease inhibitors, especially ritonavir). The potent P450 inhibition by ritonavir is intentionally used to advantage in combination drug regimens for the treatment of HIV-1 infection.

Object drugs for which clinically important toxicity has resulted from these inhibitory cytochrome P450-mediated drug interactions include theophylline, warfarin, carbamazepine, phenytoin, benzodiazepines, psychotropic drugs, cyclosporine, tacrolimus, terfenadine, cisapride, hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase inhibitors, calcium channel

 Copyright 2004, Elsevier (USA). All rights reserved.

blockers, and ergotamine. Furanocoumarins in grapefruit juice irreversibly inhibit CYP 3A4 in the intestinal mucosa, which may markedly enhance the bioavailability and pharmacologic effects of coadministered drugs that are substrates of this particular enzyme.

See Also the Following Articles

Alcohol Metabolism • Hepatotoxicity, Drug-Induced • Hepatotoxins

Further Reading

Dresser, G., Spence, D., and Bailey, D. (2000). Pharmacokinetic–pharmacodynamic consequences and clinical relevance of cytochrome P450 3A4 inhibition. *Clin. Pharmacokinet.* **38**, 41–57.

Gregg, C. (1999). Drug interactions and anti-infective therapies. *Am. J. Med.* **106**, 227–237.

Guengerich, F. (1997). Role of cytochrome P450 enzymes in drug–drug interactions. *Adv. Pharmacol. (New York)* **43**, 7–35.

Hansten, P., and Horn, J. (2001). Pharmacokinetic drug interaction mechanisms and clinical characteristics. *In* "Drug Interactions: Analysis and Management" (P. Hansten and J. Horn, eds.) pp. PM-11–PM-20. Facts and Comparisons Publ. Group, St. Louis.

Cytomegalovirus

RICHARD C. G. POLLOK
St. George's Hospital Medical School, London

allogeneic Refers to intraspecies subjects with different genetic constitution.

highly active antiretroviral therapy Combination of three or more anti-human immunodeficiency virus agents, usually including a protease inhibitor.

inflammatory bowel disease Idiopathic inflammatory condition of the bowel.

Human cytomegalovirus is a β herpesvirus that produces a characteristic cytopathology on biopsy. Approximately 50% of individuals in the developed world are seropositive for cytomegalovirus antibodies by adulthood. Primary infection is often mild or asymptomatic but may be responsible for an infectious mononucleosis syndrome. In immunocompromised individuals, primary infection or reactivation of latent cytomegalovirus infection may cause a range of severe clinical manifestations throughout the body, including the gastrointestinal tract, eye, respiratory tract, and central nervous system. Groups at risk include individuals with primary or secondary immunodeficiencies, particularly patients with human immunodeficiency virus, patients undergoing transplant with immunosuppression, oncology patients receiving chemotherapy, and congenitally infected neonates.

GASTROINTESTINAL INFECTION IN THE IMMUNOCOMPROMISED

Cytomegalovirus (CMV) gastrointestinal tract infections in patients with human immunodeficiency virus (HIV) were common in the era preceding use of highly active antiretroviral therapy (HAART); affected patients usually had CD4 cell counts below 100 cells/μl. CMV infection in allogeneic transplant patients has also been well described. In this group of patients, interstitial pneumonitis is the predominant problem, but gastrointestinal (GI) disease is well recognized and anti-CMV prophylaxis for patients with CMV viremia substantially reduces end organ disease.

Clinical Features

Gastrointestinal infection of the esophagus, stomach, small bowel, and colon occurs in 5–15% of patients during the course of HIV infection. Although CMV may infect any part of the GI tract, the most common site of infection is the colon. There is a wide spectrum of symptoms associated with CMV colitis. Chronic or

Copyright 2004, Elsevier (USA). All rights reserved.

blockers, and ergotamine. Furanocoumarins in grapefruit juice irreversibly inhibit CYP 3A4 in the intestinal mucosa, which may markedly enhance the bioavailability and pharmacologic effects of coadministered drugs that are substrates of this particular enzyme.

See Also the Following Articles

Alcohol Metabolism • Hepatotoxicity, Drug-Induced • Hepatotoxins

Further Reading

Dresser, G., Spence, D., and Bailey, D. (2000). Pharmacokinetic–pharmacodynamic consequences and clinical relevance of cytochrome P450 3A4 inhibition. *Clin. Pharmacokinet.* **38**, 41–57.

Gregg, C. (1999). Drug interactions and anti-infective therapies. *Am. J. Med.* **106**, 227–237.

Guengerich, F. (1997). Role of cytochrome P450 enzymes in drug–drug interactions. *Adv. Pharmacol. (New York)* **43**, 7–35.

Hansten, P., and Horn, J. (2001). Pharmacokinetic drug interaction mechanisms and clinical characteristics. *In* "Drug Interactions: Analysis and Management" (P. Hansten and J. Horn, eds.) pp. PM-11–PM-20. Facts and Comparisons Publ. Group, St. Louis.

Cytomegalovirus

RICHARD C. G. POLLOK
St. George's Hospital Medical School, London

allogeneic Refers to intraspecies subjects with different genetic constitution.

highly active antiretroviral therapy Combination of three or more anti-human immunodeficiency virus agents, usually including a protease inhibitor.

inflammatory bowel disease Idiopathic inflammatory condition of the bowel.

Human cytomegalovirus is a β herpesvirus that produces a characteristic cytopathology on biopsy. Approximately 50% of individuals in the developed world are seropositive for cytomegalovirus antibodies by adulthood. Primary infection is often mild or asymptomatic but may be responsible for an infectious mononucleosis syndrome. In immunocompromised individuals, primary infection or reactivation of latent cytomegalovirus infection may cause a range of severe clinical manifestations throughout the body, including the gastrointestinal tract, eye, respiratory tract, and central nervous system. Groups at risk include individuals with primary or secondary immunodeficiencies, particularly patients with human immunodeficiency virus, patients undergoing transplant with immunosuppression, oncology patients receiving chemotherapy, and congenitally infected neonates.

GASTROINTESTINAL INFECTION IN THE IMMUNOCOMPROMISED

Cytomegalovirus (CMV) gastrointestinal tract infections in patients with human immunodeficiency virus (HIV) were common in the era preceding use of highly active antiretroviral therapy (HAART); affected patients usually had CD4 cell counts below 100 cells/μl. CMV infection in allogeneic transplant patients has also been well described. In this group of patients, interstitial pneumonitis is the predominant problem, but gastrointestinal (GI) disease is well recognized and anti-CMV prophylaxis for patients with CMV viremia substantially reduces end organ disease.

Clinical Features

Gastrointestinal infection of the esophagus, stomach, small bowel, and colon occurs in 5–15% of patients during the course of HIV infection. Although CMV may infect any part of the GI tract, the most common site of infection is the colon. There is a wide spectrum of symptoms associated with CMV colitis. Chronic or

Copyright 2004, Elsevier (USA). All rights reserved.

intermittent diarrhea in association with abdominal pain is the most common manifestation of colonic infection. Colonic infection may present with mild or severe rectal bleeding, or abdominal pain without diarrhea and fever. Pain may precede the development of toxic megacolon and intestinal perforation, which is a rare but life-threatening occurrence. CMV infection of the colon may occur in association with infection elsewhere in the GI tract. Such sites include the esophagus, usually resulting in dysphagia and odynophagia, and the pancreatico-biliary tree, manifesting as pain in the upper abdomen that results from acquired immunodeficiency syndrome (AIDS)-related pancreatico-cholangiopathy or pancreatitis. Importantly, GI infection may also herald CMV retinitis, and careful retinal assessment is therefore essential in this group of patients.

Diagnosis

Definitive diagnosis of CMV enterocolitis requires intestinal biopsy. A spectrum of both upper and lower GI manifestations of CMV infection has been described. In a prospective study of HIV patients with CMV colitis, chronic diarrhea and abdominal pain occurred in 80 and 50% of patients, respectively; 9% presented with lower GI bleeding with a previous history of diarrhea. Endoscopic appearances were heterogeneous. Three main categories were identified: colitis associated with ulceration (40%), ulceration alone (40%), or colitis alone (20%). Subepithelial hemorrhage was a common manifestation in all groups; 9% of patients had disease proximal to the splenic flexure without distal involvement.

The histological diagnosis of CMV enterocolitis is largely dependent on identifying characteristic cytomegalovirus inclusion bodies, and these are best seen in biopsies obtained from the base of CMV ulcers. In addition, specific immunoperoxidase staining for CMV is also useful in identifying GI disease. Positive staining is more likely at the edge of ulcers and some clinicians claim greater sensitivity using immunostaining compared to conventional histology. The value of viral culture and the polymerase chain reaction in evaluating intestinal biopsies is limited because of the lack of specificity. Monitoring CMV antigenemia or CMV DNA levels during treatment indicates that levels broadly correspond to the clinical response.

Treatment

Treatment of CMV enterocolitis requires parenteral therapy with either ganciclovir (5 mg/kg, twice daily) or foscarnet (90 mg/kg, twice daily), both of which may be associated with severe side effects. Ganciclovir may cause severe bone marrow depression with resultant anemia and neutropenia. Foscarnet may cause severe renal impairment, although a concomitant normal saline infusion largely seems to diminish the risk of renal damage. In an open-label randomized study comparing a 2-week course of ganciclovir with foscarnet, at doses of 5 and 90 mg/kg, twice daily, respectively, no significant difference in response of GI CMV disease was found between the two therapies. Around 75% of patients had good clinical and endoscopic responses, with disappearance of inclusion bodies. Relapse occurs in at least 50% of patients within around 10 weeks, and survival without HAART in HIV-infected patients is around 20 weeks. Encouragingly, the advent of HAART has led to a marked reduction in mortality and morbidity associated with opportunistic infection among HIV patients, including CMV. HAART results in a reconstitution of anti-CMV immunity, with a resultant significant and progressive decline in CMV viremia in the absence of specific anti-CMV treatment.

GASTROINTESTINAL INFECTION IN THE IMMUNOCOMPETENT

CMV colitis may rarely occur in the immunocompetent individual and has been associated with significant morbidity (hemorrhage and perforation) and occasionally mortality. Patients with inflammatory bowel disease treated with immunomodulatory drugs such as azathioprine or cyclosporine are at increased risk of CMV colitis, which may contribute to the course of seemingly refractory severe colitis in these patients.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Gastric Infection (non-*H. Pylori*)

Further Reading

Crumpacker, C. S. (2000). Cytomegalovirus. *In* "Principles and Practise of Infectious Diseases" (G. L. Mandell, R. G. Douglas, J. E. Bennett, and R. Dolin, eds.), 5th Ed., Vol. 2, pp. 1586–1600. Churchill Livingstone, Philadelphia.

Goodrich, J. M., Mori, M., Gleaves, C. A., *et al.* (1991). Early treatment with ganciclovir to prevent cytomegalovirus disease after allogeneic bone marrow transplantation. *N. Engl. J. Med.* **325**, 1601–1607.

Pollok, R. C. (2001). Viruses causing diarrhoea in AIDS. Gastroenteritis viruses. *Novartis Found. Symp.* **238**, 276–283.

Defecation

ARNOLD WALD
University of Pittsburgh Medical Center

biofeedback A training technique by which an individual is taught to gain control over an important body function using information received from an instrument to guide their response.

enteric nervous system The nerve cell bodies and their processes that are found in the wall of the gastrointestinal tract and that act to influence motor activity and other important gut functions.

outlet dysfunction A constellation of disorders in which normal defecation is disturbed by organic or functional abnormalities of the anorectum.

pelvic floor dyssynergia A condition in which normal defecation is impeded by the unconscious contraction of the puborectalis and external anal sphincter muscles.

In humans, the colon and rectum function to store fecal material and to expel it when appropriate. As fecal material passes through the colon, water is removed from it so that a semisolid to solid stool is formed by the time it reaches the distal colon. The critical mechanisms that control defecation are examined in this article.

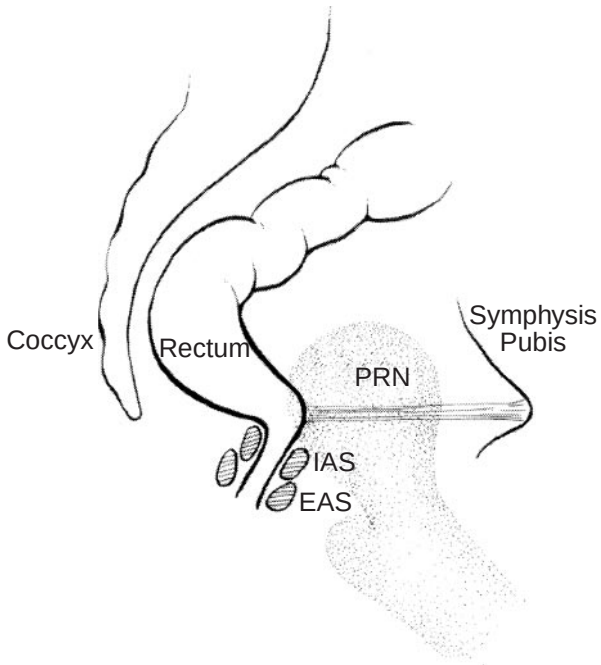


FIGURE 1 Sagittal view of the anorectum at rest. PRM, puborectalis muscle; IAS, internal anal sphincter; EAS, external anal sphincter. The anorectal angle is formed by the tone of the puborectalis muscle. Anal canal pressure is the sum of the IAS and EAS.

INTRODUCTION

The main functions of the colon and rectum in humans are to store fecal material and to expel it when socially appropriate. During the passage of fecal material through the colon, dehydration occurs so that stool is semisolid to solid and formed by the time it arrives at the distal colon. In addition to the volume, consistency, and speed with which stool arrives at the distal colon and rectum, the critical mechanisms that control continence and defecation include the reservoir function of the rectum, the anorectal angle created by the puborectalis sling muscle, and the internal and external anal sphincters, which help to regulate the passage of fecal wastes through the anal canal (Fig. 1).

The stimulus to initiate defecation appears to be distension of the rectum, which normally occurs when colonic peristaltic contractions propel stool into the rectum. Whether defecation occurs is the result of a variety of factors. The nature of rectal contents can generally be distinguished by epithelial nerve endings in the rectum and anal canal. Rectal distension is perceived via sensory afferents in the perirectal spaces and perhaps in the rectal wall itself. In most cases, defecation may be suppressed or allowed through complex cortical influences on anorectal components.

NORMAL DEFECATION

Normally, rectal distension induces relaxation of the internal anal sphincter. This is associated with transient contraction of the external anal sphincter to maintain continence and allow sampling of rectal contents by anal sensory nerves to distinguish gas, liquids, and solids. If the decision to defecate is made, the subjects assume the squatting position, which helps to straighten the

 Copyright 2004, Elsevier (USA). All rights reserved.

alignment between the rectum and anal canal. Rectal contraction increases intralumenal pressures and contraction of the rectus abdominus muscles and diaphragm during the performance of the Valsalva maneuver increases intra-abdominal pressures to provide a propulsive force. The pelvic floor descends and relaxation of the puborectalis muscle and external anal sphincter decreases resistance to the passage of the fecal mass through the anal canal (Fig. 2). After defecation, there is rebound contraction of the puborectalis and external anal sphincter (EAS) and the anal canal is closed (Fig. 1). In most healthy subjects, there is substantial emptying of the left colon as well as the rectum.

In contrast to defecation, expulsion of flatus is not associated with straightening of the anorectal angle. In this circumstance, propulsive forces move intralumenal gas past the anorectal angle and through the anal canal.

ABNORMAL DEFECATION (OUTLET DYSFUNCTION)

In certain individuals, behavioral or physiologic factors may produce defecatory dysfunction with resultant constipation. This constellation of disorders is known as outlet dysfunction, which may be divided into various subtypes.

FIGURE 2 During normal defecation, the puborectalis muscle is relaxed to widen the anorectal angle while relaxation of the EAS reduces resistance pressures in the anal canal (arrows).

Neurologic Disorders

Neurologic disorders that affect the anorectum may impair the act of defecation in a variety of ways. Defecation is triggered by an urge to defecate, a sensory signal that activates the sequence of voluntary and involuntary movements in the anorectum and pelvic floor. The sensory stimulus may be lacking when sensory fibers from the pelvic floor and anorectum are disconnected from the higher cortical centers; these diseases may occur anywhere from the peripheral nerves to the cerebral cortex.

The voluntary and coordinated activation of the thoracoabdominal muscles, the levator/puborectalis muscles, and the EAS may be impaired or absent in the presence of spinal cord lesions or bilateral cortical damage. Lesions above T10 invariably affect all three muscle groups but in lesions below T10, thoracic and abdominal muscle activity is preserved, thus allowing defecation to occur, although perhaps with some delay.

Normally, the resting activity of the puborectalis muscle and EAS is maintained by spinal sacral reflexes that are inhibited during the act of defecation. In neurologic diseases that affect the suprasacral spinal cord but with intact spinal sacral reflexes, the descending inhibition of these muscles is lost or impaired, which in turn may interfere with defecation.

Enteric Nervous System Disorders

The prime example in this group is Hirschsprung's disease, which is characterized by the congenital absence of ganglion cells in the submucosal and myenteric plexus, beginning at the internal anal sphincter and extending proximally for varying distances. At least seven different genes are associated with this disorder, but with variable penetrance, which influences the expression of the disorder.

The loss of intrinsic innervation in the distal colon results in loss of the inhibiting neurotransmitter nitric oxide and overexpression of extrinsic parasympathetic nerves. This results in unopposed contraction of the aganglionic segment. As a result of this functional obstruction, the normal innervated bowel proximally becomes dilated.

Pelvic Floor Dyssynergia

Normal defecation involves the coordinated relaxation of the puborectalis and external anal sphincter muscles together with increased intra-abdominal pressure and inhibition of colonic segmenting activity. In patients with pelvic floor dyssynergia, ineffective defecation is associated with failure to relax, or

FIGURE 3 In pelvic floor dyssynergia, there is unconscious contraction of the PRM and EAS during attempted defecation. This narrows the anorectal angle and increases anal canal pressures (arrows) to impede defecation.

inappropriate contraction of the puborectalis and external anal sphincter muscles. This narrows the anorectal angle and increases pressure of the anal canal so that evacuation is less effective (Fig. 3). Because relaxation of these muscles involves cortical inhibition of the spinal reflex during defecation, this pattern may represent a conscious or unconscious act.

The diagnosis of pelvic floor dyssynergia is generally established by diagnostic testing in a laboratory facility using techniques that attempt to duplicate normal defecation. There is recent evidence that pelvic floor dyssynergia is often an artifact of laboratory testing and may be overdiagnosed. In one study, 80% of patients who exhibited dyssynergia in the laboratory showed appropriate sphincter relaxation when tested at home. Others have demonstrated dyssynergia in approximately 25% of controls and a similar percentage of patients with fecal incontinence. Thus, more rigorous criteria for the diagnosis are required in clinical practice and for clinical trials.

A diagnosis of pelvic floor dyssynergia should be suspected only when there is impaired expulsion of a water-filled balloon or artificial stool from the rectum when attempted in privacy. It should then be confirmed by at least two of the following studies: anorectal manometry, anal sphincter electromyography (EMG), and defecography.

Weak Expulsion Forces

In this condition, expulsion forces are inadequate to expel stool from the rectum. This occurs in several scenarios such as neuromuscular disorders, painful conditions that inhibit the Valsalva maneuver, lax rectus abdominus muscles, and megarectum, which disperses expulsion forces into a capacious rectum rather than toward the anal canal.

Misdirection of Expulsion Forces

Rarely, a large rectocele (anterior in women, posterior in men) will cause misdirection of expulsion forces as stool is directed toward areas of less resistance rather than through the anal canal. However, the large majority of rectoceles are of no physiologic significance and constipation may be caused by other mechanisms such as dyssynergia.

Unexplained Rectal Dysfunction

Unexplained rectal dysfunction is characterized by difficulty expelling rectal contents in the absence of a demonstrable mechanism. This diagnosis should be made only if there is considerable delay or inability to expel an artificial stool in the absence of demonstrable organic or functional abnormalities. It should be emphasized that the expulsion of artificial stool does not mimic exactly normal defecation as the latter is associated with propulsive contractions of the distal colon.

TREATMENT OF ABNORMAL DEFECATION

Ideally, treatment for patients with outlet dysfunction should be based on presumed pathophysiologic mechanisms. Broadly speaking, treatments include stimulant and osmotic laxatives and enemas, behavioral approaches such as biofeedback, and surgery in highly selected patients.

Behavioral Approaches

One behavioral approach is the use of biofeedback training to modify striated muscle activity during simulated defecation, in the laboratory or at home. A second approach is practicing simulated defecation of a water-filled or air-filled balloon or other substances that are used as artificial stool. The last has been combined with diaphragmatic muscle training by some investigators.

Biofeedback is performed with an EMG sensor or manometry catheter, which is placed into the anorectum. The patient watches a visual display of pressure traces or EMG activity. The patient is instructed to squeeze to produce an increase in pressure or EMG activity and then is asked to strain as if to expel the catheter from the rectum. The patient is shown what a normal response is and then is asked to normalize his (her) recording using trial and error efforts. Biofeedback sessions continue until normal defecation efforts are consistently made. The patient is instructed to use these techniques when attempting to defecate at home and may be given a home trainer to practice. Therapeutic endpoints are subjective and objective (frequency, straining, sense of complete evacuation, etc.). The ideal target population is made up of adult patients with pelvic floor dyssynergia.

A modification of the biofeedback technique is simulated defecation in which a balloon attached to a catheter is inserted into the rectum and inflated with 20 cc of air or water. The patient is asked to push the balloon out slowly without straining. Relaxation of the external anal sphincter is necessary to do this easily. The process is repeated with the patient straining; the ease with which this is done is the feedback for the patient, as squeezing the EAS will make it more difficult. Phases 2 and 3 are identical to the biofeedback program above.

Surgical Approaches

In Hirschsprung's disease, surgical resection of the aganglionic bowel with anastomosis of the normal colon to the anal canal is the optimal approach. In patients with short segment disease, internal sphincterotomy with extension to the normal rectum may be effective.

Rectocele repairs should be restricted to those occasions when a rectocele is believed to be of physiologic importance. Surgery should be preceded by behavioral correction of functional abnormalities of defecation, if present.

Pharmacologic Approaches

Botulinum toxin is a potent neurotoxin that binds to presynaptic cholinergic nerve terminals to inhibit the release of acetylcholine at the neuromuscular junction. When injected into muscle, paralysis occurs for several months until there is axonal regeneration and formation of new nerve terminals. Some studies suggest that botulinum toxin injected into the puborectalis muscle may be helpful in patients with pelvic floor dyssynergia who fail to respond to behavioral treatments.

See Also the Following Articles

Anal Canal • Anal Sphincter • Constipation • Enteric Nervous System • Rectum, Anatomy

Further Reading

Bassotti, G., Germani, U., and Morelli, A. (1996). Flatus-related colorectal and anal motor events. *Digest. Dis. Sci.* **41**, 335–338.

Chiotakakou-Faliakou, E., Kamm, M. D., Roy, A. J., *et al.* (1998). Biofeedback provides long-term benefit for patients with intractable slow and normal transit constipation. *Gut* **42**, 517–521.

Corazziari, E., Badiali, D., and Inghilleri, M. (2001). Neurologic disorders affecting the anorectum. *Gastroenterol. Clin. North Am.* **30**, 253–268.

Duthie, H. L., and Bartolo, D. C. C. (1992). Anismus: The cause of constipation? Results of investigation and treatment. *World J. Surg.* **16**, 831–835.

Enck, P. (1993). Biofeedback training in disordered defecation: A critical review. *Digest. Dis. Sci.* **38**, 1953–1960.

Gordon, P. H. (2001). Anorectal anatomy and physiology. *Gastroenterol. Clin. North Am.* **30**, 1–13.

Joo, J. S., Agachan, F., Wolff, B., *et al.* (1996). Initial North American experience with botulinum toxin type A for treatment of anismus. *Dis. Colon Rectum.* **39**, 1107–1111.

Kapur, R. P. (1999). Hirschsprung disease and other enteric dysganglionoses. *Crit. Rev. Clin. Lab. Sci.* **36**, 225–273.

Lubowski, D. Z., Meagher, A. P., Smart, R. C., *et al.* (1995). Scintigraphic assessment of colonic function during defecation. *Int. J. Colorectal Dis.* **10**, 91–93.

Rao, S. S. C. (2001). Dyssynergic defecation. *Gastroenterol. Clin. North Am.* **30**, 97–114.

Voderholzer, W. A., Neuhaus, D. A., Klauser, A. G., *et al.* (1997). Paradoxical sphincter contraction is rarely indicative of anismus. *Gut.* **41**, 258–262.

Wald, A. (2001). Outlet dysfunction constipation. *Curr. Treat. Options Gastroenterol.* **4**, 293–297.

Whitehead, W. E., Wald, A., Diamant, N. E., *et al.* (2000). Functional disorders of the anus and rectum. *In* "The Functional Gastrointestinal Disorders" (D. A. Drossman, ed.), 2nd Ed., pp. 483–532. Degnon Associates, McLean, VA.

Development, Overview

ROBERT K. MONTGOMERY
Children's Hospital Boston

endoderm One of the three germ layers formed by the process of gastrulation. Gives rise to the epithelial lining of the gastrointestinal tract and its derivatives.

epithelial/mesenchymal interaction Fundamental process in development. Interaction may occur via soluble factors, produced by either cell type, binding to receptors on the surface of the other, via direct cell–cell contact, or via the basal lamina produced between the cell layers, to which the epithelial cells are attached.

forkhead-related proteins Family of transcription factors structurally similar to the product of the *Drosophila forkhead* gene. Also known as "winged helix" factors because of their three-dimensional structure. In mice and humans, the family has multiple members. It is now systematically known as the Fox family.

GATA factors Family of transcription factors that bind to DNA on the nucleic acid sequence -G-A-T-A-.

mesoderm One of the three germ layers formed by the process of gastrulation. Mesodermal cells give rise to mesenchymal tissue and the muscle layers of the gastrointestinal tract.

sonic hedgehog One member of a family of signaling proteins originally described in *Drosophila*. Sonic and Indian hedgehog proteins are key regulators of gastrointestinal morphogenesis.

Gastrulation, during which the axes of the embryo are determined and formation of the gastrointestinal tract is initiated, is an essential early step in development of all multicellular organisms. The endoderm layer generated during gastrulation gives rise to the epithelial lining of the gastrointestinal tract. Initially, the endoderm forms a simple tube surrounded by mesoderm. Regionalization and development of specialized organs along the tube appear early in evolution, suggesting that the mechanisms regulating gut tube formation are likely to be very early evolutionary developments and are similar in most organisms. Current research suggests that the mechanisms governing these processes are indeed highly conserved throughout evolution. Therefore, data from model organisms are directly relevant to human development.

INTRODUCTION

There are three major developmental milestones in formation of the gastrointestinal tract. First is the initial specification of the endoderm. Second is formation and patterning of the gut tube, which establishes the anterior–posterior axis and the boundaries between different organs. Third is the initiation of formation of organs that are outgrowths of the gut tube, such as liver and pancreas. Experiments in model organisms have identified families of genes involved in endoderm specification that are highly conserved in evolution, whereas other genes appear to be specific to vertebrate gut development. Conservation of form and function is supported by experiments in which mouse genes substitute for the homologous genes in frogs and fruit flies. Morphogenesis and functional development have been well described in other texts and reviews. The focus here is on current understanding of the molecular basis of the major milestones in gastrointestinal development and the roles of the best understood genes.

SPECIFICATION OF THE ENDODERM

Specification of the endoderm can be traced to the earliest stages of embryo formation. Classical experiments demonstrated that explants of chick embryos prior to gastrulation were capable of gastrointestinal development, indicating that their fate had already been specified. Evidence is accumulating in support of the hypothesis that the original patterning of the endoderm is cell autonomous, but that full development of the organs requires a reciprocal interaction between the endoderm and mesoderm. Six gene families that act to specify endoderm have now been identified in a number of model organisms. One class of genes encodes transcription factors that directly activate target genes. A second class encodes signaling molecules that mediate cellular interactions. At least some of the transcription factors involved in specification of the endoderm continue to be expressed in the gastrointestinal (GI) tract throughout development, such as the forkhead-related factors. Signaling pathways, such as hedgehog/bone morphogenetic protein (BMP), act at different times and in different locations to regulate GI development.

From its earliest stages, the endoderm is in close apposition to mesoderm throughout the GI tract. Tissue

 Copyright 2004, Elsevier (USA). All rights reserved.

recombination experiments have shown that patterning of the endoderm and its differentiation into separate organs result from signaling between the mesoderm and the endoderm. The earliest identified step in anterior/posterior patterning in mouse endoderm requires signaling from mesoderm to endoderm by fibroblast growth factor-4 (FGF-4). Other members of the FGF family and their receptors are critical in liver development. Three other important gene families mediating mesoderm/endoderm signaling are sonic hedgehog, the BMPs, and the *hox* genes.

It remains unclear if a single "master gene" initiates the formation of the endoderm, setting in motion the process of gastrointestinal development. In some of the model systems, genes have been identified that appear to be both necessary and sufficient to specify endoderm. In others, genes have been identified that are necessary, but may not be sufficient.

Two GATA transcription factor genes are essential in specification of the cells that give rise to the intestinal epithelium of *Caenorhabditis elegans*, and a *Drosophila* GATA factor is encoded by the gene *serpent*, previously demonstrated to be required for differentiation of gut endoderm. Three members of the GATA family, initially discovered as key regulators of erythroid differentiation, are expressed in vertebrate intestine. Distinct functions for GATA-4, -5, and -6 in intestinal epithelial cell proliferation and differentiation have been suggested, but their role in early development remains unclear. In addition to the GATA factors, members of the forkhead-related (Fox) family and members of the wnt/Tcf signaling pathway are regulators of endoderm formation. In addition to these factors, members of the transforming growth factor-β (TGF-β) superfamily critical in the initiation of endoderm formation have been identified in vertebrates. One of the effector molecules in this pathway, Smad2, has also been shown to be critical for endoderm formation. The evidence indicates that elimination of the genes for these factors blocks endoderm formation, whereas ectopic expression induces formation of endoderm from cells that would not normally do so.

Many transcription factors initially identified as liver specific have key roles in the intestine. When analyzed in model systems, several of these transcription factors have been found to be expressed in patterns suggesting that they may also regulate intestinal development. For example, hepatic nuclear factor 3β (HNF3β; now designated FoxA2) has been shown to be critical for the earliest differentiation of the gastrointestinal tract. Furthermore, it continues to be expressed in the adult progeny of the endoderm. Homozygous null mutants of HNF3β do not form a normal primitive streak, which gives rise to the gut tube and other structures. HNF3β is critical to formation of the foregut and midgut, but not the hindgut. Multiple members of this family have been identified, some of which display intestine-enriched or intestine-specific expression. One of the family members, normally expressed in the intestinal mesoderm, is a critical mediator of epithelial/mesenchymal interactions. Its elimination leads to abnormal epithelial cell proliferation and aberrant intestinal development. Thus, it appears likely that during intestinal development, multiple members of the HNF3β/forkhead family interact in a complex mechanism that remains to be elucidated.

Several mouse homeobox genes related to *Drosophila caudal* are expressed specifically in the intestine. One, *Cdx-1*, is restricted to the adult intestine but is expressed widely in the developing embryo. Another caudal-related gene, *Cdx-2*, is expressed in visceral endoderm of the early embryo but is restricted to the intestine at later stages. Forced expression of *Cdx-2* will induce differentiation in an intestinal cell line that does not normally differentiate. *Cdx-2* is clearly a critical intestine-specific differentiation factor, but its role in early development of the intestine remains unclear.

FORMATION OF THE GUT TUBE

Formation of the gut tube follows specification of the endoderm. In frogs, completion of gastrulation produces the primitive gut tube. In birds and mammals, the gut tube is formed from a layer of endoderm by a process of folding that begins at the anterior and posterior ends of the embryo. Reciprocal signaling between endoderm and mesoderm continue to be critical to the developmental process.

A key mechanism that has emerged as a mediator of endoderm/mesoderm interactions in the organization of the gastrointestinal tract involves the sonic hedgehog (Shh) and Indian hedgehog (Ihh) signaling proteins. Both Shh and Ihh play critical roles in anterior/posterior patterning and concentric patterning of the developing GI tract, at least in part through their role in development of muscle from the mesoderm. One target of this signaling pathway is a second family of signaling molecules, the bone morphogenetic proteins, members of the TGF-β superfamily.

Shh is first detectable in the primitive endoderm of the embryo, later in the endoderm of the anterior and posterior intestinal portals, and subsequently throughout the gut endoderm and in the adult crypt region. BMP-4 is expressed in the mesoderm adjacent to the intestinal portals and can be induced ectopically in the visceral mesoderm by Shh protein. The endoderm

of the intestinal portals is the source of Shh; the portal regions can act as polarizing centers if transplanted. Shh also induces the expression of *hox* genes. Knockout of a forkhead family member producing abnormal epithelial cell proliferation later in development likely has its effect through reduced expression of BMP-2 and BMP-4. Null mice display foregut anomalies such as esophageal atresia and tracheoesophageal fistula and hindgut anomalies such as persistent cloaca, indicating that Shh is a critical regulator of both foregut and hindgut development.

ORGAN DEVELOPMENT

Patterning

In *Drosophila*, the large family of homeotic genes is expressed in the body in a precise anterior to posterior order. The homeotic genes encode transcription factors, incorporating a conserved homeobox sequence, that regulate segmentation and pattern formation. Vertebrates have homologous *hox* genes that play important roles in the formation of distinctly delineated regions of the brain and skeleton. The four copies of the set of vertebrate genes, *hoxa*, *hoxb*, *hoxc*, and *hoxd*, form groups of paralogues, e.g., *hoxa-1*, *hoxb-1*, and *hoxd-1*. Within each group, the genes are expressed in the embryo in an anterior to posterior sequence of regions with overlapping boundaries, e.g., *hoxa-1* in the occipital vertebrae to *hoxa-11* in the caudal vertebrae.

A detailed study of the developing chick hindgut has demonstrated a correlation between the boundaries of expression of *hoxa-9*, *hoxa-10*, *hoxa-11*, and *hoxa-13* in the mesoderm and the location of morphologic boundaries. Regional differences in expression of homeobox genes in the developing mouse intestine have also been demonstrated. Interference with the expression of specific *hox* genes produces organ-specific gastrointestinal defects. Disruption of *hoxc-4* gives rise to esophageal obstruction due to abnormal epithelial cell proliferation and abnormal muscle development. Alteration of the expression pattern of *hox 3.1* (now *hoxc-8*) to a more anterior location causes distorted development of the gastric epithelium. Mice with disrupted *hoxd-12* and *hoxd-13* genes display defects in formation of the anal musculature. Expression of the human homologues of a number of homeobox genes has also been shown to be region specific. These data indicate that the *hox* genes are critical early regulators of proximal-to-distal, organ-specific patterning in gastrointestinal development. The *caudal* genes are members of a divergent homeobox gene family and regulate the anterior margins of *hox* gene expression as well as having GI-specific roles.

Regional Specification

Organs such as the stomach are first identifiable by thickening in the mesodermal layer. Early in the process of patterning, BMP-4 is expressed throughout the mesoderm. Sonic hedgehog is expressed in the endoderm and is an upstream regulator of BMP-4. The patterning of BMP-4 expression in the mesoderm regulates growth of the stomach mesoderm and determines the sidedness of the stomach. Location of the pyloric sphincter is dependent on the interaction of BMP-4 expression and inhibitors of that expression. Patterning of the concentric muscle layer structure is dependent on Shh signaling that induces formation of lamina propria and submucosa, while inhibiting smooth muscle and enteric neuron development near the endoderm.

Development of Organs from Outgrowths

Liver

The liver diverticulum emerges from the most caudal portion of the foregut just distal to the stomach. It is first detectable as a thickening in the endoderm of the ventral duodenum. Hepatogenesis is initiated through an instructive induction of ventral foregut endoderm by cardiac mesoderm. A series of elegant experiments have identified a number of signaling pathways involved in the complex process of development of the liver. The immediate signal is provided by cardiac mesoderm fibroblast growth factors that bind to specific receptors in the endoderm. The appearance of mRNA for the liver-specific protein albumin in endodermal cells of the liver diverticulum is one of the earliest indications of hepatocyte induction. Endothelial precursor cells provide another critical factor for hepatogenesis. After formation of the liver bud, hepatocyte growth factor (HGF) is required for continued hepatocyte proliferation. The hepatic diverticulum grows into the septum transversum and gives rise to the liver cords, which become the hepatocytes. During this process, a combination of signals from the cells of the septum transversum, including BMP, is necessary for liver development.

Pancreas

Development of the pancreas has provided one of the classic examples of epithelial–mesenchymal interactions. Previous investigations have shown that growth and differentiation of the pancreas requires the presence of mesenchyme, although both endocrine and exocrine cells develop from the foregut endoderm. Analysis of the development of separated endoderm and mesenchyme under different conditions indicates that the "default pathway" of pancreatic differentiation leads

to endocrine cells, whereas combinations of extracellular matrix and mesenchymal factors are required for complete organogenesis.

The dorsal pancreatic bud arises in an area where Shh expression is repressed by factors from the notochord. One of the earliest signs of pancreas development is expression of the *pdx-1* gene in cells of the pancreatic bud. This gene is a homeobox gene, related to the *Drosophila* gene *antennapedia*. The encoded protein has been found to be expressed in the epithelium of the duodenum immediately surrounding the pancreatic buds, as well as in the epithelium of the buds. Examination of an initial *pdx-1* knockout mouse indicates that, although development of the rest of the gastrointestinal tract and the rest of the animal is normal, the pancreas does not develop. A second study by another group of researchers, who independently made a *pdx-1* null mouse, found that the dorsal pancreas bud does form, but its development is arrested. The defect due to the *pdx* knockout is restricted to the epithelium, as indicated by the maintenance of normal developmental potential by the mesenchymal cells. In addition, the most proximal part of the duodenum in the null mice is abnormal, forming a vesicle-like structure lined with cuboidal epithelium, rather than villi lined by columnar cells, indicating that *pdx-1* influences the differentiation of cells in an area larger than that giving rise to the pancreas, consistent with the earlier delineated domain of expression. A case of human congenital pancreatic agenesis has been demonstrated to result from a single nucleotide deletion in the human *pdx-1* gene.

CONCLUSION

Key regulators of gastrointestinal development have been identified. Some of the genes critical in epithelial/mesenchymal interaction, long known to be a fundamental developmental process, are now known. Analysis of the expression pattern of the *hox* genes suggests that they act to pattern the gastrointestinal tract. The hedgehog proteins mediate several aspects of early development, but inhibition experiments suggest that after organ formation, their role is largely complete. Targeted disruption of several genes that regulate intestinal growth indicate that bone morphogenetic protein secretion has a key developmental role in cell proliferation and villus morphology. Most of the signaling pathways identified are short range. With the exception of epidermal growth factor (EGF), there is little compelling evidence for a role of any circulating or luminal growth factor in the development of the intestine.

Microarray analysis of gene expression profiles indicates that the organs of the adult gastrointestinal tract display distinct patterns. Furthermore, the analysis identifies common regulatory elements, including those for HNF1 and GATA factors, in the 5′ flanking sequences of groups of genes expressed in specific regions, suggesting organ-specific regulation. A combination of work on critical individual genes with examination of cell-and organ-specific developmental gene expression profiles should provide a deeper understanding of the regulation of gastrointestinal development.

See Also the Following Articles

Biliary Tract, Development • Esophagus, Development • Large Intestine, Development • Liver, Development • Pancreas, Development • Peritoneum, Anatomy and Development • Small Intestine, Development

Further Reading

Clements, D., Rex, M., and Woodland, H. R. (2001). Initiation and early patterning of the endoderm. *Int. Rev. Cytol.* **203**, 383–446.

Kedinger, M., Simon-Assmann, P., Bouziges, F., and Haffen, K. (1988). Epithelial–mesenchymal interactions in intestinal epithelial differentiation. *Scand. J. Gastroenterol. Suppl.* **151**, 62–69.

Kim, S. K., and Hebrok, M. (2001). Intercellular signals regulating pancreas development and function. *Genes Dev.* **15**, 111–127.

Montgomery, R. K., Mulberg, A. E., and Grand, R. J. (1999). Development of the human gastrointestinal tract: Twenty years of progress. *Gastroenterology* **116**, 702–731.

Roberts, D. J. (2000). Molecular mechanisms of development of the gastrointestinal tract. *Dev. Dyn.* **219**, 109–120.

Stainier, D. Y. (2002). A glimpse into the molecular entrails of endoderm formation. *Genes Dev.* **16**, 893–907.

Traber, P. G., and Silberg, D. G. (1996). Intestine-specific gene transcription. *Annu. Rev. Physiol.* **58**, 275–297.

Zaret, K. S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nat. Rev. Genet.* **3**, 499–512.

Diabetes Mellitus

MANISHA CHANDALIA AND NICOLA ABATE
University of Texas Southwestern Medical Center, Dallas

diabetes ketoacidosis A syndrome consisting of hyperglycemia, ketosis, and acidemia (an arterial pH <7.3, a bicarbonate value <15 mEq/liter, and a blood glucose level >250 mg/dl with a variable degree of ketonemia and ketonuria).

hyperglycemia The state of having a higher than normal plasma glucose concentration. Normal values of plasma glucose differ depending on the whether the subject is in the fasting or fed state.

hypoglycemia The state in which plasma glucose concentration is decreased sufficiently to produce symptoms, which improve with restoration of normal plasma glucose.

insulin resistance Inability of insulin to lower plasma glucose effectively, with resultant higher plasma insulin levels.

insulin secretagogues Drugs that improve insulin secretion: sulfonylurea and metiglinides.

insulin sensitizers Drugs that improve insulin action: biguanides (metformin) and thiazolidinediones (rosiglitazone and pioglitazone).

nonketotic hyperosmolar state A syndrome associated with little or no ketoacid accumulation in blood, hyperglycemia (plasma glucose often exceeding 1000 mg/dl), increased plasma osmolality (as high as 400 mosm/kg), and neurological abnormality.

Diabetes mellitus is a complex metabolic disorder characterized by hyperglycemia and long-term complications affecting the heart, kidney, eyes, and other organs including the gastrointestinal tract. The main metabolic feature of diabetes mellitus consists of a relative or absolute impairment of insulin secretion, along with varying degrees of peripheral resistance to the biological activity of insulin. This is a pathophysiologically heterogeneous group of disorders, with the common denominator being hyperglycemia. The symptoms, manifestations, and management options vary depending on the underlying pathogenesis and type of disease.

CLASSIFICATION

The American Diabetes Association (ADA) revised its classification of diabetes mellitus in 1997 according to etiologic differences and to move away from descriptions based upon age at onset or type of treatment (Table I).

In its classic form, two common types of diabetes are recognized: type 1 (formerly called insulin-dependent diabetes mellitus) and type 2 (formerly called non-insulin-dependent diabetes mellitus). The majority of patients with type 1 diabetes have autoimmune destruction of the pancreatic beta cells as the underlying cause, have an absolute requirement for insulin therapy, and will develop diabetic ketoacidosis if not given insulin. Type 2 diabetes, on the other hand, is characterized by association with obesity and absence of an absolute dependence on insulin. In addition to these two primary types of diabetes and gestational diabetes, a final group includes a large number of conditions with either secondary forms of diabetes or those associated with specific known gene mutations and is classified as diabetes of other types.

DIAGNOSIS

The diagnosis of diabetes mellitus is easy to establish when a patient presents with classic symptoms of hyperglycemia (polydipsia, polyuria, weight loss, visual blurring) and has a fasting blood glucose concentration of at least 126 mg/dl (7.0 mmol/liter) or a random value of at least 200 mg/dl (11.1 mmol/liter), confirmed on another occasion. The American Diabetes Association revised the criteria for diagnosis of diabetes in 1997 (Table II). The new diagnostic criteria strongly suggest that the diagnosis of diabetes be made on the basis of fasting blood glucose only. The blood glucose criteria for the diagnosis of gestational diabetes are lower and include values 1, 2, and 3 h after 100 g of oral glucose (Table II). The use of glycosylated hemoglobin values for screening and identification of diabetes is also considered by many investigators. Though a sensitive means of assessing glycemic control in patients with diabetes, since the assays for glycosylated hemoglobin are not yet standardized, the diagnosis of diabetes mellitus should not be made on the basis of glycosylated hemoglobin values alone. One other caution is that the values mentioned in Table II are for venous plasma. Capillary whole-blood values normally measured by patients at home are lower in the basal state and are considered abnormal if they exceed 120 mg/dl.

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Diabetes Mellitus

Type 1 diabetes
- A. Immune-mediated
- B. Idiopathic

Type 2 diabetes

Gestational diabetes mellitus

Other specific types
- A. Genetic defects of beta-cell function
  1. Chromosome 12, HNF-1-α (formerly MODY3)
  2. Chromosome 7, glucokinase (formerly MODY2)
  3. Chromosome 20, HNF-α (formerly MODY1)
  4. Mitochondrial DNA
- B. Genetic defects in insulin action
  1. Type 1 insulin resistance
  2. Leprechaunism
  3. Lipoatrophic diabetes
- C. Diseases of exocrine pancreas
  1. Pancreatitis
  2. Trauma/pancreatectomy
  3. Neoplasia
  4. Cystic fibrosis
  5. Hemochromatosis
  6. Fibrocalculous pancreatopathy
- D. Endocrinopathies
  1. Cushing's syndrome
  2. Acromegaly
  3. Pheochromocytoma
  4. Glucagonoma
  5. Hyperthyroidism
  6. Primary hyperaldosteronism
  7. Somatostatinoma
- E. Drugs or chemicals
  1. Glucocorticoids
  2. Thiazide diuretics
  3. Nicotinic acid
  4. Pentamidine
  5. Others
- F. Infections
  1. Congenital rubella
  2. Cytomegalovirus
  3. Others
- G. Uncommon forms of immune-mediated diabetes
  1. "Stiff man" syndrome
  2. Anti-insulin receptor antibodies
- H. Other genetic syndromes associated with diabetes
  1. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD)
  2. Down's syndrome
  3. Klinefelter's syndrome
  4. Others: Turner's syndrome, Freidrich's ataxia, myotonic dystrophy

Note. Data are from the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997). *Diabetes Care* **20**, 1183–1197.

PATHOGENESIS

Type 1 Diabetes

Type 1 diabetes mellitus results from autoimmune destruction of the insulin-producing beta cells in the islets of Langerhans in the pancreas. This process occurs in genetically susceptible subjects, is triggered by one or more environmental agents, and usually progresses over many months or years. Genes in both the major histocompatibility complex and elsewhere in the genome appear to be involved. The lifelong risk of type 1 diabetes is markedly increased in close relatives of patients with type 1 diabetes, averaging approximately 6% in offspring, 5% in siblings, and 30% in identical twins (versus 0.4% in subjects with no family history). The major susceptibility gene for type 1 diabetes is in the human leukocyte antigen (HLA) region on chromosome 6p and appears to account for approximately 40% of familial clustering of type 1 diabetes. In particular, more than 90% of patients with type 1 diabetes carry HLA-DR3,DQB1*0201 or -DR4,DQB1*0302 and approximately 30% carry both haplotypes, which confers the greatest susceptibility. On the other hand, the HLA allele DQB1*0602 confers protection against the development of type 1 diabetes.

The lack of 100% concordance of type 1 diabetes in monozygotic twins underscores the importance of environmental factors, such as viruses, as triggers in the development of the disease. One such virus is congenital rubella. Other viral infections (coxsackie virus, Epstein-Barr virus, mumps, or cytomegalovirus) and other exposures, such as to cow's milk protein and certain drugs, have been suggested as triggers but no conclusive data are available at this time.

Cytoplasmic islet cell antibodies are identified by immunofluorescence in 70 to 80% of patients with newly diagnosed type 1 diabetes and in prediabetic subjects. The presence of islet cell antibodies can precede the development of overt diabetes by more than a decade and can be used to identify persons at high risk. Several autoantigens have been recognized within the pancreatic beta cells that may play important roles in the initiation or progression of autoimmune islet injury. One of the first and most important autoantigens against which antibodies are detected is the enzyme glutamic acid decarboxylase (GAD) and antibodies to GAD are found in approximately 70% of patients with type 1 diabetes at the time of diagnosis.

Type 2 Diabetes

Type 2 diabetes is the most prevalent endocrine disease. It is estimated to affect more than 15 million

TABLE II Diagnosis of Diabetes Mellitus

Criteria for diagnosis of diabetes mellitus
Classic symptoms of diabetes such as polyuria, polydipsia, ketonuria, and rapid weight loss with random plasma glucose > 200 mg/dl
or
Fasting plasma glucose concentration > 126 mg/dl on more than one occasion
or
Glucose concentration at 2 h of > 200 mg/dl during an oral glucose tolerance test
Criteria for diagnosis of gestational diabetes mellitus
Two or more of the following plasma glucose concentrations after fasting and a 100 g oral glucose tolerance test:
Fasting serum glucose concentration > 95 mg/dl (5.3 mmol/liter)
1 h serum glucose concentration > 180 mg/dl (10 mmol/liter)
2 h serum glucose concentration > 155 mg/dl (8.6 mmol/liter)
3 h serum glucose concentration > 140 mg/dl (7.8 mmol/liter)

people in the United States, one-third of whom are undiagnosed. Prevalence is higher among Hispanics, Native Americans, African Americans, and Asian Indians, reaching as high as 50% among adult Pima Indians living in the United States. Development of type 2 diabetes is strongly influenced by genetic factors and environmental factors including obesity and decreased physical activity. Family history of diabetes increases the risk for development of this disease.

Type 2 diabetes mellitus is characterized by hyperglycemia, insulin resistance, and relative impairment in insulin secretion. Type 2 diabetes is often accompanied by other conditions, including hypertension, high serum triglyceride concentrations, and low serum high-density lipoprotein (HDL) cholesterol concentrations, that, like type 2 diabetes, can increase cardiovascular risk. This constellation of clinical conditions is referred to as "metabolic syndrome" and has insulin resistance as the main common pathogenetic factor. Although the mechanisms leading to insulin resistance are still a matter of investigation, evidence that this abnormality is the consequence of an interaction between genetic and environmental factors is accumulating. Obesity, particularly truncal obesity, and lack of physical activity have been shown to promote insulin resistance and, through this mechanism, are probably the most powerful risk factors for type 2 diabetes. However, the degree of fat accumulation and lack of physical activity necessary to induce insulin resistance seem to be modulated by genetic factors. This is evident from studies in families and in ethnic groups, such as the Asian Indians, who develop severe insulin resistance even with mild accumulation of total body fat. Once insulin resistance develops, the development of diabetes is determined by the degree of beta-cell compensation for the defective biological activity of insulin. If adequate hyperinsulinemia is maintained, the insulin-resistant individual may not develop hyperglycemia and the clinical picture of the metabolic syndrome will manifest without diabetes. On the other hand, the presence of a genetic predisposition to beta-cell decompensation seems to play a role in determining the onset of hyperglycemia and the complete clinical onset of type 2 diabetes. Again, acquired factors, i.e., obesity and insulin resistance, interact with genetic predisposition to determine beta-cell decompensation, a necessary step in the pathophysiology of type 2 diabetes. Both insulin resistance and beta-cell decompensation are needed for the development of this disease. The relative contribution of these two mechanisms to the pathogenesis of type 2 diabetes may vary in different individuals. The results will be increased hepatic glucose production and decreased insulin-mediated glucose utilization in the skeletal muscle. This leads to progressively worsening hyperglycemia. The constellation of metabolic abnormalities that accompany hyperglycemia in type 2 diabetes, i.e., dyslipidemia, hypercoagulable state, and hypertension, are mainly related to insulin resistance and precede the onset of hyperglycemia. This explains the common rapid progression of atherosclerotic disease in these patients. Microvascular complications, on the other hand, are mainly determined by hyperglycemia with mechanisms similar to those in type 1 diabetic patients.

MANAGEMENT OF DIABETES

Diabetes is a chronic disease and the principles of therapy are twofold, maintaining euglycemia and preventing or delaying complications of diabetes. Diet, weight reduction, and exercise can all be used to improve glycemic control. Stress management training might also be useful, although its effect is modest. In type 1 diabetes, insulin is the mainstay of therapy and is essential for survival. However, in type 2 diabetes, various other modalities of treatment may be used as either monotherapy or combination therapy.

Nonpharmacological Therapy

This involves diet, exercise, and weight reduction/maintenance. Both type 1 and type 2 patients benefit from this and many patients with type 2 may be managed with lifestyle changes alone. The diet for a diabetic individual is to be planned depending on age, level of

physical activity, obesity, other complications such as hyperlipidemia or hypertension, and cultural acceptability. The general principle of dietary therapy in type 2 diabetic patients is a low-calorie, low-fat, and complex carbohydrate diet. The controversy in the field of diabetic diet is due to potential problems with each of the dietary constituents: too much carbohydrate might worsen hyperglycemia, too much fat might increase the already high risk of atherosclerosis, and too much protein might promote the development of diabetic nephropathy. The traditional recommendation is that 55 to 60% of total calories should be derived from carbohydrate and most of the dietary carbohydrate should be in the form of unrefined, high-fiber foods (up to 40 g/day) with only modest amounts of sucrose and other refined sugar. Insoluble fiber, such as that found in whole wheat bread and brown rice, has a major impact on gastrointestinal transit time and fecal bulk. There is, however, very little alteration in the rate of blood glucose increase after a meal. In comparison, soluble fiber, such as that contained in vegetables, fruit, and especially legumes, slows the postprandial rise in blood glucose, often leading to a reduction in insulin requirements. Twelve to 20% of total calories should be derived from proteins (approximately 0.8 g/kg/day in adults). Higher levels are required in growing children and pregnant or lactating women and lower levels may be beneficial in patients with diabetic nephropathy. Less than 30% of total calories should be derived from fat, most of which should be mono- or polyunsaturated. Saturated fats and cholesterol intake should be less than 10% of total calories and less than 300 mg/day, respectively. Some data have questioned these standard dietary recommendations in patients with type 2 diabetes. In particular, a diet slightly lower in carbohydrate and higher in monounsaturated fat may result in a better metabolic profile. Therefore, the ADA has also recommended a diet high in monounsaturated fat for type 2 diabetic patients as an alternative to a high-complex-carbohydrate diet.

Some form of regular exercise is likely to be beneficial in most patients with diabetes, even those with advanced, long-standing disease. However, for those over age 35 years who have had diabetes for more than 10 years, a complete physical examination and exercise stress test should be performed before an exercise program is begun. The exercise should be performed regularly (at least three times per week) and preferably at the same time in relation to meals and insulin injections in patients treated with insulin. The duration and intensity of exercise should be increased gradually as tolerated by the patient. The type of exercise may have to be modified depending on preexisting complications, such as hypertension or neuropathy.

Lifestyle modifications (diet and exercise) have recently been shown to prevent the onset of type 2 diabetes in subjects with impaired glucose tolerance. Lifestyle modification also has a significant preventive effect on cardiovascular mortality. Thus, a diet and exercise program should be actively recommended and reinforced in all diabetic management plans.

Pharmacological Therapy

Various classes of oral agents are available for the management of hyperglycemia in type 2 diabetic patients. These include sulfonylureas and meglitinides, which are insulin secretagogues, biguanides and thiazolidinediones, which are insulin sensitizers, α-glucosidase inhibitors, which act primarily in the gut, and insulin therapy. Long-term studies on the natural progression of type 2 diabetes show a gradual worsening of beta-cell function, eventually requiring insulin therapy in a majority of patients. However, depending on the degree of insulin resistance, many patients with type 2 diabetes may not achieve glycemic control with insulin alone. Thus, combination therapy either with insulin secretagogues and insulin sensitizers or with insulin and insulin sensitizers is recommended. One major side effect with all pharmacological therapy, except monotherapy with metformin, is hypoglycemia. The incidence of hypoglycemia was higher in major clinical trials with intensive therapy to achieve good glycemic control. Another side effect with all agents except metformin is weight gain, though it may be variable with different agents (more with thiazolidinedione than with other agents). Metformin is recommended as monotherapy or as combination therapy with sulfonylurea and thiazolidinedione as well as insulin. One major side effect of metformin is diarrhea, which may be severe enough to require discontinuation of therapy. Another infrequent but serious side effect is lactic acidosis. To avoid lactic acidosis, metformin is contraindicated in patients with renal failure, liver failure, severe congestive heart failure, and significant lung disease. Thiazolidinediones have liver abnormalities as one major adverse effect. It can involve mild elevation of transaminases to fulminant liver failure. Periodic routine monitoring is necessary when patients are first started on these agents and prompt discontinuation of therapy if serum transaminases increase to more than three times baseline is warranted to prevent severe hepatotoxicity.

No pharmacological agent except insulin is recommended in the treatment of type 1 diabetes. Usually a

combination of short-acting and long-acting insulin is given by multiple injections to mimic normal physiologic insulin secretion. The mode of delivery can be either multiple subcutaneous injections or continuous subcutaneous insulin infusion by a programmable pump. Patients generally monitor their capillary blood glucose and adjust the dose of insulin to achieve optimal glycemic control. The choice of insulin and mode of delivery depend on the degree of hyperglycemia, the patient's willingness to monitor blood glucose closely, and the competency and trainability of the patient.

COMPLICATIONS

Diabetes is a chronic disease with significant late complications of the kidneys, eyes, and heart developing over many years or decades after onset of hyperglycemia. The rate of onset and development of these complications is not only dependent upon glycemic control but also on other genetic and environmental factors. One important complication is diabetic nephropathy, which can be recognized clinically as persistent proteinuria, which subsequently progresses to end-stage renal failure. Though evidence points to genetic susceptibility as a determinant, uncontrolled diabetes (hyperglycemia) appears to be a necessary factor for the development of nephropathy. Since genetic markers to identify individuals at risk are not yet available, tight glycemic control and early intervention with angiotensin-converting enzyme inhibitors or angiotensin receptor blockers in all patients with diabetes are a useful strategy in preventing and/or delaying the onset of nephropathy. Another serious late complication is retinopathy, a significant cause of blindness. Both type 1 and type 2 diabetes patients develop background retinopathy within the first 5 to 15 years of diabetes. These patients are then susceptible to develop proliferative retinopathy with risk of vision loss. The actual incidence of proliferative retinopathy is related to glycemic control. A large clinical trial demonstrated a substantial benefit of tight glycemic control in the primary prevention of diabetic retinopathy: at 9 years, the incidence of new retinopathy was 12% with intensive therapy versus 54% with conventional therapy. Another trial demonstrated a similar benefit of tight glycemic control in the prevention of retinopathy in type 2 diabetes. Furthermore, routine screening for retinopathy to detect early changes to prevent vision loss is also a very effective strategy. The benefit of intensive therapy to achieve tight glycemic control was also noted in diabetic neuropathy patients.

The most notable complication and cause of mortality in patients with type 2 diabetes is coronary heart disease. There are conflicting data on the importance of glycemic control on the development of macrovascular disease in patients with type 2 diabetes. In one large study on type 2 diabetes, it was noted that cardiovascular disease was associated with traditional cardiovascular risk factors (including hypertension, smoking, age, and serum total/HDL cholesterol ratio) and with the duration of diabetes and not associated with the degree of glycemic control. However, a subanalysis suggested that reducing the HbA1c value by 1% was associated with an 18% reduction in myocardial infarction and a 15% reduction in stroke. Therefore, vigorous cardiac risk reduction in patients with type 2 diabetes with optimal lipid and blood pressure control, lifestyle modification, cessation of smoking, treatment with aspirin and angiotensin-converting enzyme inhibitor or angiotensin receptor blocker agent is essential for prevention of coronary heart disease.

GASTROINTESTINAL COMPLICATIONS OF DIABETES

Diabetes mellitus can affect many organs in the gastrointestinal tract. Dysphagia, nausea, vomiting, diarrhea, constipation, and fecal incontinence are common complaints among diabetic patients, although some of these symptoms can be due to the side effects of pharmacological agents taken for treatment of diabetes. Differentiating causes of these symptoms can be difficult and should not be attributed to diabetes without proper evaluation for other possible causes.

Although the pathogenetic mechanisms causing diabetic gastrointestinal symptoms remains poorly understood, the most common explanation is that autonomic visceral neuropathy causes most of the symptoms. Patients with gastrointestinal (GI) complications very often have additional manifestations of autonomic neuropathy such as orthostatic hypotension, urinary bladder dysfunction, and impotence. Other mechanisms, such as hyperglycemia and electrolyte imbalance affecting nerve function, altered production of various gut hormones affecting motility, and diabetic microangiopathy, have also been suggested.

Esophagus

Esophageal motor dysfunction in diabetics has been extensively studied by techniques such as barium study, nuclear methods, and manometry. However, there is not a close correlation between results of these studies and

clinical symptoms. The most typical symptoms are heartburn, dysphagia, and chest pain. Diabetics with dysphagia should be completely evaluated to rule out other conditions, such as gastroesophageal reflux disease, benign or malignant stricture, or esophagitis. Patients with chest pain should be assessed for coronary artery disease.

Stomach

The most common gastric disorder of diabetes is gastroparesis, which is manifested by postprandial fullness, epigastric pain, nausea, vomiting, and anorexia. A complete evaluation to exclude gastric outlet obstruction, peptic ulcer disease, and side effects of pharmacological agents must be performed in all diabetic patients with complaints of gastroparesis. Barium studies in patients with diabetic gastroparesis usually show an elongated stomach with sluggish, ineffective, or absent peristalsis. Because of impaired gastric emptying, these patients are prone to the formation of bezoars. Due to fluctuation in absorption from the small intestine, glucose control becomes erratic in patients with diabetic gastroparesis and frequent small meals should be recommended to improve glycemic control. Pharmacological prokinetic agents are used for treatment. If a patient fails prokinetic therapy (e.g., with metoclopramide) and anti-emetic therapy, other options are percutaneous endoscopic gastrostomy and jejunostomy that allows that allows gastric decompression and adequate nutritional support. Gastric pacing is another option that has been shown to improve gastric emptying in some patients.

Small Intestine

The incidence of diarrhea in diabetic patients is ~20%. Usually the diarrhea is chronic and intermittent, alternating with either normal bowel movement or constipation. Despite chronicity of the symptoms, weight loss is uncommon. Etiology of diarrhea is multifactorial, including autonomic neuropathy, bacterial overgrowth, and pancreatic insufficiency. Evaluation of the diabetic patient with diarrhea involves ruling out other causes including celiac disease (see below) and infection. Stool output should be quantified and a 72 h fecal fat measurement should be performed to assess for exocrine pancreatic insufficiency and sprue. A large number of diabetic patients fall in the category of idiopathic diarrhea and may benefit from a trial of conventional anti-diarrheal agents, clonidine, or psyllium.

Celiac Disease

The incidence of celiac disease is reported to be greater in diabetics than in the general population. The histocompatible antigens HLA-DR3 and HLA-DQB1*0201 have been linked to both celiac disease and diabetes, suggesting a genetic predisposition to both conditions. The clinical manifestation of celiac disease is similar to idiopathic diabetic diarrhea; however, it may present before onset of diabetes and is usually is accompanied by steatorrhea. A biopsy of the small bowel showing flattened villi and increased cellularity in the lamina propria establishes the diagnosis. Serum immunoglobulin A (IgA) anti-endomysial and IgA anti-gliadin antibodies have 80 to 100% sensitivity and specificity for celiac disease and can be used for diagnosis. A gluten-free diet improves both the symptoms and histology. With successful therapy, titers of these antibodies decrease or even disappear and thus can be used for monitoring patients with celiac disease.

Large Intestine

The most common GI complaint in diabetics is constipation (20 to 60%). The pathogenesis of constipation is poorly understood. Neurogenic dysfunction is considered to be an important factor in diabetic constipation. The evaluation of constipation should be based on age, severity of symptoms, and family history, thus ruling out other causes including colon cancer in high-risk candidates. Dietary fiber, stool softeners, and occasionally laxatives are used for the treatment of constipation in diabetics.

Fecal incontinence is another symptom that usually is associated with diarrhea and autonomic neuropathy. Therapy of fecal incontinent is difficult; however, most patients show improvement with control of diarrhea. If it fails, biofeedback and sphincter control improvement techniques can be used but usually are not very successful.

Gallbladder Disease

Diabetic patients have twice the incidence of gallstones compared to the general population. However, obesity and dyslipidemia are confounding factors and when corrected for these factors, diabetics do not seem to have a greater incidence of cholelithiasis than the general population. Several older studies had reported that when gallstones become symptomatic, morbidity and mortality is higher in diabetic patients than in the general population due to infectious complications. However, it is not clear whether diabetics with

asymptomatic gallstones are at a greater risk of becoming symptomatic than the general population. At present, diabetic patients with cholecystitis or biliary colic should be advised to undergo cholecystectomy as soon as possible. Asymptomatic gallstones in diabetics should be followed rather than treated by prophylactic cholecystectomy.

Liver Disease

The incidence of liver dysfunction in diabetic patients is high. It may range from a mild elevation of serum transaminases to (rarely) liver failure. Viral hepatitis is associated with diabetes, especially chronic hepatitis C. Oral hypoglycemic agents such as sulfonylurea and thiazolidinediones may cause abnormality of liver functions in diabetics. Thiazolidinediones may cause liver failure. Routine monitoring of serum transaminase levels is recommended for patients taking a thiazolidinedione. Stopping the agent usually leads to clinical and biological resolution.

One condition that is seen frequently in diabetic patients is hepatic steatosis. Fat accumulation in the liver occurs in up to 50% of patients with diabetes. Obesity is associated with hepatic steatosis as well. Intake of excessive dietary fat and carbohydrate leads to high levels of fatty acids that are stored in hepatocytes. Accumulation of fatty acids may provide a source of oxidative stress and lead from steatosis to steatohepatitis to cirrhosis. Only a small percentage of patients progress from steatosis to steatohepatitis and eventually to cirrhosis. This whole spectrum of conditions is now referred to as nonalchoholic fatty liver disease. Diagnosis of nonalchoholic fatty liver disease requires exclusion of alcohol abuse, absence of other causes such as viral infections, autoimmune diseases, drugs, and toxins. Liver biopsy is the best test for confirming the diagnosis; however, it may not be required in many straightforward cases. Management of nonalcoholic fatty liver disease is gradual weight loss and glycemic and lipid control. Patients should be advised to keep away from hepatotoxic agents such as alcohol.

See Also the Following Articles

Celiac Disease • Constipation • Diabetic Gastroparesis • Diabetic Neuropathies • Diarrhea • Gallstones, Pathophysiology of • Hepatitis C • Non-Alcoholic Fatty Liver Disease • Obesity, Treatment of

Further Reading

Kahn, R. (2001). Diabetes mellitus. *In* "Principles and Practice of Endocrinology and Metabolism" (K. L. Becker, ed.), 3rd ed., pp. 1296–1468. Lippincott Williams & Wilkins, Philadelphia, PA.

Kahn, R., and Weir, G. C. (eds.) (1994). Joslin's Diabetes Mellitus, 13th ed. Lea & Febigen, Malvern, PA.

Powers, A. C. (2001). Diabetes mellitus. *In* "Harrison's Principles of Internal Medicine" (E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and J. L. Jameson, eds.), 15th Ed., pp. 2109–2137. McGraw-Hill, New York.

Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997). *Diabetes Care* **20**, 1183–1197.

Verne, G. N., and Sninsky, C. A. (1998). Diabetes and the gastrointestinal tract. *Gastroenterol. Clin. North Am.* **27**, 861–874.

Diabetic Gastroparesis

LAWRENCE R. SCHILLER
Baylor University Medical Center and University of Texas Southwestern Medical Center, Dallas

gastroparesis Chronic condition characterized by delayed gastric emptying, resulting in gastric retention of ingested material; may be idiopathic or due to drugs or an underlying condition, such as diabetes mellitus.
prokinetic drug Agent that accelerates transit through one or more segments of the gastrointestinal tract.

Diabetes mellitus is sometimes complicated by dysfunction of the stomach, characterized by slow gastric emptying that produces gastric retention and chronic nausea and vomiting. This condition was first reported in detail at the Mayo Clinic in 1945 by R. W. Rundles, who attributed the condition to neuropathy. The term "gastroparesis diabeticorum" was coined in 1958 by P. Kassander and the condition has been called diabetic gastroparesis since that time. Several hypotheses to account for this problem have been elaborated and therapies to manage the condition have been developed.

CLINICAL PRESENTATION

The initial description of diabetic gastroparesis emphasized the association of nausea and vomiting with long-standing, poorly controlled diabetes complicated by diabetic neuropathy, preponderantly in young men. It is clear now that patients with diabetes may have delayed gastric emptying and thus delayed glucose absorption without nausea or vomiting, and that a substantial number of women also are affected. Symptoms, if present, are chronic and variable, with exacerbations and remissions. Gastroparesis may be recognized unexpectedly during the investigation of dyspepsia by the radiographic or endoscopic finding of gastric retention of food.

PATHOPHYSIOLOGY

Little progress was made in understanding diabetic gastroparesis until the 1970s, when physicians gained a deeper appreciation of gastric physiology and accurate radioisotope gastric emptying studies became available. The most marked abnormality in patients with gastroparesis is impaired emptying of solids; liquid emptying is relatively well preserved. There is evidence of vagal dysfunction that may contribute to these findings in many patients with diabetic gastroparesis, thus confirming Rundles' original idea of the pathogenesis of gastroparesis. Loss of the vagally mediated interdigestive migrating motor complex appears to be critical. Recent studies have emphasized the key role that hyperglycemia plays in triggering symptoms and additional problems of gastric accommodation and pyloric function in patients with diabetes.

DIAGNOSIS

Diagnosis is based on recognition of clinical symptoms, including nausea and vomiting typically hours after meals, dyspepsia, early satiety, bloating, or weight loss in a diabetic patient with poor control of blood sugar. Most patients have a long history of diabetes (>10 years), but this is not always the case. Findings of diabetic peripheral and autonomic neuropathy, such as numbness or pain in the feet, abnormal sweating, Marcus–Gunn pupils, diarrhea or constipation, and fecal incontinence, typically are present. An abnormal radioisotope gastric emptying study showing slow emptying of solids over 4 hours can confirm the diagnosis.

MANAGEMENT

Treatment consists of efforts to control hyperglycemia, because maintenance of blood sugar levels <200 mg% is associated with reduced symptoms. "Prokinetic" drugs, such as metoclopramide, may reduce symptoms, but do not always speed gastric emptying. Erythromycin given intravenously before meals is the most effective available agent for speeding gastric emptying in diabetics with gastroparesis. This drug is a motilin-receptor agonist and induces increased propulsive activity in the stomach and upper intestine. Oral erythromycin has less consistent efficacy. Nonantibiotic macrolide motilin-receptor agonists are under development. Patients who are not responsive to drugs can be outfitted with

 Copyright 2004, Elsevier (USA). All rights reserved.

a gastrostomy tube for gastric drainage and a jejunostomy tube for feeding. Innovative treatments for diabetic gastroparesis include injection of botulinum toxin into the prepyloric antrum, gastric electrical stimulation, and pancreatic or islet cell transplantation. Initial reports suggest that botulinum toxin injection and gastric electrical stimulation may reduce symptoms, but gastric emptying may not be improved. Reversal of diabetes with successful pancreatic or islet cell transplantation can cure this condition.

See Also the Following Articles

Diabetes Mellitus • Diabetic Neuropathies • Emesis • Gastric Emptying • Nausea

Further Reading

Camilleri, M. (2002). Advances in diabetic gastroparesis. *Rev. Gastroenterol. Disord.* **2**, 47–56.

Malagelada, J. R. (1995). Diabetic gastroparesis. *Semin. Gastrointest. Dis.* **6**, 181–186.

Diabetic Neuropathies

KAREN E. HALL

Ann Arbor Veterans Affairs Healthcare System and University of Michigan

apoptosis A mechanism of cell death that is controlled by specific metabolic cell death pathways and that results in fragmentation and condensation of the cell nucleus, without significant inflammation in the surrounding tissue.

autonomic Pertaining to the internal systems in the body (cardiovascular, pulmonary, gastrointestinal) that receive innervation that is not under voluntary control.

diabetic neuropathy Short-term and long-term nerve dysfunction in diabetes characterized by generalized slowing of conduction and increased threshold for excitation in peripheral and autonomic nerves.

gastrointestinal motility The process of contraction of the gastrointestinal wall that results in movement of food and secretions along the length of the gastrointestinal tract.

gastroparesis Delayed emptying of the stomach following a meal, sometimes associated with accumulation of nondigestible fibers and debris in the stomach.

neurodegenerative disease Slowly progressive disease in the central or peripheral nervous system resulting from neuronal damage and/or neuronal death.

Diabetic neuropathy encompasses a variety of nervous system impairments that occur in the setting of diabetes. Neuropathy can develop early in both human and animal models of this disease. The sensory nervous system is most often affected in both type I and type II diabetes and much of the research into the mechanisms underlying diabetic neuropathy has focused on sensory impairment. There is evidence that similar physiologic problems may underlie autonomic neuropathy. Several metabolic pathways have been implicated in the pathogenesis of diabetic neuropathy, some of which likely interact via common signal molecules. Hyperglycemia is a central and critical component of the pathogenesis of sensory and autonomic neuropathy. The intensity and extent of the functional and anatomical abnormalities of diabetic neuropathy parallel the degree and duration of hyperglycemia. Intensive treatment with insulin in the Diabetes Control and Complications Trial decreased clinical neuropathy by 60%, whereas trials of other agents, such as aldose reductase, were far less effective.

INTRODUCTION

Insulin treatment has a beneficial effect on both sensory and autonomic neuropathy. The most common manifestations of autonomic neuropathy are hemodynamic effects, such as impaired control of blood pressure and heart rate, and gastrointestinal symptoms. Autonomic

 Copyright 2004, Elsevier (USA). All rights reserved.

a gastrostomy tube for gastric drainage and a jejunostomy tube for feeding. Innovative treatments for diabetic gastroparesis include injection of botulinum toxin into the prepyloric antrum, gastric electrical stimulation, and pancreatic or islet cell transplantation. Initial reports suggest that botulinum toxin injection and gastric electrical stimulation may reduce symptoms, but gastric emptying may not be improved. Reversal of diabetes with successful pancreatic or islet cell transplantation can cure this condition.

See Also the Following Articles

Diabetes Mellitus • Diabetic Neuropathies • Emesis • Gastric Emptying • Nausea

Further Reading

Camilleri, M. (2002). Advances in diabetic gastroparesis. *Rev. Gastroenterol. Disord.* **2**, 47–56.

Malagelada, J. R. (1995). Diabetic gastroparesis. *Semin. Gastrointest. Dis.* **6**, 181–186.

Diabetic Neuropathies

KAREN E. HALL

Ann Arbor Veterans Affairs Healthcare System and University of Michigan

apoptosis A mechanism of cell death that is controlled by specific metabolic cell death pathways and that results in fragmentation and condensation of the cell nucleus, without significant inflammation in the surrounding tissue.

autonomic Pertaining to the internal systems in the body (cardiovascular, pulmonary, gastrointestinal) that receive innervation that is not under voluntary control.

diabetic neuropathy Short-term and long-term nerve dysfunction in diabetes characterized by generalized slowing of conduction and increased threshold for excitation in peripheral and autonomic nerves.

gastrointestinal motility The process of contraction of the gastrointestinal wall that results in movement of food and secretions along the length of the gastrointestinal tract.

gastroparesis Delayed emptying of the stomach following a meal, sometimes associated with accumulation of nondigestible fibers and debris in the stomach.

neurodegenerative disease Slowly progressive disease in the central or peripheral nervous system resulting from neuronal damage and/or neuronal death.

Diabetic neuropathy encompasses a variety of nervous system impairments that occur in the setting of diabetes. Neuropathy can develop early in both human and animal models of this disease. The sensory nervous system is most often affected in both type I and type II diabetes and much of the research into the mechanisms underlying diabetic neuropathy has focused on sensory impairment. There is evidence that similar physiologic problems may underlie autonomic neuropathy. Several metabolic pathways have been implicated in the pathogenesis of diabetic neuropathy, some of which likely interact via common signal molecules. Hyperglycemia is a central and critical component of the pathogenesis of sensory and autonomic neuropathy. The intensity and extent of the functional and anatomical abnormalities of diabetic neuropathy parallel the degree and duration of hyperglycemia. Intensive treatment with insulin in the Diabetes Control and Complications Trial decreased clinical neuropathy by 60%, whereas trials of other agents, such as aldose reductase, were far less effective.

INTRODUCTION

Insulin treatment has a beneficial effect on both sensory and autonomic neuropathy. The most common manifestations of autonomic neuropathy are hemodynamic effects, such as impaired control of blood pressure and heart rate, and gastrointestinal symptoms. Autonomic

 Copyright 2004, Elsevier (USA). All rights reserved.

neuropathy is a significant underlying cause of gastroparesis and colonic dysmotility in diabetic patients. Gastroparesis in diabetes has been demonstrated to be due to a combination of reversible changes related to hyperglycemia and metabolic products of glucose and to irreversible neuronal damage and loss. Noninvasive studies of gastric electromyographic function in normal human subjects indicate that acute hyperglycemia can precipitate significant dysmotility in the stomach, accompanied by subjective feelings of nausea and distension. There is a synergy between hyperglycemia and gastrointestinal function as gastroparesis can, in turn, worsen glycemic control. This is particularly problematic in insulin-dependent individuals, as the time course of their insulin administration may be poorly correlated with delivery of nutrients to the small intestine. For these reasons, aggressive control of glucose is beneficial in treating diabetic gastroparesis. Colonic dysmotility is also common in diabetes. Both constipation and diarrhea can occur and are often refractory to pro-kinetic agents, due to the combination of neuronal loss and myopathic changes in the gastrointestinal muscle. In addition to glycemic control, a bowel regimen including stimulant laxatives and enemas may be required for refractory constipation. Diabetic diarrhea is often socially debilitating, but can be controlled with the judicious use of anti-diarrheal medications. Diabetes also impairs pancreatic function. In addition to the primary effects of diabetes on pancreatic beta-cell survival, pancreatic exocrine secretion appears to be diminished in patients with documented autonomic neuropathy involving the parasympathetic system.

DIABETIC NEUROPATHY IS A NEURODEGENERATIVE PROCESS

Studies of neuron survival, function, and gene expression suggest that diabetes more closely mirrors degenerative neuronal disease, rather than an acute injury model, such as axotomy. Slow nerve conduction in sensory and motor neurons is observed within 2–4 months following the onset of diabetes in the streptozotocin-induced and BioBred/Wistar spontaneous rat models of diabetes. At this early stage, sensory neurons do not demonstrate detectable alterations in their size, number, or expression of mRNA for a variety of signal molecules involved in neuro-transmission and survival [CGRP (α and β), substance P, vasoactive intestinal peptide, galanin, somatostatin, heat shock protein 27, c-jun, SNAP 25, p75, TrkA, TrkB, TrkC]. This observation is in sharp contrast to the rapid up-regulation of similar signal pathways in sensory and motor neurons subjected to axotomy or other injuries. One issue that needs to be addressed in future studies is the effect of age on the progression and severity of neuropathy. Most of the data obtained from animal models of diabetes (whether chemical or spontaneous) have utilized youthful animals. Retrograde labeling of extrinsic and intrinsic innervation of the gastrointestinal tract in the Fisher 344 rat demonstrates significant neuronal loss, which increases in severity from proximal to distal gut and progresses linearly with increasing age. There appears to be a trophic effect of vagal afferent innervation, although the mechanism by which this occurs remains to be determined.

METABOLIC DERANGEMENTS

Significant alterations in metabolic pathways occur in both sensory and autonomic diabetic neuropathies. Abnormalities that occur early in the course of diabetes include increased production of sorbital by aldose reductase, increased formation of glycosylation end products, decreased production of myoinositol, and impaired regulation of cellular calcium. Increased calcium influx through voltage-activated membrane calcium channels in diabetic neurons results in excessive cytosolic calcium responses to a variety of stimuli. Diabetic neurons release more calcium from cytosolic pools and demonstrate impaired ability to re-sequester calcium, compared to nondiabetic controls. Similar abnormalities have been observed in neurons from normal animals subjected to hyperglycemia, suggesting that impaired calcium regulation is an early and potentially reversible metabolic derangement in diabetes. A modest improvement in peripheral nerve conduction has been demonstrated in human studies utilizing L-type calcium channel antagonists; however, the effect is likely limited by the fact that L-type currents account for only 20–30% of the total calcium current in sensory neurons. Blockade of the predominant N-type channel by ω-conotoxin is impractical, due to toxicity. Activation of protein kinase C (PKC) isoforms, particularly PKC α and PKC βII, has been implicated in the pathogenesis of diabetic sensory and autonomic neuropathy and retinopathy. Trials of PKC inhibitors, such as WAY151003 and cremophor EL, have demonstrated improved nerve conduction velocity and endoneurial blood flow in animal models.

With increasing duration of diabetes, additional structural and metabolic effects are observed in both human and animal models of diabetic neuropathy. Loss of specific neuronal populations does occur, although the effect is relatively modest. As outlined earlier, there may be a synergistic effect of aging on

neuronal survival. Significant anatomic changes in the synaptic structures and distortion of the nodes of Ranvier have been implicated in the progression of slowing of nerve conduction observed in long-term diabetes. Expression of a variety of neuropeptides and neuronal growth factors is decreased. Exogenous delivery of growth factors such as insulin-like growth factor I, neurotrophin-3, and nerve growth factor (NGF) has been shown to improve nerve conduction and neuronal survival in animal models. Despite the promise of these studies, a recent randomized controlled trial of recombinant human NGF was not effective in restoring nerve function in patients with diabetes. The lack of efficacy may have been due to the mechanism of delivery of NGF, changes in the formulation between trials, or a requirement for additional factors. Cross-talk or synergistic actions by multiple growth and survival factors have been observed in models of neuronal survival, leading to the suggestion that simultaneous delivery of several factors may be more effective. In addition to improving nerve conduction by indirect actions on cell swelling and calcium regulation, insulin may act as a trophic factor in diabetes. Studies in which insulin was injected unilaterally near an affected nerve resulted in unilateral improvement in nerve conduction and small fiber sprouting.

NEURONAL DEATH

Many authors agree that some neuronal loss occurs during a prolonged period of diabetes; however, the mechanism remains controversial. Apoptotic cell death has been described in hyperglycemic *in vitro* culture models; however, this phenomenon may reflect effects of extreme hyperglycemia. Apoptosis has also been described in cultured neuronal cells exposed to diabetic serum. Although the noxious factor or factors involved in promoting cell death in this model have not been definitely identified, IgG class immunoglobulins have been implicated based on the results of staining and adsorption studies. *In vivo* apoptosis rates are considerably lower than those observed in culture models; however, careful counting of a large number of sensory neurons in several murine diabetic models indicates a small ($\sim$1.5%) but significant increase in apoptosis rates, when compared to nondiabetic controls ($<$0.3%). The neuronal loss in diabetes appears to occur slowly over a period of years. Thus, a small increase in the number of apoptotic neurons present at any one time could translate over time into a significant neuronal loss. These mechanisms have been demonstrated to occur in models of autonomic neuropathy, suggesting that neuronal loss may also contribute to impaired gastric and colonic function in diabetes.

TREATMENT

In addition to the use of conventional analgesics, painful sensory neuropathy is increasingly treated with tricyclic anti-depressant drugs and topical capsaicin. In a comparison study, amitriptyline was more effective at relieving pain than desipramine, but was also more likely to cause significant anti-cholinergic side effects. Topical capsaicin causes a prolonged depletion of Substance P from sensory nerves and is a useful adjunct to other analgesics. Acupuncture and trans-cutaneous electrical stimulation of cutaneous nerves have also been used with some effect in mild to moderate neuropathy. These interventions are useful in the treatment of sensory neuropathy, but have been less helpful in modulating the severity of autonomic neuropathy.

Increasing knowledge of the metabolic derangements occurring in diabetic neuropathy has led to trials of other agents. Control of hyperglycemia remains the only intervention to date that retards clinical progression of both sensory and autonomic neuropathies in human diabetes. Despite early promise in animal studies, the use of aldose reductase inhibitors such as sorbinil and tolrestat results in minimal improvement in sensory and autonomic nerve conduction velocity and symptoms. Pro-kinetic medications such as metoclopramide, domperidone, and cisapride, which treat gastrointestinal dysmotility, have all been demonstrated to improve diabetic gastroparesis, but have a minimal effect on colonic motility. The latter may reflect the pathophysiologic finding of more severe neuronal injury and impairment in the distal gut. Unfortunately, the availability of these agents is limited. Domperidone is not available in the United States and cisapride has been withdrawn due to problems with cardiac arrhythmias. Erythromycin acts on motilin receptors to initiate coordinated motility in the stomach and upper small bowel. Administered four times daily, it is an effective and inexpensive treatment for gastroparesis. Because they act on myenteric neurons, the utility of pro-kinetic drugs is ultimately limited by the function of the residual innervation present in the gut. Finally, metabolic derangements appear to precede structural changes in diabetic neuropathy. Therefore, it is likely that future trials designed to prevent or retard the progression of diabetic neuropathy will be directed toward patients presenting earlier in the course of the disease, in an effort to intervene prior to the development of irreversible structural damage.

See Also the Following Articles

Autonomic Innervation • Brain–Gut Axis • Diabetic Gastroparesis • Gastric Motility • Sensory Innervation

Further Reading

Biessels, G., and Gispen, W. H. (1996). The calcium hypothesis of brain aging and neurodegenerative disorders: Significance in diabetic neuropathy. *Life Sci.* **59**, 379–387.

The Diabetes Control and Complications Trial Research Group (1983). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986.

Diemel, L. T., Brewster, W. J., Fernyhough, P., and Tomlinson, D. R. (1994). Expression of neuropeptides in experimental diabetes: Effects of treatment with nerve growth factor or brain-derived neurotrophic factor. *Brain Res. Mol. Brain Res.* **21**, 171–175.

Erbas, T., Varoglu, E., Erbas, B., Tastekin, G., and Akalin, S. (1993). Comparison of metoclopramide and erythromycin in the treatment of diabetic gastroparesis. *Diabetes Care.* **16**, 1511–1514.

Funakoshi, H., Frisen, J., Barbany, G., Timmusk, T., Zachrisson, O., Verge, V. M., and Persson, H. (1993). Differential expression of mRNAs for neurotrophins and their receptors after axotomy of the sciatic nerve. *J. Cell Biol.* **123**, 455–465.

Hall, K. E., Liu, J., Sima, A. A. F., and Wiley, J. W. (2001). Impaired inhibitory G protein function contributes to increased calcium currents in rats with diabetic neuropathy. *J. Neurophysiol.* **86**, 760–770.

Hall, K. E., Sheng, H. C., Srinivasan, S., Spitsbergen, J. M., Tuttle, J. B., Steers, W. D., and Wiley, J. W. (2001). Treatment of aged rat sensory neurons in short-term, serum-free culture with nerve growth factor reverses the effect of aging on neurite outgrowth, calcium currents, and neuronal survival. *Brain Res.* **888**, 128–137.

Hall, K. E., Sima, A. A. F., and Wiley, J. (1995). Voltage-dependent calcium currents are enhanced in dorsal root ganglion neurones from the Bio Bred/Wistar diabetic rat. *J. Physiol.* **486**, 313–322.

Hall, K. E., Sima, A. A. F., and Wiley, J. W. (1996). Opiate-mediated inhibition of calcium signaling is decreased in dorsal root ganglion neurons from the diabetic BB/W rat. *J. Clin. Invest.* **97**, 1165–1172.

Kostyuk, E., Svichar, N., Shishkin, V., and Kostyuk, P. (1999). Role of mitochondrial dysfunction in calcium signalling alterations in dorsal root ganglion neurons of mice with experimentally-induced diabetes. *Neuroscience* **90**, 535–541.

Phillips, R. J., and Powley, T. L. (2001). As the gut ages: Timetables for aging of innervation vary by organ in the Fischer 344 rat. *J. Comp. Neurol.* **434**, 358–377.

Pittenger, G. L., Liu, D., and Vinik, A. I. (1997). The apoptotic death of neuroblastoma cells caused by serum from patients with insulin-dependent diabetes and neuropathy may be Fas-mediated. *J. Neuroimmunol.* **76**, 153–160.

Russell, J. W., Sullivan, K. A., Windebank, A. J., Herrmann, D. N., and Feldman, E. L. (1999). Neurons undergo apoptosis in animal and cell culture models of diabetes. *Neurobiol. Dis.* **6**, 347–363.

Singhal, A., Cheng, C., Sun, H., and Zochodne, D. W. (1997). Near nerve local insulin prevents conduction slowing in experiment diabetes. *Brain Res.* **763**, 209–214.

Srinivasan, S., Stevens, M., Sheng, H., Hall, K. E., and Wiley, J. W. (1998). Serum from type 2 diabetics with neuropathy induces complement-independent, calcium-dependent apoptosis in cultured neuronal cells. *J. Clin. Invest.* **102**, 1454–1462.

Srinivasan, S., Stevens, M., and Wiley, J. W. (2000). Diabetic peripheral neuropathy: Evidence for apoptosis and associated mitochondrial dysfunction. *Diabetes* **49**, 1932–1938.

Zochodne, D. W., Verge, V. M. K., Cheng, C., Sun, H., and Johnston, J. (2001). Does diabetes target ganglion neurons? *Brain* **124**, 2319–2334.

Diaphragmatic Hernia

WILLIAM HARFORD
Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

Bochdalek hernia Diaphragmatic defect in the postero-lateral chest at the lumbocostal junctions.
mixed diaphragmatic hernia Combination of sliding and paraesophageal hernia.
Morgagni hernia Diaphragmatic defect in the anterior chest at the sternocostal junctions.
paraesophageal hernia Defect in which part or all of the stomach protrudes through the esophageal hiatus alongside the esophagus.
posttraumatic diaphragmatic hernia Rent in the diaphragm due to trauma.
sliding hiatal hernia Defect in which the gastroesophageal junction and some portion of the stomach are displaced above the diaphragm, but the axis of the stomach axis remains unchanged.

A diaphragmatic hernia is an abnormal protrusion of an organ or structure into an opening in the diaphragm. In some cases, the protrusion can spontaneously resolve and is termed reversible. In other cases, the protrusion cannot spontaneously resolve and is termed incarcerated. The organ in a diaphragmatic hernia is termed strangulated when the vascular supply becomes compromised, causing ischemia or necrosis.

PATHOGENESIS

The frequency of sliding hiatal hernias increases with age and may be due to age-related deterioration of the phrenoesophageal membrane, which anchors the gastroesophageal junction to the diaphragm. As the membrane becomes lax, the normal positive intra-abdominal pressure and normal upward traction of the esophagus upon the stomach during swallowing may cause protrusion of the stomach above the diaphragm.

In addition to the stomach, the omentum, colon, or spleen may also enter a paraesophageal hernia. The stomach often twists on its axis, causing gastric volvulus. The gastroesophageal junction remains in a normal position at the level of the diaphragm. Paraesophageal hernias may be due to a congenital defect in the diaphragmatic hiatus anterior to the esophagus.

The diaphragm is derived from the septum transversum (separating the peritoneal and pericardial spaces), the mesentery of the esophagus, the pleuroperitoneal membranes, and the muscle of the chest wall. Congenital diaphragmatic hernias, such as Morgagni and Bochdalek hernias, occur where fusion of these components fails (see Fig. 1).

Posttraumatic diaphragmatic hernias are caused by blunt trauma (such as motor vehicle collisions) in approximately 80% of cases and by penetrating trauma (such as stab wounds or gunshots) in the remainder. Blunt trauma may cause large rents, whereas penetrating injuries often cause only small lacerations. Herniation of abdominal contents into the chest is thus more common after blunt trauma than after penetrating trauma. Approximately 70% of diaphragmatic injuries from blunt trauma occur on the left side. The right hemidiaphragm is somewhat protected by the liver. A small diaphragmatic defect may not lead to immediate herniation, but with time, normal negative intrathoracic pressure may cause gradual enlargement of the defect and protrusion of abdominal contents. A variety of viscera may be found in posttraumatic hernias, including stomach, omentum, colon, small bowel, spleen, and even kidney.

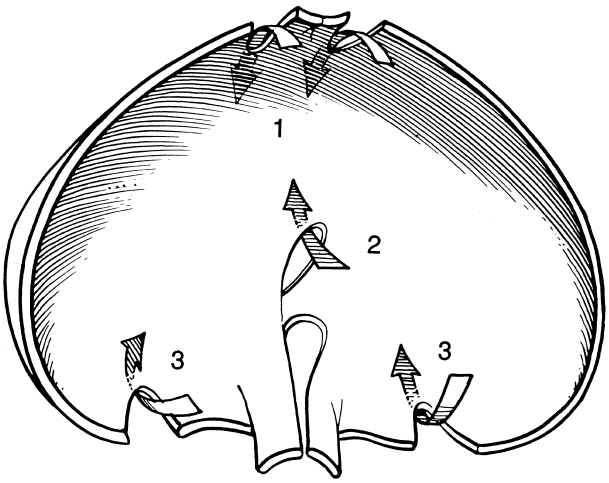


FIGURE 1 Congenital hernias of the diaphragm. Diagram of the diaphragm viewed from below with areas of potential herniation indicated. (1) Sternocostal foramina of Morgagni. (2) Esophageal hiatus. (3) Lumbocostal foramina of Bochdalek. Reprinted from Harford and Jeyarajah (2002), with permission.

 Copyright 2004, Elsevier (USA). All rights reserved.

INCIDENCE AND PREVALENCE

Hiatal hernias are common among adults undergoing upper gastrointestinal barium X rays. Sliding hiatal hernias make up approximately 90 to 95% of diaphragmatic hernias found by X ray, whereas almost all of the rest are paraesophageal or mixed. Most sliding hiatal hernias thus discovered are small and of no clinical significance. Patients with symptomatic paraesophageal or mixed hernias are most often middle-aged to elderly.

Clinically significant congenital hernias are rare, occurring in approximately 0.1 to 0.5 per 1000 births. Bochdalek hernias most often present in newborns. Only a few first present in adulthood. Approximately 80% of Bochdalek hernias occur on the left. Morgagni hernias are thought to be congenital, but usually present in adults. They make up approximately 2–3% of surgically treated diaphragmatic hernias. In contrast to Bochdalek hernias, Morgagni hernias occur on the right side in 80 to 90% of cases. Diaphragmatic injury occurs in approximately 5% of patients who suffer multiple traumatic injuries, but the exact incidence of posttraumatic diaphragmatic hernia is unknown.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most patients with small sliding hiatal hernias have no symptoms. However, sliding hiatal hernias, particularly large hernias, contribute to the pathogenesis of gastroesophageal reflux. In addition to symptoms related to gastroesophageal reflux, large sliding hiatal hernias may cause dysphagia or discomfort in the chest or upper abdomen. Cameron ulcers or erosions may develop in sliding hiatal hernias. These mucosal lesions usually occur on the lesser curve of the stomach at the level of the diaphragm, suggesting that mechanical factors may play a role in their pathogenesis, although ischemia and peptic injury may also contribute. Up to 5% of patients with large hernias undergoing endoscopy may be found to have Cameron ulcers. Acute or chronic upper gastrointestinal bleeding may occur with Cameron ulcers.

Hiatal hernias may be seen as a soft-tissue density posterior to the heart on chest X ray, but are usually diagnosed on upper gastrointestinal barium X-ray studies.

In contrast to patients with small sliding hiatal hernias, patients with paraesophageal and mixed hiatal hernias are rarely completely asymptomatic. Approximately half of such patients have symptoms of gastroesophageal reflux. Dysphagia, chest pain, vague postprandial discomfort, shortness of breath, and chronic gastrointestinal blood loss are also reported. Paraesophageal hernias may be complicated by gastric volvulus.

On chest X ray, paraesophageal or mixed hiatal hernias may be seen as an abnormal soft-tissue density (often with a gas bubble) in the posterior chest. The best diagnostic study is upper gastrointestinal barium X ray. On upper endoscopy, paraesophageal hernias are usually obvious, but the paraesophageal component of a large mixed hernia may be missed, and the anatomy may be confusing, particularly if the hernia is associated with gastric volvulus.

Congenital diaphragmatic hernias are associated with a wide variety of clinical presentations, from death in the neonatal period to an asymptomatic incidental finding in adults. Bochdalek hernia may cause a syndrome of severe respiratory distress in newborns. Pulmonary hypoplasia occurs on the side of the hernia, but some degree of hypoplasia may also occur in the contralateral lung. Pulmonary hypertension is common. Serious chromosomal anomalies (most often trisomy 13, 18, and 21) are found in 30 to 40% of newborns with Bochdalek hernias. Respiratory failure and other congenital anomalies are the major causes of mortality in such infants. In older children and adults, a Bochdalek hernia may present with symptoms due to herniation of the stomach, omentum, colon, or spleen. Gastric volvulus may occur, and approximately half of adult patients present with acute incarceration. Patients may also be asymptomatic or have chronic nonspecific intermittent symptoms, including chest discomfort, shortness of breath, dysphagia, nausea, vomiting, or constipation.

The differential diagnosis of Bochdalek hernia includes mediastinal or pulmonary cyst or tumor as well as pleural effusion or empyema. Prenatal diagnosis may be made at sonography by visualizing stomach or loops of bowel in the chest. In older children and adults, the diagnosis may be suspected when a soft tissue mass in the posterior mediastinum is discovered on chest X ray. The diagnosis is usually confirmed by barium upper gastrointestinal X ray or by computerized tomography with oral contrast.

In contrast to Bochdalek hernias, Morgagni hernias are most likely to present in adult life. They may contain omentum, stomach, colon, or liver. As with Bochdalek hernias, patients may also be asymptomatic or have chronic nonspecific intermittent symptoms. They may also present with acute gastric, omental, or intestinal incarceration with obstruction and/or ischemia.

The differential diagnosis of Morgagni hernias is similar to that of Bochdalek hernias and the diagnosis of Morgagni hernias is often made in the same manner.

In contrast to Bochdalek hernias, Morgagni hernias are found in the anterior mediastinum, usually on the right. It is important to confirm the contents of the hernia with barium X rays or computerized tomography with oral contrast.

Rupture of the diaphragm is often masked by other serious injuries after serious trauma. It may go undetected even at exploratory laparotomy. Posttraumatic diaphragmatic hernias usually cause respiratory and/or abdominal symptoms. Such symptoms occurring several days to weeks after injury should suggest the possibility of a posttraumatic hernia. Positive-pressure ventilatory support after trauma may prevent herniation through a diaphragmatic tear, but upon attempted ventilator weaning, herniation may then occur. Symptoms may also present long after injury, up to 10 years or more later, obscuring the relationship between the acute illness and remote trauma.

Diaphragmatic injury should be considered in any case of serious trauma involving the area between the fourth intercostal space and the umbilicus. The chest X ray aids the diagnosis in only half of the cases. Rapid helical computerized tomography with sagittal reconstruction is very useful in diagnosis.

TREATMENT

Small sliding hiatal hernias do not require treatment. Surgery should be considered in cases of symptomatic giant sliding hiatal hernias, paraesophageal hernias, and mixed diaphragmatic hernias. Even among patients with asymptomatic paraesophageal hernias, approximately 30% will develop acute complications if left untreated, so many surgeons recommend repair when such hernias are discovered.

Treatment of all diaphragmatic hernias includes repair of the diaphragmatic defect either by suture approximation or by insertion of mesh. When the stomach has herniated, it should be fixed in the abdomen. Since there is a high prevalence of gastroesophageal reflux and esophageal motility disorders among patients with paraesophageal hernia, a fundoplication is usually included in the repair. Sliding hiatal or paraesophageal hernias may be associated with a shortened esophagus, making it difficult to fix the gastroesophageal junction below the diaphragm. In such cases an extra length of neo-esophagus from the proximal stomach is created by the surgeon (Colles-Nissen procedure). Both open and laparoscopic techniques are used for diaphragmatic hernia repair. Laparoscopic repair is associated with less blood loss, fewer overall complications, and shorter hospital stay than open repair. Return to normal activities is faster. Potential surgical complications include esophageal and gastric perforation, pneumothorax, and liver laceration. Long-term results are probably equal with either approach. Long-term complications may include dysphagia or the gas-bloat syndrome. Gastroesophageal reflux may occur if the fundoplication breaks down or migrates into the chest. Recurrence rates are approximately 10%.

Cameron ulcers or erosions are usually treated with anti-secretory medication, just as other gastric ulcers. However, in approximately one-third of patients these lesions may persist or recur despite medication, in which cases surgical repair of the associated hernia may be needed.

Infants with Bochdalek hernias usually require ventilatory support, but barotrauma may occur with standard ventilators. The advent of new techniques for ventilatory support such as high-frequency oscillation and extracorporeal membrane oxygenation has improved the prognosis of these infants, allowing stabilization before diaphragmatic repair. Most infants survive repair of the diaphragmatic hernia, but many have long-term neurological and musculoskeletal problems. Up to 50% develop gastroesophageal reflux.

A variety of techniques have been used for repair of congenital hernias, including open chest and/or abdominal approaches, as well as thoracoscopic and/or laparoscopic techniques. Acute diaphragmatic ruptures may be approached from the abdomen or through the chest. Since chronic posttraumatic diaphragmatic hernias may be associated with extensive adhesions and lack of a peritoneal hernia sac, they are best repaired through the chest or by a combined thoracoscopic/abdominal approach.

See Also the Following Articles

Gastric Volvulus • Hernias • Hiatal Hernia • Laparoscopy • Volvulus

Further Reading

Harford, W. V., and Jeyarajah, R. (2002). Abdominal hernias and their complications, including gastric volvulus. *In* "Gastrointestinal Disease, Pathophysiology, Diagnosis, and Management" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 7th Ed., Chap. 20, pp. 369–385. W. B. Saunders Co., Philadelphia, PA.

Langer, J. C. (1998). Congenital diaphragmatic hernia. *Chest Surg. Clin. North Am.* **8**, 295–314.

Naunheim, K. S. (1998). Adult presentation of unusual diaphragmatic hernias. *Chest Surg. Clin. North Am.* **8**, 359–369.

Nyhus, L. M., and Condon, R. E. (1989). *Hernia.* 3rd Ed. J. B. Lippincott, Philadelphia, PA.

Schumpelick, V., Steinau, G., Schluper, I. *et al.* (2000). Surgical embryology and anatomy of the diaphragm with surgical applications. *Surg. Clin. North Am.* **80**, 213–239.

Diarrhea

MIHIR PATEL AND JOSEPH H. SELLIN
University of Texas Medical School, Houston

acute diarrhea Abnormal stool production that lasts for less than a month.
chronic diarrhea Abnormal stool production that lasts for more than a month.
inflammatory diarrhea Abnormal stool production associated with the presence of white cells or blood.
osmotic diarrhea Abnormal stool production associated with an osmotic gap of more than 100 mOsm/kg.
secretory diarrhea Abnormal stool production associated with an osmotic gap of less than 50 mOsm/kg.
steatorrhea Presence of more than 7 g of fecal fat excretion on a diet containing 100 g fat/day over a period of 48–72 hours. Steatorrhea can be broadly divided into three categories: intraluminal maldigestion, mucosal malabsorption, and postmucosal malabsorption related to obstruction of the lymphatics. Maldigestion is defective hydrolysis of nutrients and malabsorption is defective mucosal absorption.
stool osmotic gap The difference between luminal osmolality (equal to body fluid osmolality, approximately 290mOsm/kg) and luminal content osmolality, calculated by electrolyte measurements.

Diarrhea, derived from the Greek "to flow through," is a common health problem. More than 450,000 hospital admissions each year in the United States (1% of adult hospitalizations) are due to diarrheal illnesses. Based on a commonly used definition of chronic diarrhea (liquid stools for more than 1 month), a reasonable approximation is that chronic diarrhea affects approximately 5% of the population throughout the world.

INTRODUCTION

Diarrhea is a symptom, not a disease. It represents the end product of multiple gastrointestinal dysfunctions. The definition of diarrhea has traditionally been based on the volume, frequency, and/or consistency of stools. Patients often consider increased fluidity of stool as diarrhea. However, because it is difficult to measure stool consistency, most investigators have used stool frequency or stool weight to better define this condition. In Western countries, three or more bowel movements per day are considered to be abnormal. For the United States population, fecal weight of more than 200 g/day is also considered to be consistent with diarrhea. However, diarrhea should not be defined using only one criterion. Some individuals may have increased fecal weight but have normal stool consistency and no complaint of diarrhea. For example, Asian individuals consuming a high-fiber diet may well have a stool weight greater than 200 g and still be considered "normal." Other individuals have normal fecal weight, yet complain of diarrhea because their stools are loose and watery. Thus, a good working definition is three or more loose or watery stools per day or a definite decrease in consistency and increase in frequency of stools based on an individual baseline.

CLASSIFICATIONS

A classification scheme provides a framework for understanding the pathophysiology and treatment of diarrhea:

- Acute vs. chronic
- Large volume vs. small volume
- Motility vs. epithelial function
- Secretory vs. osmotic
- Watery vs. fatty vs. inflammatory

These classifications are not mutually exclusive. A diarrhea may be acute, inflammatory, and secretory in nature, or alternatively, chronic, osmotic, and small volume. Clinical context provides a focus for sifting through multiple potential etiologies. By considering these classifications simultaneously, the clinician can narrow the range of diagnostic possibilities.

PATHOPHYSIOLOGY AND DIAGNOSIS

Diarrhea, a complex clinical scenario, frequently represents a protective response to a variety of intestinal insults and assaults. It can also be a manifestation of an underlying systemic disease process. Normally 8 to 10 liters of fluid per day enters the small intestine. Of this load, approximately 2 liters is derived from the diet and the remainder comes from the secretions of the salivary glands, stomach, pancreas, liver, and the intestine. The small intestine normally absorbs all but 1.5 liters of this

 Copyright 2004, Elsevier (USA). All rights reserved.

fluid, which eventually enters colon. The colon can absorb approximately 90–95% of the remaining fluid, leaving 100 to 150 ml/day in the stool. The small intestinal fluid absorption is generally linked to nutrient absorption, whereas the colon primarily transports electrolytes to conserve water and regulate stool consistency. Despite a reserve capacity for absorption, an imbalance between secretion and absorption can readily lead to diarrhea.

Acute vs. Chronic Diarrhea

Acute Diarrhea

Diarrheal illness lasting for less than 4 weeks can be defined as an acute diarrhea. It represents the second most common cause of death worldwide and is the leading cause of childhood death. Etiological factors can vary. The most helpful clue is the clinical setting. Infectious causes are more common in developing countries. Contaminated water systems, extreme poverty, rapid urbanization, crowded substandard housing with inadequate sewage disposal, and poor education are responsible for the frequent occurrence of diarrhea in developing countries. In the United States, most acute diarrheas last only a few days and resolve without obtaining a specific diagnosis. They are usually labeled "viral gastroenteritis." Foreign travel was once thought to be a prerequisite for "exotic" infectious diarrheal illness. However, amebiasis, giardiasis, and other infectious agents may underlie diarrhea in individuals who have never visited any foreign countries. Acute diarrhea can also be seen in specific settings that lead to consideration of discrete diagnoses. Acute diarrhea occurring in hospitalized patients, for example, may be due to tube feedings, medications, ischemia, or antibiotic treatment (*Clostridium difficile* diarrhea).

Patients with acute onset of diarrhea should be questioned for passage of blood or mucus in the stool, duration and frequency of diarrhea, and symptoms of dehydration. Diarrheal illness lasting for more than 48 hours with fever and signs of dehydration or peritonitis needs urgent medical attention. Older and immunocompromised patients need close observation.

The diagnostic approach in acute diarrhea includes fecal leukocytes, occult blood, and stool cultures. Endoscopy is not commonly needed in the initial diagnosis of acute diarrhea. It may, however, be helpful in distinguishing inflammatory bowel disease from infectious diarrhea, in diagnosing *C. difficile* colitis, and in patients in whom ischemic colitis is suspected.

Management of patients with acute diarrhea depends on the causative agent. Signs and symptoms of dehydration should be treated immediately. Most of the cases of infectious diarrhea are self-limited. Empiric antibiotic therapy has not been shown to alter the course of illness. Patients should be carefully questioned about risk factors for *C. difficile*, such as recent hospitalization or antibiotic use. Antimotility agents should be used cautiously in acute infectious diarrhea. Specific antimicrobial therapy is recommended if an organism is isolated (see Fig. 1).

Chronic Diarrhea

Chronic diarrhea needs careful evaluation; it may have a complex pathogenesis and can be a frustrating problem, not only for patients but also for the physicians treating them. Many patients cannot maintain a good quality of life, and in some cases, chronic diarrhea jeopardizes their livelihood. Patients may feel socially isolated and view their lives as nothing more than being tethered to the toilet.

Initial evaluation should be focused on separating functional from organic causes. A detailed and thorough medical history and clinical assessment is extremely helpful. Careful questioning about prescription and over-the-counter medications may give useful hints. A patient's usual intake of fruits and of beverages containing high concentrations of sugar or caffeine should be estimated. Functional causes are more likely in the absence of significant weight loss (<5 kg), longer duration of symptoms (>1 year), and in the absence of nocturnal diarrhea. Once a functional etiology is ruled out, attempts should be made to determine specific etiological factors. This is where the classification framework becomes paramount in steering through diagnostic and therapeutic options. This classification allows the physician to isolate key characteristics of the patient's history. It is extremely important to recognize fecal incontinence. Many patients do not report this symptom voluntarily, but complain of diarrhea instead. If incontinence is a frequent problem, then diagnostic and therapeutic efforts should be directed toward incontinence rather than diarrhea (see Fig. 2).

Large-Volume vs. Small-Volume Diarrhea

The rectosigmoid area acts a storage reservoir. Clinically, intact "reservoirs" are critically important factors in diarrhea. When that storage capacity is compromised by inflammatory or motility disorders, frequent small-volume bowel movements occur. Painful defecation associated with urgency and small volume is highly suggestive of left colon or rectal involvement. Fewer and larger bowel movements are usually associated with right colon or small intestine disease processes.

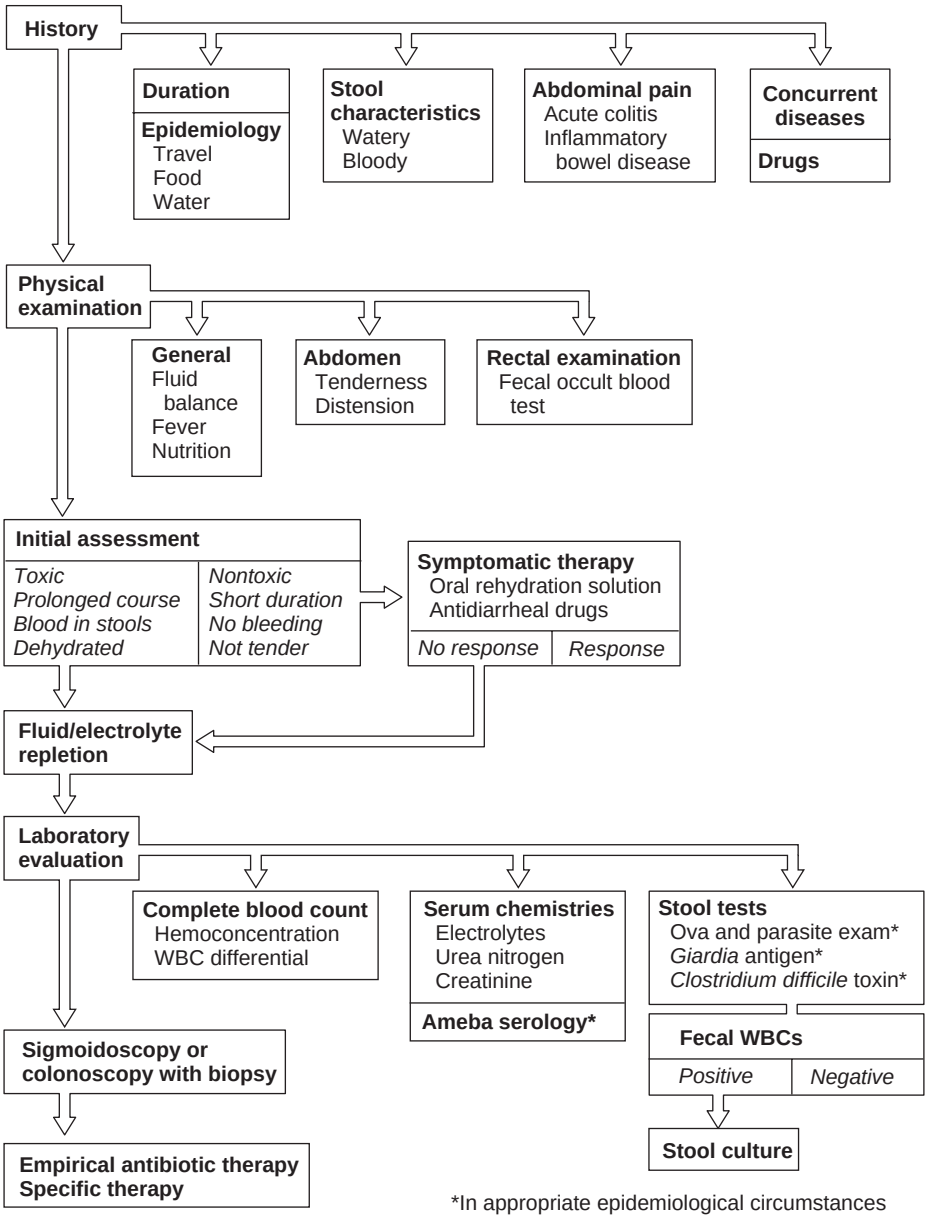


FIGURE 1 Flowchart for the evaluation of patients with acute diarrhea. Reproduced from Sellin (2002), with permission.

Motility vs. Epithelial Function

Most patients suffering from diarrhea, especially those in the Western world, would assume their symptoms are due to disordered motility, with luminal contents moving too quickly through the length of the gastrointestinal (GI) tract. However, slow transit, such as seen in scleroderma, can also lead to the diarrheal state, in part due to associated bacterial overgrowth causing epithelial disruption and altered electrolyte transport. Dysregulation of the enteric nervous system (due to surgical vagotomy, sympathectomy, or neuropathy secondary to systemic diseases) and the effects of certain mediators like vasoactive intestinal peptide (VIP) and nitric oxide may affect smooth muscle function in the gut. Dysregulation may also alter the dynamics of intestinal fluid absorption and epithelial functions. Surprisingly, there is little basic research directed to motility disorders in diarrhea. Instead, considerable emphasis has been placed on how the epithelial cells

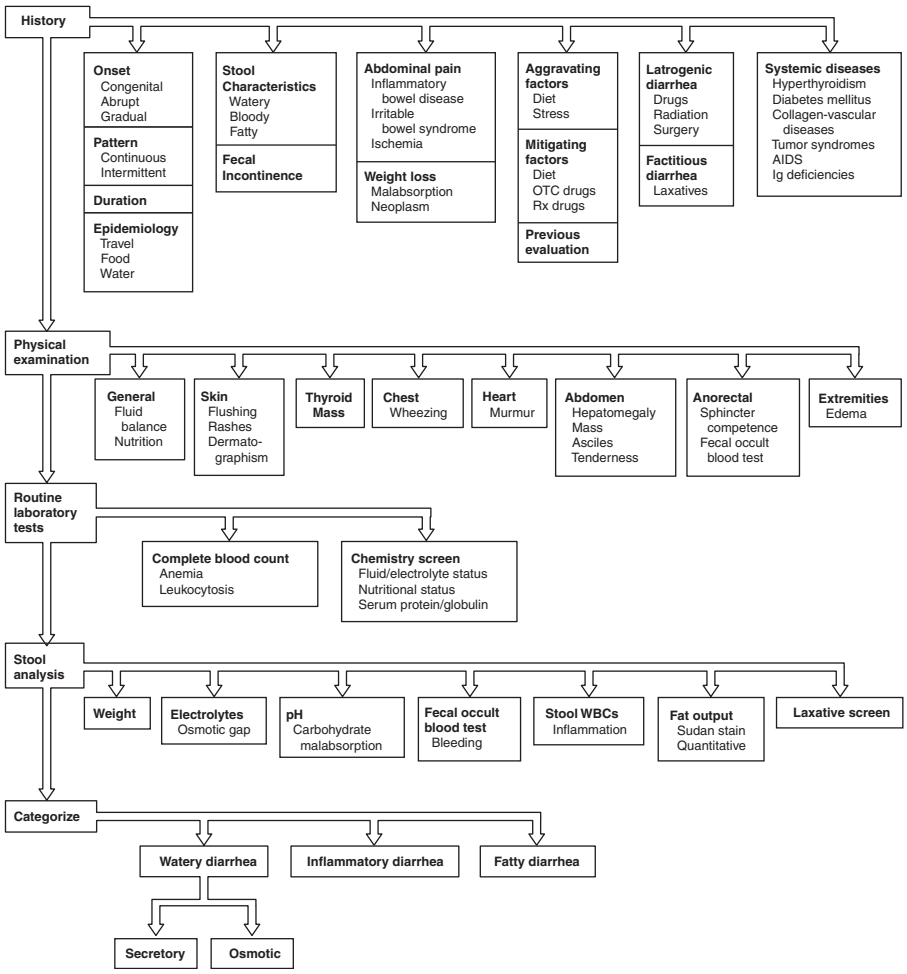


FIGURE 2 Flowchart for the initial evaluation of patients with chronic diarrhea. Reproduced from Sellin (2002), with permission.

lining the gut transport ions and nutrients and how they may be disrupted during diarrheal states. This focus on the epithelial functions has led to the differentiation of osmotic vs. secretory diarrhea.

Secretory vs. Osmotic Diarrhea

Water movement in the gut generally follows passively the movement of electrolytes and nutrients. The epithelial cells of the gastrointestinal tract have an elaborate system of pumps, carriers, and channels that provides orderly movement of ions and solutes through the cell (transcellular) and around the cell (paracellular). Sodium absorptive pathways provide the major driving forces for fluid absorption, whereas chloride movement into the gut drives secretion. In osmotic diarrhea, poorly absorbed solutes remaining in the gut lumen pull fluid from the submucosal space around the cells (paracellular). In secretory diarrhea, the transport machinery of the epithelial cells is taken over and either absorption is inhibited and/or secretion is stimulated.

Common examples of osmotic diarrhea are lactose malabsorption or diarrhea due to ingestion of magnesium-containing laxatives or unabsorbable solutes such as polyethylene glycol (PEG), used in colonic lavages and preparations for colonoscopies. Patients with lactose intolerance are lactase deficient. Lactose is usually metabolized to glucose and galactose, which are absorbable. In the setting of lactose intolerance, the unabsorbable lactose is not metabolized, leading to a failure to reabsorb fluid. Similarly, with magnesium or PEG, excess fluid entry into intestinal lumen leads to osmotic diarrhea. Other examples of osmotic diarrhea are mentioned in Table I.

The essential characteristic of osmotic diarrhea is that it disappears or improves with withdrawal of offending substances. Electrolyte absorption is unaffected by osmotically active substances and stool water has

TABLE I Major Causes of Diarrhea

Watery diarrhea			
Secretory	**Osmotic**	**Steatorrhea**	**Inflammatory diarrhea**
Osmotic laxative abuse	Nonosmotic laxative abuse	Malabsorption—mucosal diseases	Inflammatory bowel disease
	Congenital chloridorrhea	Short bowel syndrome	Infections
Magnesium-containing antacids	Bile acid malabsorption	Postresection diarrhea	Ischemic colitis
Carbohydrate malabsorption	Inflammatory bowel disease	Bacterial overgrowth	Radiation colitis
Short bowel syndrome	Collagenous colitis	Mesenteric ischemia	Neoplasia
	Vasculitis	Maldigestion	Microscopic colitis
	Postvagotomy diarrhea	Pancreatic insufficiency	Hypersensitivity
	Diabetic diarrhea	Emulsification problems	
	Hyperthyroidism	Problems with micelle formation	
	Addison's disease	Pancreas—cibal asynchrony	
	Bacterial toxins		
	Drugs		
	Neoplasia		
	Neuroendocrine tumors		

relatively low concentration of sodium and potassium. This is the basis for the calculation of the "fecal osmotic gap," which is the difference between luminal osmolality (equal to body fluid osmolality, approximately 290 mOsm/kg) and the measured luminal content osmolality:

$$\text{Stool osmolality} = 2(\text{Na} + \text{K})$$

$$\text{Fecal osmotic gap} = 290 - 2(\text{Na} + \text{K}).$$

A large osmotic gap (>50 mOsm/kg) is suggestive of the presence of osmotic diarrhea. When the osmotic gap is small, electrolytes account for most of the luminal osmolality and a secretory diarrhea is said to be present (see Fig. 3).

A number of conditions can lead to secretory diarrhea. These include many small intestine mucosal diseases and infections. Diseases affecting the colon alone (such as ulcerative colitis or microscopic colitis) may also produce secretory diarrhea. Bacterial and viral infections account for the vast majority of secretory diarrheas. Infections with microorganisms may produce enterotoxins that interact with the receptors modulating intestinal epithelial transport, leading to anion secretion through chloride channels. Enterotoxins may also block specific absorptive pathways, i.e., Na^+/H^+ exchange, thus inhibiting fluid absorption (see Fig. 4). Although other secretory conditions such as endocrine tumors and conditions such as congenital chloridorrhea and congenital sodium diarrhea are often considered in the listing of possible diagnoses, they are extremely uncommon.

Watery vs. Fatty vs. Inflammatory Diarrhea

Characteristics of the stool may help in differentiating diarrheas resulting from variety of causes. Watery diarrhea without blood, pus, or fat implies the presence of an osmotically active substance or unabsorbed/secreted electrolytes causing retention of excess of water in the lumen. Examples of watery diarrhea are bile acid malabsorption, carbohydrate malabsorption, and functional watery diarrhea due to irritable bowel syndrome.

Fatty diarrhea suggests the presence of conditions reducing fat absorption or digestion in the small intestine. Fatty diarrhea can be broadly divided into three categories: intraluminal maldigestion, mucosal

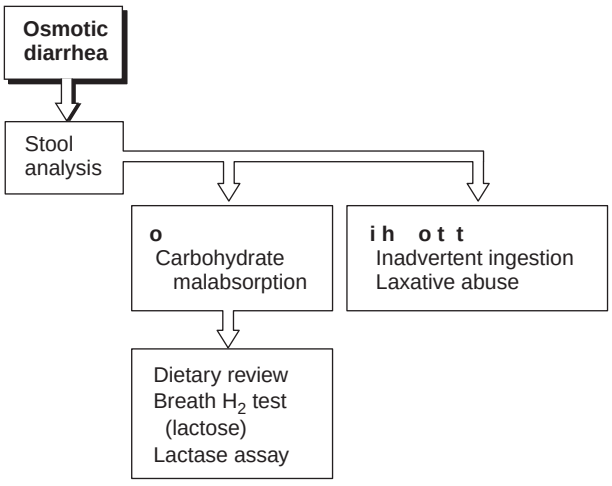


FIGURE 3 Flowchart for the evaluation of chronic osmotic diarrhea. Reproduced from Sellin (2002), with permission.

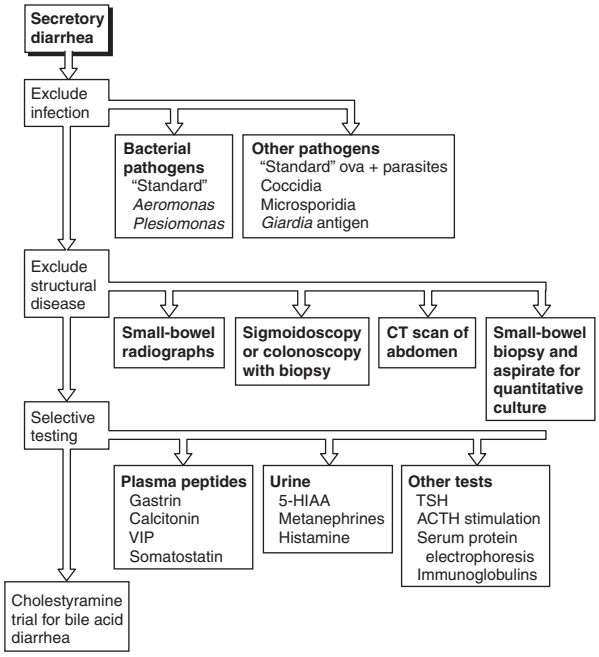


FIGURE 4 Flowchart for the evaluation of chronic secretory diarrhea. CT, Computed tomography; VIP, vasoactive intestinal peptide; 5-HIAA, 5-hydroxyindole acetic acid; TSH, thyroid-stimulating hormone; ACTH, adrenocorticotropic hormone. Reproduced from Sellin (2002), with permission.

malabsorption, and postmucosal malabsorption related to obstruction of lymphatics. Malabsorption can be global (carbohydrate, protein, and fat) or nutrient specific (vitamin B_{12}, lactose, or trehalose). It is important to identify differences between malabsorption and maldigestion. Malabsorption refers to impaired absorption of nutrients. It can result from congenital or acquired diseases affecting the intestinal epithelial surface. Examples are celiac disease, Whipple disease, abetalipoproteinemias, and lymphoma. Maldigestion refers to impaired hydrolysis and digestion of nutrients within the lumen. Examples of maldigestion include pancreatic insufficiency, hepatobiliary diseases with decreased micelle formation, or postsurgical complications leading to decreased emulsification and/or pancreas–cibal asynchrony (cibal, meaning "food intake," from the Latin, *cibus*).

Patients with chronic diarrhea with the presence of white cells or blood in the stool are classified as having inflammatory diarrhea. Basic pathologic characteristics are mucosal disruption and inflammation. Inflammatory bowel disease, radiation enteritis, enterocolitis, ischemia, and neoplasia are examples of inflammatory diarrhea. Common causes of infections causing inflammatory diarrhea are *C. difficile*, *Shigella*, cytomegalovirus, amebiasis, and tuberculosis. In these conditions, specific peptide mediators released by either bacteria or immune cells can alter epithelial function.

Most clinically significant diarrheas are complex and have several underlying pathophysiologic mechanisms rather than just one. An example of this complexity is cholera. Cholera is often cited as the paradigm of a secretory diarrhea. Cholera toxin, via the second messenger cyclic AMP, activates target apical chloride channels in epithelial cells, causing secretion. However, this same toxin also stimulates neural and endocrine cells that reinforce direct secretion by enterocytes. Studies have shown that cholera toxin also causes changes in intestinal motility and affects tight junctions, altering the permeability of intestinal mucosa. Thus, there may be added complexity in even the most straightforward example of secretory diarrhea. It is important to understand the multiple mechanisms that mediate a diarrheal illness. A single modulator may have multiple effects at the cellular and paracellular pathways and also at the muscular level. A full appreciation of the pathophysiology of diarrhea requires consideration of paracrine, immune, neural and endocrine modulators, a regulatory system that can be abbreviated by the acronym "PINES." Efforts to modulate this system with drugs may produce new classes of antidiarrheal agents in the future (see Fig. 5).

DIAGNOSTIC APPROACH AND MANAGEMENT

The initial evaluation of patients with diarrhea should include a detailed history and physical examination. The presence and absence of systemic diseases and

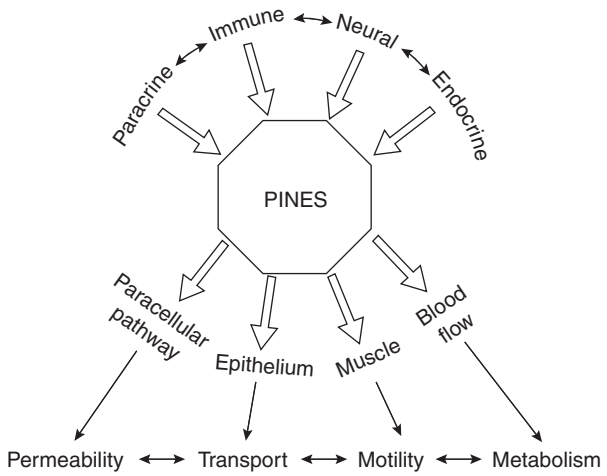


FIGURE 5 Paracrine/immune/neural/endocrine (PINES) regulatory system in the gut. Reproduced from Sellin (2002), with permission.

nutritional status should be carefully studied. Diagnostic evaluation of patients with chronic diarrhea should be directed at the most likely possibilities (Table I). Focused evaluation will help classifying diarrhea as being watery (osmotic vs. secretory), fatty, or inflammatory. Initial stool evaluation for the presence of occult blood (guaiac testing), white blood cells (Wright's stain or fecal lactoferrin assay), and fat (Sudan stain) can be helpful. Sudan staining of spot samples has a sensitivity of 78% and a specificity of 70% for detection of steatorrhea. Quantitative collection of stool over 48–72 hours provides a better idea of stool output and fat excretion but is not always essential. Fecal fat excretion of less than 7g/day on a 100-g/day fat diet is usually considered normal. Other than stool examination, routine laboratory tests (complete blood count and differential, liver function tests, prothrombin time, electrolytes, and thyroid-stimulating hormone) and sprue serology should be obtained.

As mentioned previously, analysis of stool electrolytes allows the physician to categorize diarrhea as osmotic or secretory. Although diagnosing an osmotic diarrhea produces a short list of suspects, finding a secretory diarrhea leaves many possibilities. Measurement of stool pH (less than 6) can be helpful in diagnosing conditions such as carbohydrate malabsorption. In patients with suspected laxative abuse, stool can be analyzed for laxatives by chemical or chromographic methods.

Structural diseases of the small intestine can be excluded by using noninvasive tests (small bowel radiography, breath tests) or, if needed, invasive tests such as small bowel biopsy. Some tests can be applied during initial screening and some can later separate out specific etiologies. One of the examples of a noninvasive test is the D-xylose test, which can differentiate malabsorption due to small intestinal mucosal disease from maldigestion due to pancreatic insufficiency by utilizing its natural physical properties. D-Xylose is a pentose monosaccharide that can be absorbed both by passive paracellular and transcellular pathways, reflecting the function of the proximal small intestine. Thus, this test is a measure of the permeability of the proximal small intestine. Absorption of D-xylose is normal in pancreatic insufficiency because pancreatic enzymes are not required for xylose absorption. Other noninvasive tests used for carbohydrate malabsorption rely on the fermentation of undigested carbohydrates by the intestinal bacteria, leading to increased hydrogen production.

Although jejunal cultures from the small bowel aspirate are considered a gold standard for diagnosing bacterial overgrowth, noninvasive breath tests (particularly the ^{14}C-labeled D-xylose test) are still recommended. In this test, D-xylose is catabolized by intestinal microflora. This action will lead to release of radioactive $^{14}CO_2$ that, after absorption, can be detected in breath samples.

Several invasive and noninvasive tests are available to evaluate pancreatic function and/or pancreatic structure as an indication of pancreatic insufficiency. Tests used to measure pancreatic function can be divided into two categories: direct and indirect. Direct tests are considered the gold standard. In the secretin stimulation test, pancreatic secretions are collected after stimulation by secretin, cholecystokinin, or the combination of both. Indirect tests measure pancreatic secretion after indirect stimulation, such as with a defined test meal. Both direct and indirect tests require duodenal intubation to collect pancreatic secretion samples and can thus be classified as "tube tests." Tubeless indirect tests include fecal fat assays, fecal chymotrypsin assays, fecal elastase measurements, and the pancreolauryl test. The choice of which pancreatic function test to use depends on the clinical question and availability of the test. It is important to understand that malabsorption does not occur until functional capacity is reduced to 5–10% of the normal. Thus, most of the tests may not show abnormal results unless moderate to severe pancreatic insufficiency is present.

Radiologic investigations such as computed tomography or endoscopic retrograde pancreatography may show changes in the duct and parenchyma at a later stage. A newer modality is endoscopic ultrasound, the role of which is still investigational but which has shown promising results in evaluating structural changes related to chronic pancreatitis. Most disorders that produce chronic diarrhea and gross histological changes in the colon can be diagnosed with sigmoidoscopic examination and biopsies, but some lesions exclusive to the right colon (e.g., infections in AIDS) may be missed without a complete colon examination. Testing for secretogogue-induced diarrhea should be selective. Rarely, measurement of plasma peptide concentration is required, based on the degree of suspicion and findings of radiological tests.

Specific treatment can be considered when a diagnosis is strongly suspected in the presence of supporting evidence. Examples include a patient with diarrhea due to limited resection of terminal ileum (<100 cm) in whom bile acid malabsorption is a likely cause and thus a trial of a bile acid-binding agent may be helpful. A patient with bacterial overgrowth may benefit from antibiotic therapy. A patient with suspected lactose intolerance may improve with avoidance of lactose-containing products.

Nonspecific symptomatic treatment may be considered in a patient whose diagnostic workup is pending or when the complexity of a problem is such that a definitive treatment is unavailable. A variety of medications are available. Different antidiarrheal agents may have different mechanisms of action but it is also important to understand specific side effects while using them. Loperamide oxide, bismuth subsalicylate, anticholinergics, and bulk-forming agents are relatively safe. Opiates have a potential for addiction. Bismuth subsalicylate in large doses, particularly in the presence of renal dysfunction, may precipitate salicylate toxicity. It also produces black stool that can easily be confused with melena. Anticholinergics should be used carefully in pediatric and elderly populations. The most common side effect of bulk-forming agents is bloating and flatulence.

Of all the therapeutic modalities available to treat diarrheal disease, none has been more important than the discovery and development of the oral rehydration solution (ORS), one of the major advances in public health in the twentieth century. ORS development is a paradigm of translational research, in which basic observations on the mechanisms of toxin-induced diarrhea were utilized to design an improved therapy. The oral rehydration solutions take advantage of our understanding of the pathophysiology of secretory diarrheal diseases such as cholera. In cholera, there is increased chloride secretion and inhibition of some, but not all, sodium absorptive pathways. While electroneutral sodium absorption is blocked, absorption of Na coupled with nutrients is unaffected. Therefore, supplying glucose and other small nutrient molecules via an ORS into the gut lumen can stimulate sodium absorption. The initial goal of the ORS is to offset the high morbidity and mortality associated with dehydration in geographic areas where intravenous fluids are unavailable or too expensive. The main constituents of ORSs are water, salt, and sugar. The ratio of the solutes was initially fixed, but over the past decade many modifications have been made. Currently, these modifications have allowed ORS use as a maintenance therapy for chronic diarrhea such as high output ostomy diarrhea and diarrhea due to short bowel syndrome. In the United States, ORS is underused. It is a misconception that simply drinking fruit juices and isotonic fluids such as Gatorade can prevent dehydration. A significant number of hospital admissions for dehydration might be avoided by wider utilization of ORS, as is done in developing countries.

The goal of the physician in evaluating patients with diarrhea is to make a definitive diagnosis as quickly and inexpensively as possible. The approach may be different in adult patients and in children. The majority of the patients can be diagnosed and treated with the help of careful history and physical examination, coupled with certain focused laboratory examinations. A better understanding of the pathophysiology of diarrhea and ongoing efforts to improve its diagnosis and management will undoubtedly reduce the economic burden and human cost throughout the world.

See Also the Following Articles

Antibiotic-Associated Diarrhea • Anti-Diarrheal Drugs • Bacterial Toxins • Carbohydrate and Lactose Malabsorption • Diarrhea, Infectious • Diarrhea, Pediatric • Gastroenteritis • Malabsorption • Malnutrition • Pancreatic Function Tests • Traveler's Diarrhea

Further Reading

American Gastroenterological Association Guidelines. (1999). Evaluation and management of chronic diarrhea. *Gastroenterology* **116**, 1464.

Christopher, H., and Hammer, H. F. (2002). Maldigestion and malabsorption. *In* "Gastrointestinal and Liver Disease" (M. H. Sleisenger and J. S. Fordtran, eds.), 7th Ed., pp. 1751–1781. W. B. Saunders, Philadelphia.

Powell, D. (1999). Diarrhea. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., pp. 858–909. Lippincott Williams & Wilkins, Baltimore.

Schiller, L. R., and Sellin, J. H. (2002). Diarrhea. *In* "Gastrointestinal and Liver Disease" (M. H. Sleisenger and J. S. Fordtran, eds.), 7th Ed., pp. 131–154. W. B. Saunders, Philadelphia.

Sellin, J. H. (2002). Intestinal electrolyte absorption and secretion. *In* "Gastrointestinal and Liver Disease" (M. H. Sleisenger and J. S. Fordtran, eds.), 7th Ed., pp. 1693–1715. W. B. Saunders, Philadelphia.

Diarrhea, Infectious

SILVIA RESTA-LENERT
University of California, San Diego

food-borne infection Infection acquired via contaminated food.
incidence Rate of occurrence of an event.
traveler's diarrhea Infectious diarrhea acquired while traveling in endemic areas.
waterborne infection Infection acquired via contaminated water.

"Diarrhea" derives from the Greek words δια ρεω, meaning "flow through." Diarrhea occurs when the volume of the colonic fluid is greater than the absorptive capacity of this segment, as a result of impaired absorption and/or increased secretion. In infectious diarrheas, the abnormal function is brought about by microorganisms that colonize the intestinal mucosa and subvert normal gut physiology either directly or via enterotoxins.

INTRODUCTION

Diarrheal diseases are a major cause of morbidity and mortality around the world, especially in developing countries where children suffer the greatest brunt of infectious diarrhea, malnutrition, and death. Annually, approximately 5 million children and infants die worldwide due to diarrheal diseases. In North America, the rate per year is still 0.9 diarrheal episodes per child, and in special circumstances (daycare centers, institutions), the incidence is as high as 5 episodes per year. Fourteen hospital admissions per 1000 children younger than 12 months, per year, result from acute diarrhea. Among the adult population, most patients developing acute diarrhea are managed as outpatients or will not seek medical attention. However, 0.5 million hospital admissions per year, or 1.5% of all adult hospital admissions annually, are due to diarrhea. In developing countries, inadequate water supply, inefficient or nonexistent sewage removal systems, chronic malnutrition, and lack of access to oral rehydration are responsible for the high incidence of infectious diarrheal diseases. In the industrialized world, acute diarrhea is still one of the most frequent diagnoses in general practice and children, elderly, and immunocompromised patients are the most vulnerable individuals and account for the majority of these cases.

Regardless of the etiology, diarrhea is defined clinically as the occurrence of three or more episodes of loose stool or any loose stool with blood during a 24 h time period. Symptoms lasting less than 14 days represent acute diarrhea, whereas persistent diarrhea lasts more than 14 days but less than 4 weeks, and chronic diarrhea is defined by a duration of symptoms greater than 4 weeks.

Infectious diarrheas are miserable illnesses of overwhelming impact on the general survival of entire populations. Throughout history, thousands-strong armies have been defeated by raging diarrheal diseases: from the Greeks and Macedons under Alexander (Tucidides), to the Romans in the campaigns against the Gauls (Julius Caesar), to the Hundred Years War in 13th century Europe, to Napoleon, the Civil War in America, World War II, and the Vietnam War. Scores of previously healthy men suffered and died from the scourge of diarrhea and dysentery in all of these conflicts.

EPIDEMIOLOGY

Twenty years ago, 800 million to 1 billion episodes of infectious diarrhea and nearly 5 million deaths occurred per year worldwide, primarily in developing countries. Ten years later, survival had improved, but the incidence was virtually unchanged despite greater knowledge of the pathophysiology of diarrhea and greater intervention by the World Health Organization (WHO). Approximately 100 million episodes of acute diarrhea occur in the United States yearly, with an incidence of 1.2 to 1.5 diarrheal episodes per person-year. Medical costs/analyses show that 8.0 million Americans sought physician care for diarrhea yearly and 250,000 required hospitalization. Hospitalization and medical costs approached $560 million, whereas lost productivity totaled $200 million. Approximately another 8 million people sought physician care but were not hospitalized. These patients incurred $690 million in medical costs and $2 billion in loss of work hours. An estimated 90 million cases occurred in people who

 Copyright 2004, Elsevier (USA). All rights reserved.

did not seek physician care, costing nearly $20 billion in lost productivity. Approximately 90% of all these cases were presumably of infectious origin. Thus, the total cost estimate for diarrheal diseases exceeds $23 billion annually in the United States alone.

Although the elderly have an increased risk for death from diarrhea, death from diarrhea is rare among young children in industrialized countries. In fact, of all pediatric admissions for diarrhea, 0.05% resulted in death, compared with 3% in patients older than age 80. Increased age was the most important risk factor for death with an odds ratio of 52.6 (95% confidence interval, 37.0 to 76.9) for age 70 or older versus children >5 years. The national mortality figures for the 9-year period 1979–1988 in the United States show 51% of diarrheal deaths occurring in individuals older than age 74.

Acute infectious diarrhea is transmitted mostly through the fecal–oral route and by ingestion of contaminated water and food. Infection via the fecal–oral route occurs by direct contact with index cases, especially under conditions of crowding, such as daycare centers or nursing homes. Waterborne and food-borne outbreaks are another important source of disease transmission and result from general and/or individual failures in proper standards for the safe handling of foods. In most developing nations, acute diarrhea is endemic due to poor sanitation. Furthermore, epidemics of significant proportions often result from natural disasters in areas where water and food supplies are already chronically jeopardized. In some areas of the world, such as Asia, Africa, and Latin America, certain infectious diarrheas (e.g., cholera) have become ongoing pandemics lasting several decades, notwithstanding WHO efforts at eradication.

In most parts of the world, a definite seasonality is recognized in the incidence of acute diarrhea. In industrialized nations, the highest incidence of hospital admissions for diarrhea occurs in August and September and in the winter months. In developing nations with warmer climates and endemic conditions, variations in incidence occur from year to year in relation to precipitation indices and crop failures.

PATHOPHYSIOLOGY

Infectious diarrheas may be classified according to various criteria: duration, underlying mechanism, clinical presentation, etiology, and history. Table I summarizes the various criteria for classifying diarrheas in general and infectious diarrheas in particular.

In this section, infectious diarrheas are described according to the duration of the main gastrointestinal symptom.

Acute Infectious Diarrhea

Acute diarrheas last, by definition, less than 4 days and the majority are due to infectious agents. Most of these infections are self-limited and generally do not require medical intervention, unless severe dehydration and toxicity develop. However, immunocompromised patients, the elderly, and the very young may develop complications from enteric pathogens that warrant prompt and decisive medical intervention. A list of the major organisms involved in the etiology of acute infectious diarrheas is presented in Table II. Not listed is a type of acute enteritis, waterborne and of presumed infectious origin, that has been responsible for several outbreaks of traveler's diarrhea, known as Brainerd diarrhea. The etiologic agent of this disease still escapes definition.

Many of the acute infectious diarrheas observed worldwide are diagnosed in the course of local or epidemic outbreaks. Three major situations may be encountered: (1) waterborne infections; (2) Food-borne diarrhea; and (3) traveler's diarrhea. Whereas food-borne diarrhea is often associated with residual microbial toxins, waterborne and traveler's diarrheas are more often caused by active infection via the fecal–oral route. Table III summarizes the most common causes in these epidemiological situations.

A successful enteric pathogen possesses well-developed abilities to colonize, grow, and compete for nutrients in a crowded environment and to interact effectively with the host's enterocytes, inducing changes in the balance between absorption and secretion of water and electrolytes. In most gut infections, a pathogen enters via the oral route and colonizes an area of the

TABLE I Classification of Infectious Diarrheas According to Various Criteria

Duration	Mechanism	Clinical findings	Etiology	Patient's History
Acute	Secretory	Inflammatory	Bacteria	Age
Persistent	Nonsecretory	Noninflammatory	Viruses	Travel
Chronic			Unicellular parasites	Immunocompetence
			Worms	Food-/waterborne
				Postinfectious

TABLE II Common Causes of Acute Infectious Diarrhea

Bacteria	Viruses	Unicellular parasites	Worms
Salmonella	Rotavirus	*Giardia lamblia*	*Strongyloides stercoralis*
Shigella	Norwalk virus	*Entamoeba histolytica*	*Anchilostoma duodenalis*
Escherichia coli[a]	Calicivirus	*Cryptosporidium parvum*	*Necator americanus*
Yersinia enterocolitica	Adenovirus	*Cyclospora cayetanensis*	*Hymenolepsis nana*
Vibrio spp.	Astrovirus	*Microsporidia*[b]	*Heterophyes heterophyes*
Campylobacter	Coronavirus	*Isospora belli*	
Staphylococcus aureus	Herpes simplex virus	*Blastocystis hominis*	
Bacillus cereus	Cytomegalovirus	*Balantidium coli*	
Listeria monocytogenes		*Dientamoeba fragilis*	
Clostridium perfringens			
Clostridium difficile			
Aeromonas			
Plesiomonas			

[a] EIEC, enteroinvasive *E. coli*; ETEC, enterotoxigenic *E. coli*; EPEC, enteropathogenic *E. coli*; EAEC, enteroadhesive *E. coli*.
[b] The phylum includes *Microsporidium, Encephalitozoon, Pleistophora, Trachipleistophora, Nosema, Vittaforma, Brachiola.*

intestine. Exceptions to this paradigm are the ingestion of preformed toxins. Pathogens produce diarrhea by three basic mechanisms: (1) enterotoxins that induce active intestinal secretion (*Vibrio cholerae*, *Staphylococcus aureus*, *Bacillus cereus*, *Clostridium botulinum*, rotavirus); (2) cytotoxic mediators (most bacteria, parasites); and (3) invasins promoting endocytosis, with subsequent tissue invasion and mucosal injury [*Shigella, Salmonella*, enteroinvasive *Escherichia coli* (EIEC)]. In addition to direct effects by microorganisms and their products, enteropathogens induce intestinal damage indirectly via the mucosal inflammatory response, which involves secretion of various powerful mediators of secretion and apoptosis. A summary of the current knowledge about the pathogenesis of the most common acute infectious diarrheal syndromes is shown in Fig. 1.

TABLE III Agents Associated with Outbreaks of Acute Infectious Diarrheas

Waterborne	Foodborne	Traveler's
Vibrio cholera	*Campylobacter*	*Escherichia coli*
Campylobacter	*Salmonella*	*Campylobacter*
Salmonella	*E. coli*	*Salmonella*
Shigella	*Shigella*	*Shigella*
E. coli	*Staphylococcus aureus*	*Aeromonas/ Plesiomonas*
Giardia	*Clostridia*	*Giardia*
Entamoeba hystolytica	*Vibrio parahaemolyticus*	*Cryptosporidium*
Cryptosporidium	Caliciviruses	*Cyclospora*
Cyclospora	Norwalk virus	*Isospora*
Microsporidia	*Giardia*	*Rotavirus*
Enteroviruses	*Cryptosporidia*	

On the basis of these three mechanisms, acute infections present as watery, noninflammatory diarrheal syndromes or inflammatory diarrheal syndromes. The majority of watery, noninflammatory diarrhea cases are self-limited diseases characterized by low-grade fever, nausea, vomiting, large-volume diarrhea, and the absence of blood and leukocytes in the stools. This presentation is typically reported in patients infected with enterotoxigenic *Escherichia coli*, *V. cholerae*, clostridial and staphylococcal food poisoning, rotavirus, Norwalk virus agent, *Giardia lamblia*, and *Cryptosporidium*. On the other hand, the inflammatory diarrheal syndrome is characterized by frequent, small-volume stools that may contain blood and leukocytes, tenesmus, fever, and severe abdominal pain. The most common microorganisms causing this syndrome include *Salmonella*, *Shigella*, *Campylobacter*, enterohemorrhagic *E. coli*, EIEC, *Clostridium difficile*, *Entamoeba histolytica*, and *Yersinia*. Table IV describes the basic biologic, pathophysiologic, and clinical characteristics pertinent to the most common enteric pathogens.

Persistent Infectious Diarrhea

Persistent diarrhea is emerging as a major world health problem. Children are more likely to develop persistent diarrhea and suffer malnutrition, wasting, and immunocompromise as a consequence. Persistent diarrhea is defined by loose–soft stools occurring at increased frequency and lasting for more than

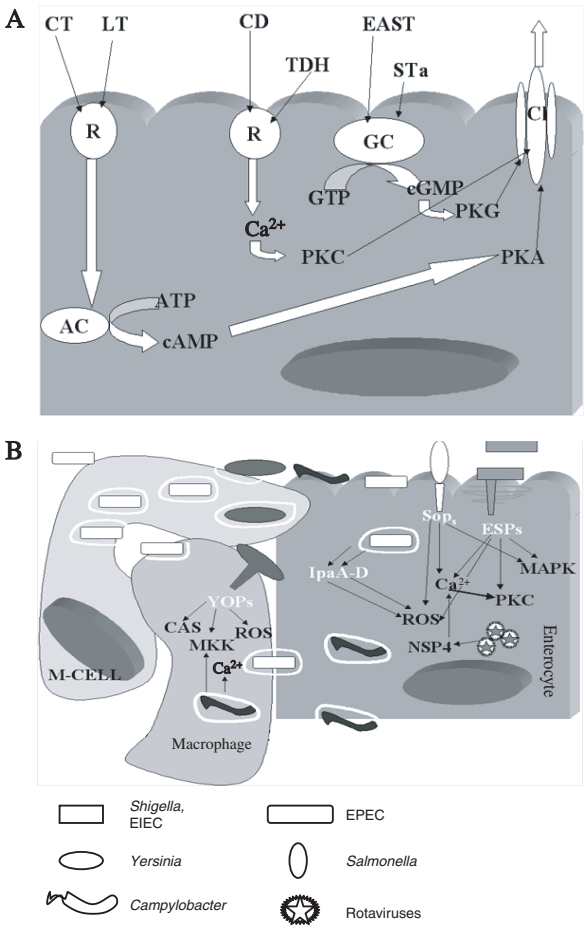


FIGURE 1 Infectious diarrhea: mechanisms of action of major enteric bacteria and viruses. Enteric pathogens can induce intestinal injury with consequent diarrhea in three ways: (1) by producing enterotoxins that interact with receptors located on the gut epithelial cells and evoke anion secretion, such as *V. cholera*, EPEC, EAEC, STEC, *C. difficile*, and *S. aureus* (A); (2) by invading the gut epithelium and M cells, thus altering the cell cytoskeleton and activating intracellular pathways through virulence factors. Organisms that lead to diarrhea through these mechanisms include EIEC, *Shigella*, EPEC, *Salmonella*, and rotaviruses (B); (3) by invading mucosal macrophages and inducing inflammatory responses leading to intestinal epithelial damage and anion secretion. *Campylobacter* and *Yersinia* use this mechanism (B).

2 weeks after the end of an acute episode of gastroenteritis. Persistent infectious diarrhea may result from multiple repeated infections, or persistent infection by the original organism, or as the so-called postgastroenteritis syndrome. Overall, the incidence of persistent infectious diarrhea is equally distributed in industrialized countries, including the United States, and developing nations. Table V lists the most common infectious agents associated with persistent diarrhea.

Postinfectious persistent diarrhea is a poorly defined syndrome that occurs as a sequela of an acute episode with definite infectious etiology. Patients may develop mild to severe degrees of malabsorption, from lactose intolerance to inability to absorb proteins, fat, and sugars, as well as permanent blunting of villi as assessed by histopathology. The condition is characterized by watery, malodorous stools and progressive wasting.

Chronic Infectious Diarrhea

Chronic infectious diarrhea occurs mostly in immunocompromised patients. After an acute infectious episode, patients sometimes develop chronic symptoms that are independent of the etiologic agents of acute diarrhea (irritable bowel syndrome with diarrhea, or, occasionally, ulcerative colitis). Table VI lists the most common agents isolated from cases of chronic infectious diarrhea.

By definition, chronic diarrhea lasts more than 4 weeks and patients developing this syndrome quite often are hospitalized and have undergone antibiotic therapy for other reasons. Elderly, human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS), transplant, and cancer patients are easy targets for reinfections or reactivation of only partially subdued infectious organisms. In addition to the causes listed above, bacterial overgrowth can occur in areas of bowel stasis or impaired bowel motility. Postsurgery patients, diabetics, posttrauma patients, and intensive care patients are more likely targets of chronic infectious diarrheas from bacterial overgrowth.

SPECIAL HOSTS

The Elderly

Infectious diarrhea causes high morbidity and mortality among the aging population worldwide. Multiorgan complications from an acute episode of infectious diarrhea are also more frequent among the elderly. Life expectancy in the United States has risen from an average of 45 years in the 20th century to 75 years at present. By the year 2025, 22% of the U.S. population will be older than age 65. Gastrointestinal physiology and gut colonization change constantly with aging and contribute in a significant way to increasing the susceptibility of elderly people to enteric infections. Furthermore, the gastric acid barrier in the elderly is impaired. The most frequently isolated organisms and most deadly in elderly patients with diarrhea are *C. difficile*, *Salmonella*, and toxigenic *E. coli*. These three agents top the list of

TABLE IV Acute Infectious Diarrhea: Biology, Pathophysiology, and Clinical Findings by Etiologic Agent

Organism	Microbiology	Pathophysiology	Epidemiology	Clinical findings
Bacteria				
Salmonella: S. enteritidis; S. typhimurium; S. typhi	Invasive, gram-negative rod; 2000 serotypes	Nontoxigenic; fimbriae, SPI-1 gene-encoded effectors (inv, spa, sic, sip, TTSS, etc.), plasmid-encoded effectors (sop A-E2, hsp)	Salmonellosis in USA: 1.4 million cases/year; >500 death/year; Typhoid fever in USA: 400 cases/year; worldwide: 16 million cases with 600,000 death/year	Salmonellosis: fever, abdominal cramps, diarrhea; typhoid fever: fever, headache, malaise, vomiting; uncommon diarrhea
Shigella	Invasive, gram-negative rod; 4 species	Toxigenic; pili, flagella, TTSS, Mxi-Spa, IpaA-C, IpgD effectors	USA: 450,000 cases/year (*S. sonnei*), worldwide: *S. flexneri* and *S. dysenteriae*; fatality rate 5–15%	Fever, abdominal pain, malaise, watery or bloody diarrhea
Escherichia coli	Gram-negative rod			
EIEC	Invasive	Pili, TTSS, IpaC, Esps, adhesin		Watery or bloody diarrhea
EPEC	Noninvasive or limited invasion; typical and atypical strains	TTSS, Bfp, intimin, EspF	USA: unknown Worldwide: unknown	Watery diarrhea, nausea, vomiting
EAEC	Adherent, limited invasion	Fimbriae, TTSS, EAST, cytotoxin		Watery diarrhea
ETEC	Noninvasive, adherent	TTSS, Cfs, LT, ST	USA: 80,000 cases/year	Watery diarrhea
EHEC	Noninvasive, adherent	Toxigenic; Intimin, Stx 1 and 2	USA: 70,000 cases/year; 61 deaths	Bloody diarrhea
Yersinia enterocolitica	Gram-negative rod	TTSS, Ysc, Yop effectors	USA: 1 case/100,000/year (culture-confirmed)	Fever, abdominal pain, bloody diarrhea
Vibrio:	Non invasive, gram-negative sickle-shaped			
V. cholera		CtxA, ctxB, zot, ace, tcpA effectors, toxR, tcpP	USA: 0–5 cases/year; pandemic in Asia, Africa, Latin America	Profuse watery diarrhea, vomiting
V. parahaemolyticus/ V. vulnificus		TxA/B	USA: 3000/95 cases/year; 7/35 deaths/year	Watery diarrhea, abdominal cramps
Campylobacter jejuni	Invasive, gram-negative	Type IV secretion	USA: 2.4 million cases/year; 124 deaths/year	Fever, abdominal cramps, diarrhea (often bloody)
Staphylococcus aureus	Noninvasive, gram-positive cocci	Staph enterotoxin	USA: true incidence unknown	Nausea, vomiting, watery diarrhea
Bacillus cereus	Rod-shaped, spore-forming	Stable emetic toxin, heat- and acid-labile enterotoxin	USA: 2% food-borne outbreaks/year;	Emetic syndrome and diarrheal syndrome
Listeria monocytogenes	Invasive, gram-positive	LLO, ActA	USA: 2500 cases/year; 500 deaths/year	Fever, abdominal pain, watery diarrhea
Clostridium perfringens	Noninvasive, gram-positive, spore-forming	CpA, cpE	USA: 2% of all acute infectious diarrheas	Nausea, vomiting, diarrhea

Clostridium difficile	Noninvasive, gram-positive, spore-forming	NeuroTx A, cytoTxB	USA: 25% of all antibiotic-associated diarrheas	Watery diarrhea, fever, anorexia, abdominal pain
Aeromonas/Plesiomonas	Gram-negative rod	Cytotoxic enzymes	USA: rare	Watery or bloody diarrhea, abdominal cramps
Viruses				
Rotavirus	*Reoviridae*, dsRNA	NSP4 enterotoxin	USA: 3 million cases/year; worldwide: l million deaths/year	Vomiting, watery diarrhea
Norwalk virus	*Caliciviridae*, ssRNA	Unknown	30% of all cases of diarrheas in children >1 year	Nausea, vomiting, diarrhea
Calicivirus	*Caliciviridae*, ssRNA	Unknown	1.5% of all cases of viral gastroenteritis	Nausea, diarrhea
Adenovirus	*Adenoviridae*, dsDNA, type 40 and 41	Unknown	<1% of all cases of viral gastroenteritis	Fever, vomiting, diarrhea
Astrovirus	*Astroviridae*, ssRNA	Unknown	1.5% of all cases of viral gastroenteritis	Watery diarrhea
Coronavirus	*Coronaviridae*, ssRNA	Unknown	<1% of all cases of viral gastroenteritis	Vomiting, diarrhea
Herpes simplex virus	*Alphaherpesvirinae*, dsDNA	Unknown		Fever, tenesmus, watery or bloody diarrhea
Cytomegalovirus	*Betaherpesvirinae*, dsDNA	Unknown	Rare in immunocompetent; 16% in solid organ transplants; 5% in HIV/AIDS	Fever, malaise, abdominal tenderness, diarrhea
Parasites, unicellular				
Giardia lamblia	Diplomonadida, cysts and throphozoites, 5 chromosomes (5K genes)	VSP (analogy with sarafotoxins)	USA: 2.5 million cases/year; endemic in developing countries	Diarrhea, flatulence, abdominal cramps, malabsorption
Entamoeba histolytica	Entamoebidae, cysts and throphozoites, 14 chromosomes	Cysteine proteinases	USA: infrequent; worldwide: 400 million infections/year, 100,000 deaths/year	Asymptomatic, mild gastroenteritis, or bloody dysentery
Cryptosporidium parvum	Alveolata, oocysts and throphozoites, ongoing genome sequencing	Peptidases	USA: 2% of the general population; worldwide: unknown	Asymptomatic or watery diarrhea
Cyclospora cayetanensis	Alveolata, oocysts and sporozoites, ongoing genome sequencing	Unknown	USA: unknown, outbreak related to contaminated berries; worldwide: unknown, endemic in Guatemala and Peru	Fever, watery diarrhea, fatigue
Microsporidia	Microsporidia, spores and schizontes	Unknown	Unknown	Asymptomatic or watery diarrhea
Isospora belli	Alveolata, oocysts and throphozoites	Unknown	Unknown	Asymptomatic or watery diarrhea
Blastocystis hominis	Stramenopiles	Unknown	Unknown	Asymptomatic or watery diarrhea
Balantidium coli	Alveolata, cysts and throphozoites	Unknown	USA: rare	Asymptomatic to bloody diarrhea

continues

TABLE IV Acute Infectious Diarrhea: Biology, Pathophysiology, and Clinical Findings by Etiologic Agent (*continued*)

Organism	Microbiology	Pathophysiology	Epidemiology	Clinical findings
Dientamoeba fragilis	Parabasalidea, throphozoites, no cysts	Peptidases	USA: infrequent	Nausea, malaise, mucous diarrhea, abdominal pain
Worms				
Strongyloides stercoralis	Helminths, nematodes, filariform larvae, can complete life cycle in humans	Organism effectors and host responses	USA: 4% prevalence in Appalachian States; worldwide: 100 million cases/year, 60% prevalence in tropical countries	Mild to severe diarrhea, malaise, fatigue, malnutrition
Anchilostoma duodenalis, Necator, americanus	Helminths, nematodes, filariform larvae and eggs	Organism effectors and host responses	USA: uncommon; worldwide: in tropical countries prevalence is increasing due to climate changes	Mild to severe diarrhea, abdominal pain, weight loss
Hymenolepsis nana, H. diminuta	Cestodes, cysticercoids and adult worms in humans (fleas and beetles intermediate hosts)	Organism effectors and host responses	USA: rare; worldwide: Latin America	Abdominal cramps, mucous diarrhea upon rupturing of villus by cysticercoids
Heterophyes heterophyes	Trematodes, metacercariae and eggs (fish and snails intermediate hosts)	Organism effectors and host responses	USA: rare; worldwide: endemic in Egypt, Middle East, and Far East	Asymptomatic to severe mucous diarrhea, intestinal wall granulomas

Note. SPI, *Salmonella* pathogenicity island; TTSS, type 3 secretory system; Tx, toxin; Cfs, cytotoxic factors; EAST, enteroaggregative heat-stable toxin; LT, heat-labile toxin; ST, heat-stable toxin.

TABLE V Agents Associated with Persistent Infectious Diarrhea

Bacteria	Parasites
Salmonella	*Cryptosporidium parvum*
Shigella	*Cyclospora cayetanensis*
Escherichia coli[a]	*Giardia lamblia*
Yersinia	*Entamoeba hystolytica*
Campylobacter	*Balantidium coli*
Clostridium	*Dientamoeba fragilis*

[a] EIEC, enteroinvasive *E. coli*; EAEC, enteroadherent *E. coli*; ETEC, enterotoxigenic *E. coli*; EPEC, enteropathogenic *E. coli*.

outbreaks in long-term and short-term care facilities and *Salmonella* by itself accounts for more than 50% of cases and more than 80% of deaths in food-borne outbreaks in nursing homes.

HIV/AIDS

More than 50% of HIV/AIDS patients in the United States experience infectious diarrhea and this estimate may approach 100% in developing countries where the HIV epidemic is currently raging unchecked. These patients are more likely to develop persistent or chronic diarrhea after an acute episode because of their impaired immunity, with a significant increase in morbidity and mortality. Table VII lists the most common causes of infectious diarrhea in AIDS patients.

The American Gastroenterological Association (AGA) has published a set of general guidelines for the management of chronic diarrhea in AIDS patients. At least three sets of stool samples should be secured for common enteric bacteria and parasites, including microsporidia, cryptosporidia, and *C. difficile*. Febrile patients with diarrhea should have blood cultures for common enteric bacteria. Patients with CD4 lymphocyte counts of <100 cells/mm are at high risk for disseminated mycobacterial infection.

TABLE VI Agents Associated with Chronic Infectious Diarrhea

Bacteria	Parasites
Campylobacter	*Amoeba*
Mycobacterium tuberculosis	*Cryptosporidium*
Aeromonas	*Giardia lamblia*
Plesiomonas	*Isospora*
Salmonella	*Cyclospora*
Clostridium difficile	*Strongyloides*
	Trichuris
	Schistosoma

CLINICAL AND LABORATORY FINDINGS

The most important finding in patients presenting with acute diarrhea is the degree of volume depletion, i.e., dehydration. Postural changes in blood pressure are a reliable sign of dehydration. Fever, abdominal tenderness, increased bowel sounds, or blood on rectal examination should alert the physician to acute infectious diarrhea.

Microscopic examination of a stool sample or rectal swab is a traditional and helpful tool in the rapid, bedside investigation of diarrheal illness. The specimen is placed on a glass slide and mixed thoroughly with two drops of methylene blue. The presence of ova, cysts, and/or leukocytes may point directly to a diagnosis. The AGA guidelines on managing acute diarrhea indicates empiric antimicrobial therapy in the case of positive fecal leukocytes in a febrile patients.

Endoscopy has limited utility in the investigation of acute infectious diarrhea and is not cost-effective. It may have a place, however, in cases of persistent or chronic diarrhea.

PREVENTION

Preventative measures against infectious diarrhea must include improvements in sanitation (water supply, sewer systems, housing), education of the general population and, where applicable, vaccination campaigns. Unfortunately, no effective vaccines are available for the organisms that cause infectious diarrheas, with the exception of typhoid fever.

TREATMENT

Most acute diarrheal illnesses are self-limited and no specific therapy is required. Water and electrolyte loss can be prevented or treated with oral fluid–electrolyte solutions. Intravenous saline–glucose solutions are recommended in cases of moderate to severe dehydration. Glucose in the intestinal lumen facilitates the absorption of sodium and the cotransport mechanism for these solutes appears to be unhampered by infection with microorganisms or by their toxins.

Antimotility therapy should be reserved for severe cases and chronic diarrheas and avoided in infants and children. Antibiotic or antiviral treatment should be considered in moderate to severe cases in which a microbiological diagnosis is obtained or strongly

TABLE VII Agents Associated with Diarrhea in AIDS Patients

Bacteria
Shigella
Salmonella
Escherichia coli
Campylobacter
Yersinia enterocolitica
Clostridium difficile
Clostridium perfringens
Staphylococcus aureus
Aeromonas
Plesiomonas
Bacillus cereus
Vibrio parahemolyticus
Mycobacterium avium complex
Treponema
Viruses
Cytomegalovirus
Adenovirus
Herpes simplex virus
Fungi
Hispolasma capsulatum
Blastocystis hominis
Parasites
Giardia lamblia
Entamoeba histolytica
Cryptosporidium
Isospora
Cyclospora
Enterocytozoon bieneusi
Encephalitozoon intestinalis
Balantidium coli

suspected. In immunocompromised patients with febrile diarrheas, empirical antibiotics should be promptly initiated after securing adequate culture specimens to define an etiology.

See Also the Following Articles

AIDS, Gastrointestinal Manifestations of • Anti-Diarrheal Drugs • *Campylobacter* • Cholera • *Cryptosporidium* • Cytomegalovirus • Diarrhea • Foodborne Diseases • Food Poisoning • Food Safety • Giardiasis • Rotavirus • *Salmonella* • *Shigella* • Traveler's Diarrhea

Further Reading

Anonymous (2001). Diagnosis and management of foodborne illnesses: A primer for physicians. *MMWR* **50**(RR-2), 1–60.

Barbut, F., and Petit, J. C. (2001). Epidemiology of *Clostridium difficile*-associated infections. *Clin. Microbiol. Infect.* **7**, 405–410.

Cezard, J. P., Duhamel, J. F., Meyer, M., *et al.* (2001). Efficacy and tolerability of racecadotril in acute diarrhea in children. *Gastroenterology* **120**, 799–805.

Guerrant, R. L., Van Gilder, T., Steiner, T. S., *et al.* (2001). Practice guidelines for the management of infectious diarrhea. *Clin. Infect. Dis.* **32**, 331–351.

Mandell, G. L., Bennett, J. E., and Dolin, R. (eds.) (2000). "Principles and Practice of Infectious Diseases," 5th Ed. Churchill Livingstone, Philadelphia, PA.

Morris, A. P., and Estes, M. K. (2001). Microbes and microbial toxins: Paradigms for microbial–mucosal interactions. VIII. Pathological consequences of rotaviral infection and its enterotoxin. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G303–G310.

Okhuysen, P. C. (2001). Traveler's diarrhea due to intestinal protozoa. *Clin. Infect. Dis.* **33**, 110–114.

Oldfield, E. C., and Wallace, M. R. (2001). The role of antibiotics in the treatment of infectious diarrhea. *Gastroenterol. Clin. North Am.* **30**, 817–836.

Ramaswamy, K., and Jacobson, K. (2001). Infectious diarrhea in children. *Gastroenterol. Clin. North Am.* **30**, 611–624.

Schiller, L. R. (2000). Diarrhea. *Med. Clin. North Am.* **84**, 1259–1274.

Stephen, J. (2001). Pathogenesis of infectious diarrhea. *Can. J. Gastroenterol.* **15**, 669–683.

Diarrhea, Pediatric

JEFFREY A. RUDOLPH* AND PAUL A. RUFO†
*Children's Hospital Medical Center, Cincinnati and †Children's Hospital Boston

acute diarrhea A diarrheal illness of less than 14 days duration. Acute diarrheal disease in children is most often the result of self-limited viral infections. Management includes prompt assessment and repletion of hydration status. Evaluation for an etiologic process is generally not warranted unless there is an associated finding such as blood in the stool or systemic symptoms.

chronic diarrhea A diarrheal illness of greater than 14 days duration. Chronic diarrhea in children can be due to either infectious or noninfectious processes. Evaluation for a specific etiology is indicated. Management of comorbid conditions such as poor growth or malnutrition is essential.

colitis Any inflammatory process affecting the colon. Colitis usually presents clinically as bloody diarrhea, abdominal cramping, and tenesmus.

congenital diarrhea A group of diarrheal illnesses that are present from birth. Congenital diarrhea can be the result of either a specific genetic defect in a secretory or absorptive pathway or abnormal intestinal development.

gastroenteritis A diarrheal process that affects the upper gastrointestinal tract and presents most typically as an acute watery diarrhea. Gastroenteritis usually denotes an acute diarrhea that is infectious and self-limiting.

hemolytic uremic syndrome A sequela of *Escherichia coli* O157:H7 colitis. This toxin-mediated microangiopathy results in a triad of hemolytic anemia, thrombocytopenia, and renal failure. The occurrence of the syndrome is generally limited to children under 10 years of age.

inflammatory diarrhea A diarrheal illness in which the predominant pathologic finding is an invasion of the intestinal epithelium by immunocytes. This type of diarrhea can be the result of either a normal immune response to an abnormal environment, as in infection, or an abnormal immune response to a normal environment, as in inflammatory bowel disease.

IPEX syndrome (immunodysregulation, polyendocrinopathy, and enteropathy: X-linked) An inherited X-linked syndrome that results from a mutation in the FOXP3 gene in humans. It is characterized by autoimmune enteropathy and multiple endocrinological abnormalities including diabetes mellitus, hypothyroidism, and hemolytic anemia.

osmotic diarrhea A diarrheal illness that is driven by osmotic forces that promote a net flux of water out of the interstitium and into the intestinal lumen. A stool sodium level of mEq/liter and an osmotic gap of greater than 100 mosm/liter suggest an osmotic diarrhea.

secretory diarrhea A diarrheal illness that is driven by the active secretion of salt and water by intestinal epithelial cells. A stool sodium level of >70 mEq/liter, an osmotic gap of less than 100 mosm/liter, and a failure of the diarrhea to respond to a controlled fast suggest a secretory diarrhea.

Diarrheal diseases continue to contribute significant morbidity and mortality to pediatric populations in developed and developing countries around the world. The prevalence of diarrheal illness across cultures is inversely proportional to the availability of public sanitation, clean water supply, and adequate medical care. As such, it is not surprising that the incidence of diarrheal disease is much higher in developing societies and can approach 10 episodes per child per year in children under 5 years of age. In these areas, aggregate mortality can reach 3 to 5 million deaths per year. In the United States and other developed nations, both the incidence (1–2 episodes per year) and mortality (approximately 400–500 deaths annually) are considerably decreased. Nonetheless, the burdens placed on Western health care systems by pediatric diarrheal disease are considerable and approximately 20% of all pediatric ambulatory visits and 10% of all inpatient hospital admissions in children under 3 years of age are for the evaluation and treatment of these disorders and their complications.

The frequency and consistency of stool can vary considerably from individual to individual, as well as in the same individual over time. There has therefore remained a lack of a consensus as to how diarrheal illness should be defined. Investigators have employed a number of qualitative and quantitative dimensions of stool output to address this issue in the past. For the most part, children pass between one and three stools, or approximately 5–10 ml of stool per kilogram of body weight, per day. As such, investigators have begun to use these benchmarks as the upper limits of normal in their identification of subjects in studies addressing acute or chronic diarrheal disease.

 Copyright 2004, Elsevier (USA). All rights reserved.

REGULATION OF INTESTINAL FLUID SECRETION AND ABSORPTION

The mucosa lining the gastrointestinal tract must reconcile daily a seemingly contradictory array of physiologic tasks. These conflicting responsibilities include the maintenance of a tight barrier against potentially virulent bacterial and viral pathogens in the intestinal lumen, while at the same time presenting a selectively permeable interface through which to carry out immune surveillance and nutrient absorption. In this context, intestinal fluid secretion can serve both defensive (flushing away pathogens and toxins) and homeostatic (maintenance of mucosal hydration necessary to facilitate enzymatic digestion) purposes.

Stool output in humans is a composite of ingested, secreted, and absorbed fluid intermixed with residual dietary matter and cellular debris. Adults typically ingest approximately 2 liters of fluid per day and produce an additional 9 liters in the form of salivary, gastric, small intestinal, and pancreato-biliary secretions, to complete the process of digestion. The small intestine and colon have evolved highly efficient intercellular and transcellular pathways for the reabsorption of the vast majority (approximately 99%) of this intestinal fluid, and the average adult will pass only approximately 200 g of stool per day. This balance between fluid secretion and absorption is therefore quite tight. Any microbiologic, dietary, pharmacologic, or hormonal input that affects cell membrane transporters and/or the intercellular tight junctions responsible for fluid absorption can tip this net fluid balance in favor of secretion (or reduced absorption) and thereby trigger the increased stool output observed in patients with diarrheal illnesses.

The cellular basis for salt and water secretion in the intestine, as well as in other hydrated mucosal surfaces in the body, depends upon a vectorial transport of Cl^- ions by specialized epithelial cells. Intestinal crypt epithelial cells use basolateral membrane Na^+/K^+-ATPase pumps as well as the Na^+- and K^+-coupled cotransporter NKCC1 to accumulate Cl^- ions above their electrochemical gradient (Fig. 1). The subsequent opening of Cl^- channels located in the apical membrane of enterocytes permits sequestered Cl^- ions to move down their electrochemical gradient and into the intestinal lumen. The parallel activation of plasma membrane K^+ channels conducts K^+ outside, thereby sustaining the inside-negative cell membrane potential that is necessary to initiate and maintain a Cl^- secretory response.

Fluid secretion in the intestine is tightly regulated by endocrine as well as neuroenteric mechanisms that utilize either cyclic nucleotides [3′,5′-monophosphate (cAMP) or cyclic GMP (cGMP)] or Ca^{2+} as second messengers. Cyclic nucleotide-dependent agonists initiate Cl^- secretion through the parallel activation of the apical membrane Cl^- channel CFTR (the cystic fibrosis transmembrane receptor) as well as the basolateral membrane K^+ channel KCNQ1/KCNE3. In contrast, agonists utilizing Ca^{2+} as a second messenger activate the apical membrane Cl^- conductance CaCC in concert with the basolateral membrane K^+ channel IK1 (KCNN4). The net movement of Cl^- ions into the intestinal lumen imparts a transiently negative charge to this extracellular compartment and positively charged Na^+ ions move via paracellular pathways in response. The osmotic force generated by transported Cl^- and Na^+ ions pulls water molecules along to effect net fluid secretion. The activity of CFTR is regulated primarily by cAMP- and cGMP-dependent protein kinases. In contrast, Ca^{2+}-dependent Cl^- secretion in the intestine conducted by CLCA appears to be limited by the generation of the intracellular down-regulatory intermediates inositol-3, 4,5,6-tetrakisphosphate, and phosphorylated extracellular signal-regulated kinase.

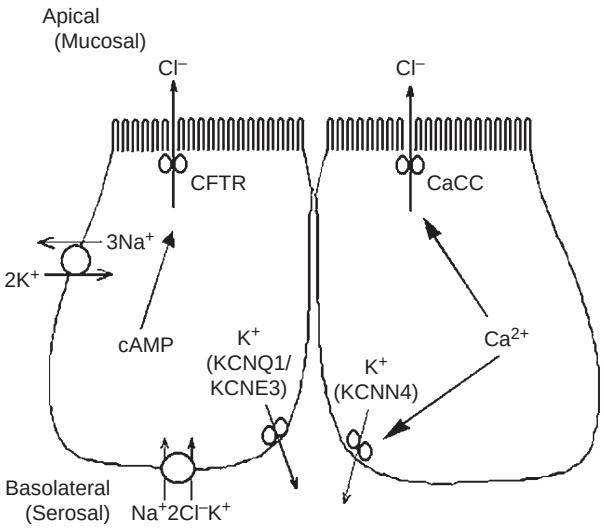


FIGURE 1 Intestinal crypt epithelial cells use basolateral membrane Na^+/K^+-ATPase pumps as well as the Na^+- and K^+-coupled cotransporter NKCC1 to accumulate Cl^- ions above their electrochemical gradient. The subsequent opening of Cl^- channels located in the apical membrane of enterocytes permits sequestered Cl^- ions to move down their electrochemical gradient and into the intestinal lumen. The parallel activation of plasma membrane K^+ channels conduct K^+ outside, thereby sustaining the inside-negative cell membrane potential that is necessary to initiate and maintain a Cl^- secretory response.

Whereas Cl^- secretion drives intestinal fluid secretion, fluid absorption is mediated primarily by the vectorial transport of Na^+ ions out of the intestinal lumen and into the interstitium. Na^+ transport can be

electrogenic (as in the case of apical Na^+ channels), Na^+-coupled, or electroneutral. The accumulation of absorbed Na^+ ions in the tissue interstitium favors the subsequent movement of Cl^- ions and water molecules out of the intestinal lumen via transcellular and paracellular pathways, thereby effecting salt and water uptake. Na^+ channels have been identified in the apical membrane of gut epithelium. By acting in a coupled fashion with basolateral membrane Na^+/K^+-ATPase pumps, these channels permit lumenal Na^+ ions to move down their electrochemical gradient and into the cell. The favorable Na^+ gradient established by Na^+/K^+ pumps has also been exploited by the small intestine to promote nutrient absorption. SGLT1 is the Na^+-coupled glucose transporter expressed along the apical membrane of enterocytes. Similarly, Na^+ uptake in the small intestine is effected through Na^+-coupled amino acid transporters that are present along the enterocyte brush border. Finally, the Na^+/H^+ exchanger NHE-3, expressed in the apical membrane of enterocytes, appears to mediate electroneutral Na^+ transport in the intestine.

The tasks of intestinal fluid secretion and absorption are separated spatially along the length of the crypt–villus axis through a segregation of relevant plasma membrane channels and transporters. Cells newly differentiated at the crypt base display a primarily secretory phenotype and express high levels of CFTR. As these cells mature and migrate up the axis to take up more villous positions, they express increasing numbers of absorptive proteins including NHE-3 and Na^+-coupled glucose and amino acid transporters. Stool output is therefore the net product of intestinal fluid secretion originating in crypt cells (which occupy approximately one-third of the crypt–villus axis) and fluid absorption from villus cells (which take up the remaining two-thirds of the crypt–villus axis). Any disorder damaging surface villi, and thereby decreasing the villus/crypt ratio, will selectively decrease mucosal absorptive potential and cause increased stool output. This explains the increased stool output observed in patients with celiac disease, postviral syndromes, and giardiasis.

APPROACH TO THE CHILD WITH DIARRHEA

Diarrhea can be classified on the basis of several descriptive factors (acute versus chronic, inflammatory versus noninflammatory, infectious versus noninfectious, secretory versus osmotic) that aid in the diagnostic approach. These include the duration of the illness, the existence of a secretory or osmotically driven mechanism, the presence or absence of a pathogen, and the degree of mucosal inflammation. Although the pathogenesis of diarrheal disease can be explained by a discrete process in some patients, increased stool output is more often the result of a combination of factors. As such, patients with inflammatory diarrhea can present with a secretory component due to the local release of endogenous secretagogues. Clinical diagnosis rests on an understanding of the close interplay between environmental and host factors in these patients.

Central to the diagnosis of a diarrheal illness is the clinical context in which it presents. Characteristics of the individual, such as age, are often the first clue in determining an etiology. This is most apparent in the case of congenital diarrheas, which present exclusively within the first few days of life. Components of the child's overall health, such as atopy or immunodeficiency, can also suggest a particular etiology. Environmental factors, including diet, must also be taken into consideration in the diagnostic approach to the pediatric patient with diarrhea. In the setting of infectious diarrhea, an exposure history such as an ill contact at home or in daycare, a recent travel history, or contact with a pet or animal, can sometimes provide useful epidemiologic information when attempting to understand how a pathogen may have been acquired.

The character of the stool itself is often helpful when arriving at a specific diagnosis. Stool that is both watery and voluminous in nature suggests an abnormality in the absorptive or secretory function of the small intestine. In contrast, crampy abdominal pain, tenesmus, and the presence of frank blood in the stool suggest colitis or large bowel disease.

Several aspects of diarrheal disease in children merit special consideration. Children, and most especially infants, are more susceptible to dehydration than their adult counterparts. This is due both to their greater overall body surface area relative to their weight and to a dependence on caregivers, who may be less likely to offer fluids to or feed a child who is vomiting or appears ill. Poor growth and malnutrition can also become a factor in children when diarrhea is chronic in nature. During infancy and early childhood, a large proportion of caloric intake is devoted to growth. Diarrheal disease, resulting in inadequate intake or poor nutrient absorption during this critical developmental period, can alter weight gain and, in severe cases, result in stunted linear growth.

The scope of the remaining article will discuss the causes, evaluation, and treatment of diarrheal disease in infants and children. By convention, the discussion will be segregated into infectious causes and noninfectious causes with a special reference to age of onset where appropriate.

INFECTIOUS DIARRHEA IN CHILDREN

Infectious diarrhea is usually of acute onset in a previously healthy child. Fortunately, most causes of infectious diarrhea are self-limited and require only symptomatic care. However, if left untreated, acute diarrheal illness can progress to chronic diarrhea in some patients. Fever is a common associated symptom of infectious diarrhea and vomiting is not unusual, especially if the infection occurs in the upper gastrointestinal tract (i.e., gastroenteritis). In general, infectious diarrheas are secretory or mixed secretory/osmotic in character. Toxin production, pathogen adherence, or frank tissue invasion all can contribute to increased Cl^- secretion by affected epithelial cells. When pathogenic invasion of the epithelium occurs, there is usually an inflammatory component to the diarrhea as well. Pathogens that cause diarrhea can be viral, bacterial, or parasitic.

Viruses are the most common cause of acute infectious gastroenteritis in children (Table I, part A). There are several reasons for the preponderance of cases of viral diarrheas. The naive immune system of an infant has not been exposed to many of the viral pathogens present in the environment. In addition, daycare provides group settings that facilitate the transmission of enteric and respiratory viral diseases.

TABLE I Etiology of Pediatric Diarrhea

Infectious diarrhea	Noninfectious diarrhea
A. Viral pathogens	**D. Inflammatory**
Rotavirus	Inflammatory bowel disease
Adenovirus	Celiac disease
Norwalk agent	Allergic enteropathy
Calicivirus	Autoimmune enteropathy
Astrovirus	Graft-versus-host disease
Coronavirus	**E. Noninflammatory**
B. Bacterial pathogens	**Congenital diarrheas**
Campylobacter spp.	Congenital chloride diarrhea
Salmonella spp.	Congenital sodium diarrhea
Shigella spp.	Microvillus inclusion disease
Escherichia coli	Tufting enteropathy
Enterotoxigenic	Carbohydrate transporter defects
Enteropathogenic	Dissacharidase deficiency
Enterohemorrhagic (shigatoxin producing)	Amino acid transporter defects
Enteroadherent	Pancreatic insufficiency
Enteroinvasive	Bile acid transport defects
Yersinia spp.	Abetalipoproteinemia
Vibrio spp.	**Acquired diarrheas**
Aeromonas spp.	Toddler's diarrhea
Plesiomonas spp.	Short bowel syndrome
Clostridium difficile	Small bowel overgrowth
C. Parasitic pathogens	Antibiotic-associated diarrhea
Giardia lamblia	Münchausen's syndrome
Cryptosporidia	Secondary lactase deficiency
Cyclosporidia	
Entamoeba histolytica	
Nematodes	
Cestodes (tapeworms)	
Trematodes	

Rotavirus is the most common viral pathogen. All children exposed to rotavirus, regardless of whether or not they manifest symptomatic diarrhea, will develop circulating antibodies to this pathogen. The decreasing incidence of rotavirus in adults is thought to be due to the protective effect of these antibodies. Rotaviruses are small, wheel-shaped viruses approximately 70 nm in diameter. Of the four major groups (A, B, C, and D), type A viruses are the most important in children. The virus invades the epithelium and promotes an inflammatory response that ultimately contributes to the destruction of the villous surface. However, the frequency and severity of stool output in these patients does not correlate closely with the degree of intestinal damage observed endoscopically or histologically. This has led to the speculation that there are other pathogenic mechanisms that contribute to the malabsorption and net fluid losses observed in these patients. Although villous destruction can be severe in rotaviral disease, recovery is rapid in most patients and symptoms typically resolve in 2 to 7 days.

Caliciviruses, including the Norwalk and Norwalk-type agents, are the second leading cause of pediatric viral diarrheas. This group of viruses presents in a similar fashion to rotavirus, with the exception that the diarrhea is usually milder. Astroviruses are similar to calciviruses and are a common cause of diarrheal illness. Adenovirus (serotypes 40 and 41) is a well-established cause of viral diarrhea and has a slightly longer incubation period and a longer course than rotaviral disease. More recently, Torovirus has been implicated as a potential cause of diarrhea in children. However, more definitive epidemiologic data concerning this pathogen are currently lacking.

Bacterial infections can also cause diarrheal disease in infants and children (Table I, part B). As in the case of viral diarrhea, the onset of bacterial illness is usually acute and presents with fever and sometimes vomiting. Because the most common forms of bacterial diarrhea are invasive, bloody diarrhea is often reported in these patients. Specific types of bacterial illness have been reported to occur more commonly in specific age groups. *Campylobacter jejuni*, for instance, has a bimodal distribution of onset with the first peak occurring in children from 1 to 5 years old and a second peak in adolescents. Nontyphoid *Salmonella enteritidis* can cause a bacteremia in infants and in

immunocompromised hosts. *Shigella* species can be found in the toddler age group, but is not a commonly isolated pathogen in the United States. *Clostridium difficile*, an important cause of antibiotic-associated diarrhea in adults, is not usually a pathogen in infants. *C. difficile* toxin can be found in up to 10% of healthy newborns and is even more prevalent in neonatal intensive care units. The reason for the inability of this organism to cause diarrhea in infants remains unclear. Based on animal studies, it is thought that the receptor for this toxin is developmentally regulated and absent in early infancy. *Vibrio cholerae* causes a prototypical bacterial secretory diarrhea. It produces a toxin composed of two subunits. The B, or binding, subunit displays a pentameric form that binds selectively to the ganglioside GM_1. The A, or active, subunit is internalized by intestinal epithelia, alters signal transduction, and leads to increased production of cAMP and Cl^- secretion. Other forms of toxin-producing organisms include enterotoxigenic *Escherichia coli*, the pathogen responsible for traveler's diarrhea, and organisms responsible for acute food poisoning such as *Staphylococcus aureus* and *Bacillus cereus*. *E. coli* O157:H7 is an important pathogen in children. This enteropathic *E. coli* adheres to the intestinal lumen and produces a toxin that is absorbed and causes the hemolytic–uremic syndrome.

Parasitic disease causing diarrhea is far less common in industrialized countries (Table I, part C). One notable exception is *Giardia lamblia*, which is especially prevalent in the daycare setting. *Giardia* can present as an acute diarrheal illness or as a more chronic process. The mechanism by which this organism causes diarrhea is not fully understood. There is no gross alteration in intestinal architecture or evidence of a significant immunologic response. There are multiple other parasites that can cause diarrheal disease in children. However, these occur much less commonly and will not be discussed further.

NONINFECTIOUS DIARRHEA IN CHILDREN

Occasionally, a child will present with a diarrheal illness that is not self-limiting. Fever may or may not be present and other co-morbidities, such as growth failure and malnutrition may be prominent. Stool cultures are negative. The etiology of diarrheal disease in these patients can be broadly classified as being inflammatory or noninflammatory in nature, based on clinical history, physical examination, and biochemical workup. Similar to patients with infectious diarrhea, the increased stool output observed in these patients is typically the result of a combination of pathogenic mechanisms.

Inflammatory Diarrhea

The intestine displays a tremendous capacity to generate an immune response based on the presence of numerous effector immunocytes that lie within the intestinal mucosa and submucosa. More recent data have demonstrated that intestinal epithelial cells themselves also possess the ability to process lumenal antigens and present them to the underlying immune cells. The intestinal epithelium is in constant contact with the external environment. It is subsequently in a constant state of low-grade inflammation (often referred to as "physiologic inflammation") that is the result of the epithelium playing its role in the surveillance of and response to the broad array of dietary, microbiologic, and toxigenic stimuli present within the intestinal lumen. When the degree of mucosal inflammation is severe enough to affect the absorptive and secretory function of the intestine, diarrhea ensues.

A number of immune defects or imbalances can affect the intestine (Table I, part D). Inflammatory bowel disease is example of an inflammatory diarrhea that is likely the result of a genetically driven dysregulated immune response to the lumenal environment. It is also likely that genetic predisposition may leave some individuals vulnerable to an exaggerated immune response to dietary antigens that are usually not perceived to be a threat to intestinal function. This may explain the incidence of allergic enteropathies in some children. In patients with celiac disease, or gluten-sensitive enteropathy, there is an immune-mediated response to a protein present in wheat and related grain products. Although these patients can show marked diarrhea, they more commonly present with a failure to thrive precipitated by the introduction of wheat-containing solid foods between 6 and 9 months of age.

Autoimmune disease can target the intestinal epithelium itself and antibodies directed against enterocytes contribute to the severe inflammation and tissue destruction observed histologically in these patients. The IPEX syndrome is an X-linked autoimmune enteropathy that is associated with polyendocrinopathy and results in high morbidity and mortality. The gene defect is thought to lie within the FOXP3 gene and it has been shown to encode the protein scurfin, a regulator of T-cell function in mice. The important role played by lymphocytes in maintaining intestinal barrier function can be appreciated in the context of bone-marrow transplant recipients. Diarrhea is a major feature of graft-versus-host disease, a clinical condition in which

donor lymphocytes recognize host intestinal epithelial cells as being foreign. Activated immunocytes subsequently initiate a destructive process that is manifest histologically as increased epithelial cell apoptosis and clinically as a secretory or inflammatory diarrhea.

Noninflammatory Diarrheas

Children can also suffer from diarrhea that is neither infectious nor inflammatory in nature. These diarrheal illnesses can be broadly categorized into congenital or acquired forms (Table I, part E). Congenital diarrheas are most often the result of abnormal gene expression, resulting in a clinical presentation within the first week of life. Congenital chloride diarrhea is caused by a mutation in the Down-Regulated in Adenoma (DRA) gene, thought to be a colonic chloride transporter. This disease presents uniformly *in utero* with polyhydramnios. Severe diarrhea and abdominal distension appear shortly after birth and profound electrolyte disturbances can occur in these patients if not resuscitated promptly. In contrast, the cause of congenital sodium diarrhea is not known but is thought to be due to a functional uncoupling of sodium and hydrogen exchange in the intestine. No mutations have been described in the known Na^+/H^+ exchangers in the intestine to date. The clinical presentation of congenital sodium diarrhea is similar to congenital chloride diarrhea with the exception that stool chloride levels in these patients are typically lower and the stool pH tends to be more alkaline. In addition to defects in ion transporters, there have been a number of diseases that have been described with altered transport of glucose, galactose, and amino acids. Gastrointestinal symptoms vary from defect to defect. Amino acid transport defects often have extraintestinal manifestations whose consequences far outweigh changes in bowel patterns.

Congenital diarrheas can also be caused by genetic defects that result in the malabsorption of the products of digestion such as carbohydrates and fat. Congenital disaccharidase deficiencies are rare and result in an osmotically driven diarrhea. Much more common are the transient and secondary deficiencies in mucosal disaccharidase levels that result from small intestinal injury or inflammation. Fat malabsorption can also present with diarrhea of variable severity. Congenital fat malabsorption can be the result of pancreatic insufficiency, seen in patients with cystic fibrosis, or due to specific genetic defects such as abetalipoproteinemia. Fat malabsorption is characterized by varying degrees of greasy and malodorous stools. Finally, congenital disorders of the intestinal architecture can lead to diarrhea. Microvillus inclusion disease is a rare autosomal recessive disease that is characterized by severe watery diarrhea at birth. Diagnosis is based on a histologic demonstration of marked or complete villous atrophy and electron microscopic evidence of intracellular microvillus inclusions and absent or rare microvilli.

There are multiple acquired forms of pediatric diarrhea that can be characterized as being noninfectious and noninflammatory in nature. Often, these diarrheas result from a predisposing insult that diminishes the ability of the intestinal mucosa to absorb nutrients, thereby contributing to an osmotic diarrhea. The most common example of this is toddler's diarrhea or chronic nonspecific diarrhea of childhood. There is no underlying inflammatory or biochemical abnormality that drives the increased stool output seen in these young children. In many cases, these patients will respond to a reduced dietary intake of fruit juices. Because many of these juices contain large amounts of sorbitol, an indigestible carbohydrate, they can induce an osmotic diarrhea. As such, the diarrhea will resolve in most patients within a few days after removal of the offending juice. Other examples of acquired and primarily noninflammatory diarrheas that fall into this category include antibiotic-associated diarrhea, short bowel syndrome, and small bowel bacterial overgrowth. Additionally, Münchausen's syndrome-by-proxy must always be considered in children with diarrhea and no predisposing factors.

LABORATORY EVALUATION OF DIARRHEA

Laboratory evaluation of the pediatric patient with diarrhea varies with the suspected cause and is dictated by the clinical picture. Any suspicions about potential inflammation or bacterial infection should be addressed immediately. Evaluation of acute diarrhea is usually limited to cases in which a given patient is presenting with systemic symptoms or co-morbidities. Chronic diarrhea must always be evaluated, especially in the context of poor growth or malnutrition. An evaluation that proceeds in a logical and stepwise manner generally results in the most expedient and cost-effective diagnosis.

The first step in the evaluation process is to determine whether or not the presenting patient's symptoms are most consistent with an inflammatory or noninflammatory process. This can be done by an examination of the stool for gross or occult blood or the presence of fecal leukocytes. Previous studies have also demonstrated the sensitivity and specificity of biochemical assays for fecal lactoferrin, a constituent of neutrophil granules. Patients with infectious or inflammatory diarrhea will

typically present with rectal bleeding or overt (positive fecal leukocyte smear) or biochemical evidence (lactoferrin) of fecal white blood cells. In contrast, these studies should be negative in patients with noninflammatory (viral, osmotic, or secretory) diarrheal disease. Nonetheless, although these markers may increase the yield of sending stool cultures, they do not exclude intestinal inflammation and any final decision about pursuing an infectious workup must be made on clinical grounds.

If there is clinical or biochemical evidence of an inflammatory process, then routine stool cultures remain the gold standard in the search for a bacterial cause of diarrhea. Most hospital-based laboratories have a standard panel of cultures associated with common pathogens including *Campylobacter*, *Shigella*, *Salmonella*, and *Yersinia enterocolitica*. Many hospitals also routinely screen for *E. coli* O157:H7. The identification of some pathogens relies on the detection of a particular toxin that is produced by the bacteria and released into the stool. *C. difficile* is perhaps the best recognized pathogen in this class.

The diagnosis of parasitic disease is most often made by a close microscopic evaluation of the stool for ova and parasites. The identification of *Giardia* and Cryptosporidia has been further facilitated by the development of enzyme-linked immunosorbent assay (ELISA)-based stool tests. It is imperative to know a specific laboratory's capabilities and limitations prior to interpreting the results of any stool, toxin, or parasitic studies.

Most "noninflammatory" diarrheal disease is viral in nature. However, routine evaluation of stool for viral pathogens is not often useful because of the self-limiting nature of the disease process in the vast majority of patients, the specialized nature of obtaining viral cultures, and the expense of detecting specific viral pathogens. One notable exception is the rotavirus stool antigen test. This commercially available ELISA-based test provides relatively rapid results that can assist both in patient care and in making decisions about the need for isolation of hospitalized patients. Other viral stool tests include polymerase chain reaction-based screening for viral DNA in the case of adenovirus. However, these more costly and specialized tests are typically reserved for the evaluation of immunocompromised patients, in whom targeted supportive or antiviral therapy is much more critical.

Characterization of the stool can be helpful for determining the nature of noninflammatory diarrheal illness. Stool evaluation for fat, pH, and reducing substances is important in determining whether or not there is an underlying malabsorptive process. The presence of "neutral" fat in the stool suggests some deficiency in the production or delivery of pancreatic (lipase) or hepatic (bile acid) secretions into the intestinal lumen. An increase in "split" fat in the stool indicates a primary inability of enterocytes to perform fat absorption. Reducing substances are the result of undigested carbohydrates making their way into the large intestine. The presence of these fecal sugars can be readily assessed with commercially available colorimetric strips or test solutions. It must be remembered that sucrose is a nonreducing sugar. As such, stool must first be pretreated with an acid solution to make this nonreducing sugar detectable. Undigested carbohydrates, as well as dietary fiber, are consumed by bacteria in the large bowel and generate short-chain fatty acids. Carbohydrate malabsorption can therefore also be assessed by a fall in stool pH.

Stool electrolytes can help to determine whether or not a diarrheal process is secretory in nature. In general, a stool Na^+ concentration of greater than 70 mEq/liter is indicative of a secretory process. The stool osmotic gap, calculated by

$$([Na^+] + [K^+]) \times 2 - \text{stool osmolarity}$$

$$[Na^+] = \text{concentration of Na}$$

$$[K^+] = \text{concentration of K}$$

is useful in distinguishing between osmotic and secretory diarrheal disease. An osmotic gap greater than 100 mosm/liter suggests an underlying osmotic process. Similarly, whereas osmotic diarrhea will typically respond to a dietary fast, secretory diarrheal diseases are driven by processes that are independent of exogenous (dietary or pharmaceutical) factors.

The ability to study the large and small intestine of patients using videoscopic endoscopy has greatly advanced the ability to diagnose and treat diarrheal disease in pediatric and adult patients. Clinicians are now able to assess the gross appearance of the lining of the small and large intestine, obtain biopsy samples for histologic examination, measure directly mucosal disaccharidase levels, collect pancreatic and biliary secretions, and sample fluid from the small intestine for quantitative culture.

Blood tests can often prove to be useful adjuncts to stool studies. Peripheral eosinophilia may point to an underlying allergic disease. Decreased serum albumin levels can suggest malnutrition or a protein-losing enteropathy. Specialized serum tests such as the detection of antibodies directed against tissue transglutaminase are highly predictive of celiac disease. However, for most patients, blood work plays a

supportive role in the workup of diarrheal disease. Results from serologic studies most often suggest an etiology that will need to be confirmed by more definite stool or endoscopic studies.

TREATMENT OF DIARRHEAL DISEASE

The treatment of pediatric diarrheal disease can be divided into symptomatic and curative therapies. First and foremost in the treatment of any child with diarrhea is a prompt assessment of hydration status. For most cases of mild to moderate diarrhea, oral rehydration solutions are the first line of therapy. When oral intake is limited secondary to an altered mental status or when severe dehydration or shock is present, intravenous replacement of fluid and electrolytes can be lifesaving. Once the patient is adequately hydrated, the diet may be readily advanced. The provision of adequate calories is critical to maintain an anabolic state that will provide the metabolic fuel necessary to promote epithelial restitution. The advantages of enteral supplementation should not be overlooked as lumenal contents have been shown to be trophic to the intestinal epithelium. A transient lactose intolerance may occur in either acute or chronic diarrhea. This can be addressed using soy, rice-based, or lactose-free milk products. High-fructose and sorbitol-containing drinks are palatable, but should be avoided due to the increased osmotic load they place on an already compromised epithelial lining. Other supportive measures that have been used include antisecretory agents, anti-motility agents, and resin binders. These agents decrease overall stool output by slowing intestinal transit. Although clinically beneficial in most cases, clinicians must be wary of the possibility that these agents can contribute to third-spacing of body fluid in distended and pharmacologically atonic intestinal loops.

Specific therapies that are designed to treat the underlying cause of diarrhea can be employed. This includes antibiotic use in certain forms of infectious diarrheas. In general, however, antibiotics should be avoided in patients with diarrheal disease unless there are systemic consequences of the diarrhea, such as that observed with *Salmonella* infections in infants and the elderly. Inappropriate antibiotic use can lead to resistant organisms or prolong the carrier state. Notable exceptions include infectious diarrheas that may become chronic if left untreated, such as diarrhea caused by *C. difficile* and *G. lamblia*.

Other specific therapies for diarrheal disease in pediatric patients include the following: immunosuppression in the immunologically mediated diarrheas such as inflammatory bowel disease or autoimmune enteropathy; specific replacement of electrolytes in the case of the congenital chloride and sodium diarrheas; or enzyme replacement therapy in patients with pancreatic insufficiency or lactose intolerance. Removal of an offending agent, such as gluten-containing foods in celiac disease, lactose in lactase deficiency, or specific dietary antigens in congenital or acquired protein intolerances, can be critical in certain diarrheal illnesses.

SUMMARY

The intestine is a site of competing physiologic processes including salt and water secretion, nutrient absorption, and immune surveillance. Stool output is subsequently the net product of opposing secretory and absorptive capacities that are separated geographically along the length of the intestine as well as along the length of the crypt–villus axis. Any disruption of these tightly regulated homeostatic processes can lead to altered stool formation and the development of pathologic diarrhea. In most cases, these illnesses are self-limited in nature and respond favorably to supportive measures. Nonetheless, pediatric diarrheal diseases remain a significant cause of morbidity and mortality worldwide.

The diagnostic approach to diarrheal disease in children differs substantially from that pursued in other age groups. Consideration must be given to congenital or developmental etiologies not seen in adult populations. Moreover, because children are still growing, the impact of chronic diarrheal processes on linear growth and physical development must also be addressed. Evaluation of diarrheal disease in pediatric patients should proceed in a stepwise fashion that begins with an in-depth clinical history and includes a limited number of microbiologic and biochemical tests. Physicians with a firm grasp of the epidemiology and pathogenesis of diarrheal illness in children will be better positioned to pursue a rational approach to the diagnosis and management of their pediatric patients with these common and potentially debilitating illnesses.

See Also the Following Articles

Anti-Diarrheal Drugs • Bacterial Toxins • Carbohydrate and Lactose Malabsorption • Colitis, Ulcerative (Pediatric) • Diarrhea • Diarrhea, Infectious • Gastroenteritis • Malabsorption • Malnutrition

Further Reading

American Academy of Pediatrics. Provisional Committee on Quality Improvement, Sub-committee on Acute Gastroenteritis (1996). Practice parameter: The management of acute gastroenteritis in young children. *Pediatrics* **97**, 424–435.

Corrigan, J. J., and Boineau, F. G. (2001). Hemolytic–uremic syndrome. *Pediatr. Rev.* **22**, 365–369.
Fuller, C. M., Ji, H. L., Tousson, A., Elble, R. C., Pauli, B. U., and Benos, D. J. (2001). Ca(2+)-activated Cl(−) channels: A newly emerging anion transport family. *Pflug. Arch.* **443**, S107–S110.
Guandalini, S. (2000). Acute diarrhea. *In* "Pediatric Gastrointestinal Disease" (W. A. Walker, P. R. Drurie, J. R. Hamilton, and J. B. Watkins, eds.), pp. 28–38. B. C. Decker, Lewiston, NY.
Jensen, B. S., Strobaek, D., Olesen, S. P., and Christophersen, P. (2001). The Ca^{2+}-activated K^+ channel of intermediate conductance: A molecular target for novel treatments? *Curr. Drug Targets* **2**, 401–422.
Keely, S. J., and Barrett, K. E. (2000). Regulation of chloride secretion: Novel pathways and messengers. *Ann. N. Y. Acad. Sci.* **915**, 67–76.
Ramaswamy, K., and Jacobson, K. (2001). Infectious diarrhea in children. *Gastroenterol. Clin. North Am.* **30**, 611–624.
Rudolph, J. A., and Cohen, M. B. (1999). New causes and treatments for infectious diarrhea in children. *Curr. Gastroenterol. Rep.* **1**, 238–244.
Sandhu, B. K. (2001). Practical guidelines for the management of gastroenteritis in children. *J. Pediatr. Gastroenterol. Nutr.* **33**, S36–S39.
Schroeder, B. C., Waldegger, S., Fehr, S., Bleich, M., Warth, R., Greger, R., and Jentsch, T. J. (2000). A constitutively open potassium channel formed by KCNQ1 and KCNE3. *Nature* **403**, 196–199.
Sellin, J. H. (1993). Intestinal electrolyte absorption and secretion. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Sharshmidt, and M. H. Sleisenger, eds.), pp. 1451–1471. W. B. Saunders, Philadelphia, PA..
Sicherer, S. H. (2002). Food allergy. *Lancet* **360**, 701–710.
Vanderhoof, J. A. (1998). Chronic diarrhea *Pediatr. Rev.* **19**, 418–422.
Velázquez, F. R., Matson, D. O., Guerrero, M. L., Shults, J., *et al.* (2000). Serum antibody as a marker of protection against natural rotavirus infection and disease. *J. Infect. Dis.* **182**, 1602–1609.
Wildin, R. S., Smyk-Pearson, S., and Filipovich, A. H. (2002). Clinical and molecular features of the immunodysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome *J. Med. Genet.* **39**, 537–545.

Diet and Environment, Role in Colon Cancer

MARÍA ELENA MARTÍNEZ AND ELIZABETH T. JACOBS
University of Arizona, Tucson

colorectal adenocarcinoma A malignant neoplasm of the colon or rectum.
colorectal adenoma A benign neoplasm of the colon or rectum with malignant potential.
colorectal polyp A growth protruding from the mucosal surface into the bowel lumen.
dietary fiber The components of plant material that are resistant to digestive secretions produced by humans.

Significant progress has been made over the past decade in identifying factors that modify the risk of colorectal cancer. Large international variation in colorectal cancer incidence and mortality rates, and the prominent increases in the incidence of colorectal cancer in groups that migrated from low to high incidence areas, provided important evidence that lifestyle factors influencethe development of this malignancy. These observations formed the basis for various hypotheses of lifestyle factors in the etiology of colorectal neoplasia. The focus of this article is on the role of diet in relation to colorectal cancer.

FAT AND RED MEAT

Rates of colon cancer are strongly correlated with national per capita disappearance of animal fat and meat, with correlation coefficients ranging between 0.8 and 0.9. A sharp increase in colon cancer incidence rates in Japan in the decades following World War II coincided with a 2.5-fold increase in fat intake. Intake of animal or saturated fat or red meat has been shown to be associated with colon cancer risk; however, some studies do not support these associations. Data from earlier epidemiologic studies provided some evidence for a positive association between dietary fat and increased risk of several cancers, including the colorectum. However, an overall review of studies

 Copyright 2004, Elsevier (USA). All rights reserved.

Corrigan, J. J., and Boineau, F. G. (2001). Hemolytic–uremic syndrome. *Pediatr. Rev.* **22**, 365–369.
Fuller, C. M., Ji, H. L., Tousson, A., Elble, R. C., Pauli, B. U., and Benos, D. J. (2001). Ca(2+)-activated Cl(−) channels: A newly emerging anion transport family. *Pflug. Arch.* **443**, S107–S110.
Guandalini, S. (2000). Acute diarrhea. *In* "Pediatric Gastrointestinal Disease" (W. A. Walker, P. R. Drurie, J. R. Hamilton, and J. B. Watkins, eds.), pp. 28–38. B. C. Decker, Lewiston, NY.
Jensen, B. S., Strobaek, D., Olesen, S. P., and Christophersen, P. (2001). The Ca^{2+}-activated K^+ channel of intermediate conductance: A molecular target for novel treatments? *Curr. Drug Targets* **2**, 401–422.
Keely, S. J., and Barrett, K. E. (2000). Regulation of chloride secretion: Novel pathways and messengers. *Ann. N. Y. Acad. Sci.* **915**, 67–76.
Ramaswamy, K., and Jacobson, K. (2001). Infectious diarrhea in children. *Gastroenterol. Clin. North Am.* **30**, 611–624.
Rudolph, J. A., and Cohen, M. B. (1999). New causes and treatments for infectious diarrhea in children. *Curr. Gastroenterol. Rep.* **1**, 238–244.
Sandhu, B. K. (2001). Practical guidelines for the management of gastroenteritis in children. *J. Pediatr. Gastroenterol. Nutr.* **33**, S36–S39.
Schroeder, B. C., Waldegger, S., Fehr, S., Bleich, M., Warth, R., Greger, R., and Jentsch, T. J. (2000). A constitutively open potassium channel formed by KCNQ1 and KCNE3. *Nature* **403**, 196–199.
Sellin, J. H. (1993). Intestinal electrolyte absorption and secretion. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Sharshmidt, and M. H. Sleisenger, eds.), pp. 1451–1471. W. B. Saunders, Philadelphia, PA..
Sicherer, S. H. (2002). Food allergy. *Lancet* **360**, 701–710.
Vanderhoof, J. A. (1998). Chronic diarrhea *Pediatr. Rev.* **19**, 418–422.
Velázquez, F. R., Matson, D. O., Guerrero, M. L., Shults, J., *et al.* (2000). Serum antibody as a marker of protection against natural rotavirus infection and disease. *J. Infect. Dis.* **182**, 1602–1609.
Wildin, R. S., Smyk-Pearson, S., and Filipovich, A. H. (2002). Clinical and molecular features of the immunodysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome *J. Med. Genet.* **39**, 537–545.

Diet and Environment, Role in Colon Cancer

MARÍA ELENA MARTÍNEZ AND ELIZABETH T. JACOBS
University of Arizona, Tucson

colorectal adenocarcinoma A malignant neoplasm of the colon or rectum.
colorectal adenoma A benign neoplasm of the colon or rectum with malignant potential.
colorectal polyp A growth protruding from the mucosal surface into the bowel lumen.
dietary fiber The components of plant material that are resistant to digestive secretions produced by humans.

Significant progress has been made over the past decade in identifying factors that modify the risk of colorectal cancer. Large international variation in colorectal cancer incidence and mortality rates, and the prominent increases in the incidence of colorectal cancer in groups that migrated from low to high incidence areas, provided important evidence that lifestyle factors influencethe development of this malignancy. These observations formed the basis for various hypotheses of lifestyle factors in the etiology of colorectal neoplasia. The focus of this article is on the role of diet in relation to colorectal cancer.

FAT AND RED MEAT

Rates of colon cancer are strongly correlated with national per capita disappearance of animal fat and meat, with correlation coefficients ranging between 0.8 and 0.9. A sharp increase in colon cancer incidence rates in Japan in the decades following World War II coincided with a 2.5-fold increase in fat intake. Intake of animal or saturated fat or red meat has been shown to be associated with colon cancer risk; however, some studies do not support these associations. Data from earlier epidemiologic studies provided some evidence for a positive association between dietary fat and increased risk of several cancers, including the colorectum. However, an overall review of studies

 Copyright 2004, Elsevier (USA). All rights reserved.

recently published indicates that there is no strong evidence for the association of colon cancer with high fat consumption per se and that the association may partly be explained by animal fat or red meat consumption. In addition, whether this association is independent of total energy intake is unclear. Results of combined data from 13 case-control studies of colon cancer showed a positive association between energy intake and no association for various fat components in the diet independent of energy intake.

Results of some prospective studies of colon cancer have shown positive associations with fat or red meat consumption but the data are less compelling for total fat than for red meat. Other cohort studies have shown statistically significant or suggestive positive associations for intake of processed meats and risk of colon cancer. Although published reports support a possible or probable increased risk of colon cancer in relation to red meat, an expert workshop convened in Australia did not support this association.

FIBER, FRUIT, AND VEGETABLES

High consumption of fruit and vegetables has been shown to be associated with a decreased risk of colorectal neoplasia, particularly for vegetable consumption. Foods high in fiber have also been shown to be inversely associated with colon cancer risk in most, but not all studies. Conversely, results of large prospective studies have shown weak or nonexistent inverse associations for fiber intake and risk of colon cancer. As with case-control studies, when sources of fiber were examined separately, a reduced risk appears to be stronger for vegetable sources than for other fiber components. However, in a large prospective study examining the role of fiber on risk of colorectal neoplasia in female nurses, no association was observed between colorectal cancer and fiber intake. Additionally, when data from the Nurses' Health Study and the Health Professionals Follow-up Study cohorts were combined for a follow-up of over 1,700,000 person-years to yield 937 cases of colon and 244 of rectal cancer, no protective effect was observed for fruit and vegetable intake and colorectal cancer.

Given the scientific and public health interest related to fiber and fruit and vegetable consumption and colorectal cancer risk, two separate, large U. S. trials were conducted to test these associations. The interventions included a diet high in fiber, high in fruit and vegetables, and low in fat versus a usual diet and a high (13.5 g/day) versus low (2.0 g/day) wheat bran fiber intervention. No statistically significant difference in adenoma recurrence rate was shown for either intervention.

MICRONUTRIENTS

Calcium

The role of calcium in colorectal neoplasia has been investigated in a variety of study settings including animal studies, international correlational studies, case-control and cohort studies, and intervention studies of adenoma recurrence. It is hypothesized that calcium might reduce colon cancer risk by binding secondary bile acids and ionized fatty acids to form insoluble soaps in the lumen of the colon; this in turn reduces the proliferative stimulus of these compounds on colon mucosa.

Results of analytic epidemiological studies that have examined calcium as a risk factor for colorectal cancer have been inconsistent. Data from large cohort studies show weak, nonsignificant inverse associations with no evidence of a dose–response relationship. Results from the Nurses' Health Study, in which data from three dietary questionnaires were collected prospectively over 6 years, did not support a major inverse association between calcium intake and risk of colorectal cancer over a 6-year period. The modest effect of calcium intake on risk of colorectal neoplasia observed in epidemiologic studies is consistent with findings from a recent adenoma recurrence intervention trial. In this trial, calcium supplementation (1200 mg of elemental calcium versus placebo) among 913 participants who underwent adenoma removal was associated with a statistically significant reduction in the risk of adenoma recurrence.

Folic Acid

In addition to animal data, an increasing epidemiologic body of evidence shows a potential role for folate in reducing the risk for colorectal cancer. An additional study using plasma folate also supports this finding. However, data from a large case-control study (1993 cases and 2410 controls) showed no important inverse association between dietary folate and colon cancer risk. Perhaps the strongest evidence is derived from the Nurses' Health Study cohort, where increased consumption of supplemental folic acid, after a period of 15 or more years, was associated with a 75% reduction in risk of colon cancer. In addition, studies of colorectal adenomas support a protective effect of folic acid in colorectal neoplasia.

Further evidence of a role for folate is that inherited variation in the activity of methylenetetrahydrofolate reductase (MTHFR), a critical enzyme in the production of the form of folate that supplies the methyl group for methionine synthesis, influences the risk of colon cancer. In this proposed pathway, key nutrient and nonnutrient components are involved and different endogenous forms of folate, 5-methyltetrahydrofolate and 5,10-methylenetetrahydrofolate, are essential for DNA methylation and DNA synthesis, respectively. DNA hypomethylation is among the earliest events observed in colon carcinogenesis; however, it is unclear whether this process directly influences the carcinogenic process. Additional micronutrients involved in the DNA methylation process include vitamins B6 and B12. Furthermore, since alcohol is known to influence folate metabolism and methyl group availability, its interaction with the key micronutrients must be considered in this process.

Other Micronutrients

Several additional micronutrients have been implicated in relation to colorectal neoplasia; however, the evidence for their specific role is unclear. Among the proposed micronutrients, those with antioxidant potential are thought to be important in protection against the development of colorectal cancer. These include β-carotene, selenium, and vitamins C and E and have been shown to be inversely associated with colon cancer risk, perhaps through their effects on cell proliferation. Among participants in the American Cancer Society's Cancer Prevention Study II cohort, regular use of vitamin C or E supplements was not associated with colorectal cancer mortality.

Selenium, an essential trace mineral found in cereal grains and seafood, has been shown to be inversely related to colorectal cancer. Correlational data show higher cancer mortality rates in low-selenium areas compared to those of high-selenium regions. Perhaps the most provocative finding to date derives from secondary analyses of the Nutritional Prevention of Skin Cancer study, where a greater than 50% reduction in colorectal cancer incidence was shown with a selenium intervention of 200 μg/day compared with placebo. When these results were updated to include an additional 3 years of participant follow-up, a similar magnitude of reduction was observed, with a Hazard Ratio of 0.46 (95% CI = 0.21−1.02; $P = 0.057$). Since the results were based on secondary end-point data among a population in selenium-deficient areas of the United States, additional large trials will be needed.

To test the antioxidant hypothesis on colorectal cancer etiology, a randomized controlled trial with adenoma recurrence was conducted. Participants were randomized to one of four arms: 25 mg of β-carotene; 1 g of vitamin C and 400 mg of vitamin E; 25 mg of β-carotene, 1 g of vitamin C, and 400 mg vitamin E; or placebo. Recurrence rates after 4 years of follow-up between the placebo group and the intervention groups were similar, suggesting a lack of effect of chemopreventive properties by these antioxidant nutrients. In addition, results of a secondary analysis of the Alpha-Tocopherol, Beta-Carotene Prevention (ATBC) Study showed no significant protective effect of β-carotene, α-tocopherol, or both nutrients combined, on colorectal cancer.

PHYSICAL ACTIVITY

Although not a dietary factor per se, a sedentary lifestyle is linked to obesity and overall caloric imbalance. Results of prospective and retrospective studies support an inverse association between physical activity and risk of colon cancer, but not rectal cancer. The results are consistent across studies, in men and women, and whether assessing occupational or leisure-time activity. In a study that assessed the joint effect of physical activity and body mass index (BMI), the highest risk of colon cancer was shown among those both physically inactive and with high BMI levels.

GENE−NUTRIENT INTERACTIONS

Major advances in characterization of new genomes have provided tremendous excitement in the scientific community and the area of cancer prevention is no exception. Low-penetrance susceptibility genes occur commonly and lead to sporadic disease; these susceptibility genes aggregate with disease and may interact with environmental factors or other genes to increase the risk of cancer. Genetic polymorphisms, defined as changes in the nucleotide sequence (mutations) that are present in at least 1% of the population, have been explored in the recent literature in relation to cancer etiology, including colorectal cancer. Examples from recent reports include markers of folate status and mutations in the *MTHFR* gene, well-cooked meat (a source of heterocyclic amines), and polymorphisms in the cytochrome P450 genes as well as the *N*-acetyltransferase genes 1 and 2, and cruciferous vegetable consumption (source of dietary isothiocyanates) and polymorphisms in the glutathione-*S*-transferase family of genes. This

complex area of research continues to evolve; however, proposed complex interactions can be adequately addressed only in the setting of large epidemiological studies with sufficient statistical power to generate meaningful results.

DIETARY AND LIFESTYLE GUIDELINES

Dietary Guidelines from the American Cancer Society are based on the mass of the published data (Table I). Of these guidelines, the most consistent support for any one variable from the scientific literature is found for physical activity and lower risk of colon cancer. Although these guidelines target cancer prevention, they have a great deal in common with guidelines aimed at the prevention of other chronic diseases. Recommendations from a large, international panel of experts were similar to those of the American Cancer Society.

SUMMARY AND FUTURE CHALLENGES

The past decade has been filled with a plethora of literature related to the primary prevention of colorectal cancer. Although the precise mechanisms have not been clarified, several lifestyle factors are likely to have a major impact on colorectal cancer development. Physical inactivity and, to a lesser extent, excess body weight are consistent risk factors for colon cancer. Although not addressed in this article, exposure to tobacco products early in life is associated with a higher risk of developing colorectal neoplasia. Diet and nutritional factors are also clearly important. Excess alcohol consumption, probably in combination with a diet low in some micronutrients such as folate and methionine, appears to increase risk. Diets high in red meat may also increase risk. The role of fiber, although recently challenged, continues to be supported, given the overall general findings from numerous studies. Furthermore, additional support on the role of folic acid awaits the results of ongoing intervention trials of adenoma recurrence.

TABLE I American Cancer Society Guidelines on Diet, Nutrition, and Cancer Prevention

Choose most of the foods you eat from plant sources:
Eat five or more servings of fruit and vegetables each day
Eat other foods from plant sources, such as breads, cereals, grain products, rice, pasta, or beans several times each day
Limit your intake of high-fat foods, particularly from animal sources
Choose foods low in fat
Limit consumption of meats, especially high-fat meats
Be physically active: Achieve and maintain a healthy weight
Be at least moderately active for 30 min or more on most days of the week
Stay within your healthy weight range
Limit consumption of alcoholic beverages, if you drink at all

See Also the Following Articles

Calcium, Magnesium, and Vitamin D Absorption, Metabolism, and Deficiency • Cancer, Overview • Colorectal Adenocarcinoma • Colorectal Adenomas • Dietary Fiber • Dietary Reference Intakes (DRI): Concepts and Implementation • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Alberts, D. S., Martinez, M. E., Roe, D. J., Guillen-Rodriguez, J. M., Marshall, J. R., van Leeuwen, B., Reid, M. E., Ritenbaugh, C., Vargas, P. A., Bhattacharyya, A. B., Earnest, D. L., Sampliner, R. E., Parish, D., Koonce, K., and Fales, L. (2000). Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. *N. Engl. J. Med.* **342**, 1156–1162.

Armstrong, B., and Doll, R. (1975). Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int. J. Cancer* **15**, 617–631.

Baron, J. A., Beach, M., Mandel, J. S., van Stolk, R. U., Haile, R. W., Sandler, M. D., Rothstein, R., Summers, R. W., Snover, D. D., Beck, G. J., Bond, J. H., and Greenberg, E. R., for the Calcium Polyp Prevention Study Group (1999). Calcium supplements for the prevention of colorectal adenomas. *N. Engl. J. Med.* **340**, 101–107.

Byers, T., Nestle, M., McTiernan, A., Doyle, C., Currie-Williams, A., Gansler, T., Thun, M., and Committee, American Cancer Society 2001 Nutrition and Physical Activity Guidelines Advisory (2002). American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing risk of cancer with healthy food choices and physical activity. *CA Cancer J. Clin.* **52**, 92–119.

Chief Medical Officer's Committee on Medical Aspects of Food (1998). "Nutritional Aspects of the Development of Cancer," No. 48, Her Majesty's Stationery Office, London.

Clark, L. C., Combs, G. F., Turnbull, B. W., Slate, E. H., Chalker, D. K., Chow, J., *et al.* (1996). Effects of selenium supplementation for cancer prevention for cancer prevention in patients with carcinoma of the skin: A randomized controlled trial. *J. Am. Med. Assoc.* **276**, 1957–1963.

Fuchs, C. S., Giovannucci, E. L., Colditz, G. A., Hunter, D. J., Stampfer, M. J., Rosner, B., Speizer, F. E., and Willett, W. C. (1999). Dietary fiber and the risk of colorectal cancer and adenoma in women. *N. Engl. J. Med.* **340**, 169–176.

Martínez, M. E., and Willett, W. C. (1998). Calcium, vitamin D, and colorectal cancer: A review of the epidemiologic evidence. *Cancer Epidemiol. Biomarkers Prev.* **7**, 163–168.

Rose, D. P., Boyar, A. P., and Wynder, E. L. (1986). International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. *Cancer* **58**, 2263–2271.

Schatzkin, A., Lanza, E., Corle, D., Lance, P., Iber, F., Caan, B., Shike, M., Weissfeld, J., Burt, R. W., Cooper, M. R.,

Kikendall, J. W., and Cahill, J. (2000). Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. *N. Engl. J. Med.* **342**, 1149–1155.

Thune, I., and Furberg, A. S. (2001). Physical activity and cancer risk: Dose–response and cancer, all sites and site-specific. *Med. Sci. Sports Exerc.* **33**, S530–S550.

Willett, W. C. (2001). Diet and cancer: One view at the start of the millennium. *Cancer Epidemiol. Biomarkers Prev.* **10**, 3–8.

Willett, W. C. (1998). Dietary fat intake and cancer risk: A controversial and instructive story. *Semin. Cancer Biol.* **8**, 245–253.

World Cancer Research Fund (1997). "Food, Nutrition and the Prevention of Cancer: A Global Perspective," pp. 497–500.

Dietary Fiber

DAVID N. MOSKOVITZ AND YOUNG-IN KIM
University of Toronto and St. Michael's Hospital

adenoma A benign epithelial tumor in which the cells form recognizable glandular structures.

colorectal adenocarcinoma A malignant new growth confined to the colon or rectum, made up of epithelial cells tending to infiltrate the surrounding tissue and giving rise to metastases.

constipation A bowel movement every 3 to 4 days or less with the presence of straining, hard stools, feelings of incomplete evacuation, or two or fewer bowel movements per week.

dietary fiber The components of plant material that are resistant to digestive secretions produced by humans.

diverticulitis Inflammation of a diverticulum that may undergo perforation with abscess formation.

diverticulosis Gastrointestinal disease characterized by the presence of diverticula (pouches or sacs created by the herniation of the lining mucous membrane through a defect in the muscular coat of a tubular organ).

inflammatory bowel disease Consists of ulcerative colitis and Crohn's disease, both of which are chronic, relapsing disorders of unknown etiology affecting primarily the gastrointestinal tract.

irritable bowel syndrome A functional bowel disorder in which abdominal pain is associated with defecation or a change in bowel habit, with features of disordered defecation and distension.

resistant starch That portion of ingested starch that escapes digestion in the small intestine.

short-chain fatty acids Organic acids produced by anaerobic fermentation of undigested carbohydrates within the colonic lumen.

Dietary fiber consists of the components of plant material that are resistant to digestive secretions produced by humans. There is no internationally accepted definition or method for determining the dietary fiber content of foods. The roles that dietary fiber plays in a number of gastrointestinal diseases are examined in this article.

DIETARY FIBER

The U. S. Expert Panel on Dietary Fiber defined dietary fiber as the endogenous components of plant materials in the diet that are resistant to digestion by enzymes produced by humans. The components consist of nonstarch polysaccharides and lignin as well as other associated substances (Table I). The polysaccharide constituents of fiber are composed of cellulose, β-glucans, hemicelluloses, pectin, gums and mucilages, and seaweed extract. These components themselves are made up of primarily glucose, galactose, and mannose chains as well as secondary xylose, fucose, and galactiose chains. Depending on its solubility in water and buffer solution, dietary fiber can be further analytically classified as soluble (some hemicelluloses, pectins, gums, and mucilages) or insoluble (most hemicelluloses, celluloses, and lignins). When the effect of dietary fiber on the colon is being considered,

 Copyright 2004, Elsevier (USA). All rights reserved.

Kikendall, J. W., and Cahill, J. (2000). Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. *N. Engl. J. Med.* **342**, 1149–1155.

Thune, I., and Furberg, A. S. (2001). Physical activity and cancer risk: Dose–response and cancer, all sites and site-specific. *Med. Sci. Sports Exerc.* **33**, S530–S550.

Willett, W. C. (2001). Diet and cancer: One view at the start of the millennium. *Cancer Epidemiol. Biomarkers Prev.* **10**, 3–8.

Willett, W. C. (1998). Dietary fat intake and cancer risk: A controversial and instructive story. *Semin. Cancer Biol.* **8**, 245–253.

World Cancer Research Fund (1997). "Food, Nutrition and the Prevention of Cancer: A Global Perspective," pp. 497–500.

Dietary Fiber

DAVID N. MOSKOVITZ AND YOUNG-IN KIM
University of Toronto and St. Michael's Hospital

adenoma A benign epithelial tumor in which the cells form recognizable glandular structures.

colorectal adenocarcinoma A malignant new growth confined to the colon or rectum, made up of epithelial cells tending to infiltrate the surrounding tissue and giving rise to metastases.

constipation A bowel movement every 3 to 4 days or less with the presence of straining, hard stools, feelings of incomplete evacuation, or two or fewer bowel movements per week.

dietary fiber The components of plant material that are resistant to digestive secretions produced by humans.

diverticulitis Inflammation of a diverticulum that may undergo perforation with abscess formation.

diverticulosis Gastrointestinal disease characterized by the presence of diverticula (pouches or sacs created by the herniation of the lining mucous membrane through a defect in the muscular coat of a tubular organ).

inflammatory bowel disease Consists of ulcerative colitis and Crohn's disease, both of which are chronic, relapsing disorders of unknown etiology affecting primarily the gastrointestinal tract.

irritable bowel syndrome A functional bowel disorder in which abdominal pain is associated with defecation or a change in bowel habit, with features of disordered defecation and distension.

resistant starch That portion of ingested starch that escapes digestion in the small intestine.

short-chain fatty acids Organic acids produced by anaerobic fermentation of undigested carbohydrates within the colonic lumen.

Dietary fiber consists of the components of plant material that are resistant to digestive secretions produced by humans. There is no internationally accepted definition or method for determining the dietary fiber content of foods. The roles that dietary fiber plays in a number of gastrointestinal diseases are examined in this article.

DIETARY FIBER

The U. S. Expert Panel on Dietary Fiber defined dietary fiber as the endogenous components of plant materials in the diet that are resistant to digestion by enzymes produced by humans. The components consist of nonstarch polysaccharides and lignin as well as other associated substances (Table I). The polysaccharide constituents of fiber are composed of cellulose, β-glucans, hemicelluloses, pectin, gums and mucilages, and seaweed extract. These components themselves are made up of primarily glucose, galactose, and mannose chains as well as secondary xylose, fucose, and galactiose chains. Depending on its solubility in water and buffer solution, dietary fiber can be further analytically classified as soluble (some hemicelluloses, pectins, gums, and mucilages) or insoluble (most hemicelluloses, celluloses, and lignins). When the effect of dietary fiber on the colon is being considered,

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Dietary Fiber

Classification
Nonstarch polysaccharides
Celluloses
Noncelluloses: hemicelluloses, pectins, gums, mucilages
Nonpolysaccharides: lignins
Classification based on solubility
Soluble: pectins, gums, mucilages, hemicelluloses
Insoluble: celluloses, lignins, hemicelluloses
Minor components
Phylates, cutins, saponins, lectins, protein, waxes, silicon
Related components
Resistant starch and protein
Lignans

the classification of fiber as fermentable (i.e., metabolized by colonic bacteria) or nonfermentable is also useful. Some starch escapes small bowel digestion and reaches the colon; this starch is referred to as "resistant starch." There are three categories of resistant starch: physically enclosed starch, ungelatinized crystallite granules, and retrograded amylose. Resistant starch escapes digestion in the small bowel, but enters the colon, where it can be fermented by bacteria. As a result, resistant starch is similar to dietary fiber. Resistant starch represents 2 to 5% of the average starch ingested in the Western diet.

Analytical methods that estimate the dietary fiber content of foods exist. These assays likely underestimate the actual dietary fiber content in foods. The methods are based on the assumption that all starch is digested in the small bowel and that other complex carbohydrates are completely undegraded; dietary fiber is therefore considered to include all plant polysaccharides except starch and nonpolysaccharides. Currently available assays that account for resistant starch estimate the amount of polysaccharides that reach the colon to be in the range of 15 to 25 g/day. Even with the inclusion of resistant starch in the assessment of dietary fiber intake, currently available assays account for less than one-third of total dietary fiber that reaches the colon. Therefore, the assays that are currently available to estimate nonstarch polysaccharides and resistant starch do not accurately measure dietary intake and underestimate the amount of dietary fiber that reaches the colon.

Dietary fiber is found mainly in legumes, vegetables, whole grain cereals, nuts, and fruits. In the Third National Health and Nutrition Examination Survey (NHANES III, 1988 to 1994), mean dietary fiber intake in the U. S. adult population (>19 years old) is 17 g/day for males and 13 g/day for females. The median intakes of dietary fiber for adults were 12−14 g, indicating that a large portion of adults do not meet the currently recommended amounts for good health.

Short-chain fatty acids (SCFA) are organic acids produced by anaerobic fermentation of undigested carbohydrates within the colon. Acetate, propionate, and butyrate account for 90−95% of SCFA in the colon, with isobutyrate, valerate, iosovalerate, and caproate constituting the rest. Once inside the colonocyte, SCFA are an important energy source for the cell. Butyrate is the preferred SCFA to meet colonic energy requirements. The SCFA could account for up to 80% of the energy requirements of the colon and for 5−10% of total body energy requirements. Butyrate is the preferred SCFA to meet colonic energy requirements and is the best energy source for the colon.

Different types of dietary fiber exert their physiological effects in various parts of the digestive tract. The effect of dietary fiber on the small intestine is to slow transit and regulate digestion. Fiber also influences viscosity, water-binding capacity, and cation exchange in the small intestine. Most of the action of dietary fiber takes place in the colon. Approximately 40−95% of fiber is fermented by intestinal flora, mainly anaerobic bacteria. The fermentation products have physiological effects. Fermentable products provide substrates for bacterial growth, which contributes to fecal bulk. Acids produced by bacteria acidify the colon, modify lipid metabolism, and affect mucosal growth. Large-fiber products provide a physical resistance against water absorption, resulting in decreased fecal density, which prevents impaction and constipation. The interest in dietary fiber is due to its role in various diseases. In 1975, Trowel and Burkitt proposed that a low-fiber diet was associated with a number of diseases including colon cancer, diabetes, gallstones, obesity, and diverticular disease. Based on the evidence of the effect of fiber on health, it is recommended that the daily intake of carbohydrate should consist of 55−60% of the daily energy intake from a variety of sources with an increase in the form of complex carbohydrates. Any change in dietary intake of fiber should be gradual.

COLORECTAL CANCER

Introduction

It is estimated that 1,284,900 new cancers will be diagnosed in 2002 and that 550,000 Americans will die of cancer this year. There were 148,300 estimated

new cases of colorectal cancer (CRC) (72,600 male; 75,700 female) in 2002. It is estimated that 56,600 people will die from CRC in 2002. CRC is the third most common cancer and the second most common cause of cancer death after lung cancer in the United States. CRC results from an interaction between environmental and genetic factors. Dietary and lifestyle factors are among the most important environmental factors implicated. It has been estimated that 35% (10 to 70%) of all cancers are attributable to diet and that 50 to 75% of CRC in the United States may be preventable through dietary modifications. Dietary factors implicated in colorectal carcinogenesis include consumption of red meat, animal and saturated fat, refined carbohydrates, and alcohol, as well as total caloric (energy) intake. The intake of dietary fiber, vegetables, fruits, antioxidant vitamins, calcium, and folate is negatively associated with the development of CRC. Epidemiological studies suggest that the risk of developing CRC is related to overall diet and lifestyle patterns. However, because of inherent limitations associated with study design, a cause-and-effect relationship between dietary or other lifestyle factors and CRC is difficult to establish. Epidemiological, animal, and interventional studies examining this relationship often produced conflicting results. Observations that suggest that overall diet and lifestyle, rather than individual factors, play the more important role underscore the importance of as yet undetermined interactions among dietary components and lifestyle factors in the development of CRC.

Dietary fiber is one of several factors whose role in colorectal carcinogenesis has been extensively studied. The relationship between fiber intake and CRC risk, however, has not been clearly established.

Evidence

The inverse relationship between dietary fiber intake and the risk of CRC has been investigated by four types of human epidemiological studies: correlation (or ecological) studies, case-control studies, prospective studies, and interventional trials.

Correlation Studies

Correlation studies examine the relationship between the per capita consumption of a dietary factor and the prevalence, incidence, or mortality of CRC in the population. Of 28 published international, within-country correlation studies of CRC and fiber, vegetables, grains, fruits, and cereals, 23 (82%) showed either a strong or a moderate protective effect of dietary fiber on the development of CRC. Four studies found no evidence for a protective effect of fiber. One study showed that a significant excess risk of CRC was associated with high intake of fiber-rich foods. Limitations exist when interpreting data generated from correlation studies. Studies are based on intake of crude fiber, which greatly underestimates total dietary fiber levels; correlation studies fail to correct for unmeasured confounding factors that may be responsible for the observed association and they do not control for other dietary variables.

Case-Control Studies

Case-control studies compare prior consumption of a dietary factor in subjects with CRC and matched control subjects without CRC. Limitation in retrospective studies is the accuracy with which intake of dietary factors or supplementation can be established: individuals may misreport their habitual past diets. Furthermore, control individuals often have another disease that may be related to diet. Another problem associated with case-control studies is selection bias because of the absence of patients who do not survive long enough to be enrolled in the study. Three analyses have critically evaluated the bulk of case-control studies that address the role of dietary fiber in CRC. In a meta-analysis by Trock *et al*., 15 (65%) of 23 studies demonstrated either a strong or moderate protective effect of dietary fiber and vegetables. Only 2 studies (9%) lacked support for a protective effect of fiber. With fiber-rich diets, a 43% reduction in CRC risk was observed [odds ratio (OR), 0.57; 95% confidence interval (CI), 0.50−0.64] when the highest and lowest quartiles of intake were compared. Howe *et al*. performed a combined analysis of data from 13 case-control studies. The individual data records for 5287 case subjects and 10,470 control subjects were pooled for a common analysis. The risk of CRC was shown to decrease incrementally as dietary fiber intake increased. Consumption of more than 31 g of fiber per day was associated with a 47% reduction in the risk of CRC compared with diets incorporating less than 10 g of fiber per day (95% CI, 0.47−0.61). When all of the studies were combined and adjusted for total energy intake, age, and sex, individuals who consumed 27 g fiber per day had a 50% reduction in the risk of developing CRC compared with those who consumed less than 11 g fiber per day [relative risk (RR), 0.51; 95% CI, 0.44−0.59]. Friedenreich *et al*. examined the study design features and data collection methods from the 13 case-control studies to determine whether they influenced the results obtained from a pooled analysis. Their results show that subjects consuming >27 g fiber per day had a 50% reduction

in the risk of developing CRC compared with those taking <11 g fiber per day (OR, 0.49; 95% CI, 0.37−0.65).

Several case-control studies have investigated the relationship between dietary fiber or fiber-rich foods and the risk of colonic adenomas, well-established precursors of adenocarcinoma. Data suggest that there exists an inverse relationship between fiber intake and the development of colonic adenomas. The magnitude of the reduction in the risk ranged from 10 to 60%. Some studies show a dose-dependent inverse association between colorectal adenoma risk and dietary intake of fiber but other studies suggest that the protective effect associated with dietary fiber is evident in women or for large (>1 cm) adenomas.

In summary, most of the published case-control studies show either a strong or a moderate protective effect of dietary fiber, consistent with the fiber hypothesis. Studies conducted in meta-analysis format provide strong support for the protective effect of dietary fiber on colorectal carcinogenesis. The strongest argument for the fiber hypothesis that can be made from case-control studies is the protective effect of dietary fiber among studies conducted in populations with different patterns of diet and CRC. The combined analysis and meta-analyses of case-control studies suggest, on average, a 50% reduction in the risk of developing CRC in individuals with the highest dietary fiber intake compared with those with the lowest fiber intake. Most case-control studies show a significant inverse dose-dependent relationship between dietary fiber intake and the risk of CRC and colorectal adenomas. Several shortcomings associated with case-control studies limit the interpretation of the results. Limitations associated with analytical methods of determining fiber content in diet, the effect of other potential anti-carcinogens present in fiber-rich foods, and questionnaires that have not been validated are all factors affecting study results.

Prospective Studies

Prospective studies assess the diets of a large group of healthy individuals and include follow-up over time. Prospective studies can control and correct confounding factors more adequately than correlation and case-control studies. They provide the opportunity to obtain repeated assessments of diet at regular intervals. Earlier prospective studies investigated the relationship between dietary fiber intake and CRC mortality. One study from Japan investigated the relationship between diet and other lifestyle variables and major causes of deaths in 265 subjects, ages ≥ 40 years, followed up for 13 years. This study observed that CRC mortality rate decreased as rice and wheat consumption increased with an RR of 0.6 in those with the highest intake ($>720\,cm^3$ of rice and wheat per day) compared with those with the lowest intake ($<180\,cm^3$/day). A Dutch study with a follow-up of 13 years involving 871 middle-aged men showed a threefold reduction in cancer mortality in men in the highest quintile of dietary fiber intake compared with men in the lowest quintile. Another study involving 25,493 white California Seventh Day Adventists followed up for 21 years showed no protective effect of cereal or green salad intake on CRC mortality. Data from Cancer Prevention Study II, an ongoing prospective mortality study, support the protective role of dietary fiber in colorectal carcinogenesis. Multivariate analyses showed that the risk of fatal colon cancer decreased with more frequent consumption of vegetables and high-fiber grains (P trend $= 0.031$ in men and 0.0012 in women).

More recently, several well-designed prospective studies have examined the relationship between dietary fiber intake and the risk of CRC and adenomas; the findings, however, are not consistent. In the Nurses' Health Study, 121,700 female registered nurses between 34 and 59 years of age in the United States completed a mailed questionnaire on known and suspected risk factors for cancer. Every 2 years thereafter, follow-up questionnaires were sent to identify new cases of cancer. A subgroup of these women show that energy-adjusted intakes of crude and total dietary fiber were both inversely associated with the risk of colon cancer. These results were not statistically significant. Only fiber from fruit was associated with any appreciable reduction in risk. This trial, extended for 16 years of follow-up, identified 787 new cases of CRC among the 88,757 eligible women. Total dietary fiber intake was not significantly associated with the incidence of CRC. The relative risk for the quintile group with the highest (median 24.9 g/day) compared with the lowest (median 9.8 g/day) total dietary fiber intake was 0.95 (95% CI, 0.73 to 1.25) and no dose-dependent inverse association was observed (P trend $= 0.59$).

The Health Professional Follow-up Study is a prospective study of heart disease and cancer among 51,529 U. S. male health professionals between the ages of 40 and 75 years. Among 47,949 men who were free of diagnosed cancer in 1986, 205 new cases of colon cancer were diagnosed and confirmed between 1986 and 1992. Age, family history of CRC, obesity, physical activity, cigarette use, alcohol consumption, and other confounding factors were adjusted for analysis. No clear association between total dietary fiber intake and risk

of colon cancer was observed. In this study, however, dietary fiber was inversely associated with risk of adenoma (P trend <0.0001); RR for men in the highest (>28.3 g/day) vs the lowest (<16.6 g/day) quintile was 0.36 (95% CI, 0.22−0.60).

The Iowa Women's Health Study included 98,030 postmenopausal women ages 55 to 69 years. A statistically nonsignificant inverse association was observed between dietary fiber intake and the risk of colon cancer.

In summary, published large prospective studies have produced equivocal findings. Data from earlier prospective studies examining the relationship between dietary fiber intake and CRC mortality proved to be inconsistent. Recent large prospective studies show a significant inverse relationship with a 30% reduction of CRC mortality in subjects consuming the highest amount of dietary fiber compared with those consuming the lowest amount. Published prospective studies of the relationship between dietary fiber intake and the risk of CRC or adenomas have demonstrated a protective effect of dietary fiber against distal colon and rectal adenomas in men. The quality of recent studies is impressive. Recent prospective studies involved a large number of patients and controlled for confounding factors. Limitations exist in that some studies failed to take into account changing dietary fiber patterns. In such cases, it is impossible to delineate the period of exposure to dietary fiber and the risk of developing CRC. The imprecise estimation of dietary fiber limits study results. Moreover, fiber values assigned to individual foods have errors. Finally, the applicability of the results from highly selected subpopulations to the general population is questionable.

Intervention Trials

Randomized intervention studies in humans should provide definitive support for the purported cause-and-effect relationship between a dietary factor and CRC. Intervention studies are often difficult to carry out because of the slowly progressive nature of neoplastic transformation and the large number of subjects necessary to achieve an adequate statistical power. In many studies, intermediate biomarkers of CRC rather than occurrence or recurrence have been used as endpoints. Biomarkers include adenoma, proliferation markers, mitotic index, DNA aneuploidy aberrant crypts, mucins, and alterations of several molecular biological markers. Intermediate biomarkers have limitations and most have not been validated in clinical studies. Several randomized or single-arm intervention studies using a high-fiber diet as a component of chemopreventive strategies against the development of CRC have been conducted. Subjects with familial adenomatous polyposis who had undergone total colectomy and ileorectal anastomosis at least 1 year before entry into the trial were randomized to receive a low-fiber supplement (2.2 g/day) plus placebo (control group), a low-fiber supplement (2.2 g/day) plus ascorbic acid (4 g/day) and α-tocopherol (400 mg/day), or a high-fiber supplement (22.5 g/day) plus ascorbic acid (4 g/day) and α-tocopherol (400 mg/day). Those who had consumed 11 g of supplemental fiber in addition to their usual dietary fiber intake had a significant reduction in polyp occurrence in the rectal stump and polyp number decreased incrementally as the amount of ingested, prescribed fiber increased. One study from the Arizona Cancer Center was a single-arm study that investigated the effect of supplemental wheat bran fiber on a proliferation marker ([^{3}H]thymidine labeling index) in patients who had undergone resection for colon or rectal cancer. In this study, 13.5 g of supplemental wheat bran per day significantly reduced colorectal epithelial proliferation. The same investigators completed a double-blind, randomized study to determine the effects of wheat bran (2.0 or 13.5 g/day) and calcium carbonate (250 or 1500 mg/day) supplementation on [^{3}H]thymidine labeling index in rectal mucosal biopsies and fecal bile acid concentrations at 3 and 9 months. The results of this study showed that neither wheat bran fiber nor calcium treatment significantly decreased the labeling index. In the Australian Polyp Prevention Project, a diet high in fiber and low in fat was shown to prevent recurrence of large adenomas (>10 mm). The Toronto Polyp Prevention Study reported on 201 subjects with adenomatous colorectal polyps. Patients were randomized after polypectomy to receive counseling on a diet low in fat (<50 g/day or 20% of energy) and high in fiber (50 g/day) or to follow a normal Western diet. After 2 years of follow-up with colonoscopy, an intention to treat analysis showed no significant difference between groups with regard to the recurrence of adenomatous polyps. It was, however, shown that women who ate the low-fat, high-fiber diet showed a nonsignificant 50% reduction in polyp recurrence (RR 0.5; 95% CI 0.2−1.9) associated with a reduced concentration of fecal bile acids.

Three recent intervention studies did not support the use of dietary fiber in reducing CRC risk. In the Polyp Prevention Trial, 2079 subjects with resected colonic adenomas were randomized to a diet low in fat (20% fat calories), high in fiber (18 g/1000 kcal daily), and enriched with vegetable and fruits (5−8 servings daily) or to a control group given a standard brochure in healthy eating. The end-point was the development of recurrent colorectal adenomas. The

results show that a diet low in fat and high in fiber, fruits, and vegetables does not affect the rate of recurrence of colorectal adenomas. In the Arizona Colon Cancer Prevention Trial, Alberts *et al*. randomly assigned 1303 subjects to either a diet high in wheat bran fiber (13.5 g/day) or a diet low in fiber (2 g/day); the primary end-point was the presence or absence of new adenomas at the time of follow-up colonoscopy. The relative risk of recurrence of adenoma in the high-fiber group compared with the low-fiber group was 0.99 (95% confidence interval 0.71−1.36; $P = 0.93$). The investigators concluded that a dietary supplement of wheat bran fiber does not protect against recurrence of colorectal adenomas. Bonithon-Kopp *et al*. for the European Cancer Prevention Organization Intervention Study showed that fiber supplemented as ispaghula husk may have adverse effects on colorectal adenoma recurrence especially in patients with high calcium intake. In their study, 665 patients with a history of colorectal adenomas were randomly assigned to one of three treatment groups (calcium gluconolactate and carbonate, fiber, or placebo). The primary end-point was adenoma recurrence.

In summary, nine intervention studies in humans have been completed and published (Table II). One study was uncontrolled and eight were randomized and placebo controlled. Six studies used adenoma recurrence or regression as the end-point of the trial and the other three used less well-established intermediate biomarkers of CRC. Four studies showed a moderate protective effect of dietary fiber supplements. The other five studies showed no effect on labeling index or adenoma recurrence. Weaknesses of earlier studies are short follow-up, small numbers of subjects, poor compliance with dietary interventions, high drop-out rates, and use of less well-established intermediate biomarkers. Results of these intervention studies should therefore be interpreted with caution. This is emphasized by the recent results of three intervention studies detailed earlier showing either a null effect or an increased risk associated with fiber intervention. Other limitations are associated with intervention trials in humans. Blind or double-blind trials are difficult to perform with foods or dietary macronutrients that are recognizable. Often, in nonblind studies of foods, subjects in the control group may adopt the dietary behavior of the treatment group if they think the treatment diet is beneficial. Moreover, the time between a change in the level of a dietary factor and any expected change in the incidence of cancer is uncertain. Finally, patients at higher risk for the development of CRC tend to be underrepresented in trials as studies often enroll motivated health-conscious people. Studies in which intervention shows no effect do not indicate that the agents are irrelevant or harmful in the context of whole diets or among normal healthy populations. Conversely, trials in which intervention shows beneficial effects are good evidence that the agents used are protective.

Biological Mechanisms

Burkitt's initial hypothesis was that dietary fiber increases stool bulk. This had the effect of diluting potential carcinogens and decreasing transit time. As a result, there would be less contact time between potential carcinogens in the lumen and the gut mucosa (Table III).

Dietary fiber can bind carcinogens as well as bile acid. Fecal bile acids have been shown to be cytotoxic to or act as mitogens on colonic epithelial cells in animal and *in vitro* studies. The mechanism by which dietary fiber may modulate carcinogenesis involves cytokinetics of the colonic mucosa. Conversion of primary to secondary bile acids by bacterial enzymes may be prevented if the dietary fibers to which the bile acids or bile salts are bound are undegraded in the colon. Bound bile acids or bile salts pass out of the alimentary tract in the feces.

Epidemiological studies have shown that human populations with lower fecal pH have lower rates of colon cancer. Direct experimental acidification of the colon contents in animal models has not always led to a reduction in tumorigenesis. Dietary fiber decreases fecal pH, resulting in reduced solubility of free bile acids. Furthermore, the activity of the colonic bacterial enzyme 7-dehydroxylase, which converts primary bile acids to secondary bile acids, is inhibited at a pH of <6 to 6.5. Acidification of colonic contents increases the availability of calcium for binding to free bile and fatty acids, thereby inhibiting their effects on the colonic mucosa.

Another potential mechanism of dietary fiber relates to alterations in colonic microflora. Dietary fibers modulate colonic bacterial enzyme activity; however, the relationship between colonic bacterial enzyme activity and development of human CRC has not been elucidated. Bacteria are the major water-holding component of feces. Dietary fibers degraded in the colon have been shown to increase fecal bulk by a stimulation of bacterial growth. Increased fecal bulk and reduced transit time resulting from increased bacterial growth reduce interactions of carcinogens with the colonic mucosa. Dietary fiber decreases the numbers of anaerobes, resulting in a decrease in secondary bile acids.

McKeown-Eyssen and Giovannucci have put forward a unifying hypothesis explaining how diet and

TABLE II Summary of Intervention Studies Using High Fiber

Study	Location (year)	Case diagnosis	Sample size	Type of study	Intervention	Duration	Primary end-point	Outcome
De Cosse *et al.*	USA (1989)	FAP, total colectomy, ileorectal anastomosis	58	RCT	Low-fiber supplement (2.2 g/day) + Vitamin C (4 g/day) + Vitamin E (400 mg/day) vs High-fiber supplement (22.5 g/day) + Vitamin C (4 g/day) + Vitamin E (400 mg/day) vs Placebo	4 years	Adenoma regression occurrence	High fiber protective only if >11 g/day Vitamins C and E: 22% decrease compared to baseline trend toward protection
Alberts *et al.*	USA (1990)	Previous CRC	17	Single-arm uncontrolled	Fiber supplement (wheat bran, 13.5 g/day)	8 weeks	Proliferation labeling index [^{3}H]thymidine	
Alberts *et al.*	USA (1997)	Previous colorectal adenomas	100	RCT	2 × 2 Factorial Fiber (wheat bran): high (13.5 g/day), low (2.0 g/day); Calcium: high (1500 mg/day), low (250 mg/day)	9 months	Proliferation labeling index [^{3}H]thymidine—no effect; total fecal bile acids 52% decrease with high fiber; fecal deoxycholic bile acids—36% decrease with high fiber	Proliferation labeling index [^{3}H]thymidine—no effect; total fecal bile acids 52% decrease with high fiber; fecal deoxycholic bile acids—36% decrease with high fiber
Toronto Polyp Prevention Group	Canada (1994)	Previous colorectal adenoma	201	RCT	Dietary counseling to achieve 20% fat calories and 50 g fiber/day vs Placebo	2 years	Adenoma recurrence	No effect
Australian Polyp Prevention Project	Australia (1995)	Previous colorectal adenomas	424	RCT	2 × 2 × 2 Factorial: <25% fat calories; 25 g wheat bran/day; β-Carotene (20 mg/day)	4 years	Adenoma recurrence	Low-fat, high-fiber decreased recurrence of >10 mm adenomas
Arizona Cancer Center Polyp Prevention Study	USA (2001)	Previous colorectal adenomas	1400	RCT	High-fiber supplement (13.5 g wheat bran/day)	3 years	Adenoma recurrence	No effect
Polyp Prevention Trial	USA (2001)	Previous colorectal adenomas	2079	RCT	20% fat calories/day; 18 g fiber/1000 kcal/day; 5–9 servings of vegetables and fruits/day vs Typical North American diet	4 years	Adenoma recurrence	No effect
European Cancer Prevention Organization Study	Europe (2001)	Previous colorectal adenomas	656	RCT	3.8 g ispaghula husk/day vs 2 g/day calcium vs Placebo	3 years	Adenoma recurrence	Increased incidence of colonic adenomas

Note. FAP, familial adenomatous polyposis; RCT, randomized controlled trial.

TABLE III Mechanisms of Action of Dietary Fiber

Increased stool bulk
Dilution of potential carcinogens
Decrease in transit time
Binding to potential carcinogens
Binding to bile acids
Decrease in fecal bile acid concentrations
Prevention of conversion of primary to secondary bile acids
Lower fecal pH
Reduced solubility of free bile acids
Inhibition of 7 α-dehydroxylase
Inhibition of bacterial degradation of normal fecal constituents to potential carcinogens
Altered colonic microflora
Changes in bacterial species
Stimulation of bacterial growth
Inhibition of microbial enzymes
Prevention of insulin resistance and hyperinsulinemia
Fermentation of fecal flora to SCFA
Induction of apoptosis
Inhibition of proliferation
Modulation of gene expression

lifestyle factors modulate colorectal carcinogenesis. Epidemiological studies have shown that insulin-like growth factor-I (IGF-I) is positively associated with the risk of CRC. Experimental studies have shown that IGF-I has mitogenic and anti-apoptotic actions on CRC cells. The hypothesis suggests that dietary and lifestyle factors associated with CRC risk cause insulin resistance and hyperinsulinemia. It is hyperinsulinemia that may in turn stimulate the growth of colorectal tumors. Insulin is an important growth factor for colonic mucosal cells and colon cancer tissue has both insulin and IGF-I receptors. Insulin receptors can be bound by IGF-I and a binding protein from IGF-I inhibits the growth of colon cancer cells *in vitro*. IGF-I is regulated by six binding proteins (IGFBP-1 to IGFBP-6). Insulin inhibits the production of IGFBP-1. Chronically elevated fasting insulin levels may lead to a decrease in circulating IGFBP-1 concentrations and an increase in IGF-I bioavailability. This increase in bioavailability over time may increase the risk of CRC. Interestingly, subjects with acromegaly, characterized by chronic growth hormone and IGF-I hypersecretion, have an increased risk of developing CRC. It has been postulated that the stimulation of IGF-I receptors by IGF-I or IGF-II in subjects with acromegaly promotes colorectal carcinogenesis. Two recently published large prospective studies indicate a modest increase in CRC risk in subjects with type 2 diabetes compared with nondiabetic control subjects. As soluble fiber affects glycemia and insulinemia, the insulin hypothesis could be a mechanism by which dietary fiber can modulate colorectal carcinogenesis.

Mechanisms relating to SCFA will be addressed in the next section.

Short-Chain Fatty Acids/Resistant Starch

SCFA produced by fermentation of dietary fiber and resistant starch by colonic bacteria play a role in colorectal carcinogenesis. Butyrate inhibits growth in colonic tumor cell lines and induces differentiation and apoptosis. Butyrate also has been shown to alter the binding of regulatory transacting proteins to specific DNA sequences that control the expression of the gene. Butyrate has been shown to inhibit histone deacetylase, resulting in hyperacetylation of histones and increased accessibility of DNA to factors controlling gene expression. Hyperacelylation of histones disrupts ionic interactions with the adjacent DNA backbone, creating less densely packed chromatin, allowing transcription factors to activate specific genes. It has been postulated that the link between histone hyperacetylation-induced transcriptional regulation and growth inhibition may lie in the effects on specific cell cycle regulators. Butyrate increases apoptosis in various cell lines. Butyrate induces apoptosis through a histone hyperacetylation-mediated pathway that results in the conversion of capase-3 from its proenzyme form to the catalytically active protease. Another theory postulates that SCFA significantly inhibit colon cancer cell invasion by inhibiting urokinase plasminogen activator secretion and stimulating tissue inhibitor matrix metalloproteinase, which protects the basement membrane against degradation.

A report by Vernia *et al*. indicated that fecal acetate was lower in cancer patients than in those with polyps or in controls. Weaver *et al*. showed that there were no consistent differences in SCFA between patients with various gastrointestinal diseases.

SCFA are an important energy source for the colonocytes. Fermentation of dietary fiber and resistant starch by colonic bacteria generates SCFA. Butyrate and other SCFA have been shown to have anti-carcinogenic properties. Further studies are needed to delineate the effects of SCFA on colorectal carcinogenesis.

Resistant starch, defined as that portion of ingested starch that escapes digestion in the small intestine, has been shown to increase stool bulk, decrease fecal pH, alter the colonic microflora, decrease secondary bile acid concentrations and cytotoxicity of fecal water, decrease colonic mucosal proliferation, increase

colonic fermentation, and contribute to the synthesis of SFCA, especially butyrate. Data from studies are inconclusive. A published international correlation study supports the protective role of resistant starch in the development of CRC. In this study, there was a strong inverse association between starch consumption and CRC (correlation coefficient, $r = 0.70$; $P < 0.001$). In three different studies using chemical rodent models of CRC, resistant starch was observed to be protective, have no effect, or enhance tumorigenesis. Resistant starch was shown to significantly increase small bowel tumors in a study using a knockout murine model of the adenomatous polyposis coli gene (Apc1638N).

Fructooligosaccharides

Interest in fructooligosaccharides (FOS) as a health-promoting agent is on the rise. FOS constitute a group of linear fructose oligomers with a degree of polymerization ranging from one to five oligosaccharides. FOS occur in a number of plants such as chicory, onions, asparagus, and wheat. FOS escape digestion in the human upper intestinal tract and reach the colon, where they are completely fermented mostly to lactate, short-chain fatty acids, and gas. As a consequence of their fermentation, their caloric value is 2 kcal/g. Interest regarding the health benefits of FOS has increased as the role of prebiotics and probiotics on gut health continues to evolve. Prebiotics are nondigestible food ingredients that affect the host by selectively stimulating the growth or activity, or both, of one or a limited number of bacteria in the colon. FOS are the only products that meet all criteria allowing classification as prebiotics. An important property of FOS is the stimulation of bifidobacterial growth coupled with the suppression of the growth of potentially harmful pathogens. FOS are associated with a decrease in fecal pH, an increase in fecal organic acids, a decrease in the production of nitrogenous end products in urine and stool, and a decrease in fecal bacterial enzymatic activity. Several studies to date have examined the role of FOS in health and disease. Delzenne *et al*. in 1995 showed that FOS enhanced the bioavailability of calcium, magnesium, and iron in rats. Cherbut *et al*. conducted a study investigating the effect of FOS on colitis in rats. They found that FOS reduced intestinal inflammatory activity by decreasing lactic acid bacteria counts in the intestine. These results have promoted investigations into the effect of FOS on human diseases of chronic inflammation. Olesen *et al*. have recently demonstrated that in patients with irritable bowel syndrome, long-term use of FOS does not affect symptoms or disease outcome. Studies examining the effects of FOS on colon cancer, inflammatory bowel disease, and the absorption of specific nutrients are currently under way. The importance of FOS on disease outcomes therefore remains to be proven.

Summary

The strongest evidence that supports the fiber hypothesis is witnessed in the protective effect of dietary fiber among correlation and case-control studies. However, large prospective studies conducted in specific populations in the United States do not support the protective effect of dietary fiber on the development of CRC. Moreover, human interventional studies have failed to confirm a positive role for fiber in the reduction of CRC. Currently available evidence from epidemiological, animal, and intervention studies does not unequivocally support the protective role of fiber against the development of CRC. However, when the whole body of evidence from these studies is analyzed critically, the overall conclusion supports an inverse association between dietary fiber intake and CRC risk.

Whereas most studies have adjusted for potential confounding factors, it is difficult to delineate the effect associated with dietary fiber from other potential anticarcinogens. It is possible that undetermined interactions among anti-carcinogens present in fiber-rich foods and fiber are responsible for the observed protective effect of dietary fiber on the development of CRC. Three combined analyses or meta-analyses of case-control studies suggest a 50% reduction in the risk of developing CRC in subjects with the highest dietary fiber intake compared with those with the lowest intake. Although it is difficult to estimate accurately the magnitude of CRC risk reduction attributable solely to dietary fiber or fiber-rich foods, there appears to be a significant degree of reduction.

Animal studies suggest that insoluble and less fermentable fibers and wheat bran are the most effective in reducing CRC rates. Recent analysis in human studies suggests that only total dietary fiber, fruit fiber, and soluble fiber are significantly associated with decreased risk of colonic adenomas. Although the role of resistant starch in colorectal carcinogenesis has recently received much attention, convincing epidemiological evidence is lacking except for one international correlation study that showed a strong inverse association between starch and resistant starch consumption and CRC risk.

CRC mortality rates may be reduced through dietary means by an estimated 50–75%. CRC risk reduction associated specifically with dietary fiber intake has been estimated at 50% when subjects with the highest dietary intake are compared to those with the lowest dietary intake. The maximum daily dose of dietary fiber associated with a significant degree of CC risk reduction has yet to be established. Two combined studies of case control subjects suggest a 50% reduction in CRC risk in individuals consuming 27 g/day compared with those consuming less than 11 g/day of fiber. Prospective studies suggest that intakes between 28 and 33 g/day of dietary fiber are necessary to show a significant reduction in CRC risk. Any fiber intervention with the goal of reducing mortality due to CRC should be initiated at least 10–20 years before the peak age incidence of CRC. This is based on the principle that dietary fiber has its full impact on preventing CRC decades after initiation of a modified fiber-rich diet. Clearly, individuals at high risk for developing CRC, such as those with familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer, will achieve the greatest benefit from dietary fiber. However, extrapolating data from these patients suggests that the general population will benefit from dietary modification.

It is difficult to advise patients regarding the role of fiber and CRC. It is reasonable to recommend total fiber intake of at least 20 to 30 g/day. Dietary fiber should be from all sources, including five to seven servings of vegetables and fruits daily and generous portions of whole grain cereals as recommended by the World Health Organization and the National Cancer Institute. The guidelines can be used in conjunction with the dietary fiber recommendations (Table IV).

TABLE IV Dietary Recommendations

Daily carbohydrate intake: 55–60% of daily energy intake
Total fiber intake: 20–30 g/day
Eat each of the five food groups daily (meat, dairy products, grains, fruits, and vegetables)
Reduce total fat intake to less than 25 to 30% of total calories and saturated fat to less than 10% of total calories
Eat 5 or more servings of fresh vegetables and fruits daily (raw better than cooked; include deep yellow vegetables and dark green cruciferous vegetables)
Eat red meat infrequently (substitute chicken or fish without skin)
Eat more fiber-rich foods such as whole-grain cereals, fruits, and vegetables (daily total of 20 to 30 g fiber)
Avoid obesity
Eat salt-cured, smoked, and nitrite-cure foods in moderation
Keep alcohol consumption moderate
Participate in daily physical activity
Do not smoke

DIVERTICULAR DISEASE

Introduction

Diverticular disease encompasses two disease entities: diverticulosis and diverticulitis. Disease incidence increases from less than 5% at age 40, to 30% by age 60, to 65% by age 85. Although a male preponderance was noted in early series, more recent studies have suggested either an equal distribution or a female preponderance. Geographic variations exist in both the prevalence and pattern of diverticulosis. In some Westernized nations, prevalence rates of 5 to 45% exist with a tendency toward left-sided disease. In Africa and Asia, where the prevalence is less than 0.2%, diverticulosis is predominantly right-sided. Diverticulosis or diverticular disease of the colon is due to outpouchings of colonic mucosa through points of weakness in the colonic wall where the blood vessels penetrate the muscularis propria. Diverticulitis represents micro- or macroscopic perforation of a diverticulum. These diverticula are prone to infection or "diverticulitis," presumably because they trap feces with bacteria. Among all patients with diverticulosis, 70% remain asymptomatic, 15 to 25% develop diverticulitis, and 5 to 15% develop some form of diverticular bleeding. Diverticula and vascular ectasias cause the majority of lower gastrointestinal bleeding. Severe hemorrhage occurs in 3–5% of all patients with diverticulosis. The bleeding is arterial and caused by medial thinning of the vasa recta as it courses over a diverticulum. Patients with acute diverticulitis present with increasing left lower quadrant pain and fever, often with constipation and lower abdominal obstructive symptoms such as bloating and distension. The diagnosis of acute diverticulitis is made on the basis of the history and the physical examination. Physical examination reveals localized tenderness in the left lower quadrant and, with severe infection and an abscess, there may be peritoneal signs in the left lower quadrant. In some cases, a palpable mass can be identified over the sigmoid colon (the most common site of diverticulitis). Computer tomographic (CT) scanning has become the optimal method of investigation in patients suspected of having acute diverticulitis. After resolution of an episode of acute diverticulitis, the colon requires full evaluation by colonoscopy, barium enema, or both, to establish the extent of disease.

From a clinical perspective, the diagnosis of diverticular disease is controversial. Confusion exists

regarding the differences between diverticulitis and diverticulosis with irritable bowel disease. The distinction between the two entities is important not only from a clinical perspective but also from a research perspective. Classification of patients into the proper disease entity is critical in achieving good research outcomes. The confusion between diverticulitis and diverticulosis with irritable bowel syndrome (IBS) symptoms lies in the clinical presentation of symptoms. The classic features of diverticulitis include pain, fever, bowel obstruction, absence of blood in stool, and bloating. These symptoms clearly overlap with the symptoms of diverticulosis overlapping those of IBS. The distinction between the two entities therefore lies in clinical acumen as well as investigations. Patients with diverticulitis can present with an elevated white count, plain X rays suggestive of an obstruction, or a CT scan showing a perforation or an abscess. These findings are typically absent in diverticulosis with IBS. Future studies therefore need to include both clinical and laboratory investigations when classifying the various diverticular diseases.

Etiology

It has been proposed that low dietary fiber predisposes to the development of diverticular disease. In one study, Burkitt and Painter demonstrated that individuals in the United Kingdom eating a Western diet low in fiber had colonic transit times of 80 h and a mean stool weight of 110 g/day. In comparison, Ugandans eating very-high-fiber diets had transit times of 34 h and greater stool weights, >450 g/day. Longer transit times and smaller stool volumes were considered to contribute to the development of diverticular disease increases in intralumenal pressures. Such changes can lead to diverticular herniation.

The ultimate etiology of diverticular disease is unknown. Leading theories suggest that altered colonic motility plays a major role in the development of diverticulum. Higher resting, postprandial, and neostigmine-stimulated pressures in diverticular patients suggest that a delay in transport with augmentation of water reabsorption could cause excessively high pressures, forcing mucosa to herniate. Observations that diverticular disease is less common in vegetarians than nonvegetarians is compatible with a role for dietary fiber, since vegetables and fruits are important sources of fiber. A recent report evaluating a cohort of over 47,000 men provided strong evidence for the role of dietary fiber. Total dietary fiber intake was noted to be inversely associated with the risk of symptomatic diverticular disease after adjustment for age, energy-adjusted total fat intake, and physical activity. Results from the Health Professionals Follow-up Study suggest an etiologic role of fiber in diverticular disease. The study followed 51,529 male health professional over a 6-year period. A significant inverse association was found between insoluble dietary fiber and the risk of developing symptomatic diverticular disease. The greatest benefit was seen in those consuming an average of 32 g/day of total fiber.

Other dietary factors that might contribute to the pathogenesis of diverticular disease have been examined. There is no substantially increased risk associated with smoking, caffeine, or alcohol. Other studies suggest an association between obesity in men under 40 years of age and acute diverticulitis. This finding is consistent with observations that the risk of symptomatic diverticular disease is increased (relative risk 2.35 to 3.32, 95% CI) by a diet characterized by a high intake of total fat or red meat and a low intake of dietary fiber.

Treatment

The mainstay of treatment for preventing diverticulosis is a diet high in fruit and vegetable fiber. This suggestion is based on observations that low-fiber diets are associated with colonic diverticulosis. Recommendations are also based on results from the aforementioned Health Professionals Follow-up Study.

The management of uncomplicated diverticulosis involves a diet high in fruit and vegetable fiber. The majority of patients with diverticular disease will remain asymptomatic. There are no data to support any therapeutic recommendations in this group of patients. Multiple uncontrolled studies demonstrating the effect of fiber in patients with diverticulosis exist; however, the lack of a placebo group complicates the results. The first randomized controlled trial of a high-fiber diet in patients with symptomatic diverticular disease was conducted in 1977 by Brodribb. Results were reported on 18 patients at 1 and 3 months, with results indicating a statistically significant decrease in bowel symptoms in those in the treatment arm. Conversely, a study by Ornstein *et al*. failed to show a difference in symptoms between patients taking bran versus placebo after 4 months. Other well-conducted studies have reported conflicting results regarding the use of fiber. Despite conflicting data, it appears safe to recommend a diet high in fiber for patients with uncomplicated diverticular disease.

Complicated diverticulosis includes diverticulitis and hemorrhage. Patients with mild diverticulitis are often treated as outpatients with broad-spectrum

oral antibiotics. Patients may expect symptoms of increasing pain, fever, or inability to tolerate oral foods. Such patients are treated with a clear liquid diet. Patients with more severe diverticulitis admitted to hospital should be placed on bowel rest with clear liquids or nothing by mouth. Intravenous fluid therapy is required. Intravenous antibiotics should be initiated to target colonic anaerobes and gram-negative rods. Improvement is expected within 2–4 days, at which point the diet can be advanced. Fifteen to 30% of patients hospitalized with acute diverticulitis will require surgery. The likelihood of recurrence and the role of elective surgical resection are important to prevent further attacks as the risk of recurrence after an acute attack ranges from 7 to 62%. Recurrent attacks are less likely to respond to medical and diet therapy and often require emergent surgery.

TABLE V Differences between Crohn's Disease and Ulcerative Colitis

	Crohn's disease	Ulcerative colitis
Symptoms		
Diarrhea	+	++
Abdominal pain	++	+
Weight loss	++	−
Fever	+	+
Fistula	+	−
Stricture	+	−
Bleeding	+	++
Histology		
Granulomas	++	−
Skip lesions	++	−
Transmural inflammation	++	−
Rectal sparing	++	−

CROHN'S DISEASE

Introduction

Inflammatory bowel disease refers to two disorders, ulcerative colitis and Crohn's disease, both chronic, relapsing disorders of unknown etiology.

Crohn's disease is a chronic inflammatory disorder that occurs throughout the world, with a prevalence of 10 to 100 cases per 10^5 people. Although the disorder can begin at any age, its onset most often occurs between 15 and 30 years of age. There appears to be a familial aggregation of patients with Crohn's disease such that 20–30% of patients with Crohn's disease have a family history of inflammatory bowel disease. Epidemiological studies suggests that the disorder occurs most frequently among people of European origin. Crohn's disease is three to eight times more common among Jews than among non-Jews and is more common among whites than nonwhites.

The typical clinical manifestations for most patients with Crohn's disease include diarrhea, abdominal pain, weight loss, and fever (Table V). The transmural nature of the inflammatory process leads to fibrotic strictures that can lead to episodes of small bowel or colonic obstruction. Transmural inflammation is also associated with the development of sinus tracts that can lead to serosal penetration and bowel wall perforation. Penetration of the bowel wall often presents not as an acute abdomen but as an indolent process related to the production of fistulas. Weight loss and fever are the primary systemic symptoms in Crohn's disease.

Patients with Crohn's disease can be divided into those with small bowel disease alone (30%), those with both small and large bowel involvement (50%), and those with disease involving only the colon (20%). The hallmark of Crohn's disease is inflammation, which may or may not be accompanied by noncaseating granulomas, extending through all layers of the gut wall. Microscopic examination reveals (1) hyperplasia of perilymphatic histiocytes, (2) diffuse granulomatous infiltration, (3) discrete noncaseating granulomas in the submucosa and lamina propria, (4) edema and lymphatic dilation of all layers of the gut, and (5) monocytic infiltration within lymph nodules and Peyer's patches on the serosal surface of the bowel.

The impact of dietary fiber in Crohn's disease focuses on (1) the role of fiber in the etiology of disease and (2) the role of fiber in the management of disease. The role of management will be discussed in the next section.

Several studies have investigated the role of dietary fiber in Crohn's disease. One study in 1979 by Thornton *et al.* interviewed 30 patients newly diagnosed with Crohn's disease. Patients were asked about their habitual pre-illness diet. When compared to healthy controls matched for age, sex, social class, and marital status, it was found that the Crohn's disease patients ate more refined sugar and slightly less fiber, raw fruit, and vegetables. It was postulated that a diet high in refined sugar and low in raw fruit and vegetables may predispose to Crohn's disease. In a review by Riemann *et al.* in 1984, it was shown that patients with Crohn's disease have a significantly increased consumption of refined carbohydrates compared to controls. Persson *et al.* conducted a population-based case control study of inflammatory bowel disease and dietary habits in Stockholm, Sweden. The relative risk for Crohn's disease was increased in

subjects who had a high intake of sucrose and decreased for subjects who had a high intake of fiber. The difficulty with the studies investigating the role of dietary fiber's causal effect in Crohn's disease is the retrospective nature of the studies, the inability to control for confounding factors, and the small sample sizes. Therefore, one cannot draw a conclusion about any causative role for fiber in Crohn's disease.

Treatment

In patients with narrowing of the lumen, the rationale for restricting dietary fiber is apparent. Fiber intake can be reduced by avoiding coarse whole grain breads and cereals, nuts, and most fruits and vegetables. Fruit and vegetables may not pass through strictures and may cause a bolus obstruction behind a stricture. Patients with symptoms suggestive of mild or partial bowel obstruction should avoid the intake of raw fruits and vegetables. Studies have investigated the role of dietary fiber in managing Crohn's disease. One study by Koga *et al.* evaluated the effectiveness of a low-residue diet in maintaining remission in people with Crohn's disease. The results suggest that drinking a low-residue drink for longer than 1 year following induction of remission was useful for the maintenance of disease remission. Kasper *et al.*, in 1979, measured the mean daily intake of dietary fiber, sugar, starch, fat, protein, and total energy in patients with Crohn's disease and controls matched for age, sex, and socioeconomic background. In the patients with Crohn's disease, the mean dietary fiber intake was greater than in controls. Crohn's disease patients also exhibited greater sugar and starch consumption with greater total energy intake. Heaton *et al.*, in 1979, treated 32 patients with Crohn's disease with a fiber-rich, unrefined carbohydrate diet in addition to conventional management. The results of the trial show fewer hospital admissions in the diet-treated patients than in those receiving only conventional management. In 1985, Jones *et al.* investigated 20 patients with Crohn's disease in which remission was maintained by either an unrefined carbohydrate fiber-rich diet or a strict exclusion diet. Their results show greater remission rates in patients on exclusion diets than in those on the carbohydrate fiber-rich diet. In a study by Levenstein *et al.* in 1985, there was no difference in hospitalization, need for surgery, new complications, nutritional status, or postoperative recurrence between patients on a low-residue diet and those following a normal Italian diet. In a recent study by Stange *et al.* in 1990, patients with Crohn's disease received either an exclusion diet or a diet low in refined carbohydrates and rich in fiber. A total of 26 patients were observed for 1 year. There was no difference between diets with respect to Crohn's disease activity index.

ULCERATIVE COLITIS

Introduction

Ulcerative colitis is an inflammatory disease of unknown etiology affecting the colonic mucosa from the rectum to the cecum. The diagnosis of ulcerative colitis rests on discovery of a combination of clinical and pathological criteria, investigation of the extent and distribution of lesions, and exclusion of other forms of inflammatory colitis caused by infectious agents (*Entamoeba histolytica, Clostridium difficile, Campylobacter jejuni, Escherichia coli*, and *Shigella dysenteriae*). The inflammatory state in ulcerative colitis is confined to the mucosa. Polymorphonuclear cells accumulate in the crypt abscesses and frank necrosis of the surrounding crypt epithelium occurs. Several crypt abscesses may coalesce to produce ulceration visible on the mucosal surface. Following mucosal destruction, highly vascular granulation tissue develops in denuded areas, resulting in friability and bleeding. Diarrhea and rectal bleeding, two of the most prominent symptoms of ulcerative colitis, are related both to the extensive mucosal damage that renders the colon less capable of absorbing electrolytes and water and to the highly friable vascular granulation tissue.

Treatment

A low-residue diet is recommended during acute phases of colitis because insoluble particles irritate the bowel and make diarrhea worse. Recently, however, the rarity of ulcerative colitis in developing countries together with the ability of dietary fiber to affect colonic function and its bacterial content have suggested that a low intake of fiber might be a factor in causing ulcerative colitis. A high-fiber diet can be prescribed for patients during quiescent phases of the disease. In one open-label, parallel group, multicenter randomized trial, the effect of *Plantago ovata* seeds compared with mesalamine in maintaining remission in ulcerative colitis was investigated. The primary end-point was maintenance of remission for 12 months. Results show that treatment failure occurred in 40% of the *P. ovata* seed group, 35% of the mesalamine group, and 30% of the *P. ovata* plus mesalamine group. *P. ovata* seeds might be as effective as mesalamine in maintaining remission in ulcerative colitis. Hallert *et al.* investigated the efficiency of ispaghula husk in relieving gastrointestinal symptoms in patients with ulcerative colitis in remission compared with placebo. Results show that

ispaghula can be helpful in the management of symptoms secondary to ulcerative colitis. It has been shown that germinated barley foodstuff results in significant clinical and endoscopic improvement in patients with ulcerative colitis.

SCFA have been proposed as a topical treatment for active distal ulcerative colitis. Colonic biopsy specimens from patients with ulcerative colitis reveal impaired utilization of butyrate as measured by carbon dioxide production. The rationale behind using SCFA in ulcerative colitis is that supraphysiologic lumenal SCFA concentrations may be able to overcome the partial metabolic defect of the colonic mucosa to oxidize SCFA. Moreover, SCFA stimulate colonic cell proliferation, provide a more effective barrier between mucosa and the intralumenal contents, dilate the resistance arteries of the colon, and increase blood flow and mucosal oxygen uptake. Four uncontrolled studies investigating the effect of SCFA on distal active ulcerative colitis have been published. In all four trials, 50–78% of the patients had clinical improvement, some with complete remission, during the study periods ranging from 4 to 6 weeks. Disease activity indices and endoscopic appearance all improved. The trials to date are limited by the small number of subjects, varying combinations and concentrations of SCFA, and differences in study design and patient populations. In a study by Breuer *et al*. in 1997, 103 patients with distal ulcerative colitis were entered into a 6-week, double-blind, placebo-controlled trial of rectal SCFA twice daily. Of the 91 patients completing the trial, more patients in the SCFA-treated group than in the placebo-treated group improved (33% vs 20%, $P = 0.14$, NS). In summary, SCFA irrigation can effectively treat refractory distal ulcerative colitis that has failed to respond to standard topical, systemic, or combination therapies. As a first-line treatment of mild to moderate distal ulcerative colitis, SCFA are associated with a modest clinical response rate. SCFA are as effective as topical corticosteroid and 5-aminosalicylic acid in inducing clinical, endoscopic, and histologic improvement in mild to moderate distal ulcerative colitis. SCFA are free of significant side effects and are well tolerated. Unresolved issues regarding SCFA include their role in maintaining remission of distal ulcerative colitis once active disease is brought under control.

DIVERSION COLITIS

Diversion colitis is an inflammatory process that occurs in segments of the colon after surgical diversion of the fecal stream. Several studies have documented that endoscopic abnormalities and histologic changes occur in the distal colonic segment of most patients after intestinal diversion. Pathologic examination shows lymphoglandular complexes expanding the submucosa with increased lymphocytes and plasma cells. Cryptitis with abscesses, patchy neutrophil infiltration in the lamina propria, and superficial erosions overlying lymphoid follicles are common findings. The disease is characterized by bleeding from inflamed colonic mucosa, tenesmus, mucus discharge, and abdominal pain. The majority of patients remain asymptomatic. Diversion colitis is likely caused by a deficiency of SCFA, which serve as lumenal nutrients for colonocytes as earlier studies showed beneficial effects of SCFA on diversion colitis. The definitive treatment of diversion colitis is restoration of intestinal continuity. However, randomized controlled studies did not confirm a beneficial role of SCFA in the treatment of diversion colitis. These studies were hampered by small sample size and short duration of treatment. In those patients with preexisting inflammatory bowel disease, SCFA combined with anti-inflammatory drugs may be effective.

IRRITABLE BOWEL SYNDROME

Introduction

IBS is defined as "a functional bowel disorder in which abdominal pain is associated with defecation or a change in bowel habit, with features of disordered defecation and distention." Criteria for IBS have been formalized in the Rome and Manning Criteria (Table VI). The incidence of IBS is 1% per year with a prevalence of 2.9–20%. The prevalence of IBS is lower in the elderly and higher in female patients. Patients are classified into subgroups based on predominant symptom. Subgroups include constipation-predominant IBS, diarrhea-predominant IBS, and IBS with alternating bowel movements.

IBS is considered a biopsychosocial disorder resulting from a combination of psychosocial factors, altered motility and transit, and increased sensitivity of the intestine or colon. One hypothesis suggests that IBS results from altered peripheral functioning of visceral afferents and the central processing of afferent information that are important in the altered somatovisceral sensation and motor dysfunction. As infectious diarrhea precedes the onset of IBS symptoms in 7–30% of patients, it has therefore been postulated that persistent neuroimmune interactions after infectious gastroenteritis result in continuing sensorimotor dysfunction. Food allergens may play a role in IBS as symptoms often improve with dietary exclusion. Stress and

TABLE VI Criteria for the Diagnosis of Irritable Bowel Syndrome

Manning Criteria	Rome II Criteria
Pain relieved by defecation More frequent stools at the onset of pain Looser stools at the onset pain Visible abdominal distension Passage of mucus Sensation of incomplete evacuation	At least 12 weeks or more, which need not be consecutive, in the previous 12 months of abdominal pain or discomfort with 2 of 3 features: relief with defecation, onset associated with a change in the frequency of stool, onset associated with a change in (appearance of) stool

emotions affect gastrointestinal function and cause symptoms in patients with IBS.

Treatment

There are currently no evidence-based guidelines on how patients with IBS should be treated. The dietary treatment of patients with IBS has centered around bran supplementation, manipulating the dietary fiber content of the diet, and the identification of food intolerance. Dietary intervention needs to be individualized with respect to the patient's symptoms. An in-depth assessment of a patient's dietary intake is essential prior to any therapeutic dietetic intervention. Of eight studies evaluating the role of diet in IBS, five identified food intolerance as a major contributor to symptoms. Unfortunately, because of impracticalities in trial design, there are no randomized control trials. In summary, seven studies mentioned dairy products, six mentioned coffee, and five mentioned wheat. Others cite eggs, corn, potatoes, onions, fruits, and vegetables. In the majority of the studies, the patients with diarrhea responded favorably to an exclusion diet compared to other subtypes of IBS. The recommendation that elimination diets control the symptoms of IBS is based on poorly designed and incomplete trials.

The mainstay of dietary therapy for IBS has centered around the manipulation of dietary fiber. Rees *et al*. conducted a critical review of clinical trials examining the effect of dietary fiber on symptoms in patients with IBS. It was shown that six of eight investigations detected no significant difference in most of the symptoms of IBS when fiber was compared to placebo. In the eight trials, the amount of supplement varied, the form in which the fiber was administered differed, and the type of fiber supplement was different. The literature does not support any beneficial effects of increasing insoluble nonstarch polysaccharides in IBS patients. Bran is reported to be no better than placebo in relief of overall IBS symptoms and may be worse than a normal diet for symptoms of IBS caused by intralumenal distension. However, subgroup analysis suggests that increasing insoluble wheat products in the diet may be of some benefit in patients with predominantly symptoms of constipation. There appears to be significant improvement in constipation if sufficient quantities of fiber (20–30 g/day) are consumed. It is therefore common practice to start with a low dose, increasing gradually and abandoning high levels of supplementation (>30 g/day) if patients experience worsening of symptoms. It is therefore generally recognized that fiber may have a role in treating constipation with a minimal role in the relief of abdominal pain and diarrhea.

The use of dietary fiber in patients with constipation-predominant IBS was prompted by the success of fiber in controlling general symptoms of constipation. Constipation is defined as a bowel movement every 3 to 4 days or less. An international panel of experts developed a consensus definition of constipation. Components of the definition include the presence of straining, hard stools, feelings of incomplete evacuation, or two or fewer bowel movements per week. The prevalence of constipation has been estimated to be 2% of the U.S. population.

An initial approach to the management of constipation consists of a gradual increase in fiber intake. To date, only one meta-analysis investigating therapeutic trials has been conducted. Tramonte *et al*. evaluated 25 treatments in 36 randomized trials. Although there was inadequate evidence to conclude that fiber was superior to laxatives, there was agreement that both fiber and laxatives improve bowel movement frequency in adults with chronic constipation. It is well known that there is a dose response between fiber intake and fecal output.

Studies have shown that patients with IBS develop symptoms after being exposed to a sorbitol diet. Trials cannot confirm a true malabsorption of carbohydrates in patients with IBS and therefore the role of

carbohydrate and sorbitol is not routinely considered in the management of IBS. Although the number of subjects reporting lactose intolerance is higher (60% compared to 27%) than in the general population, the actual percentage of lactose intolerance in patients with IBS is the same as in healthy subjects. The inconclusive findings and association of lactose intolerance and IBS can be explained by poorly designed studies. Moreover, lactose intolerance has been associated with gastrointestinal symptoms; however, whether this is due to lactase deficiency or increased sensitivity is not known.

In summary, the goal of dietary manipulation in patients with IBS is to help patients control their symptoms. Abnormal eating practices need to be assessed in relation to the patient's symptoms. Exclusion diets should be tried only when patients complain of multiple food intolerance and single food avoidance has not helped control symptoms. Patients consuming large quantities of sorbitol should be discouraged, especially if their symptoms are predominantly pain and diarrhea. In those patients in whom dairy products are associated with symptoms, a trial milk-free or lactose-free diet should attempted. If a patient's predominant symptom is constipation, an assessment of fluid intake should be undertaken. Regular meal patterns should be encouraged. An assessment of the type and quantity of nonstarch polysaccharides consumed should be made. The addition of bran and insoluble fibers should be discouraged, with more emphasis placed on increasing the proportion of foods containing a higher concentration of soluble nonstarch polysaccharides.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Colorectal Adenomas • Constipation • Crohn's Disease • Dietary Reference Intakes (DRI): Concepts and Implementation • Diverticulosis • Irritable Bowel Syndrome

Further Reading

American Diabetic Association (2002). Position of the American Dietetic Association: Health implications of dietary fiber *J. Am. Diet. Assoc.* **102**, 993–1000.

Asano, T., and McLeod, R. S. (2002). Dietary fiber for the prevention of colorectal adenomas and carcinomas. *Cochrane Database Syst. Rev.* **3**.

Camilleri, M. (2001). Management of the irritable bowel syndrome. *Gastroenterology* **120**, 652–668.

Han, P., Burke, A., Baldassano, R., Rombeau, J., and Lichtenstein, G. (1999). Nutrition and inflammatory bowel disease. *Gastroenterol. Clin. North Am.* **28**, 423–443.

Kim, Y.-I. (2000). AGA Technical Review: Impact of dietary fiber on colon cancer occurrence. *Gastroenterology* **118**, 1235–1257.

National Center for Health Statistics (2002). Available at http://www.cdc.gov/nchs/nhanes.htm.

Schneeman, B. O., and Tietyen, J. (1999). Dietary fiber. *In* "Modern Nutrition in Health and Disease" (M. E. Shils, J. A. Olson, and M. Shike, eds.), pp. 89–100. Lippincott Williams & Wilkins, Philadelphia, PA.

Stollman, N., and Raskin, J. (1999). Diagnosis and management of diverticular disease of the colon in adults. *Am. J. Gastroenterol.* **94**, 3110–3121.

Dietary Reference Intakes (DRIs): Concepts and Implementation

JOHANNA DWYER
Tufts University and Tufts–New England Medical Center Hospital

adequate intake The nutrient intake amount that is set when experts believe that the data on the requirement are still preliminary.

estimated average requirement The amount of a nutrient that meets the requirement for a specific criterion of adequacy of half of the healthy individuals of a specific age, sex, and life stage.

estimated energy requirement The requirement for energy for individuals of normal weight with body mass indices of between 18.5 and 25 for adults.

Recommended Dietary Allowance The average daily dietary intake level that meets the nutrient requirements of nearly all (97–98%) healthy persons of a specific sex, age, life stage, or physiological condition (such as pregnancy or lactation).

tolerable upper intake level The highest level of chronic, usual daily nutrient intake that is likely to pose no risk of adverse health effects to almost everyone in the population.

total energy expenditure A set of equations that estimate the energy expenditure needed to maintain current body weight and activity levels.

This article describes the dietary reference intakes, which are reference standards for nutrient requirements that were recently completed by Health Canada and by the Food and Nutrition Board, Institute of Medicine, National Academy of Sciences in the United States. The rationale for developing a series of dietary reference intakes for nutrients is that by the early 1990s, it had become apparent that much new work on nutrient needs, the balance and interactions between nutrients, and the relationship of diet and development of chronic degenerative diseases had become available and needed to be incorporated into the thinking about dietary standards. New techniques for evaluating the biological functions of nutrients were also available and more precise estimates of energy output were at hand. Also, there was a need to evaluate whether requirements needed to be stated for many nonnutrients in food with biological activities of health significance. The development of the dietary reference intakes permitted expert groups of biomedical scientists to update data on the validity and reliability of existing dietary standards, to present a new paradigm for assessing dietary adequacy and excess using new statistical techniques, and to plan adequate, balanced intakes that were moderate in their dietary contributions. A new paradigm was needed because the existing nutrient standards could not appropriately address the uses to which they were applied. After over a decade of deliberations, by 2003 the new dietary reference intakes were nearly completed for all known nutrients and methods for applying them for assessing and planning intakes of individuals and groups are now available. This article focuses on their application to individuals for clinical purposes in the United States and Canada.

DIETARY REFERENCE INTAKE VALUES

Dietary reference intakes (DRIs) are quantitative recommendations for nutrient intakes used as reference values or standards for planning and evaluating the diets of healthy people. The DRIs include the estimated average requirement (EAR) for vitamins, minerals, and macronutrients, including protein, essential fatty acids, carbohydrate, and fiber when possible, and three other reference values, the Recommended Dietary Allowances (RDAs), the adequate intake (AI), and the tolerable upper intake level (UL). For energy, the total energy expenditure (TEE) and estimated energy requirement (EER) are presented. For energy-yielding nutrients, an acceptable macronutrient distribution range (AMDR) is also presented.

The process for establishing estimated average requirements and other dietary reference intakes used in the United States and Canada is described in detail in monographs on groups of nutrients published by the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes of the Food and Nutrition Board, Institute of Medicine, National Academy of Sciences. They replace both the 1989 Recommended Dietary Allowances that had been used previously in the United States and previous reference data used in Canada.

 Copyright 2004, Elsevier (USA). All rights reserved.

ESTIMATED AVERAGE REQUIREMENT

The EAR is the amount of a nutrient that meets the requirement for a specific criterion of adequacy of half of the healthy individuals of a specific age, sex, and life stage. Table I presents the EAR for nutrients. In setting the EAR, the evidence for each possible functional criterion that might be chosen is considered and the reason for selecting the criterion that is finally chosen is justified. The amount of the nutrient necessary to meet the appropriate criterion of adequacy varies from one individual to the next. Requirements are usually distributed normally or can be transformed to achieve a normal distribution. The EAR is not useful for estimating nutrient adequacy in individuals because it is a mean requirement for a group and since the variation around the mean is considerable, the range of error is considerable. For example, at the EAR, 50% of the individuals in a group are below their requirement and 50% are above it. Thus, a person whose usual intake is at the EAR has a 50% risk of an inadequate intake during the reporting period. An individual with an intake between the EAR and the RDA would have a risk of inadequacy between 50 and 2−3%. An individual whose usual intake is below the EAR would have a risk of inadequacy between 50 and 100%. The precise amount of a nutrient that will be adequate for any given individual is therefore unknown. It can be stated only in terms of probabilities and thus the EAR has little use in clinical practice.

RECOMMENDED DIETARY ALLOWANCES

The RDA is the average daily dietary intake level that suffices to meet the nutrient requirements of nearly all (97−98%) healthy persons of a specific sex, age, life stage, or physiological condition (such as pregnancy or lactation). It is the nutrient intake goal for planning the diets of individuals. Tables II and III present the RDA for vitamins and minerals, respectively.

To ensure that the needs of any given individual are met, a measure of variability around the EAR, usually the standard deviation, is assumed. Then, the RDA, the amount of the nutrient that covers virtually everyone in the population [defined statistically as 2 standard deviations (SD) above the EAR], is calculated assuming a normal distribution of nutrient requirements. Thus, the RDA is the EAR + 2 SD. If the SD is not available, using a coefficient of variation (CV) of 10%, the RDA is set at 1.2 times the EAR. If the CV is assumed to be 15%, then the RDA is set at 1.3 times the EAR.

As intakes fall further and further below the RDA, the risk, but not the certainty, of inadequacy increases. The RDA is a good target for individuals to aim for in planning their diets. The RDA is not appropriate for assessing the diets of individuals or for assessing or planning the diets of groups. The RDA should not be applied as a standard for the evaluation of group (population) nutrient intakes because it is overly generous; by definition, the RDA exceeds actual requirements of all but approximately 2−3% of the population. Thus, many individuals whose intakes are below the RDA may still be getting enough of the nutrient in question to be above their requirement levels.

ADEQUATE INTAKE

For some nutrients, it is not possible to calculate an EAR and set a RDA because data on the nutrient requirement (EAR) for the function of health significance that has been chosen is not available. However, there may be enough information to make quantitative recommendations about healthful nutrient intake levels. An AI is set when experts believe that data on the requirement using the criterion that has been chosen to determine the requirement are still preliminary. This signifies that more research is needed before an EAR and a RDA for that particular criterion or function can be determined. The AI is an appropriate goal or target for the nutrient intake of individuals. The AI is *not* a requirement and it should not be interpreted as such. Rather, it is an average or median intake in a group of healthy people, all of whom are assumed to be meeting their nutrient requirements. Thus, falling below the AI does not necessarily signify deficiency. Tables II and III also contain the AI for nutrients that have them.

The AI is a value based on observed or experimentally determined approximations of nutrient intakes by groups of healthy people that is used when an EAR and a RDA cannot be determined. For infants, the AI is always based on the mean intakes of groups of healthy infants. For adults, in some instances the AI is set as the mean of diets in some reference group. For other nutrients, the criteria are less precisely determined, but they are always chosen to be generous enough for good health.

TOLERABLE UPPER INTAKE LEVELS OF NUTRIENTS

Requirements for nutrients consist not only of values that ensure adequacy, but should also recognize that avoiding intakes far in excess of the RDA is important

TABLE I Dietary Reference Intakes (DRIs): Estimated Average Requirements for Groups

Life-stage group	Protein (g/day)	Vitamin A (μg/day)[a]	Vitamin C (mg/day)	Vitamin E (mg/day)[b]	Thiamine (mg/day)	Riboflavin (mg/day)	Niacin (mg/day)[c]	Vitamin B_6 (mg/day)	Folate (μg/day)[d]	Vitamin B_{12} (μg/day)	Copper (μg/day)	Iodine (μg/day)	Iron (mg/day)	Magnesium (mg/day)	Molybdenum (μg/day)	Phosphorus (mg/day)	Selenium (μg/day)	Zinc (mg/day)
Infants																		
7–12 months	10												6.9					2.5
Children																		
1–3 years	11	210	13	5	0.4	0.4	5	0.4	120	0.7	260	65	3.0	65	13	380	17	2.5
4–8 years	15	275	22	6	0.5	0.5	6	0.5	160	1.0	340	65	4.1	110	17	405	23	4.0
Males																		
9–13 years	27	445	39	9	0.7	0.8	9	0.8	250	1.5	540	73	5.9	200	26	1055	35	7.0
14–18 years	44	630	63	12	1.0	1.1	12	1.1	330	2.0	685	95	7.7	340	33	1055	45	8.5
19–30 years	46	625	75	12	1.0	1.1	12	1.1	320	2.0	700	95	6	330	34	580	45	9.4
31–50 years	46	625	75	12	1.0	1.1	12	1.1	320	2.0	700	95	6	350	34	580	45	9.4
51–70 years	46	625	75	12	1.0	1.1	12	1.4	320	2.0	700	95	6	350	34	580	45	9.4
>70 years	46	625	75	12	1.0	1.1	12	1.4	320	2.0	700	95	6	350	34	580	45	9.4
Females																		
9–13 years	28	420	39	9	0.7	0.8	9	0.8	250	1.5	540	73	5.7	200	26	1055	35	7.0
14–18 years	38	485	56	12	0.9	0.9	11	1.0	330	2.0	685	95	7.9	300	33	1055	45	7.3
19–30 years	38	500	60	12	0.9	0.9	11	1.1	320	2.0	700	95	8.1	255	34	580	45	6.8
31–50 years	38	500	60	12	0.9	0.9	11	1.1	320	2.0	700	95	8.1	265	34	580	45	6.8
51–70 years	38	500	60	12	0.9	0.9	11	1.3	320	2.0	700	95	5	265	34	580	45	6.8
>70 years	38	500	60	12	0.9	0.9	11	1.3	320	2.0	700	95	5	265	34	580	45	6.8
Pregnancy																		
≤18 years	50	530	66	12	1.2	1.2	14	1.6	520	2.2	785	160	23	335	40	1055	49	10.5
19–30 years	50	550	70	12	1.2	1.2	14	1.6	520	2.2	800	160	22	290	40	580	49	9.5
31–50 years	50	550	70	12	1.2	1.2	14	1.6	520	2.2	800	160	22	300	40	580	49	9.5
Lactation																		
≤18 years	60	880	96	16	1.2	1.3	13	1.7	450	2.4	985	209	7	300	35	1055	59	10.9
19–30 years	60	900	100	16	1.2	1.3	13	1.7	450	2.4	1000	209	6.5	255	36	580	59	10.4
31–50 years	60	900	100	16	1.2	1.3	13	1.7	450	2.4	1000	209	6.5	265	36	580	59	10.4

Sources. Reprinted, with permission, from Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC; Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids." National Academies Press, Washington, DC. (These reports are available at http://www.nap.edu.) Copyright 2002 by the National Academy of Sciences. All rights reserved.

Note. This table presents estimated average requirements (EARs), which serve two purposes: to assess the adequacy of population intakes and to calculate Recommended Dietary Allowances for individuals for those nutrients. EARs have not been established for vitamin D, vitamin K, pantothenic acid, biotin, choline, calcium, chromium, fluoride, manganese, or other nutrients not yet evaluated via the DRI process.

[a] As retinol activity equivalents (RAEs). 1 RAE = 1 μg retinol, 12 μg β-carotene, 24 μg α-carotene, or 24 μg β-cryptoxanthin. The RAE for dietary provitamin A carotenoids is twofold greater than retinol equivalents (RE), whereas the RAE for preformed vitamin A is the same as RE.

[b] As α-tocopherol. α-Tocopherol includes *RRR*-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2*R*-stereoisomeric forms of α-tocopherol (*RRR*-, *RSR*-, *RRS*-, and *RSS*-α-tocopherol) that occur in fortified foods and supplements. It does not include the 2*S*-stereoisomeric forms of α-tocopherol (*SRR*-, *SSR*-, *SRS*-, and *SSS*-α-tocopherol), also found in fortified foods and supplements.

[c] As niacin equivalents. 1 mg of niacin = 60 mg of tryptophan.

[d] As dietary folate equivalents (DFE). 1 DFE = 1 μg food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach.

TABLE II Dietary Reference Intakes: Recommended Intakes for Individuals, Vitamins

Life-stage group	Vitamin A (μg/day)[a]	Vitamin C (mg/day)	Vitamin D (μg/day)[b,c]	Vitamin E (mg/day)[d]	Vitamin K (μg/day)	Thiamine (mg/day)	Riboflavin (mg/day)	Niacin (mg/day)[e]	Vitamin B_6 (mg/day)	Folate (μg/day)[f]	Vitamin B_{12} (μg/day)	Pantothenic acid (mg/day)	Biotin (μg/day)	Choline (mg/day)[g]
Infants														
0–6 months	400	40	5	4	2.0	0.2	0.3	2	0.1	65	0.4	1.7	5	125
7–12 months	500	50	5	5	2.5	0.3	0.4	4	0.3	80	0.5	1.8	6	150
Children														
1–3 years	300	15	5	6	30	0.5	0.5	6	0.5	150	0.9	2	8	200
4–8 years	400	25	5	7	55	0.6	0.6	8	0.6	200	1.2	3	12	250
Males														
9–13 years	600	45	5	11	60	0.9	0.9	12	1.0	300	1.8	4	20	375
14–18 years	900	75	5	15	75	1.2	1.3	16	1.3	400	2.4	5	25	550
19–30 years	900	90	5	15	120	1.2	1.3	16	1.3	400	2.4	5	30	550
31–50 years	900	90	5	15	120	1.2	1.3	16	1.3	400	2.4	5	30	550
51–70 years	900	90	10	15	120	1.2	1.3	16	1.7	400	2.4[h]	5	30	550
>70 years	900	90	15	15	120	1.2	1.3	16	1.7	400	2.4[h]	5	30	550
Females														
9–13 years	600	45	5	11	60	0.9	0.9	12	1.0	300	1.8	4	20	375
14–18 years	700	65	5	15	75	1.0	1.0	14	1.2	400[i]	2.4	5	25	400
19–30 years	700	75	5	15	90	1.1	1.1	14	1.3	400[i]	2.4	5	30	425
31–50 years	700	75	5	15	90	1.1	1.1	14	1.3	400[i]	2.4	5	30	425
51–70 years	700	75	10	15	90	1.1	1.1	14	1.5	400	2.4[h]	5	30	425
>70 years	700	75	15	15	90	1.1	1.1	14	1.5	400	2.4[h]	5	30	425
Pregnancy														
≤18 years	750	80	5	15	75	1.4	1.4	18	1.9	600[j]	2.6	6	30	450
19–30 years	770	85	5	15	90	1.4	1.4	18	1.9	600[j]	2.6	6	30	450
31–50 years	770	85	5	15	90	1.4	1.4	18	1.9	600[j]	2.6	6	30	450

Lactation														
≤18 years	**1200**	**115**	**5**	**19**	75	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7	35	550
19–30 years	**1300**	**120**	**5**	**19**	90	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7	35	550
31–50 years	**1300**	**120**	**5**	**19**	90	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7	35	550

Sources. Reprinted, with permission, from Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC; Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids." National Academies Press, Washington, DC. (These reports are available at http://www.nap.edu.) Copyright 2001 by the National Academy of Sciences. All rights reserved.

Note. This table presents Recommended Dietary Allowances (RDAs) in **boldface type** and adequate intakes (AIs) in lightface type. Both RDAs and AIs may be used as goals for individual intake. RDAs are set to meet the needs of almost all (97 to 98%) individuals in a group. For healthy breast-fed infants, the AI is the mean intake. The AI for other life-stage and gender groups is believed to cover needs of all individuals in the group, but due to lack of data or uncertainty in the data the percentage of individuals covered by this intake cannot be specified with confidence.

[a] As retinol activity equivalents (RAEs). 1 RAE = 1 μg retinol, 12 μg β-carotene, 24 μg α-carotene, or 24 μg β-cryptoxanthin. To calculate RAEs from REs of provitamin A carotenoids in foods, divide the REs by 2. For preformed vitamin A in foods or supplements and for provitamin A carotenoids in supplements, 1 RE = 1 RAE.

[b] calciferol. 1 μg calciferol = 40 international units of vitamin D.

[c] In the absence of adequate exposure to sunlight.

[d] As α-tocopherol. includes *RRR*-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2*R*-stereoisomeric forms of α-tocopherol (*RRR*-, *RSR*-, *RRS*-, and *RSS*-α-tocopherol) that occur in fortified foods and supplements. It does not include the 2*S*-stereoisomeric forms of α-tocopherol (*SRR*-, *SSR*-, *SRS*-, and *SSS*-α-tocopherol), also found in fortified foods and supplements.

[e] As niacin equivalents (NE). 1 mg of niacin = 60 mg of tryptophan; for infants ages 0–6 months, preformed niacin (not NE) is meant.

[f] As dietary folate equivalents (DFE). 1 DFE = 1 μg food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach.

[g] Although AIs have been set for choline, there are few data to assess whether a dietary supply of choline is needed at all stages of the life cycle and it may be that the choline requirement can be met by endogenous synthesis at some of these stages.

[h] Because 10 to 30% of older people may malabsorb food-bound B12, it is advisable for those older than 50 years to meet their RDA mainly by consuming foods fortified with B12 or a supplement containing B12.

[i] In view of evidence linking folate intake with neural tube defects in the fetus, it is recommended that all women capable of becoming pregnant consume 400 μg from supplements or fortified foods in addition to intake of food folate from a varied diet.

[j] It is assumed that women will continue consuming 400 μg from supplements or fortified food until their pregnancy is confirmed and they enter prenatal care, which ordinarily occurs after the end of the periconceptional period— the critical time for formation of the neural tube.

TABLE III Dietary Reference Intakes: Recommended Intakes for Individuals, Elements

Life-stage group	Calcium (mg/day)	Chromium (μg/day)	Copper (μg/day)	Fluoride (mg/day)	Iodine (μg/day)	Iron (mg/day)	Magnesium (mg/day)	Manganese (mg/day)	Molybdenum (μg/day)	Phosphorus (mg/day)	Selenium (μg/day)	Zinc (mg/day)
Infants												
0–6 months	210	0.2	200	0.01	110	0.27	30	0.003	2	100	15	2
7–12 months	270	5.5	220	0.5	130	11	75	0.6	3	275	20	3
Children												
1–3 years	500	11	340	0.7	90	7	80	1.2	17	460	20	3
4–8 years	800	15	440	1	90	10	130	1.5	22	500	30	5
Males												
9–13 years	1300	25	700	2	120	8	240	1.9	34	1250	40	8
14–18 years	1300	35	890	3	150	11	410	2.2	43	1250	55	11
19–30 years	1000	35	900	4	150	8	400	2.3	45	700	55	11
31–50 years	1000	35	900	4	150	8	420	2.3	45	700	55	11
51–70 years	1200	30	900	4	150	8	420	2.3	45	700	55	11
>70 years	1200	30	900	4	150	8	420	2.3	45	700	55	11
Females												
9–13 years	1300	21	700	2	120	8	240	1.6	34	1250	40	8
14–18 years	1300	24	890	3	150	15	360	1.6	43	1250	55	9
19–30 years	1000	25	900	3	150	18	310	1.8	45	700	55	8
31–50 years	1000	25	900	3	150	18	320	1.8	45	700	55	8
51–70 years	1200	20	900	3	150	8	320	1.8	45	700	55	8
>70 years	1200	20	900	3	150	8	320	1.8	45	700	55	8
Pregnancy												
≤18 years	1300	29	1000	3	220	27	400	2.0	50	1250	60	12
19–30 years	1000	30	1000	3	220	27	350	2.0	50	700	60	11
31–50 years	1000	30	1000	3	220	27	360	2.0	50	700	60	11
Lactation												
≤18 years	1300	44	1300	3	290	10	360	2.6	50	1250	70	13
19–30 years	1000	45	1300	3	290	9	310	2.6	50	700	70	12
31–50 years	1000	45	1300	3	290	9	320	2.6	50	700	70	12

Sources. Reprinted, with permission, from Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC; Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids." National Academies Press, Washington, DC. (These reports are available at http://www.nap.edu.) Copyright 2001 by the National Academy of Sciences. All rights reserved.

Note. This table presents Recommended Dietary Allowances (RDAs) in **boldface type** and adequate intakes (AIs) in lightface type. Both RDAs and AIs may be used as goals for individual intake. RDAs are set to meet the needs of almost all (97 to 98%) individuals in a group. For healthy breast-fed infants, the AI is the mean intake. The AI for other life-stage and gender groups is believed to cover needs of all individuals in the group, but due to lack of data or uncertainty in the data the percentage of individuals covered by this intake cannot be specified with confidence.

since such intakes can disturb body functions and cause acute, progressive, or permanent disability. Therefore, caution is warranted in consuming large amounts of nutrients. Tables IV and V present the UL for vitamins and minerals, respectively. The UL of a nutrient is the highest level of chronic, usual daily nutrient intake that is likely to pose no risk of adverse health effects to almost everyone in the population. Below the UL, individuals should be able to biologically tolerate this amount of the nutrient. The UL is set for the adverse effect that is believed to be most likely to harm health and for which sufficient data are available. The UL is set by determining the level at which no observed adverse effects are noted or the lowest level of intake associated with observed adverse effects. An uncertainty factor is then applied to ensure that even very sensitive persons would not experience adverse effects at the UL dose chosen. For many nutrients, data on the adverse effects of large amounts of nutrients are unavailable or they are so limited that a UL cannot be determined. However, the lack of a UL does *not* mean that the risk of adverse effects is nonexistent from high intakes, but only that data are not yet available and the verdict on risk is unknown. The UL is *not* intended to be a recommended level of intake.

TABLE IV Dietary Reference Intakes: Tolerable Upper Intake Levels (ULs)[a] of Vitamins

Life-stage group	Vitamin A (μg/day)[b]	Vitamin C (mg/day)	Vitamin D (μg/day)	Vitamin E (mg/day)[c,d]	Vitamin K	Thiamine	Riboflavin	Niacin (mg/day)[d]	Vitamin B6 (mg/day)	Folate (μg/day)[d]	Vitamin B12	Pantothenic acid	Biotin	Choline (g/day)	Carotenoids[e]
Infants															
0–6 months	600	ND[f]	25	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
7–12 months	600	ND	25	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Children															
1–3 years	600	400	50	200	ND	ND	ND	10	30	300	ND	ND	ND	1.0	ND
4–8 years	900	650	50	300	ND	ND	ND	15	40	400	ND	ND	ND	1.0	ND
Males, Females															
9–13 years	1700	1200	50	600	ND	ND	ND	20	60	600	ND	ND	ND	2.0	ND
14–18 years	2800	1800	50	800	ND	ND	ND	30	80	800	ND	ND	ND	3.0	ND
19–70 years	3000	2000	50	1000	ND	ND	ND	35	100	1000	ND	ND	ND	3.5	ND
>70 years	3000	2000	50	1000	ND	ND	ND	35	100	1000	ND	ND	ND	3.5	ND
Pregnancy															
≤18 years	2800	1800	50	800	ND	ND	ND	30	80	800	ND	ND	ND	3.0	ND
19–50 years	3000	2000	50	1000	ND	ND	ND	35	100	1000	ND	ND	ND	3.5	ND
Lactation															
≤18 years	2800	1800	50	800	ND	ND	ND	30	80	800	ND	ND	ND	3.0	ND
19–50 years	3000	2000	50	1000	ND	ND	ND	35	100	1000	ND	ND	ND	3.5	ND

Sources. Reprinted, with permission, from Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC; Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids." National Academies Press, Washington, DC. (These reports are available at http://www.nap.edu.) Copyright 2001 by the National Academy of Sciences. All rights reserved.

[a] UL indicates the maximum level of daily nutrient intake that is likely to pose no risk of adverse effects. Unless otherwise specified, the UL represents total intake from food, water, and supplements. Due to lack of suitable data, ULs could not be established for vitamin K, thiamine riboflavin, vitamin B12, pantothenic acid, biotin, or carotenoids. In the absence of ULs, extra caution may be warranted in consuming levels above recommended intakes.

[b] As preformed vitamin A only.

[c] As α-tocopherol; applies to any form of supplemental α-tocopherol.

[d] The ULs for vitamin E, niacin, and folate apply to synthetic forms obtained from supplements, fortified foods, or a combination of the two.

[e] β-Carotene supplements are advised only to serve as a provitamin A source for individuals at risk of vitamin A deficiency.

[f] ND, not determinable due to lack of data of adverse effects in this age group and concern with regard to lack of ability to handle excess amounts. Source of intake should be from food only to prevent high levels of intake.

TABLE V Dietary Reference Intakes: Tolerable Upper Intake Levels (ULs)[a] of Elements

Life-stage group	Arsenic[b]	Boron (mg/day)	Calcium (g/day)	Chromium	Copper (μg/day)	Fluoride (mg/day)	Iodine (μg/day)	Iron (mg/day)	Magnesium (mg/day)[c]	Manganese (mg/day)	Molybdenum (μg/day)	Nickel (mg/day)	Phosphorus (g/day)	Selenium (μg/day)	Silicon[d]	Vanadium (mg/day)[e]	Zinc (mg/day)
Infants																	
0–6 months	ND[f]	ND	ND	ND	ND	0.7	ND	40	ND	ND	ND	ND	ND	45	ND	ND	4
7–12 months	ND	ND	ND	ND	ND	0.9	ND	40	ND	ND	ND	ND	ND	60	ND	ND	5
Children																	
1–3 years	ND	3	2.5	ND	1000	1.3	200	40	65	2	300	0.2	3	90	ND	ND	7
4–8 years	ND	6	2.5	ND	3000	2.2	300	40	110	3	600	0.3	3	150	ND	ND	12
Males, Females																	
9–13 years	ND	11	2.5	ND	5000	10	600	40	350	6	1100	0.6	4	280	ND	ND	23
14–18 years	ND	17	2.5	ND	8000	10	900	45	350	9	1700	1.0	4	400	ND	ND	34
19–70 years	ND	20	2.5	ND	10000	10	1100	45	350	11	2000	1.0	4	400	ND	1.8	40
>70 years	ND	20	2.5	ND	10000	10	1100	45	350	11	2000	1.0	3	400	ND	1.8	40
Pregnancy																	
≤18 years	ND	17	2.5	ND	8000	10	900	45	350	9	1700	1.0	3.5	400	ND	ND	34
19–50 years	ND	20	2.5	ND	10000	10	1100	45	350	11	2000	1.0	3.5	400	ND	ND	40
Lactation																	
≤18 years	ND	17	2.5	ND	8000	10	900	45	350	9	1700	1.0	4	400	ND	ND	34
19–50 years	ND	20	2.5	ND	10000	10	1100	45	350	11	2000	1.0	4	400	ND	ND	40

Sources. Reprinted, with permission, from Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC; Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC; Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids." National Academies Press, Washington, DC. (These reports are available at http://www.nap.edu). Copyright 2001 by the National Academy of Sciences. All rights reserved.

[a] UL indicates the maximum level of daily nutrient intake that is likely to pose no risk of adverse effects. Unless otherwise specified, the UL represents total intake from food, water, and supplements. Due to lack of suitable data, ULs could not be established for arsenic, chromium, and silicon. In the absence of ULs, extra caution may be warranted in consuming levels above recommended intakes.

[b] Although the UL for arsenic, was not determined, there is no justification for adding arsenic to food or supplements.

[c] The ULs for magnesium represent intake from a pharmacological agent only and do not include intake from food and water.

[d] Although silicon has not been shown to cause adverse effects in humans, there is no justification for adding silicon to supplements.

[e] Although vanadium in food has not been shown to cause adverse effects in humans, there is no justification for adding vanadium to food and vanadium supplements should be used with caution. The UL is based on adverse effects in laboratory animals and these data could be used to set a UL for adults but not children and adolescents.

[f] ND, not determinable due to lack of data of adverse effects in this age group and concern with regard to lack of ability to handle excess amounts. Source of intake should be from food only to prevent high levels of intake.

There is no established benefit for healthy individuals in consuming nutrient levels above the RDA or AI. The amounts of individual foods that most people eat rarely reach levels that are likely to exceed the UL. Nutrient supplements that provide more concentrated amounts of nutrients per dose have a somewhat greater potential risk of excess.

TOTAL ENERGY EXPENDITURE AND ESTIMATED ENERGY REQUIREMENT

There is no RDA or AI for energy because although there are no adverse effects from consuming slightly more minerals, protein, or energy than is necessary to meet one's requirement, small amounts of energy excess cause weight gain. That is, a RDA for energy that exceeded the energy requirements of 97 to 98% of all individuals in the population would cause weight gain. Therefore, energy needs are stated as minimal average requirements. The EER is the requirement for energy for individuals of normal weight with a body mass index of between 18.5 and 25 for adults. This reflects the energy expenditure associated with an individual's sex, age, height, weight, and physical activity level. The TEE is a set of equations that estimate the energy expenditure needed to maintain current body

TABLE VI Dietary Reference Intakes: Recommended Intakes for Individuals, Macronutrients

Life-stage group	Carbohydrate (g/day)	Total fiber (g/day)	Fat (g/day)	Linoleic acid (g/day)	α-Linolenic acid (g/day)	Protein[a] (g/day)
Infants						
0–6 months	60	ND	31	4.4	0.5	9.1
7–12 months	95	ND	30	4.6	0.5	13.5
Children						
1–3 years	**130**	19	ND	7	0.7	**13**
4–8 years	**130**	25	ND	10	0.9	**19**
Males						
9–13 years	**130**	26	ND	12	1.2	**34**
14–18 years	**130**	38	ND	16	1.6	**52**
19–30 years	**130**	38	ND	17	1.6	**56**
31–50 years	**130**	38	ND	17	1.6	**56**
51–70 years	130	30	ND	14	1.6	**56**
>70 years	**130**	30	ND	14	1.6	**56**
Females						
9–13 years	**130**	31	ND	10	1.0	**34**
14–18 years	**130**	26	ND	11	1.1	**46**
19–30 years	**130**	25	ND	12	1.1	**46**
31–50 years	**130**	25	ND	12	1.1	**46**
51–70 years	**130**	21	ND	11	1.1	**46**
>70 years	**130**	21	ND	11	1.1	**46**
Pregnancy						
14–18 years	**175**	28	ND	13	1.4	**71**
19–30 years	**175**	28	ND	13	1.4	**71**
31–50 years	**175**	28	ND	13	1.4	**71**
Lactation						
14–18 years	**210**	29	ND	13	1.3	**71**
19–30 years	**210**	29	ND	13	1.3	**71**
31–50 years	**210**	29	ND	13	1.3	**71**

Source. Reprinted, with permission, from the Standing Committee on the Scientific Evaluation of Dietary Reference intakes, Food and Nutrition Board, Institute of Medicine (2002). "*Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids.*" National Academies Press, Washington, DC. (The report is available at http://www.nap.edu.) Copyright 2002 by the National Academy of Sciences. All rights reserved.

Note. This table presents Recommended Dietary Allowances (RDAs) in **boldface type** and adequate intakes (AIs) in lightface type. Both RDAs and AIs may be used as goals for individual intake. RDAs are set to meet the needs of almost all (97 to 98%) individuals in a group. For healthy breast-fed infants, the AI is the mean intake. The AI for other life-stage and gender groups is believed to cover the needs of all individuals in the group, but due to lack of data or uncertainty in the data the percentage of individuals covered by this intake cannot be specified with confidence.

[a] Based on 0.8 g protein/kg body weight for reference body weight.

TABLE VII Acceptable Macronutrient Distribution Ranges

Macronutrient	Range (as percentage of total energy)		
	Children, 1–3 years	Children, 4–18 years	Adults
Fat	30–40%	25–35%	20–35%
n–6 Polyunsaturated fats (linoleic acid)	5–10%	5–10%	5–10%
n–3 Polyunsaturated fats[a] (α-linolenic acid)	0.6–1.2%	0.6–1.2%	0.6–1.2%
Carbohydrate	45–65%	45–65%	45–65%
Protein	5–20%	10–30%	10–35%

Source. Reprinted, with permission, from the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2002). "*Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*". National Academies Press, Washington, DC. (The report is available at http://www.nap.edu.) Copyright 2002 by the National Academy of Sciences. All rights reserved.

[a]Approximately 10% of the total can come from longer-chain *n*-3 fatty acids.

weight and activity levels. Those who are obese may need to lose weight by taking in less energy and expending more energy than specified by these values in order to achieve optimal health.

ACCEPTABLE MACRONUTRIENT DISTRIBUTION RANGE

The acceptable macronutrient distribution ranges for individuals are presented for macronutrients as a proportion of total energy intakes. These values are estimated to minimize the potential for chronic disease over the long term, to permit essential nutrients to be consumed at adequate levels, and to be associated with adequate energy intakes and physical activity to maintain energy balance.The AMDRs for adults are as follows: 20–34% of calories should come from fat (of which 5–10% of calories should come from *n*-6 polyunsaturated fatty acids and 0.6 to 1.2% of calories should come from *n*–3 polyunsaturated fatty acids), 45–65% of calories should come from carbohydrates, and 10–35% of calories should come from protein. Although AMDRs for protein and carbohydrate do not vary with age, for children the AMDRs for total fat are 30–40% for ages 1–3 years and 25–35% for ages 4–18 years. In order to know with confidence that an individual's intake falls within the AMDR, the procedure is similar to that used to determine that usual intakes exceed the AI or remain below the UL. Necessary data include the individual's average intake of the macronutrient of interest as a percentage of his energy intake over a sufficient number of days, the day-to-day standard deviation of percentage energy intake within persons of that sex and life-stage group, and the range or boundaries of the AMDR. To determine the degree of confidence that intake is above the lower end of the AMDR, the equation for the AI is used. To determine the degree of confidence that intake is below the upper end of the AMDR, the equation for the UL is used. Table VI presents DRIs for macronutrients. Table VII presents AMDRs for fat, carbohydrate, and protein and Table VIII presents additional advice on intakes.

CONCLUSION

Each of the DRIs has specific uses in dietary assessment, for making nutritional recommendations and for dietary planning for individuals. The appropriate use of the

TABLE VIII Additional Macronutrient Recommendations

Dietary cholesterol	As low as possible while consuming a nutritionally adequate diet
Trans-Fatty acids	As low as possible while consuming a nutritionally adequate diet
Saturated fatty acids	As low as possible while consuming a nutritionally adequate diet
Added sugars	Limit to no more than 25% of total energy

Source. Reprinted, with permission, from the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2002). "*Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholestrol, Protein, and Amino Acids*." National Academies Press, Washington, DC. (The report is available at http://www.nap.edu.) Copyright 2002 by the National Academy of Sciences. All rights reserved.

reference values for dietary assessment and planning for both individuals and groups is described in detail in the various publications of the Food and Nutrition Board.

See Also the Following Articles

Carbohydrate Digestion and Absorption • Cobalamin Deficiency • Dietary Fiber • Fat Digestion and Absorption • Folate Deficiency • Iron Absorption • Nutritional Assessment • Protein Digestion and Absorption of Amino Acids and Peptides • Trace Minerals; Metabolism and Deficiency (Zinc, Copper, Selenium, Manganese) • Vitamin A: Absorption, Metabolism, and Deficiency • Vitamin B12: Absorption, Metabolism, and Deficiency • Vitamin K: Absorption, Metabolism, and Deficiency • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Panel on Dietary Antioxidants and Related Compounds, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids." National Academies Press, Washington, DC.

Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academies Press, Washington, DC.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2002). "Dietary Reference Intakes: Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids." National Academies Press, Washington, DC.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2001). "Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride." National Academies Press, Washington, DC.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and Its Panel on Folate, Other B Vitamins, and Choline, Food and Nutrition Board, Institute of Medicine (1998). "Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B-6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline." National Academies Press, Washington, DC.

Subcommittee on the Uses of the Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2000). "Dietary Reference Intakes: Applications in Dietary Assessment." National Academies Press, Washington, DC.

Subcommittee on the Uses of the Dietary Reference Intakes, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2002). "Dietary Reference Intakes: Applications in Planning" National Academies Press, Washington, DC.

Digestion, Overview

JAMES J. FARRELL
University of California, Los Angeles

digestion A process in which ingested nutrients are broken down into smaller components to facilitate their absorption by the small intestine.
endopeptidase An enzyme that cleaves peptide bonds within a dietary protein that are adjacent to certain specific amino acids.
exopeptidase An enzyme that removes a single amino acid from the carboxyl-terminal of a dietary peptide or protein.

The different parts of the gastrointestinal tract act in an integrated and coordinated manner under the control of neural and humoral regulatory mechanisms to ensure that nutrients are absorbed with remarkable efficiency: less than 5% of ingested carbohydrate, fat, and protein is excreted in the stool of adults following their normal diet. Absorptive mechanisms adapt to the nature and amount of various nutrients presented to the intestinal tract. These changes occur during early development, throughout life, and at times of specific need (e.g., during pregnancy).

INTRODUCTION

The cerebral (or cephalic) phase of digestion, whether triggered by the sight, smell, or thought of food, initiates the digestive process with the salivary and gastric secretory responses mediated via the autonomic nervous system. Good mastication and mixing with saliva initiate digestion of starch by salivary amylase and of fat by gastric lipase in the stomach. Protein digestion begins in the stomach with secretion of gastric pepsinogens and their rapid conversion to pepsins by gastric acid. Gastric emptying is carefully controlled so that food enters the duodenum at a controlled rate, thus allowing efficient mixing with pancreaticobiliary secretions. This ensures that the nutrient is presented optimally to the pancreatic enzymes, which function best at neutral pH. The major digestive processes occur in the duodenum.

The gallbladder is stimulated to contract and the pancreas to secrete simultaneously in response to the presence of nutrients in the duodenal lumen. The simultaneous release of bile salts, pancreatic enzymes, and bicarbonate provides optimal conditions for further nutrient digestion. The activation of pancreatic proteolytic enzymes by enteropeptidase released from the duodenal mucosa encourages proteolysis within the duodenal lumen rather than the pancreatic duct.

Once intestinal chyme leaves the ileum and enters the colon, most nutrients have been digested and absorbed and colonic function largely resolves itself into dehydration of lumenal contents through the absorption of salt and water. Dietary fiber may be digested by colonic bacteria, with release of short-chain fatty acids, which do not usually have much nutritional significance.

DIGESTION AND ABSORPTION OF FAT

Approximately 40% of adult energy requirements are supplied by lipids, of which triglycerides form the majority. The majority of fatty acids present in dietary triglyceride are oleate and palmitate. In animal triglyceride, most fatty acids are long-chain saturated fatty acids. Polyunsaturated fatty acids, such as linoleic and linolenic acid, are derived from phospholipid of vegetable origin. Cholesterol and phospholipids constitute the remaining types of lipid.

The problem of fat insolubility in water dominates the mechanisms that have been developed to digest and absorb lipid, which occurs mostly in the upper two-thirds of the jejunum. Liberation of fatty acids from the glycerol backbone of triglycerides (lipolysis) is achieved by lipases. This process occurs initially in the stomach and then in the small intestine. First, a stable emulsion of fat droplets of such a size that they present a large surface area to the digestive enzyme is required. This stable emulsion is then presented to pancreatic lipase. The presence of co-lipase, secreted by the pancreas with lipase (molar ratio 1 : 1), is critical in approximating lipase to triglyceride. The products of lipolysis are distributed between the aqueous, oil, and intermediate phases in a number of forms prepared for transfer across the lumen to the mucosal brush border membrane. The shuttling of these products is dependent, in part, on the formation of micelles with bile salts. Although the uptake of lipid digestion products by

 Copyright 2004, Elsevier (USA). All rights reserved.

enterocytes has been accepted as a passive process, it is possible that some lipids may be taken up by enterocytes via carrier-mediated processes that are energy dependent. The brush border membrane–fatty acid-binding protein, which is likely to be concerned with the transfer of fatty acids into the cell, could provide one explanation for facilitated diffusion and the observed saturability of fat absorption.

Once within the cell, fatty acids bind to specific fatty acid-binding proteins, which are found predominantly in the jejunum and more in villous cells than in crypt cells. They may assist transfer across the cytoplasm to the endoplasmic reticulum for triglyceride resynthesis. In the endoplasmic reticulum, triglyceride is resynthesized by two processes. In the first process, monoglyceride is reesterified with absorbed fatty acid after it has been activated to form acyl coenzyme A (the monoglyceride pathway). This route, involving monoglyceride esterification, accounts for the majority of the triglyceride synthesized during the absorptive phase. During fasting, triglyceride (and phospholipid) is synthesized via the second route, which involves acylation of α-glycerophosphate with the formation of phosphatidic acid and, thence, triglyceride or phospholipid.

Once synthesized, triglyceride, cholesterol and its esters, and phospholipids are packaged for export in the form of chylomicrons and very-low-density lipoproteins (VLDLs). During fasting, VLDLs are the major triglyceride-rich lipoproteins that emerge from the epithelium; after feeding, chylomicrons predominate. Once the chylomicrons have formed in the smooth endoplasmic reticulum, they are transferred to the Golgi apparatus. Golgi-derived chylomicron vesicles are then incorporated into the basolateral membrane and secreted by exocytosis into the lymphatic circulation. During absorption, lacteals distend, and endothelial cells, which overlap one another in the fasting state, move apart and open gaps through which chylomicrons can readily pass.

DIGESTION AND ABSORPTION OF CARBOHYDRATE

Carbohydrate provides 45% of total energy requirements. Approximately half of the digestible carbohydrate is starch derived from cereals and plants, in which it is the major storage form. Starch (as either amylose or amylopectin) is made up of long chains of glucose molecules. Other major sources of dietary carbohydrate include sugar derived from milk (lactose), sugars contained within the cells of fruit and vegetables (fructose, glucose, sucrose), or sugar purified from cane or beet sources (sucrose). Processed foods form a major source of dietary sugars, particularly fructose."Nonstarch polysaccharides" form the majority of the"unavailable" carbohydrate. This includes cellulose and hemicelluloses, both of which are resistant to digestion in the small bowel because their β-1-4 bond, unlike the α bond in starch, is resistant to amylases. They are, however, broken down to some extent by colonic bacteria to yield short-chain fatty acids, which are avidly absorbed by colonic mucosa.

Salivary amylase depends for its effect on its proximity to the ingested starches and the time spent within the mouth. Pancreatic amylase is the major enzyme of starch digestion and, as with salivary amylase, produces short oligosaccharides: maltotriose, maltose, and α-limit dextrins; glucose monomer is not produced. Most of this hydrolysis occurs within the lumen, but because amylase also attaches itself to the brush border membrane of enterocytes, some digestion may occur at this site.

The terminal products of lumenal starch digestion just mentioned, together with the major disaccharides in the diet (sucrose and lactose), cannot be absorbed intact and are hydrolyzed by specific brush border membrane hydrolases that are maximally expressed in the villi of duodenum and jejunum (Table I). Lactase hydrolyzes lactose to produce one molecule of glucose and one of galactose. Sucrase–isomaltase hydrolyzes sucrose to yield one molecule of glucose and one of fructose. In addition, two other carbohydrases participate in terminal hydrolysis of starch products. Maltase acts on 1-4-linked oligosaccharides containing as many as nine glucose residues, liberating glucose monomers. α-Limit dextrinase is responsible for rapid hydrolysis of penta- and hexa-α-limit dextrins.

The three major diet-derived monosaccharides, glucose, galactose, and fructose, are absorbed by saturable carrier-mediated transport systems located in the

TABLE I Brush Border Membrane Carbohydrases

Enzyme	Substrate	Products
Lactase	Lactose	Glucose Galactose
Maltase	α-1-4 linked oligosaccharides up to 9 residues	Glucose
Sucrase–isomaltase (sucrase–α-dextrinase)		
Sucrase	Sucrose	Glucose Fructose
Isomaltase	α-limit dextrin α 1-6 link	Glucose

brush border membrane of enterocytes. The active transport of glucose and galactose is achieved by the same transport protein that acts as a sodium cotransporter. Sodium enters the cell across the apical membrane moving down its concentration gradient, bringing with it glucose or galactose in a one-to-one molar ratio. Fructose absorption occurs by facilitated diffusion; that is, transport occurs not against a concentration gradient but with a carrier protein to achieve transport rates greater than one would expect from simple diffusion.

Most hexoses are exported from the epithelial cell by way of the basolateral membrane using a facilitated diffusion (not requiring energy) via a specific carrier. Two genes that are expressed in the small intestine, GLUT2, the basolateral membrane-associated glucose transporter, and GLUT5, an apical membrane fructose transporter, encode these facilitative sugar transport proteins. Once the hexoses have entered the interstitial space, they pass onward by diffusion into the portal circulation.

Not all potentially digestible carbohydrate is absorbed in the small intestine. As much as 20% of dietary starch may escape into the colon, particularly that derived from cereals and potatoes. Most of this is metabolized by colonic bacteria and the short-chain fatty acids thus derived are readily absorbed. Hydrogen and methane are also generated and they contribute to flatus.

DIGESTION AND ABSORPTION OF PROTEIN

Dietary proteins are the major source of amino acids and provide approximately 10 to 15% of energy intake. Digestion of proteins begins in the stomach with the action of pepsins, which are released from their precursor pepsinogens by autoactivation in an acid pH. Pepsins remain active at the acid pH of gastric contents to produce a mixture of peptides with a small portion of amino acids.

Each of the pancreatic proteases is secreted as a proenzyme and thus must be activated within the lumen by enterokinase (enteropeptidase), which is liberated from the brush border membrane (Table II). It acts to convert trypsinogen to trypsin, which in turn activates the other proteases and continues to split more trypsin from trypsinogen. Trypsin, chymotrypsin, and elastase (endopeptidases) have specificity for peptide bonds adjacent to certain specific amino acids. They split peptide bonds in the protein molecule, whereas exopeptidases remove a single amino acid from the carboxyl-terminal end of the peptide. The final products of intralumenal digestion are thus produced by cooperative activity of endo- and exopeptidases and consist of a number of neutral and basic amino acids together with peptides of two to six amino acids in length.

In contrast to the absorption of carbohydrate, which is largely restricted to uptake of hexose monomers across the brush border membrane, amino acids can be absorbed either as monomers or as di- or tripeptides. However, the fact that the vast majority of the end products of protein digestion that reach the portal circulation are amino acids speaks strongly in favor of the presence of peptidases in the epithelium.

Evidence suggests that small peptides utilize a separate transporter system from those utilized by single amino acids. Di- and tripeptides are taken up by a single type of transporter and there is some stereospecificity because the length of the amino acid side chains on the di- or tripeptides is important; the longer the side chain, the more preferred the substrate for the absorption site. The L-isomers of the amino acids in dipeptides are much preferred to the D-isomers, whereas the presence of acidic and basic amino acid residues in dipeptides

TABLE II Pancreatic Proteolytic Enzymes

Enzyme	Action	Products
Trypsin	Endopeptidase; cleaves internal bonds at lysine or arginine residues; cleaves other pancreatic proenzymes	Oligopeptides
Chymotrypsin	Endopeptidase; cleaves bonds at aromatic or neutral amino acid residues	Oligopeptides
Elastase	Endopeptidase; cleaves bonds at aliphatic amino acid residues	Oligopeptides
Carboxypeptidase A	Exopeptidase; cleaves aromatic amino acids from carboxy-terminal end of protein and peptides	Aromatic amino acids and peptides
Carboxypeptidase B	Exopeptidase; cleaves arginine or lysine from carboxy-terminal end of proteins and peptides	Arginine, lysine, and peptides

reduces affinity for the transport system, compared with neutral amino acid residues. Affinity is also greater for dipeptides than for tripeptides, at least in the example of peptides containing glycine. The transporter for peptides is not dependent on sodium, but cotransport with protons may occur instead.

Although there appears to be only one type of dipeptide transporter in the brush border membrane for the 400 different possible dipeptides, there is a multiplicity of transport mechanisms for the 20 amino acids. These are sited on villous enterocytes and involve carrier-mediated active transport or facilitated diffusion processes; a small proportion may be absorbed by simple diffusion. Separate sodium-dependent, active transport processes for basic and acidic amino acids have also been demonstrated and there is some evidence to suggest that facilitated diffusion of these types of amino acids also occurs, although this is likely to be a minor pathway.

The intestinal basolateral membrane possesses a set of amino acid transport systems that are different from those in the brush border membrane. The amino acid transport systems in the basolateral membrane function to export amino acids from the enterocytes into the portal circulation during feeding. They also participate in the import of amino acids from the circulation into the enterocyte for cellular metabolism when amino acids are not available from the intestinal lumen, such as between meals. The intestinal basolateral membrane also possesses a peptide transporter system that is probably identical to that in the brush border membrane. This transport system facilitates the exit of hydrolysis-resistant small peptides from the enterocyte into the portal circulation.

VITAMINS

Water-Soluble Vitamins

With the loss of the capacity for hepatic synthesis, a specific absorptive mechanism for vitamin C has developed in humans. Folic acid consists of the complex pterin molecule conjugated to *para*-aminobenzoic acid and glutamic acid. Although much dietary folate is in the form of polyglutamates comprising at least six glutamic acid residues, much is present as formyl- and methylhydrofolate. Absorption of dietary polyglutamates depends on hydrolysis at the brush border membrane followed by transport into the cytoplasm. Uptake is achieved by a specific carrier-mediated, sodium-dependent, pH-sensitive process that is active at acid pH. Specific sodium-dependent active transport processes have also been demonstrated for thiamine, riboflavin, pantothenic acid, and biotin.

Vitamin B12 (cobalamin) exists largely as hydroxycobalamin, methylcobalamin, and adenosylcobalamin and these are found almost entirely in animal sources. Three types of binding proteins are involved in the absorption of cobalamin: one in saliva, one in gastric juice, and one in the circulation. The vitamin is released by gastric acid and pepsin from the various dietary proteins with which it is associated. The first specific binding protein secreted in saliva, the R protein, takes up the free cobalamin and binds it with strong affinity. In the duodenum, where the R protein is hydrolyzed by pancreatic enzymes, that intrinsic factor is able to bind the cobalamin that has been released. This complex resists pancreatic proteolysis, passes down the intestine to the terminal ileum, and there binds to specific receptors on ileal enterocytes. After binding to the receptor, the intrinsic factor–cobalamin complex probably enters the cell intact by translocation. Free cobalamin leaves the base of the cell, where it is immediately bound to an ileal pool of transcobalamin II, which transports it into the portal circulation.

Fat-Soluble Vitamins

Vitamin A (retinol) is absorbed in the small bowel. Transport across the apical membrane appears to occur by passive diffusion. Vitamin A leaves the mucosa largely in chylomicrons as retinyl palmitate. Most of a person's requirement for vitamin D is supplied by endogenous synthesis in the skin during exposure to sunlight and dietary intake becomes critical only when such exposure is inadequate. Like vitamin A, vitamin D absorption occurs by simple passive diffusion in the small intestine. Bile salts are unnecessary, but lumenal pH influences absorption. Absorption is reduced at neutral pH and increased in acid. Most absorbed vitamin D passes into the lymphatics unchanged. Vitamin E is absorbed passively across the intestinal mucosa. The ester form, in which many vitamin preparations are presented, is hydrolyzed prior to absorption, but the ester can be absorbed intact. Vitamin E is transported into the lymphatics largely unchanged.

Vitamin K is found in two forms. K_1, the major dietary form, is found in green vegetables, but beef liver is another good source. K_2 is produced by colonic bacteria, and although some may be absorbed from the colon this alone is an inadequate source if K_1 absorption is impaired. Absorption of K_1 from the small intestine is dependent on lumenal bile salts and uptake is achieved by a carrier-mediated process, whereas K_2 absorption is entirely passive.

MINERALS AND TRACE ELEMENTS

Various divalent ions are essential nutrients; some are absorbed in milligram amounts and are major constituents of the body, whereas others are necessary only in trace amounts. Iron, calcium, magnesium, phosphorus, and sulfur are in the former category and specialized absorptive mechanisms are involved in their assimilation. Zinc, copper, iodine, and selenium are considered trace elements and their absorption is not as well understood.

Absorption of calcium across the intestinal mucosa is achieved by two parallel processes: an active, transcellular transport process and a passive, paracellular diffusive process. Under normal dietary conditions, the duodenum is the major site for active transport, whereas passive, paracellular transfer occurs throughout the small intestine. Despite this localization of the active transport site, quantitatively more calcium may be absorbed in the jejunum and ileum than in the duodenum because of the relative amounts of time the lumenal contents spend in these regions of the intestine. The human jejunum absorbs calcium faster than the ileum and absorption rates in both are increased by treatment with vitamin D. The rate-limiting step in the absorption process is the intracellular calbindin concentration, which is regulated by a metabolite of vitamin D, 1,25-dihydroxyvitamin D_3, produced in the kidneys from 25-hydroxyvitamin D_3 converted by the liver from absorbed vitamin D.

In contrast to calcium, magnesium absorption in the basal state is greater in the human ileum than in the jejunum. Jejunal magnesium absorption is increased by vitamin D, whereas ileal absorption is not. Ileal transport involves both a paracellular, diffusive pathway and a transcellular, carrier-mediated, saturable process. There is some competition from calcium for the diffusive pathway but not for the saturable, presumably carrier-mediated process.

Most iron absorption occurs in the proximal small intestine and the ferrous (Fe^{2+}) form is absorbed better than the ferric (Fe^{3+}) form. The latter is insoluble at pH values above 3 and gastric acid and some sugars and amino acids render it more available for absorption. The presence of some anions, such as oxalate, phosphate, and phytate, precipitate iron out of solution and reduce its absorption. The iron in hemoglobin and myoglobin is well absorbed, and although these apparently compete with inorganic iron for absorption, the organic molecules are absorbed intact by a separate process. An iron transport protein, called divalent metal ion transporter 1 (DMT-1, also called DCT-1 or Nramp2), that is responsible for dietary iron absorption in the villus cells of the small intestine has been identified. The stimuli for increased iron absorption include iron deficiency, hypoxia, increased erythropoiesis, and pregnancy, but the circulating signal to the intestine is unknown.

ADAPTATION TO CHANGES IN NEED OR LOAD

One of the most fascinating aspects of intestinal function is the phenomenon of adaptation. Two specific forms of intestinal adaptation have been identified in the intestine: (1) mucosal hypertrophy leading to a global increase in absorption of all nutrients and (2) an increase in specific transport mechanisms induced in response to specific dietary needs or availability.

Resection of more than 50% of the human intestine results in increased fecal nitrogen losses, which slowly return toward normal, thus indicating that mucosal adaptation has occurred. This is due largely to hypertrophy of intestinal mucosa, which is manifested in increases in the number of villous enterocytes and in villous height without an obvious increase in the absorption rate per individual cell. Absorption increases for all nutrients and absorptive capacity may be enhanced up to fivefold in response to intestinal resection.

The digestive capacity of pancreatic juice can be altered by changes in nutritional intake. A high-protein diet enhances proteolytic enzyme production, a high-carbohydrate diet enhances amylase secretion, and a high-fat diet stimulates lipase secretion. Furthermore, adaptive responses to changes in dietary intake influence mucosal digestive and absorptive processes. Activity of the disaccharidase enzymes sucrase and maltase increases in response to high carbohydrate intake over several days but not to manipulation of protein intake. Absorptive function also adapts to dietary manipulation. Transport activity for carbohydrates, fats, and lipids rises in response to increased dietary loads. In another example, absorptive mechanisms for a number of vitamins and trace elements are switched on by low dietary loads and switched off by a large load. Here, absorption is enhanced in nutrient deficiency but inhibited with nutrient excess, when potentially toxic effects may result.

See Also the Following Articles

Carbohydrate Digestion and Absorption • Dietary Reference Intakes (DRI): Concepts and Implementation • Fat Digestion and Absorption • Malabsorption • Pancreatic Digestive Enzymes • Protein Digestion and Absorption of Amino Acids and Peptides • Trace Minerals: Metabolism and Deficiency (Zinc, Copper, Selenium, Manganese) • Vitamin

A: Absorption, Metabolism, and Deficiency • Vitamin B12: Absorption, Metabolism, and Deficiency • Vitamin K: Absorption, Metabolism, and Deficiency • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Alpers, D. H. (1987). Digestion and absorption of carbohydrates and proteins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., p. 1469. Raven Press, New York.

Farrell, J. J. (2002). Digestion and absorption of nutrients and vitamins. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., W. B. Saunders, Philadelphia, PA.

Farrell, J. J. (2002). Overview and diagnosis of malabsorption syndrome. *Semin. Gastrointest. Dis.* **13**, 182–190.

Ferraris, R. P., and Diamond, J. M. (1989). Specific regulation of intestinal nutrient transporters by their dietary substrates. *Annu. Rev. Physiol.* **51**, 125–141.

Ganapathy, V., Brandsch, M., and Leibach, F. H. (1994). Intestinal transport of amino acids and peptides. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1782. Raven Press, New York.

Karasov, W. H., and Diamond, J. M. (1987). Adaptation of intestinal nutrient transport. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., p. 1489. Raven Press, New York.

Leibach, F. H., and Ganapathy, V. (1996). Peptide transporters in the intestine and the kidney. *Annu. Rev. Nutr.* **16**, 99–119.

Phan, C. T., and Tso, P. (2001). Intestinal lipid absorption and transport. *Front. Biosci.* **6**, 299–319.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1873. Raven Press, New York.

Disinhibitory Motor Disorder

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

achalasia Pathologic failure of relaxation of gastrointestinal sphincters and nonsphincteric smooth muscles.

interstitial cells of Cajal Pacemaker cells that generate electrical slow waves in the gastrointestinal tract.

megacolon Dilatation of the colon.

paraneoplastic syndrome Autoimmune pathology in tissues of patients with certain forms of cancer.

pseudo-obstruction Failure of propulsive motility that cannot be explained by mechanical blockage in the gastrointestinal tract.

syncytium Group of cells that behave as a unit due to some form of coupling between cells (electrical coupling in smooth muscles).

Trypanosoma cruzi Blood-borne protozoan parasite responsible for Chagas' disease.

Postnatal loss of, or congenital absence of, the enteric nervous system has pathologic effects on intestinal motility. In disinhibitory motor disorder, the pathophysiologic basis is loss or malfunction of enteric inhibitory motor neurons, to the musculature causing the spasticity and achalasia associated with a number of symptoms and diseases.

INTRODUCTION

In the enteric nervous system, a subpopulation of the motor neurons consists of inhibitory motor neurons to the musculature. Neurotransmitters released from inhibitory motor neurons suppress contractile activity in the musculature. Inhibitory motor neurons are among the missing neurons when the enteric nervous system is either destroyed by disease processes or fails to develop *in utero*. The primary pathologic effect in disinhibitory motor disorder is chronic intestinal pseudo-obstruction that is associated with hyperactive contractile activity of the circular coat of the intestinal musculature.

Copyright 2004, Elsevier (USA). All rights reserved.

A: Absorption, Metabolism, and Deficiency • Vitamin B12: Absorption, Metabolism, and Deficiency • Vitamin K: Absorption, Metabolism, and Deficiency • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Alpers, D. H. (1987). Digestion and absorption of carbohydrates and proteins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., p. 1469. Raven Press, New York.

Farrell, J. J. (2002). Digestion and absorption of nutrients and vitamins. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., W. B. Saunders, Philadelphia, PA.

Farrell, J. J. (2002). Overview and diagnosis of malabsorption syndrome. *Semin. Gastrointest. Dis.* **13**, 182–190.

Ferraris, R. P., and Diamond, J. M. (1989). Specific regulation of intestinal nutrient transporters by their dietary substrates. *Annu. Rev. Physiol.* **51**, 125–141.

Ganapathy, V., Brandsch, M., and Leibach, F. H. (1994). Intestinal transport of amino acids and peptides. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1782. Raven Press, New York.

Karasov, W. H., and Diamond, J. M. (1987). Adaptation of intestinal nutrient transport. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., p. 1489. Raven Press, New York.

Leibach, F. H., and Ganapathy, V. (1996). Peptide transporters in the intestine and the kidney. *Annu. Rev. Nutr.* **16**, 99–119.

Phan, C. T., and Tso, P. (2001). Intestinal lipid absorption and transport. *Front. Biosci.* **6**, 299–319.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., p. 1873. Raven Press, New York.

Disinhibitory Motor Disorder

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

achalasia Pathologic failure of relaxation of gastrointestinal sphincters and nonsphincteric smooth muscles.

interstitial cells of Cajal Pacemaker cells that generate electrical slow waves in the gastrointestinal tract.

megacolon Dilatation of the colon.

paraneoplastic syndrome Autoimmune pathology in tissues of patients with certain forms of cancer.

pseudo-obstruction Failure of propulsive motility that cannot be explained by mechanical blockage in the gastrointestinal tract.

syncytium Group of cells that behave as a unit due to some form of coupling between cells (electrical coupling in smooth muscles).

Trypanosoma cruzi Blood-borne protozoan parasite responsible for Chagas' disease.

Postnatal loss of, or congenital absence of, the enteric nervous system has pathologic effects on intestinal motility. In disinhibitory motor disorder, the pathophysiologic basis is loss or malfunction of enteric inhibitory motor neurons, to the musculature causing the spasticity and achalasia associated with a number of symptoms and diseases.

INTRODUCTION

In the enteric nervous system, a subpopulation of the motor neurons consists of inhibitory motor neurons to the musculature. Neurotransmitters released from inhibitory motor neurons suppress contractile activity in the musculature. Inhibitory motor neurons are among the missing neurons when the enteric nervous system is either destroyed by disease processes or fails to develop *in utero*. The primary pathologic effect in disinhibitory motor disorder is chronic intestinal pseudo-obstruction that is associated with hyperactive contractile activity of the circular coat of the intestinal musculature.

Copyright 2004, Elsevier (USA). All rights reserved.

ENTERIC INHIBITORY MOTOR NEURONS

The physiologic importance of enteric inhibitory motor neurons emerges from requirements for compatibility of neural control with the physiology of the digestive musculature. The intestinal musculature behaves like a self-excitable electrical syncytium. Action potentials triggered anywhere in the muscle will spread from muscle fiber to muscle fiber in three dimensions throughout the syncytium, which could potentially be the entire length of the small or large intestine. The action potentials, as they spread, trigger contraction of the muscle. A nonneural pacemaker system of electrical slow waves (i.e., basic electrical rhythm) accounts for the self-excitable characteristic of the electrical syncytium; this is accomplished by specialized pacemaker cells, the interstitial cells of Cajal. In the integrated system, the electrical slow waves are an extrinsic factor to which the intestinal circular muscle responds with action potentials, which in turn trigger contractions.

The cellular neurophysiology of enteric inhibitory motor neurons explains why the intestinal circular muscle generally fails to respond with action potentials and contractions during each and every slow-wave cycle. Ongoing activity of the inhibitory motor neurons also explains why action potentials and contractions do not spread from muscle fiber to muscle fiber and cause an entire length of intestine to contract after they are initiated at any site in the syncytium. Inhibitory motor neurons to intestinal circular muscle discharge continuously. Continuous discharge of action potentials in subpopulations of intestinal inhibitory motor neurons results in continuous inhibition of the autogenous activity of the circular muscle. Studies done in segments of intestine *in vitro* show that when neuronal discharge of inhibitory motor neurons is prevalent, muscle action potentials and associated contractile activity are absent or occur only at reduced levels with each electrical slow-wave cycle. When the inhibitory neuronal discharge is blocked experimentally, every cycle of the electrical slow wave triggers intense discharge of action potentials and large-amplitude contractions. Vasoactive intestinal polypeptide and nitric oxide are two of the inhibitory neurotransmitters released at the neuromuscular junctions. Continuous release of these transmitters can be detected experimentally in intestinal preparations.

PATHOPHYSIOLOGIC BASIS

The physiology of neuromuscular relations in the intestine predicts that spasticity and achalasia will accompany any condition wherein inhibitory motor neurons are destroyed. Without inhibitory control, the self-excitable syncytium of nonsphincteric regions will contract continuously and behave as an obstruction. This happens because the muscle is freed to respond to the pacemaker with contractions that propagate without any amplitude, distance, or directional control. Contractions spreading in the uncontrolled syncytium collide randomly, resulting in fibrillation-like behavior in the affected intestinal segment.

Loss or malfunction of inhibitory motor neurons is the pathophysiologic basis of disinhibitory motor disease. It underlies several forms of chronic intestinal pseudo-obstruction and failure of relaxation of tension in sphincters of the digestive tract. Enteric neurodegeneration in its earlier stages may be manifest as symptoms that can be confused with the irritable bowel syndrome.

Neuropathic forms of intestinal pseudo-obstruction arise from degenerative changes in the enteric nervous system. Failure of propulsive motility in the affected length of intestine reflects loss of the neural microcircuits that program and control the repertoire of motility patterns required for the necessary functions of that region. Intestinal pseudo-obstruction occurs in part because contractile behavior of the circular muscle is hyperactive but disorganized in the denervated regions. Hypercontractile activity is a diagnostic sign of the neuropathic form of chronic intestinal pseudo-obstruction in humans. The hyperactive and disorganized contractile behavior reflects the absence of inhibitory nervous control of the muscles that are self-excitable (autogenic) when released from the braking action imposed by inhibitory motor neurons.

Degenerative noninflammatory and inflammatory enteric neuropathies are two forms of disinhibitory motor disease that culminate in pseudo-obstruction. Noninflammatory neuropathies can be either familial or sporadic. The mode of inheritance may be autosomal recessive or dominant.

Degenerative inflammatory enteric neuropathies are characterized by a dense inflammatory infiltrate confined to enteric ganglia. Paraneoplastic syndrome, Chagas' disease, and idiopathic degenerative disease are recognizable forms of pseudo-obstruction related to inflammatory neuropathies. Paraneoplastic syndrome is a form of pseudo-obstruction in which commonalty of antigens between small-cell carcinoma of the lungs and enteric neurons leads to immune attack, resulting in loss of neurons. The majority of patients with symptoms of pseudo-obstruction in combination with small-cell lung carcinoma have immunoglobulin G autoantibodies that react with enteric neurons. Immunostaining with sera from paraneoplastic patients shows a characteristic pattern of staining in enteric neurons. The detection of

antienteric neuronal antibodies in the patient's serum is a means to a specific diagnosis.

The association of enteric neuronal loss with symptoms of pseudo-obstruction in Chagas' disease also reflects autoimmune attack on the neurons, mimicking the situation in the paraneoplastic syndrome. *Trypanosoma cruzi*, the blood-borne parasite that causes Chagas' disease, has antigenic epitopes similar to antigens expressed on the surfaces of enteric neurons. This activates the immune system to assault the gut neurons coincident with the attack on the parasite.

Idiopathic inflammatory degenerative neuropathy occurs unrelated to neoplasms, infectious conditions, or other known diseases. Small groups of patients have been identified with early complaints of symptoms similar to the irritable bowel syndrome; in these patients, the symptoms progressively worsened and were later diagnosed as idiopathic degenerative inflammatory neuropathy based on full-thickness biopsies taken during exploratory laparotomy that revealed chronic intestinal pseudo-obstruction. Each patient had inflammatory infiltrates localized to the enteric nervous system. Analysis of sera from these cases shows circulating antibodies against enteric neurons.

Recognition of the complex functions of the enteric nervous system leads to the conclusion that early neuropathic changes are expected to be manifest as functional symptoms that worsen with progressive neuronal loss. In diagnostic motility studies (e.g., manometry), degenerative loss of enteric neurons is reflected by hypermotility and spasticity because inhibitory motor neurons are included in the missing neuronal population.

See Also the Following Articles

Achalasia • Chagas' Disease • Colonic Obstruction • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Interstitial Cells of Cajal • Intestinal Pseudo-obstruction • Paraneoplastic Syndrome

Further Reading

Lennon, V. (1994). Paraneoplastic autoantibodies. *Neurology* **44**, 2236–2240.

Mayer, E., Schuffler, M., and Rotter, J. (1986). Familial visceral neuropathy with autosomal dominant transmission. *Gastroenterology* **91**, 1528–1535.

Roy, A., Bharucha, H., Nevin, N., and Odling-Smee, G. (1980). Idiopathic intestinal pseudo-obstruction: A familial visceral neuropathy. *Clin. Genet.* **18**, 292–297.

Schuffler, M., Baird, H., and Fleming, C. (1983). Intestinal pseudo-obstruction as the presenting manifestation of small cell carcinoma of the lung: A paraneoplastic neuropathy of the gastrointestinal tract. *Ann. Intern. Med.* **98**, 129–134.

Smith, V., Gregson, N., and Foggensteiner, L. (1997). Acquired intestinal aganglionosis and circulating autoantibodies without neoplasia or other neural involvement. *Gastroenterology* **112**, 1366–1371.

Stanghellini, V., Camilleri, M., and Malagelada, J.-R. (1987). Chronic idiopathic intestinal pseudoobstruction. Clinical and intestinal manometric findings. *Gut* **23**, 824–828.

Wood, J. (2000). Neuropathy in the brain-in-the-gut. *Eur. J. Gastroenterol. Hepatol.* **12**, 597–600.

Wood, J., Alpers, D., and Andrews, P. (1999). Fundamentals of neurogastroenterology. *Gut* **45**(Suppl. II), II6–II16.

Wood, J., Hudson, L., Jessel, T., and Yamamoto, M. (1982). A monoclonal antibody defining antigenic determinants on subpopulations of mammalian neurones and *Trypanosoma cruzi* parasites. *Nature (London)* **296**, 34–37.

Diverticulosis

BRUCE HENNESSY AND SHERYL PFEIL
Ohio State University

diverticula of the colon Herniations of the mucosa and submucosa though or between fibers of the major muscle layer (muscularis propria) of the colon.
diverticulosis Presence of multiple diverticula of the intestine, common in middle age; the lesions are acquired pulsion diverticula.
diverticulum Pouch or sac opening from a tubular or saccular organ, such as the gut or bladder.

Colonic diverticular disease is extremely common, particularly in the middle-aged and older populations. A colonic diverticulum is a herniation through the muscularis, generally at the site of a nutrient artery. Colonic diverticuli are rarely found in patients under the age of 30 years, and occur with increasing prevalence as a population ages. More than two-thirds of the United States population will have diverticulosis of varying degrees by age 80 years. There is no reported difference in prevalence between men and women.

EPIDEMIOLOGY

The ethnic and geographic variation in colonic diverticulosis is marked. Left-sided diverticulosis is classically associated with a Western diet, and to a lesser degree European ancestry. Although left-sided diverticulosis is still common in East Asians and Indians, right-sided diverticulosis is far more common in these patients than it is in the Western population. The prevalence of diverticulosis has steadily increased over the past 80 years. The Western diet, specifically the lack of high intake of dietary fiber, has been blamed for this change. Another theory regarding the modern increase in diverticulosis is the lower physical activity level of people in industrialized countries. Higher levels of physical activity seem to reduce the risk of diverticular disease.

PATHOPHYSIOLOGY

The formation of acquired colonic diverticulosis is thought to be secondary to increased intraluminal pressure. The lumen of the sigmoid colon narrows when the circular muscle layer enlarges, causing shortened taeniae. Pressure increases as luminal narrowing occurs, according to the law of Laplace. The luminal pressure is exaggerated by the chambering effect of colonic motility, whereby chambers are created by colonic segmentation. Diverticuli are most likely to develop at weak places in the bowel wall. The insertion of the vasa recta is thought to be such a weak point.

UNCOMPLICATED DIVERTICULOSIS

Colonic diverticuli are often diagnosed incidentally. They are often observed during screening flexible sigmoidoscopy and colonoscopy. They also may be apparent on plain abdominal films, computed tomography (CT) scans, or barium enemas. The importance of diagnosing asymptomatic diverticulosis is questionable. The mere presence of diverticulosis does not require specific therapy other than recommending a high-fiber diet. A high-fiber diet appears to be associated with a lower risk for developing diverticulosis. Treatment and prevention call for the same lifestyle changes. Although there is no evidence that dietary change will reverse diverticulosis, a high-fiber diet may decrease the risk of developing diverticular complications. Other dietary changes are controversial. The age-old advice to avoid nuts and seeds is based on their theoretical risk of becoming lodged in a diverticulum and inciting diverticulitis. This theory is unproved, and there is considerable debate with regard to the necessity of these dietary recommendations.

DIVERTICULITIS

Diverticulitis and diverticular bleeding are the most common complications of diverticulosis. Diverticulitis occurs from the erosion of the diverticular wall by focal necrosis and inflammation, resulting in perforation. Perforation occurs on two different levels, micro- or macroperforation. A microperforation is walled off by the adjacent mesentery and pericolic fat. The microperforation may resolve or it may form a focal

 Copyright 2004, Elsevier (USA). All rights reserved.

abscess, or lead to obstruction or fistula formation to an adjacent organ. Macroperforation causes free air entry and peritonitis.

Diverticulitis typically presents with left lower quadrant pain. The pain is often present for several days prior to presentation. Diverticulitis is sometimes accompanied by nausea and vomiting or constipation. Less commonly, a patient may experience diarrhea or narrow stools. Physical exam findings provide a limited but important cornerstone in the diagnosis of diverticulitis. The most common finding on examination is left lower quadrant tenderness. An uncommon but more impressive finding is that of a palpable left lower quadrant mass. Physical examination findings of general tenderness and peritonitis may reflect perforated diverticulitis. Exam findings, although important, will rarely provide enough information to make a diagnosis of diverticulitis.

Laboratory studies may be equally nonspecific. The patient will generally have a leukocytosis, indicating an infective or inflammatory process. Diverticulitis that occurs in close proximity to the bladder or that develops a colovesical fistula may cause pyuria on urinalysis and can be mistaken for a urinary tract infection. Other laboratory tests, including liver enzymes, amylase, and lipase, are generally normal. Chemistries are also commonly normal unless the patient presents with sepsis and acidosis.

The history and physical findings may prompt further evaluation with imaging studies. Generally, an acute abdominal series is obtained first. Plain abdominal films are useful to exclude macroperforation with free air perforation and colonic obstruction. Prior to giving the patient oral contrast, both obstruction and free air must be ruled out. Computed tomography is the generally accepted "gold standard" for diagnosing diverticulitis. Common CT findings of diverticulitis include bowel wall thickening, colonic diverticulosis, soft tissue density within the pericolic fat, and an abscess or phlegmon. Computed tomography has a very high sensitivity and specificity for detecting diverticulitis, which makes CT imaging a powerful tool in the diagnosis of diverticulitis. It may also provide additional information to exclude other causes of abdominal inflammation. Barium enema is generally not warranted in suspected diverticulitis secondary to the potential for barium to escape through a perforation into the peritoneal space. If performed in the face of suspected diverticulitis, the enema should consist of water-soluble contrast only. There is also no role for colonoscopy in acute diverticulitis. Colonoscopy is contraindicated because of the risk of perforation. Colonoscopy should be performed following resolution of the acute episode. Colonoscopy is performed after diverticulitis subsides to evaluate for tumors, polyps, and strictures, and to evaluate the extent of diverticular disease.

Treatment of acute diverticulitis depends on the severity of the process. A patient with mild to moderate diverticulitis can be treated as an outpatient with oral antibiotics and bowel rest. More severe disease requires inpatient therapy. Antibiotic regimens for diverticulitis must cover gut flora, including both gram-negative and anaerobic organisms. Uncomplicated diverticulitis is treated with oral antibiotics such as ciprofloxacin and metronidazole for 7–10 days. Patients should also be placed on a clear liquid diet. Patients with more complicated disease require intravenous antibiotics, bowel rest, and intravenous fluids. Diverticulitis may be complicated by abscess formation, obstruction, fistula formation, or peritonitis. Percutaneous drainage of an abscess can be performed by interventional radiologic technique. Abscess drainage and concurrent antibiotic treatment are sufficient treatment in a subgroup of patients, especially patients not considered surgical candidates. Percutaneous drainage may also allow time for bowel preparation and resolution of inflammation, thereby allowing resection and primary anastomosis with a one-step surgical procedure. Most patients with fistulas and/or obstruction do require surgery, but not emergently. Peritonitis is a more severe complication that requires immediate surgical intervention. Emergent surgery in this setting often involves a two-step procedure consisting of a resection of the diseased colon and colostomy. This is followed by colostomy takedown and reanastomosis at a later date.

Patients suffering recurrent diverticulitis may also be referred for elective surgery. Surgery involves removing the diseased portion of the colon as well as removing the sigmoid colon. The sigmoid colon is removed to eliminate the "high-pressure" area of the colon, in order to prevent further diverticulum formation and recurrent episodes of diverticulitis.

DIVERTICULAR BLEEDING

Diverticular bleeding is a significant source for lower gastrointestinal bleeding and an important cause of morbidity and mortality in those patients over age 60 years of age. Less than one-fifth of patients with diverticulosis will have at least one episode of hematochezia. Of these patients, only a few will present with massive hemorrhage and hypovolemia. Recurrent inspissation of fecal matter within the diverticulum causes a weakening of the arterial wall, following

which rupture and bleeding may occur. Bleeding is often painless and patients do not often have symptoms other than bleeding. Diverticulitis does not typically present with bleeding. Treatment consists of two important steps, resuscitation/stabilization of the patient and localization of the bleeding. Colonoscopy is the test of choice to diagnose, localize, and in some cases to treat diverticular bleeding. Failure of colonoscopy to provide a diagnosis or localization may prompt further evaluation by radionuclide imaging or angiography. Proper localization of the bleeding source provides crucial information in the event that bleeding cannot be controlled and surgical intervention is required. Fortunately for patients, three-quarters of those with diverticular bleeding will stop bleeding spontaneously and require no further therapy.

See Also the Following Articles

Colonic Obstruction • Compound Tomography (CT) • Dietary Fiber • Lower Gastrointestinal Bleeding and Severe Hematochezia • Meckel's Diverticulum

Further Reading

Almy, T. P. (1980). Medical progress. Diverticular disease of the colon. *N. Engl. J. Med.* **302**(6), 324–331.
Berman, P. M. (1972). Current knowledge of diverticular disease of the colon. *Am. J. Dig. Dis.* **17**(8), 741–759.
Simpson, J. (1999). Diverticular disease of the colon. *J. Clin. Gastroenterol.* **29**(3), 241–252.
Simpson, J. (2002). Pathogenesis of colonic diverticula. *Br. J. Surg.* **89**, 546–554.
Stollman, N. H. (1999). Diverticular disease of the colon. *J. Clin. Gastroenterol.* **29**(3), 241–252.

Dumping Syndrome

GEORGE A. SAROSI, JR.
University of Texas Southwestern Medical Center, Dallas

antrectomy Partial removal of the distal stomach encompassing the gastric antrum.
chyme Mass of partially digested food and digestive juices.
pyloroplasty Surgical division of the pyloric muscle, destroying it as a sphincter, and reconstruction of the pyloric channel to facilitate gastric emptying.
vagotomy Surgical transaction of the vagus nerve to reduce gastric innervation.

Dumping syndrome is a spectrum of symptoms thought to be a consequence of rapid gastric emptying following vagotomy and/or bypass or destruction of the pylorus. The symptoms of dumping syndrome occur after the ingestion of food and are absent in the fasting state. Dumping syndrome typically includes both gastrointestinal and vasomotor symptoms, and can be quite variable in presentation. The incidence of dumping syndrome after gastric surgery is variable, and depends to some degree on the operation performed. However, about 10% of patients will have significant dumping symptoms after gastric surgery.

DIAGNOSIS

The diagnosis of dumping syndrome is made largely on clinical grounds and hinges on the presence of typical symptoms in a patient following gastric surgery. Dumping occurs in the postprandial period and is categorized into two forms, early and late. Early dumping occurs within 30–60 minutes of eating and typically presents with both gastrointestinal and vasomotor symptoms. Gastrointestinal symptoms include fullness, bloating, crampy abdominal pain, nausea, vomiting, and explosive diarrhea. The vasomotor symptoms include diaphoresis, weakness, dizziness, flushing palpitations, and an intense urge to lie down. Late dumping occurs 2–3 hours after eating and usually presents with only vasomotor symptoms. In patients with severe dumping symptoms, weight loss and food fear may also be present. Although there are no commonly accepted tests for dumping

Copyright 2004, Elsevier (USA). All rights reserved.

which rupture and bleeding may occur. Bleeding is often painless and patients do not often have symptoms other than bleeding. Diverticulitis does not typically present with bleeding. Treatment consists of two important steps, resuscitation/stabilization of the patient and localization of the bleeding. Colonoscopy is the test of choice to diagnose, localize, and in some cases to treat diverticular bleeding. Failure of colonoscopy to provide a diagnosis or localization may prompt further evaluation by radionuclide imaging or angiography. Proper localization of the bleeding source provides crucial information in the event that bleeding cannot be controlled and surgical intervention is required. Fortunately for patients, three-quarters of those with diverticular bleeding will stop bleeding spontaneously and require no further therapy.

See Also the Following Articles

Colonic Obstruction • Compound Tomography (CT) • Dietary Fiber • Lower Gastrointestinal Bleeding and Severe Hematochezia • Meckel's Diverticulum

Further Reading

Almy, T. P. (1980). Medical progress. Diverticular disease of the colon. *N. Engl. J. Med.* **302**(6), 324–331.
Berman, P. M. (1972). Current knowledge of diverticular disease of the colon. *Am. J. Dig. Dis.* **17**(8), 741–759.
Simpson, J. (1999). Diverticular disease of the colon. *J. Clin. Gastroenterol.* **29**(3), 241–252.
Simpson, J. (2002). Pathogenesis of colonic diverticula. *Br. J. Surg.* **89**, 546–554.
Stollman, N. H. (1999). Diverticular disease of the colon. *J. Clin. Gastroenterol.* **29**(3), 241–252.

Dumping Syndrome

GEORGE A. SAROSI, JR.
University of Texas Southwestern Medical Center, Dallas

antrectomy Partial removal of the distal stomach encompassing the gastric antrum.
chyme Mass of partially digested food and digestive juices.
pyloroplasty Surgical division of the pyloric muscle, destroying it as a sphincter, and reconstruction of the pyloric channel to facilitate gastric emptying.
vagotomy Surgical transaction of the vagus nerve to reduce gastric innervation.

Dumping syndrome is a spectrum of symptoms thought to be a consequence of rapid gastric emptying following vagotomy and/or bypass or destruction of the pylorus. The symptoms of dumping syndrome occur after the ingestion of food and are absent in the fasting state. Dumping syndrome typically includes both gastrointestinal and vasomotor symptoms, and can be quite variable in presentation. The incidence of dumping syndrome after gastric surgery is variable, and depends to some degree on the operation performed. However, about 10% of patients will have significant dumping symptoms after gastric surgery.

DIAGNOSIS

The diagnosis of dumping syndrome is made largely on clinical grounds and hinges on the presence of typical symptoms in a patient following gastric surgery. Dumping occurs in the postprandial period and is categorized into two forms, early and late. Early dumping occurs within 30–60 minutes of eating and typically presents with both gastrointestinal and vasomotor symptoms. Gastrointestinal symptoms include fullness, bloating, crampy abdominal pain, nausea, vomiting, and explosive diarrhea. The vasomotor symptoms include diaphoresis, weakness, dizziness, flushing palpitations, and an intense urge to lie down. Late dumping occurs 2–3 hours after eating and usually presents with only vasomotor symptoms. In patients with severe dumping symptoms, weight loss and food fear may also be present. Although there are no commonly accepted tests for dumping

Copyright 2004, Elsevier (USA). All rights reserved.

syndrome, radioisotope gastric emptying studies demonstrating rapid gastric emptying are often used to confirm the clinical suspicion. Several clinical scoring systems have also been used to support the diagnosis of dumping. Recently, investigators have suggested a simple provocative test to confirm the diagnosis of dumping. After ingestion of 50 g of oral glucose, a heart rate rise of greater than 10 beats per minute in the first hour has been found to be a sensitive predictor of early dumping. Late dumping is confirmed with this test by documenting typical vasomotor symptoms in the postchallenge period.

PATHOPHYSIOLOGY

Early and late dumping differ in their pathophysiology. In early dumping, the rapid delivery of hyperosmolar chyme into the small intestine as a consequence of rapid gastric emptying results in several maladaptive responses. The presence of hyperosmolar proximal small intestinal contents is thought to lead to fluid shifts into the intestinal lumen, which results in intestinal distension and precipitates the gastrointestinal symptoms of early dumping. In addition, an oral glucose challenge leads to both peripheral and splanchnic vasodilatation, with both peripheral and splanchnic blood pooling, which is thought to result in the vasomotor symptoms of early dumping. Release of a variety of gastrointestinal hormones is enhanced in patients with early dumping; these include enterglucogon, vasoactive intestinal peptide (VIP), peptide YY, pancreatic polypeptide, and neurotensin. In late dumping, the rapid delivery of glucose to the small bowel followed by rapid glucose absorption is thought to result in an exaggerated insulin release. This exaggerated insulin secretion combined with rapid gastric emptying results in a excess insulin state 2–3 hours after eating and hypoglycemia with subsequent vasomotor symptoms. The exact cause of the excess insulin is unclear, but the enteric hormone glucagon-like peptide-1 (GLP-1) may play an important role in late dumping.

TREATMENT

Most cases of dumping syndrome will improve with time. The initial mainstay of therapy for dumping syndrome is dietary modification. For most patients, instituting an "antidumping" diet will effectively control dumping symptoms. An antidumping diet consists of (1) frequent small meals, ideally six or more per day, (2) low amounts of simple sugars, moderate amounts of fat, and high amounts of complicated carbohydrates, protein, and fiber, and (3) dry meals, with mealtime liquid intake delayed at least 30 minutes after solid intake. For patients with pronounced vasomotor symptoms, lying down for 30 minutes postprandially may decrease the unpleasantness of the vasomotor symptoms. Dietary fiber supplementation has been shown to help with reactive hypoglycemia. Acarbose, a competitive inhibitor of carbohydrate absorption, has also been shown to help with late dumping, although with long-term use it can cause malabsorption. In patients with severe dumping, or in those whose symptoms fail to respond to dietary modification, the next step should be the use of octreotide, the long-acting somatostatin analogue. Octreotide acts to delay intestinal transit time, causes splanchnic vasoconstriction, and suppresses the release of a wide variety of enteral hormones, including insulin. In a systematic review of seven randomized controlled trials, octreotide, when administered prior to meals, has been shown to alleviate the gastrointestinal symptoms and vasomotor symptoms of early dumping and to prevent reactive hypoglycemia. An initial dose of 50 μg, administered 15–60 minutes prior to meals, should alleviate symptoms. The dose can be increased to 100 μg as needed. Side effects of octreotide therapy include malabsorption and steatorrhea, but the steatorrhea can be managed with pancreatic enzyme supplementation. Although there is a relative paucity of published long-term results of octreotide therapy, extrapolation from intermediate results suggests it is safe and effective.

For patients with debilitating dumping symptoms that are refractory to dietary modification and octreotide, remedial surgery should be considered. In general, a cautious approach when considering surgical therapy of dumping is appropriate, because most cases of dumping will improve with time and because no remedial operation has a 100% success rate. For patients whose initial operation was a pyloroplasty, surgical reconstruction of the pylorus is the preferred approach. In patients with an antrectomy and Billroth II reconstruction, conversion to either a Billroth I reconstruction or a Roux-en-Y reconstruction is the preferred approach. For patients with a Billroth I reconstruction as their initial operation, conversion to a Roux-en-Y reconstruction is the preferred approach. In patients for whom the Roux conversion fails to resolve dumping syndrome, the addition of a 10-cm antiperistaltic segment of jejunum downstream from the gastric outlet is a very effective antidumping procedure, but can result in significant incidence of gastric outlet obstruction. For this reason, an antiperistaltic jejunal interposition should probably not be used as a primary antidumping operation.

See Also the Following Articles

Gastrectomy • Gastric Surgery • Pyloroplasty

Further Reading

Carvajal, S., and Mulvihill, S. (1994). Postgastrectomy syndromes: Dumping and diarrhea. *Gastroenterol. Clin. North Am.* **23**, 261–279.

Li-Ling, J., and Irving, M. (2001). Therapeutic value of octreotide for patients with severe dumping syndrome—A review of randomised controlled trials. *Postgrad. Med. J.* **77**, 441–442.

Vecht, J., Masclee, A., and Lamers, C. (1997). The dumping syndrome. Current insights into pathophysiology, diagnosis and treatment. *Scand. J. Gastroenterol. Suppl.* **223**, 21–27.

Duodenal Motility

WILLIAM L. HASLER
University of Michigan Medical Center, Ann Arbor

fed motor pattern The stereotypic contractile pattern that initiates soon after a caloric meal and that is responsible for mixing and propulsion of food residue for efficient digestion and absorption.

gastric emptying The process by which ingested material is passed in a controlled fashion into the duodenum, where it undergoes further mixing and propulsion.

interstitial cells of Cajal Specialized cells that are located at the interface of the longitudinal and circular muscle layers and that generate rhythmic electrical depolarizations (slow waves) that act as pacemakers for duodenal contractions.

migrating motor complex The stereotypic contractile pattern present during fasting that is responsible for the clearance of undigested debris and sloughed enterocytes during the interdigestive period.

myenteric plexus The nerve layer that resides between the longitudinal and the circular muscle layers and that regulates all local and programmed motor activities of the duodenum.

peristaltic reflex The basic local motor reflex that is responsible for caudal propulsion of intralumenal contents.

The duodenum accepts lumenal contents delivered from the stomach and processes them for absorption or distribution to other intestinal regions. Many duodenal contractile patterns are similar to those of the distal intestine, whereas some complexes are unique to the region. Duodenal motility is controlled by myogenic characteristics, extrinsic and intrinsic nerves, and circulating hormones. External influences from central nervous system activation or immune factors modulate duodenal contractile function as do intralumenal stimuli.

ANATOMIC AND PHYSIOLOGIC CONSIDERATIONS

The duodenum spans the region from the pylorus to the ligament of Treitz. The duodenal wall has two regions of muscle tissue, the muscularis externa and the muscularis mucosa. The muscularis externa consists of an outer longitudinal layer and an inner circular layer and is the major effector of mixing and propulsion. The role of the muscularis mucosa is poorly understood. Duodenal smooth muscle cells are spindle-shaped and uninucleate with no specialized regions for neuronal interaction. They are electrically active with membrane potentials of −40 to −80 mV. Duodenal smooth muscle is supplied with a rich network of intrinsic nerve tissue with modulatory influences from extrinsic innervation.

Duodenal myoelectrical activity exhibits ubiquitous membrane potential fluctuations of 3–15 mV oscillating at a frequency of 11–12 cycles per minute (cpm). This duodenal slow wave is electrically isolated from that of the distal stomach by a thick fibrous septum at

 Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Gastrectomy • Gastric Surgery • Pyloroplasty

Further Reading

Carvajal, S., and Mulvihill, S. (1994). Postgastrectomy syndromes: Dumping and diarrhea. *Gastroenterol. Clin. North Am.* **23**, 261–279.

Li-Ling, J., and Irving, M. (2001). Therapeutic value of octreotide for patients with severe dumping syndrome—A review of randomised controlled trials. *Postgrad. Med. J.* **77**, 441–442.

Vecht, J., Masclee, A., and Lamers, C. (1997). The dumping syndrome. Current insights into pathophysiology, diagnosis and treatment. *Scand. J. Gastroenterol. Suppl.* **223**, 21–27.

Duodenal Motility

WILLIAM L. HASLER
University of Michigan Medical Center, Ann Arbor

fed motor pattern The stereotypic contractile pattern that initiates soon after a caloric meal and that is responsible for mixing and propulsion of food residue for efficient digestion and absorption.

gastric emptying The process by which ingested material is passed in a controlled fashion into the duodenum, where it undergoes further mixing and propulsion.

interstitial cells of Cajal Specialized cells that are located at the interface of the longitudinal and circular muscle layers and that generate rhythmic electrical depolarizations (slow waves) that act as pacemakers for duodenal contractions.

migrating motor complex The stereotypic contractile pattern present during fasting that is responsible for the clearance of undigested debris and sloughed enterocytes during the interdigestive period.

myenteric plexus The nerve layer that resides between the longitudinal and the circular muscle layers and that regulates all local and programmed motor activities of the duodenum.

peristaltic reflex The basic local motor reflex that is responsible for caudal propulsion of intralumenal contents.

The duodenum accepts lumenal contents delivered from the stomach and processes them for absorption or distribution to other intestinal regions. Many duodenal contractile patterns are similar to those of the distal intestine, whereas some complexes are unique to the region. Duodenal motility is controlled by myogenic characteristics, extrinsic and intrinsic nerves, and circulating hormones. External influences from central nervous system activation or immune factors modulate duodenal contractile function as do intralumenal stimuli.

ANATOMIC AND PHYSIOLOGIC CONSIDERATIONS

The duodenum spans the region from the pylorus to the ligament of Treitz. The duodenal wall has two regions of muscle tissue, the muscularis externa and the muscularis mucosa. The muscularis externa consists of an outer longitudinal layer and an inner circular layer and is the major effector of mixing and propulsion. The role of the muscularis mucosa is poorly understood. Duodenal smooth muscle cells are spindle-shaped and uninucleate with no specialized regions for neuronal interaction. They are electrically active with membrane potentials of −40 to −80 mV. Duodenal smooth muscle is supplied with a rich network of intrinsic nerve tissue with modulatory influences from extrinsic innervation.

Duodenal myoelectrical activity exhibits ubiquitous membrane potential fluctuations of 3–15 mV oscillating at a frequency of 11–12 cycles per minute (cpm). This duodenal slow wave is electrically isolated from that of the distal stomach by a thick fibrous septum at

 Copyright 2004, Elsevier (USA). All rights reserved.

the pylorus. Few gastric slow waves (3 cpm) propagate into the duodenum because of this septum, but some spike potentials cross the pyloric region in patches and may form the basis for gastroduodenal coordination. The duodenal slow wave controls the direction and maximal frequency of phasic contractions, although its amplitude usually is insufficient to elicit contraction under basal conditions. Neurohumoral stimuli evoke motor activity by inducing spike potentials of brief duration (10–100 ms) but high amplitude (50 mV) in phase with the slow wave. The interstitial cells of Cajal (ICC) at the interface of the circular and longitudinal muscle layers are believed to be responsible for slow-wave generation. In cat intestine, rhythmic oscillations are generated only by tissue containing ICC. In animals with a genetic absence of ICC, slow-wave activity is undetectable. Cyclic fluctuations in intracellular calcium underlie this electrical rhythmicity. In humans, the dominant small intestinal slow-wave pacemaker is in the duodenum. This site entrains adjacent regions; however, some uncoupling occurs over the entire intestine, resulting in decreases in slow-wave frequency in the ileum (7–8 cpm). These properties are advantageous for the different roles of each region. Proximally, nutrients are propelled over a large mucosal surface area for rapid digestion and absorption while, distally, retarded propulsion permits absorption of slowly digested and absorbed substances such as fats and bile.

Duodenal smooth muscle is innervated by the vagus and splanchnic nerves that relay information to and from the extraintestinal ganglia, the spinal cord, and the central nervous system. The myenteric plexus provides the major intrinsic innervation. The number of intrinsic neurons greatly exceeds the number of extrinsic fibers. Thus, most motor activities are directed by intrinsic nerves, whereas extrinsic innervation provides a modulatory function. Low-threshold preganglionic vagal cholinergic neurons activate nicotinic receptors within enteric ganglia and excite motor activity. High-threshold preganglionic vagal neurons are inhibitory to motor activity via release of vasoactive intestinal polypeptide (VIP) and nitric oxide (NO). Preganglionic cholinergic sphlanchnic neurons project from the spinal cord to the prevertebral ganglia. Noradrenergic postganglionic neurons then project to the enteric ganglia and inhibit excitatory cholinergic transmission. Afferent fibers outnumber efferent fibers by 10-fold in the vagus and by 3-fold in the splanchnic nerves. The duodenum is richly supplied with sensory fibers that may respond to chemical, thermal, or mechanical stimulation. Mechanoreceptors in the outer muscle layers or myenteric plexus are activated by passive distension or during active contractions. Mesenteric and serosal receptors respond to tension or to forceful contraction and may mediate visceral pain perception. The density of myenteric neurons ranges from 3700 to 12,170 per square centimeter in the cat small intestine, approximating that of the spinal cord. Motor neurons typically project 1–2 mm longitudinally although some fibers extend for 30 mm. Excitatory fibers run cephalad and inhibitory fibers project caudally. Eighty to 90% of myenteric neurons contain either tachykinins (40–45%) or VIP (40–45%), with no overlap between the two groups. Tachykinin neurons containing Substance P, neurokinin A, and acetylcholine mediate most of the excitatory functions of intestinal smooth muscle. The inhibitory supply to the duodenum is provided by NO- and VIP-containing myenteric neurons. A number of other transmitters modulate the physiologic motor activity of the small intestine.

ORGANIZED MOTOR PATTERNS OF THE DUODENUM

The best characterized organized motor patterns in the duodenum are the interdigestive migrating motor complex and the fed motor pattern. Other contractile complexes are observed infrequently at irregular intervals, often during illness. In many instances, duodenal motor patterns resemble those of the jejunum and ileum; however, patterns specific to the duodenum may be observed.

Migrating Motor Complex

The migrating motor complex (MMC) serves the role of housekeeper of the small intestine by propelling undigested food residue and sloughed enterocytes. The MMC consists of four phases with a total duration of 84–112 min. Phase I is a period of motor quiescence lasting 40–60% of the cycle. Phase II, occupying 20–30% of the cycle, exhibits irregular phasic contractions. Phase III is a 5- to 10-min period of lumenally occlusive, rhythmic contractions occurring at the slow-wave frequency. Phase IV is a transitional period of irregular contractions between phase III and phase I. Seventy-one percent of phase III complexes originate in the stomach, 28% begin in the duodenum, and 1% originate in the proximal jejunum. Most MMCs terminate in the jejunum. MMC cycle duration is nearly twice as long if the previous phase III complex originated in the stomach versus the duodenum. Phase III contractions propagate over longer distances than those in phase II. The transit of inert substances is four times faster in phase III than in phase I and 50% of total flow occurs during phase III. On cineradiography, transit during phase

III is characterized by intermittent boluses of 4–5 cm in length separated by 1- to 2-cm ring contractions.

The MMC is influenced by extrinsic and intrinsic neural sources. Bilateral vagotomy, removal of the superior and inferior mesenteric ganglia, total sympathectomy, and complete extrinsic denervation do not prevent MMC cycling, although cycle duration and regularity may be altered. Isolated denervated intestinal segments exhibit spontaneous phase III activity, which propagates aborally, demonstrating that the complex is programmed into enteric nerves. Cholinergic and noncholinergic pathways participate in MMC contractions. In dogs, atropine, the ganglionic blocker hexamethonium, or the neural toxin tetrodotoxin eliminates MMC cycling. Adrenoceptor antagonists disrupt MMC cycling, whereas NO synthase inhibitors evoke premature MMC activity. Ablation of myenteric serotonergic neurons prolongs MMC periodicity and decreases propagation velocities, whereas 5-HT_3 antagonists inhibit phase III activity and prolong cycle length.

The major humoral regulator of MMC activity is motilin, which is released from specialized cells in the duodenal mucosa. Antral phase III complexes correlate temporally with plasma motilin elevations in healthy humans and premature antral phase III activity is inducible by motilin infusion. In dogs, gastroduodenal phase III is abolished for several hours after infusion of motilin antisera and is replaced by irregular phasic contractions. Resection of duodenal motilin-producing tissue alters fasting antroduodenal motility in dogs. Motilin stimulates antral phase III in part through cholinergic pathways, because atropine reduces its contractile effects. However, duodenal phase III activity is controlled by atropine-resistant noncholinergic mechanisms. The stimulus for motilin cycling is not known. Motilin fluctuations are blocked by atropine and hexamethonium, suggesting regulation by cholinergic pathways, and motilin release is evoked by vagal stimulation, bile release, cholinergic agonists, opioid agents, NO synthase inhibitors, and duodenal pH changes, but the roles of these factors in regulating motilin release are unclear. There are differences in the motilin dependence of phase III complexes in the duodenum compared to the distal intestine. Ectopic complexes in the jejunum or ileum often are not associated with motilin elevations. The effects of motilin antisera on jejunal and ileal phase III complexes are minimal and ectopic phase III complexes form in these regions after duodenal resection. Thus, "programming" of phase III in the distal intestine is a motilin-independent phenomenon that is entrained by the actions of motilin on the antrum and duodenum.

The duodenal MMC cycles in phase with biliary motor activity and secretory functions. Gastric acid and pepsin production and intestinal fluid secretion increase prior to duodenal phase III as does the release of bicarbonate, bile acids, bilirubin, pancreatic enzymes, and secretory immunoglobulin A. Duodenal phase III has been proposed to contribute to gallbladder refilling during the interdigestive period. Chronic pancreatitis disrupts the synchrony of MMC activity and pancreatic exocrine cycling, suggesting differential regulation of the two phenomena.

Fed Motor Pattern

After a meal, the duodenal MMC is replaced by a fed pattern of intermittent phasic contractions of varying amplitude that serve to mix and propel intestinal contents. Forty-four percent of fed contractions do not propagate. Of the contractions that do, 90% migrate <30 cm and 66% migrate <9 cm. In some studies, a transitional motor pattern is observed immediately after meal ingestion characterized by highly propagative contractile clusters that expose extended regions of mucosa to nutrients. The duration of the fed motor pattern is dependent on meal characteristics. In dogs, a mixed 450 kcal meal induces a fed pattern of >3 h. In humans, a 400 kcal meal with 9% fat disrupts MMC activity for 294 ± 21 min, whereas a meal with 50% fat prolongs the fed period to 410 ± 42 min.

Neural input is important for induction of the fed pattern. In humans, sham-feeding disrupts duodenal phase III, indicating a cephalic contribution. Bilateral vagotomy, splanchnicectomy, mesenteric ganglionectomy, and total extrinsic denervation do not prevent induction of the fed state, but bilateral vagotomy shortens its duration and increases the latency to the onset of contractions. If nutrients are perfused into an isolated, extrinsically innervated intestinal loop, the MMC is disrupted in the unconnected main portion of the intestine, demonstrating the importance of extrinsic supply. Intestinal transection and reanastomosis decreases the frequency and amplitude of fed contractions in dogs and decreases propagation, showing the modulatory effects of intrinsic nerves. Atropine and hexamethonium abolish fed motor activity and NO synthase inhibitors shorten its duration. The hormonal mediators of the fed motor pattern are less defined. Investigation has focused on CCK as CCK levels increase 5- to 10-fold after eating and CCK analogues increase intestinal motility. However, motilin cycling is suppressed by meals but is unaffected by CCK infusion. In rats, CCK receptor blockade prevents MMC disruption by eating. In dogs, CCK receptor an-

tagonists reduce but do not prevent the fed response or the interruption of MMC cycling after eating. Other transmitters inhibit the MMC and induce complexes similar to the fed pattern. Neurotensin has received attention based on its ability to convert the fasting to a fed pattern along the entire small intestine in rats and humans.

The mechanism for reversion to MMC cycling after completion of the fed pattern is uncertain. The first MMC after eating begins distal to the duodenum, indicating that factors that initiate normal complexes are not yet operational. The initial duodenal phase III occurs soon after completion of gastric emptying. In dogs, continuous intraduodenal nutrient perfusion induces a fed pattern for only a finite time, after which the fasting pattern returns.

Other Stereotypic Contractile Complexes

The duodenum participates in other infrequent contractile complexes that may serve distinct roles in health and disease. Giant migrating complexes (GMCs) are intense contractile waves that propagate aborally for long distances in the intestine. Most GMCs begin in the distal intestine rather than the duodenum and clear debris from the ileum. GMCs are increased in several diarrheal conditions. Discrete clustered contractions (DCCs) consist of 3–10 contractions preceded and followed by 1 min of motor quiescence and are seen in the proximal and distal intestine. A highly propagative pattern in the duodenum and proximal jejunum, called the rapidly migrating contraction, migrates for 200 cm at >30 cm/s. Retrograde peristaltic contractions (RPCs) develop in the mid small intestine after emetic agents are administered and migrate orally to the duodenum before vomiting occurs. RPCs migrate rapidly (8–10 cm/s) over distances >100 cm and serve to evacuate intestinal contents into the stomach so that they may be expelled during emesis.

Motor Patterns Specific for the Duodenum

In addition to participating in the organized motor activity of the entire small intestine, the duodenum exhibits motor patterns that are specific for the region. The duodenal cross-sectional area increases during phase II of the MMC to accommodate pancreaticobiliary secretions, indicating region-specific tonic changes. Retroperistaltic contractions in the duodenum propel lumenal contents orally. Retrograde duodenal motor activity during late phase III directed against antral phase III pressure waves may increase fasting duodenal pH and nocturnal antral pH, serving to protect both regions. Retroperistaltic activity is associated with duodenogastric reflux of bicarbonate and immunoglobulin A, which may reconstitute the antral mucosal barrier during fasting. Duodenogastric retropulsion of bile is minimized by phase III-associated closure of the sphincter of Oddi. During the fed period, 40–50% of pressure waves in the proximal duodenum are retrograde in nature, which may contribute to controlling gastric emptying. Correlative ultrasonographic studies show bursts of duodenogastric reflux prior to pyloric closure during gastric emptying, suggesting the presence of duodenal mixing and retropulsion. Other investigators have questioned the importance of retrograde duodenal activity. Using impedancometry, one group reported that only 4% of fasting and 8% of fed bolus fluid movements were retrograde. Another study observed that only 2 of 180 phase III pressure wave sequences were purely retrograde. Similarly, during duodenal lipid perfusion, most pressure waves propagate in an antegrade direction. Some have observed bidirectional contractions in the proximal duodenum during phase III, especially toward the end of the complex.

EXTRINSIC AND INTRINSIC MODULATION OF DUODENAL MOTOR ACTIVIITY

Motor patterns in the duodenum are modified by external neural and immunologic stimulation. Duodenal motility also is subject to modulation by reflex activations of other regions of the gastrointestinal tract.

Extrinsic Factors

Duodenal motility is influenced by input from the central nervous system. Sleep decreases MMC cycle length, reduces propagation velocity, and increases contractile amplitudes. Intestinal phase III is reduced by depression and stress. Stress also prolongs the intestinal fed motor pattern. Corticotropin-releasing factor (CRF) may be a physiologic mediator of stress effects on the gut. In mice, cold stress effects on intestinal transit are mimicked by intracerebroventricular CRF. MMC inhibition by CRF in dogs is associated with suppression of motilin cycling. The effects of CRF and stress are reversed by CRF antagonists. Other transmitters acting centrally also modify duodenal motility. Inhibition of NO synthase in the brain suppresses duodenal phase III in dogs via vagal pathways. Destruction of hypothalamic and locus ceruleus adrenergic pathways lengthens MMC periodicity. Central neural calcitonin, calcitonin gene-related peptide, neurotensin, and opioids evoke intestinal phase III activity, whereas neuropeptide Y delays intestinal transit.

Duodenal motor activity is modulated by immunologic stimulation. In rats antigen-sensitized by intraperitoneal egg albumin, subsequent oral albumin evokes diarrhea that is associated with MMC disruption and induction of high-amplitude clustered contractions. A role for mucosal mast cells is demonstrated by the abilities of mast cell stabilizers and degranulation inhibitors to blunt antigen-induced intestinal motor responses. Intestinal infections also disrupt gut motor function. In rats, *Yersinia enterocolitica* and *Escherichia coli* alter MMC cycling via free radical generation. Similarly, fasting myoelectric activity increases and spike potentials are generated with *Hymenolepsis diminuta* infection. Much information on the impact of the immune system on gut function has been provided by studies of the nematode *Trichinella spiralis*. In rats, *T. spiralis* disrupts intestinal motility and induces GMCs. Intestinal muscle from *T. spiralis*-infected rats exhibits increased smooth muscle responses to contractile agonists and blunted neurally mediated contractions with associated decreases in Substance P and VIP content. Both T-lymphocyte-dependent and -independent pathways are involved in these pathophysiologic responses.

Intrinsic Regulation

Duodenal motor activity is affected by stimuli within the gut. The peristaltic reflex is the most basic reflex response demonstrable in intestinal segments. Stimuli that evoke the peristaltic reflex include mucosal pinching, hypertonic saline infusion, and radial stretch. The peristaltic reflex consists of an excitatory response proximal to stimulation (ascending contraction) and a distal inhibition (descending relaxation). The ascending contraction is characterized by simultaneous circular muscle shortening and longitudinal muscle relaxation and descending relaxation involves simultaneous longitudinal contraction and circular relaxation. Ascending contractions are partly inhibited by atropine at low levels of stimulation, whereas tachykinin antagonists or antisera block contractions induced by intense radial stretching, indicating dual cholinergic and tachykinin mediation. VIP and NO likely mediate the descending relaxation as both are released during peristalsis. Serotonin is released with distension and lowers the threshold for eliciting the reflex. Mucosal stimulation initiates the reflex by activating 5-HT_4/5-HT_{1P} receptors on sensory neurons in human intestine. In marmoset intestine, 5-HT_4 receptors are activated by low serotonin concentrations and 5-HT_3 receptors are activated by higher concentrations.

Nonmechanical stimuli have modulatory influences on duodenal function as well. Intraduodenal lipid or bile perfusion decreases duodenal flow in association with increased lumenal diameter, isolated low-amplitude pressure waves, and reduced propagation distances. In contrast, hydrochloric acid and hyperosmolar solutions retard duodenal propagation by inducing tonic duodenal occlusions. Duodenal myotomy reduces the evoked duodenal contractions with enhancement of lumenal propulsion.

Stimulation of other gut segments also modulates duodenal motility. The inhibitory intestino-intestinal reflex is the inhibition of several hundred centimeters of intestine by abrupt stretching or dilation of a localized segment. In humans, the inhibitory effects of distension are more pronounced in the duodenum than in the ileum. Distension of the stomach abolishes fasting duodenal motor activity and delays intestinal transit via nonvagal NO-mediated neural pathways. In other studies, gastric fundus distension produces irregular phasic contractions mimicking the duodenal fed pattern, some of which are propagated from the antrum. Colonic nutrient perfusion inhibits duodenal fasting activity, whereas colonic distension delays MMC onset via nicotinic ganglionic receptors.

DUODENAL FEEDBACK REGULATION OF GASTRIC EMPTYING

The duodenum plays a crucial role in regulating gastric emptying. The absence of duodenal motor activity is associated with accelerated emptying, whereas continuous duodenal contractions delay liquid emptying. Enhanced liquid emptying also is noted after duodenal myotomy. Duodenal pH receptors mediate inhibition of emptying by duodenal acidification. Duodenal amino acid perfusion slows both solid and liquid emptying. Placement of a duodenal fistula to divert nutrients converts liquid emptying to rapid first-order kinetics instead of the normal curvilinear profile. Agents acting on intestinal osmoreceptors increase duodenal outflow resistance and delay emptying. Inhibition of gastric emptying by lipids requires component fatty acids, because perfusion of nonhydrolyzable fats does not inhibit gastric emptying. Osmoreceptors mediating hypertonic saline-evoked inhibition of emptying are located only in the duodenum, whereas acids and lipids have inhibitory effects on longer segments.

The mediators that regulate duodenal inhibition of gastric emptying are incompletely characterized. Intestinal capsaicin reduces the inhibition of emptying induced by hydrochloric acid, glucose, and lipid, indicating the importance of afferent pathways. Duodenal transection impairs the nutrient-evoked delay of gastric emptying, showing the participation of intrinsic

duodenal neurons. NO may facilitate gastric emptying by inhibiting obstructing pyloroduodenal motor activity. Physiologic CCK infusions inhibit liquid emptying, whereas CCK antagonists accelerate emptying of mixed meals, glucose, lipids, and radiopaque markers in most studies. The ability of the CCK antagonist devazepide to blunt the retardation of emptying evoked by peptones decreases as the caloric density increases, indicating CCK-independent factors with greater nutrient loads.

The inhibitory effects of duodenal stimulation on gastric emptying are mediated by specific enterogastric reflexes. Duodenal acid exposure induces gastric fundus relaxation via secretin-dependent and -independent pathways. Duodenal distension inhibits antral contractions, a reflex that is partly blocked by vagotomy or splanchnicectomy and abolished by both. In contrast, duodenal distension-evoked inhibition of antral activity is unaffected by duodenal transection. Intraduodenal lipids, proteins, or hydrochloric acid decreases spontaneous antral contractions, effects reduced but not abolished by vagotomy. Vagal application of capsaicin partly reverses the inhibitory effects of protein, glucose, and trypsin inhibitor on antral motor activity, showing mediation by afferent neurons. Duodenal lipids, amino acids, glucose, hypertonic saline, or hydrochloric acid elicits pyloric closure and decreases transpyloric flow. Proximal duodenal distension produces greater pyloric responses than more distal stimulation. Isolated pyloric contractions evoked by duodenal hydrochloric acid are antagonized by atropine and the ganglionic blocker hexamethonium but are not blocked by vagotomy. The ability of the topical anesthetic xylocaine to prevent acid-induced pyloric contraction is evidence of a role of mucosal receptors in this reflex. Intrinsic pathways are involved in pyloric responses as duodenal transection prevents induction of pyloric contractions by duodenal distension. The 5-HT_3 antagonist zacopride reduces pyloric motor responses to intraduodenal hydrochloric acid, indicating that serotonergic nerves play a role. CCK may mediate the pyloric response to intraduodenal lipids; however, other studies suggest that CCK plays a limited role in enteropyloric reflex activity under physiologic conditions. Mediation by opioid pathways is suggested by the ability of naloxone to block pyloric contractions induced by intraduodenal amino acids.

See Also the Following Articles

Cholecystokinin (CCK) • Gastric Emptying • Gastric Motility • Interstitial Cells of Cajal • Migrating Motor Complex • Motilin • Postprandial Motility

Further Reading

Andrews, J. M., O'Donovan, D. G., Hebbard, G. S., Malbert, C. H., Doran, S. M., and Dent, J. (2002). Human duodenal phase III migrating motor complex activity is predominantly antegrade, as revealed by high-resolution manometry and colour pressure plots. *Neurogastroenterol. Motil.* **14**, 331–338.

Castedal, M., and Abrahamsson, H. (2001). High-resolution analysis of the duodenal interdigestive phase III in humans. *Neurogastroenterol. Motil.* **13**, 473–481.

Castedal, M., Bjornsson, E., and Abrahamsson, H. (1998). Postprandial peristalsis in the human duodenum. *Neurogastroenterol. Motil.* **10**, 227–233.

Huizinga, J. D., Thuneberg, L., Kluppel, M., Malysz, J., Mikkelsen, J. B., and Bernstein, A. (1995). W/kit gene required for interstitial cells of Cajal and for intestinal pacemaker activity. *Nature* **373**, 347–349.

Kuiken, S. D., Tytgat, G. N., and Boeckxstaens, G. E. (2002). Role of endogenous nitric oxide in regulating antropyloroduodenal motility in humans. *Am. J. Gastroenterol.* **97**, 1661–1667.

Nguyen, H. N., Silny, J., Wuller, S., Marschall, H. U., Rau, G., and Matern, S. (1995). Chyme transport patterns in human duodenum, determined by multiple intraluminal impedancometry. *Am. J. Physiol.* **268**, G700–G708.

Pallotta, N., Cicala, M., Frandina, C., and Corazziari, E. (1998). Antro-pyloric contractile patterns and transpyloric flow after meal ingestion in humans. *Am. J. Gastroenterol.* **93**, 2513–2522.

Rao, S. S., Lu, C., and Schulze-Delrieu, K. (1996). Duodenum as an immediate brake to gastric outflow: A videofluoroscopic and manometric assessment. *Gastroenterology* **11**, 740–747.

Torrents, D., and Vergara, P. (2000). *In vivo* changes in the intestinal reflexes and the response to CCK in the inflamed small intestine of the rat. *Am. J. Physiol.* **279**, G543–G551.

Wilmer, A., Andrioli, A., Coremans, G., Tack, J., and Janssens, J. (1997). Ambulatory small intestinal manometry. Detailed comparison of duodenal and jejunal motor activity in man. *Digest. Dis. Sci.* **42**, 1618–1627.

Duodenal Obstruction

GEORGE A. SAROSI, JR.
University of Texas Southwestern Medical Center, Dallas

annular pancreas Congenital condition in which the pancreas does not rotate normally during embryogenesis, resulting in an extrinsic compression of the second portion of the duodenum.

duodenal atresia Congenital condition in which the embryonic cells lining the duodenum fail to completely resorb, resulting in a blockage in the duodenum (stricture).

duodeno-jejunostomy A surgical procedure in which the proximal duodenum is attached side-to-side to the proximal jejunum in order to bypass an obstructing lesion in the mid- to distal duodenum.

web A thin membrane of tissue attached to the wall of the intestine and causing a variable degree of intestinal blockage.

The duodenum is the most proximal portion of the small bowel, extending from the pylorus to the ligament of Treitz. Although a relatively uncommon site of intestinal obstruction, specific issues in duodenal obstruction can make these disorders difficult to manage. Obstruction of the first portion of the duodenum is usually due to peptic ulcer disease and is best thought of as a gastric outlet obstruction. This article focuses on duodenal obstruction occurring beyond the ampulla of Vater in both neonates and adults.

CLINICAL PRESENTATION

Duodenal obstruction presents in a similar fashion to a proximal small bowel obstruction. Adult patients report bloating, epigastric discomfort, and nausea with bilious emesis. Patients report symptoms regardless of oral intake, but they are often worsened by oral intake. Despite symptoms of proximal obstruction, the patients often continue to have normal bowel movements and flatus although they may report acholic stools. In neonates and infants, the history usually reveals only reports of poor feeding, fussiness, and bilious emesis. Many of the causes of duodenal obstruction in infants are associated with other congenital abnormalities. On examination, the patients often have epigastric fullness and tenderness, but an otherwise nondistended abdomen. In some cases, a mass may be palpable in the epigastrium. Laboratory evaluation will be consistent with vomiting and dehydration. Metabolic alkalosis is less consistent than in gastric outlet obstruction because of the loss of pancreatic bicarbonate with postampullary obstruction. Plain abdominal radiographs will sometimes demonstrate a double-bubble sign with an air-distended stomach, a closed pylorus, and an air-distended duodenum. A listing of the etiologies of duodenal obstruction broken down by patient age is presented in Table I.

TABLE I Etiologies of Duodenal Obstruction in Neonates, Infants, and Adults

Neonates and infants	Adults
Malrotation ± midgut volvulus	Malignancy
	Periampullary cancer[a]
	Lymphoma
	Renal malignancies
	Ascending colon cancer
	Metastatic cancer
Duodenal web	Inflammatory diseases
	Chronic pancreatitis
	Pancreatic pseudocyst
	Acute pancreatitis
	Crohn's disease
	Retroperitoneal fibrosis
Duodenal atresia	Vascular anomalies
	Preduodenal portal vein
	SMA or SMV syndrome
	Abdominal aortic aneurysm
Annular pancreas	Congenital malformations
	Duodenal web
	Annular pancreas
	Malrotation
Proximal jejunal atresia	Intramural or retroperitoneal hematoma
Vascular anomalies	Foreign body
Preduodenal portal vein	Bouveret's syndrome (gallstone obstruction of the duodenum)
	Duodenal diverticulum

Note. SMA, superior mesenteric artery; SMV, superior mesenteric vein.

[a] Periampullary cancer includes cancer of the head of the pancreas, the distal bile duct (cholangiocarcinoma), the ampulla of Vater, and the second portion of the duodenum close to the ampulla of Vater.

 Copyright 2004, Elsevier (USA). All rights reserved.

DIAGNOSTIC EVALUATION

Neonates

Duodenal obstruction in neonates and infants should be considered a surgical emergency until the diagnosis of malrotation with potential midgut volvulus can be ruled out. In a neonate with a presentation of duodenal obstruction, an upper gastrointestinal (GI) contrast study needs to be obtained urgently to establish the presence of a GI rotational abnormality. If malrotation cannot be ruled out by contrast study, these neonates require surgery to rule out malrotation and treat the obstruction. In neonates with normal gastrointestinal rotation and in adults, the evaluation and management of duodenal obstruction can proceed at a less urgent pace as the likelihood of bowel infarction from a midgut volvulus is low. In a neonate, the initial therapeutic effort is directed toward correcting volume deficits and any electrolyte abnormalities and toward nasogastric decompression. At this time, identification and management of additional congenital anomalies should also occur. In a neonate, no additional diagnostic evaluation beyond an upper GI contrast study is of significant preoperative value and after adequate preoperative preparation the infant should undergo surgery for correction of the source of duodenal obstruction.

Adults

In an adult patient, the initial management should consist of nasogastric decompression and correction of existing volume and electrolyte abnormalities. Because the likelihood of malrotation and midgut volvulus is low, debate exists about the next appropriate diagnostic study beyond the initial plain abdominal radiographs. Although upper GI contrast study will establish the presence of duodenal obstruction and give information as to the exact location of the obstruction, it cannot provide much information about the etiology of the obstruction. In contrast, abdominal computed tomography (CT) scan with GI and intravenous contrast will localize the obstruction and also can provide staging information for malignancies, diagnose a variety of inflammatory processes, and provide information about vascular etiologies of duodenal obstruction. For these reasons, CT would seem to be the most appropriate initial imaging modality in the evaluation of duodenal obstruction. If water-soluble GI contrast agents are used, CT does not interfere with the ability to perform subsequent upper GI endoscopy. While a diagnostic evaluation is ongoing, a careful assessment of the patient's nutrition should be undertaken, as many patients with duodenal obstruction will have a prolonged preoperative period of poor oral intake and subsequent malnutrition.

MANAGEMENT

Neonates

The therapy for duodenal obstruction will depend on the etiology of the obstruction. In neonates, essentially all of the etiologies will require surgical intervention. In the case of malrotation, an urgent laparotomy with detorsion of the small bowel, lysis of the Ladd's bands, and appendectomy is the surgical therapy of choice. In the case of a duodenal web, duodenotomy and web excision, duodeno-duodenostomy, and duodeno-jejunostomy have been used with good results. In the case of duodenal atresia and annular pancreas, bothduodeno-duodenostomyandduodeno-jejunostomy have all been used, depending on the location of the atresia and the mobility of the duodenum. Surgical therapy of duodenal obstruction in neonates is in general associated with excellent outcomes.

Adults

Medical Therapy

Treatment options for duodenal obstruction for adults are somewhat wider than for neonates, in part based on the broader differential diagnosis. Whereas some causes of duodenal obstruction can be managed with medical therapy, the majority will require surgical intervention. In the case of patients with malignant duodenal obstruction, a third option, intralumenal stenting, should be considered.

In patients with inflammatory etiologies for duodenal obstruction such as Crohn's disease, pancreatitis, and pancreatic pseudocysts, medical management of the underlying disease will occasionally result in partial or complete resolution of the obstruction. In Crohn's disease and acute pancreatitis, a 1- to 2-week course of conservative therapy is usually warranted prior to operation, although there is a paucity of outcome data to support this contention. An initial course of conservative management of duodenal obstruction secondary to intramural or retroperitoneal hematoma is warranted as several series have shown the majority to resolve over a 2-week period of observation. Duodenal obstruction secondary to peripancreatic lymphoma will often resolve in patients with a good response to chemotherapy and for this reason a conservative approach is warranted in managing these duodenal obstructions.

Surgery and Stenting

For the majority of causes of duodenal obstruction, the primary therapy will be surgical. For benign processes of an inflammatory, vascular, or congenital nature, or for foreign body obstruction, surgical extraction of the foreign body or web or a bypass of the obstruction will provide an effective and durable treatment of the obstruction with low morbidity. Many patients with duodenal obstruction will be malnourished at presentation and a 1- to 2-week course of preoperative nutrition support is often valuable for patients. The preferred operation for most causes of benign duodenal obstruction is a duodeno-jejunostomy, bypassing the obstruction from a point on the duodenum proximal to the obstruction to the proximal jejunum. In cases where the obstruction is in close proximity to the ampulla of Vater, a gastrojejunostomy may be preferable to a duodeno-jejunostomy in order to avoid inadvertent injury to the ampulla and/or common bile duct. If a gastrojeunostomy is performed for benign disease, consideration must be given either to an acid-reducing procedure, such as a parietal cell vagotomy, or to lifetime therapy with a proton pump inhibitor, as gastrojejunostomies are known to predispose patients to marginal ulcers.

In cases of malignant duodenal obstruction, surgical therapy remains the mainstay of therapy, but in cases where treatment of the obstruction is for palliation rather than cure, endolumenal stenting is emerging as an attractive alternative to surgery. The majority of malignant obstructions of the duodenum will be secondary to periampullary malignancies, often pancreatic cancers. The first goal in treating malignant duodenal obstruction is determining whether the patient is a candidate for pancreaticoduodenectomy with curative intent. In patients without evidence of metastatic disease or of locally advanced disease, the optimal surgical approach to duodenal obstruction is an attempt at curative resection. In the majority of patients, however, where resection for cure is not an option, a number of options for palliation of duodenal obstruction exist.

The least invasive option is placement of an endolumenal stent across the obstruction via either flexible endoscopy or fluoroscopy. There is a growing experience in managing duodenal obstruction utilizing self-expanding metallic stents placed across the obstruction. Primary success rates for symptomatic improvement of 85–90% have been reported in small case series. Recurrent obstruction rates of 15–20% prior to death have been reported in stented patients, but in some of these cases, repeat stent placement was possible. The reported incidence of GI perforation secondary to stent placement appears to be low.

In cases where stent placement is not possible or fails, surgical bypass of the obstruction remains the therapeutic gold standard. For malignant duodenal obstruction, a gastrojejunostomy is the procedure of choice based on its technical ease and the fact that few patients with malignant obstruction will survive long enough to develop ulcer complications of this operation. Most surgeons will also place a gastrostomy tube at the time of a bypass to allow the 10–20% of patients with poor gastric emptying after the gastrojejunostomy to avoid a long-term nasogastric tube. Gastrojejunostomy and gastrostomy tube placement can be performed using either a laparoscopic or a conventional open surgical approach with acceptable morbidity and mortality rates.

See Also the Following Articles

Gastric Outlet Obstruction • Intestinal Atresia • Malrotation • Marginal Ulcer • Neonatal Intestinal Obstruction • Webs • Volvulus

Further Reading

Chesire, N., and Glazer, G. (1997). Diverticula, volvulus, superior mesenteric artery syndrome and foreign bodies. *In* "Maingot's Abdominal Operations" (M. Zinner, S. Schwartz, and H. Ellis, eds.), 10th Ed., Vol. 1, pp. 913–940. Appleton and Lange, Stamford, CT.

Fischer, A., and Ziegler, M. (2002). Surgery of the stomach and duodenum in children. *In* "Shackleford's Surgery of the Alimenatary Tract" (G. Zuidema and C. Yeo, eds.), 5th Ed., Vol. 2, pp. 211–228. W. B. Saunders, Philadelphia, PA.

Lillemoe, K., and Pitt, H. (1996). Palliation: Surgical and otherwise. *Cancer* **78**, 605–614.

Mauro, M., Koehler, R., and Baron, T. (2000). Advances in gastrointestinal intervention: The treatment of gastroduodenal and colorectal obstructions with metallic stents. *Radiology* **215**, 659–669.

Reynolds, H., and Stellato, T. (2001). Crohn's disease of the foregut. *Surg. Clin. North Am.* **81**, 117–135.

Duodenal Ulcer

RUPERT WING-LOONG LEONG* AND FRANCIS KA-LEUNG CHAN†
*University of New South Wales, Australia and †Chinese University of Hong Kong

cyclooxygenase Enzyme that exists in two isoforms, cyclooxygenase-1 and cyclooxygenase-2. Cyclooxygenase-1 is constitutive and regulates normal function in many organs. Cyclooxygenase-2 is inducible and is expressed by inflammatory cells to mediate inflammatory responses. A variant of cyclooxygenase-1 has been found in the central nervous system.

erosion Histologically defined as a defect that does not penetrate the muscularis mucosae; endoscopically, a lesion with a diameter less than 3–5 mm and with no perceptible lesion depth.

gastric metaplasia Mucus-type gastric cell growth that replaces the normal villous surface of the duodenal mucosa.

Helicobacter pylori Spiral-shaped gram-negative bacilli that inhabit the gastric mucous layer.

mucosal-associated lymphoid tissue lymphoma Low-grade B cell lymphoma of the stomach, principally caused by infection by *Helicobacter pylori*.

nonsteroidal antiinflammatory drugs Compounds that nonselectively inhibit the cyclooxygenase enzyme and are potent antiinflammatory and antipyretic agents.

non-ulcer dyspepsia Syndrome of recurrent chronic epigastric pain with no ulceration found on gastroscopy.

trefoil factors Family of acid- and enzyme-resistant peptides secreted by the mucin-secreting cells in the gastrointestinal tract; protect the gastrointestinal mucosa and assist ulcer healing.

ulcer Defined histologically as a defect that penetrates the muscularis mucosa; the exact endoscopic definition remains controversial, with disagreement on the lesion size (usually 3–5 mm), but most agree perception of lesion depth is necessary.

urease Enzyme produced by *Helicobacter pylori*; converts urea into ammonia and assists in survival of the bacteria in the stomach. Utilized by various tests to detect the presence of the *H. pylori*.

urease breath test Method to test for *Helicobacter pylori*. A carbon isotope-labeled urea compound is swallowed; in the presence of *H. pylori* urease, the compound is metabolized into isotopic carbon dioxide (^{13}C or ^{14}C) that is detected in exhaled breath.

A duodenal ulcer refers to a break in the duodenal mucosal integrity. The histopathologic description encompasses an epithelial defect that extends down through the muscularis mucosae into the submucosa. The endoscopic definition usually requires the surface breach to be at least 5 mm in diameter and with perceivable depth, characteristics that distinguish ulcers from erosions. Because erosions are superficial, they heal without scarring and are unlikely to be complicated apart from slow blood loss. Duodenal ulcers are usually solitary, round, punched out lesions that occur most often in the proximal 2 cm of the duodenum on the anterior or posterior walls. This area corresponds to the area of highest acidity and is also the junction between the gastric antral and intestinal mucosa.

EPIDEMIOLOGY

The incidence of peptic ulceration has changed dramatically over the past century. Prior to the nineteenth century, duodenal ulceration was uncommon. Duodenal ulcer incidence then increased up until the late 1950s, after which time there was a progressive decline. The peak in the incidence of duodenal ulcer and ulcer mortality occurred 10 to 20 years after that of gastric ulcer. Since the 1980s, although there has been an increase in the use of aspirin and nonsteroidal antiinflammatory drugs (NSAIDs) that relates to increased risk of ulcers, there has been a plateau in the incidence of duodenal ulcer that correlates with the introduction of histamine-2 (H2) receptor antagonists. Peptic ulcers in the past two decades have been characterized by occurrence in an older population and incidence in a higher proportion of females. The recurrence of ulcers in the elderly in recent years may be the result of failure to recognize *Helicobacter pylori* involvement in the treatment of ulcers in the 1970s and 1980s.

The temporal trend of peptic ulcer incidence follows a birth-cohort effect. The population born between the 1870s and 1900s had the highest risk of developing peptic ulcers, and as that group died, the incidence of peptic ulcers declined. This birth-cohort effect supports an environmental etiology of ulcers as opposed to a genetic cause; the environmental factor exerted its effect at a young age, and this factor is now postulated to be *H. pylori*. The widespread acquisition of *H. pylori* occurred at the onset of the industrial revolution of the nineteenth century, which was characterized

 Copyright 2004, Elsevier (USA). All rights reserved.

by increases in urbanization, overcrowding, and poor hygiene. Subsequently, in the second half of the twentieth century, increasing sanitation standards may have resulted in the decline in the transmission rate of *H. pylori* and ultimately the declining incidence of ulcer disease. However, the within-country incidence of peptic ulcer varies despite the same rate of *H. pylori* infection, suggesting that additional factors contribute towards the pathogenesis of ulcers and the changing temporal trends. Other factors postulated to be relevant to the declining incidence of peptic ulcer include dietary changes, reduced cigarette smoking, improved treatment for peptic ulcers, and improved diagnosis that excludes cases of non-ulcer dyspepsia, which may have been previously misdiagnosed as ulcers.

Duodenal ulceration is geographically heterogeneous. In Asia, the incidence of peptic ulcers continued to climb in the latter half of the twentieth century, in contrast to the trend in the West. In China, the prevalence of duodenal ulcers in the north in patients undergoing gastroscopy is 23% compared to 9.7% in the south, and the prevalence of duodenal ulcer is twice that of gastric ulcer. Asians, compared to Caucasians, present with ulcers at a younger age and the male to female ratio is higher.

Duodenal ulcer is now more common than gastric ulcer in the West, and affects 6–15% of the population. Despite improved endoscopic and pharmacological management of ulcer hemorrhage, the mortality rate has remained at 7% because more elderly patients with comorbidities now suffer from ulcer complications. Therefore, peptic ulcer disease imposes a significant economic burden on the community and a significant risk of morbidity and mortality on the individual.

ETIOLOGY

Factors implicated to be important for the cause of duodenal ulcers have ranged from personality type, emotional stress, and genetic susceptibility. However, in recent decades, two causes stand out as the principal etiological factors of duodenal ulcers—*H. pylori* and NSAIDs.

Helicobacter pylori

Helicobacter pylori was first isolated in 1982 at Royal Perth Hospital by Robin Warren and Barry Marshall. The serendipitous growth of the spiral organism on culture plates that were left in the incubator is legendary and marked the beginnings of a landmark medical discovery. At the time, an infection as the cause of ulcer disease was contrary to the dogma that bacteria could not survive the acid milieu of the stomach. In the two decades following its discovery, the role of *H. pylori* in causing chronic gastritis was proved through Koch's postulates, and eradication of the bacteria was found to prevent ulcer recurrence. *Helicobacter pylori* has now been shown to cause 80% of duodenal ulcers.

Helicobacter pylori, a flagellated and microaerophilic gram-negative bacillus related to the genus *Campylobacter*, is 3 μm long and has a characteristic spiral shape. The bacteria inhabit the gastric antral mucus layer, often overlying inflamed gastric mucosa. Biochemically, *H. pylori* produces the enzyme urease, which allows it to survive in the hostile acidic environment of the stomach by converting urea into ammonia to neutralize acid and maintain a periplasmic pH of 6.2. This acid-neutralizing ability is limited, however, and the organism survives poorly in the corpus of the stomach due to the high concentration of acid-producing parietal cells, and preferentially colonizes the less acidic gastric antrum. Two complete genomes of *H. pylori* have been published so far.

Helicobacter pylori is the most common infection worldwide, affecting 80% of the population in developing countries, but only 20–30% of Western nations. The bacteria are transmitted from person to person and are acquired in childhood through oral–oral or fecal–oral transmission. The prevalence of infection increases with age due to the cohort effect. Risk factors for infection include domestic overcrowding, lower socioeconomic status, poor hygiene, and exposure to gastric contents of infected individuals.

Aspirin and NSAIDs

Aspirin, introduced in the early 1900s, proved to be one of the major therapeutic advances in the treatment of inflammatory conditions and in the prevention of atherosclerotic vascular diseases. However, endoscopic studies as early as the 1930s demonstrated a predisposition toward aspirin causing ulcerogenesis and bleeding in the gastrointestinal tract, with a point prevalence for ulcer of 20%. This is believed to be related to cyclooxygenase (COX), which exists in two main isoforms. COX-1 is the constitutive form that is found in the gastrointestinal mucosa and platelets; it has a housekeeping role in maintaining the supply of prostaglandin necessary for gastrointestinal mucosal integrity. COX-2, inducible in inflammatory states, is the form of the enzyme produced by leukocytes. COX-3, a COX-1 variant that is expressed in the brain, may be the enzyme that becomes inhibited in the central analgesic and antipyretic effects of NSAIDs and acetaminophen.

NSAIDs and *Helicobacter pylori* Interaction

The role of NSAIDs ingestion in patients who are infected *H. pylori* is a controversial issue. There is now good evidence from prospective randomized studies and a meta-analysis to suggest that NSAIDs and *H. pylori* have at least an independent additive effect on the risk of ulceration.

Other Risk Factors

Smoking

Smoking increases acid production and inhibits duodenal and pancreatic bicarbonate secretion. Smoking alone is often insufficient to induce duodenal ulceration but may have an additive or synergistic effect in the presence of other risk factors.

Personality

The role of stress in causing gastroduodenal ulceration is well demonstrated in animal models; ice water immersion and restriction in movement cause ulceration in mice. Stressful occupation and the "type A" personality have been traditionally linked to the development of ulcer possibly through hyperacidity and local ischemia.

Hypercalcemia

Hypercalcemia increases acid secretion in healthy volunteers. In addition, hypercalcemia increases gastrin secretion in patients with gastrinoma.

Constitutional

People who have increased parietal cell mass, which is probably genetically determined, are more prone to developing duodenal ulceration because of acid hypersecretion. Monozygotic twins show a higher concordance rate for peptic ulcers than do dizygotic twins. Blood group O, Lewis phenotype Le (a−b−), and the ABH nonsecretor trait have also been implicated as genetic risk factors.

Zollinger−Ellison Syndrome

Zollinger−Ellison syndrome (ZES) is a gastrin-secreting neuroendocrine tumor that can be solitary or multifocal. Autonomous gastrin secretion increases the basal acid output (BAO) of the parietal cells and leads to peptic ulceration. There may be other concurrent endocrinopathies as part of a multiple endocrine neoplasia (MEN) syndrome. Screening tests include serum calcium, which can identify the hyperparathyroidism component of the MEN syndromes, and fasting gastrin level >1000 pg/ml. Gastrin elevation following injection with secretin supports ZES. BAO is typically increased (>15 mEq/hour). Radiology, scintigraphy, and selective venous sampling can localize these tumors. Gastrinomas are potentially malignant and should be treated by surgical resection.

Systemic Mastocytosis

In systemic mastocytosis, there is a proliferation of functionally or cytologically abnormal mast cells that may autonomously secrete histamine, among other mediators, resulting in hyperacidity. Patients may develop abdominal pain and gastroduodenal ulceration.

Other Causes of Ulceration in the Duodenum

Infective Ulcers

Cytomegalovirus, herpes simplex, fungal, and tuberculous infections can cause ulceration, most commonly in immunosuppressed patients and in the elderly.

Crohn's Disease

Crohn's disease rarely affects the proximal duodenum. Features suggestive of Crohn's disease include atypical location of deep ulcers, nodularity, perforation, stricturing, fistulization, non-caseating granulomas, and ulcers that are resistant to acid suppression and responsive to treatment of Crohn's disease.

Stress Ulcers

Stress ulcers (Cushing's ulcers) and ulcers in burns patients (Curling's ulcers) are acute, multifocal, superficial (0.5−2 cm) ulcers and erosions that show more necrosis than inflammation on histology. They may be caused by shock and local ischemia induced by cytokine release.

Neoplasm

Malignancy of the small bowel is uncommon and may present as primary neuroendocrine tumors, lymphomas, or metastasis from melanomas; hematological cancers; or infiltration by pancreas or gallbladder cancers.

Idiopathic

Data from Western countries show that 20% of patients have ulcers that are not associated with NSAIDs or *H. pylori*. A proportion of these patients may have taken aspirin-containing compounds or NSAIDs without reporting their use, or there may have been a false negative *H. pylori* test. However, there is a growing trend for these "non-NSAID, non-*H. pylori* ulcers" to be found worldwide. Most of these patients are elderly with comorbidities and, in particular, chronic liver disease.

PATHOGENESIS

Only 10–15% of patients infected with *H. pylori* develop duodenal ulcers. Ulcerogenesis requires a combination of host and bacteria factors, as illustrated in Table I.

Antral-Predominant Gastritis

Duodenal ulceration almost never occurs concurrently with gastric cancer, despite both being related to *H. pylori*. It appears that location and extent of the *H. pylori* gastritis determine why different diseases develop. The degree of luminal acidity is of prime importance in determining the site of gastritis. Patients with high acidity tend to develop body-sparing nonatrophic antral gastritis. People with this pattern of gastritis do not develop hypochlorhydria and tend to have increased acid secretion. This pattern is more likely if *H. pylori* were acquired in later childhood. Often this pattern occurs concurrently with duodenitis with gastric metaplasia due to the effect of the increased duodenal acid load.

Gastric Secretion, Duodenitis, and Duodenal Ulcer

Several factors have been found to be important in the pathogenesis of duodenal ulcer.

Gastric Acid Hypersecretion

Patients with duodenal ulcers have on average a higher acid output (both basal and maximal gastric acid output) than do healthy individuals. Often their parietal cell mass is higher and does not decrease following cure of *H. pylori* infection, suggesting an inherent predisposing factor for the development of ulceration independent of *H. pylori* infection.

TABLE I Host and Bacterial Factors Important in the Pathogenesis of Duodenal Ulcers

Host factors	Bacterial factors
Gastric acid hypersecretion	Virulence factors, cytotoxins
Increased parietal cell mass	Mucolytic enzymes
Increased gastrin production	Urease production
Increased gastrin sensitivity	Cytokine induction
Reduced somatostatin production by D cells	
Reduced intraduodenal neutralizing effect	
Inhibition of duodenal bicarbonate release	
Loss of gastrin inhibition effect	
Neutralization of bile acids	

Gastrin

Patients who are infected with *H. pylori* have an exaggerated gastrin release in response to meals and this phenomenon reverses following eradication of the bacteria. Factors that ordinarily inhibit gastrin release, namely, antral acidification, antral distension, and duodenal fat administration, are impaired in the setting of *H. pylori* infection. Reduction of somatostatin production by the antral D cells due to antral gastritis contributes to the loss of gastrin inhibition.

Gastric Metaplasia and Duodenal Acid Load

Gastric metaplasia, which develops in response to an increased duodenal acid load, allows colonization of *H. pylori* to occur in the duodenum. Colonization of *H. pylori* in the metaplastic gastric tissue induces further metaplasia because of bacteria-activated inflammation, duodenitis, and ulceration. Increased duodenal acid load, through precipitation of bile salts, neutralizes the inhibitory effect of bile on the survival of *H. pylori*. *Helicobacter pylori* also produces a nitric oxide synthase inhibitor that impairs bicarbonate secretion in response to acidification of the duodenal bulb in duodenal ulcer patients. This normalizes following the cure of *H. pylori* infection.

Helicobacter pylori and Duodenitis

Infection of the duodenum results in chronic active duodenitis, and the inflamed mucosa renders the duodenum even more susceptible to damage by acid and pepsin. *Helicobacter pylori* produces virulence factors that induce an inflammatory response. Vacuolating cytotoxin A (vacA), an extracellular toxin found frequently in *H. pylori* strains from patients with duodenal ulcers, increases polymorph recruitment and migration. The cytotoxin-associated gene A (cagA) protein is produced by 60% of strains in the United States and by almost 100% of strains in Asia. The cagA-positive strains induce more cytokines, especially interleukin-8 (IL-8), and are significantly associated with increased chronic inflammatory activity. IL-8 promotes neutrophil recruitment and activation, leading to inflammatory injury to the epithelium. The cagA protein, which is injected into the host cell following attachment of the organism to the epithelium, interacts with a tyrosine phosphatase to mediate phosphorylation.

CLINICAL FEATURES

Symptoms and Signs

The textbook description of duodenal ulcer pain states that it is intermittent and seasonal in spring

and autumn. The pain typically is epigastric and burning in quality, occurring in the evening and early morning, waking the patient from sleep between midnight and 3 am, corresponding to times of high duodenal acid load. Eating food and antacids relieves the pain. However, many patients with duodenal ulcers are asymptomatic. Importantly, 10% of patients taking a NSAID presenting with complicated ulcers do not have antecedent abdominal pain, possibly due to the analgesic effect from the drug. Conversely, many patients who have ulcerlike pain turn out to have gastroduodenitis or non-ulcer dyspepsia and not a duodenal ulcer. Thus, the sensitivity and specificity of typical dyspepsia for the diagnosis of duodenal ulcer is low.

Vomiting is uncommon unless there is scarring and deformity of the pylorus or duodenum. Patients may present with complications such as peritonitis if the ulcer perforates or with hematemesis or melena caused by the ulcer eroding into a major blood vessel, usually a branch of the gastroduodenal artery. Anemia may result from chronic iron deficiency from blood loss.Examination of a patient with duodenal ulcer is often unrewarding. There may be nonspecific tenderness at the epigastrium. Peritonitis, especially with right-sided tenderness, predominantly suggests perforation with peritoneal contamination down the right paracolic gutter. Right-sided abdominal tenderness and peritonitis may be caused by a perforated peptic ulcer with peritoneal contamination along the right paracolic gutter.

Diagnosis

Endoscopy

Esophagogastroduodenoscopy (EGD) has revolutionized the diagnosis of upper gastrointestinal ulcers. EGD can diagnose ulcers as well as treat bleeding ulcers, using various combinations of through-the-scope therapeutic capabilities, such as injection, clipping, and thermocoagulation. For these reasons, EGD is considered to be the gold standard for the diagnosis of peptic ulcers, but should be avoided in perforated duodenal ulcers because gaseous distension by the endoscope can aggravate peritonitis or result in pneumoperitoneum.

Imaging

Plain erect chest X-ray helps to exclude perforated peptic ulcers. Double-contrast barium meals can also detect ulcers.

Tests for Helicobacter pylori

As one of the major causes of duodenal ulcers, *H. pylori* is readily diagnosed by either invasive means (with an endoscopy) or noninvasive means (without an endoscopy), as shown in Table II. Noninvasive tests are convenient, inexpensive, and sufficiently accurate to diagnose *H. pylori*. Invasive tests can be performed if an EGD is indicated for the patient.

Biopsy urease test At least three commercially available biopsy urease kits test for the presence of the *H. pylori* enzyme urease collected during endoscopic gastric biopsies. The presence of this enzyme changes the color of a pH-sensitive indicator within 24 hours (but often within 1 hour), providing an easy and fast way of diagnosis at the time of endoscopy. Luminal blood can produce a false negative urease test result and therefore histology is recommended for patients who present with recent upper gastrointestinal bleeding. The sensitivity of the biopsy urease test is 89–95% with a specificity of close to 100%.

Histology and culture Histology of antral biopsies is the gold standard for the diagnosis of *H. pylori* and with proper staining technique has a sensitivity of 90–100%. Culture of antral biopsies is usually reserved for research purposes or to identify antibiotic susceptibility after failed attempts at *H. pylori* eradication.

Urea breath test The urea breath test (UBT) detects for the presence of gastric urease. Following ingestion of carbon-labeled urea (^{13}C or ^{14}C), *H. pylori*-produced urease in the stomach metabolizes the urea into ammonia and carbon dioxide. The carbon isotope is then exhaled as a labeled carbon dioxide and can be directly measured. The sensitivity of the urea breath test is reduced by recent use of antibiotics, bismuth, and acid suppressive drugs, which impair the ability of *H. pylori* to metabolize urea. The sensitivity and specificity of UBT is approximately 90% and 96% respectively.

Serology Whole blood, serum, and finger-prick serology are inexpensive and convenient but do not prove current *H. pylori* infection. Serum immunoglobulin G (IgG) antibodies are almost universally present in patients who have *H. pylori* infection and decline slowly following eradication therapy. However, 15% of patients continue to have a positive antibody result 6 months following a cure, therefore the test should

TABLE II Investigations for *Helicobacter pylori*

Invasive (endoscopy–biospy based)	Noninvasive (endoscopy not required)
Biopsy urease test (e.g., CLO test)	Urea breath test
Histology	Serology
Culture	Urine antibody
	Fecal antigen

not be employed for confirmation of eradication in the short term.

Fecal antigen, urine and saliva antibody Urinary and salivary excretion of *H. pylori* antibody can diagnose exposure to *H. pylori* but may not differentiate between past and current infection. Fecal antigen assays detect antigen and therefore can be used to monitor for eradication of the organism.

MANAGEMENT

The management of duodenal ulcers involves healing of the ulcer, removal of the cause of the ulceration (such as ceasing smoking and NSAID use), and eradication of *H. pylori*. For high-risk patients who are on long-term NSAIDs, maintenance acid suppressive therapy may be indicated.

Acid Suppression

Proton Pump Inhibitors

Proton pump inhibitors (PPIs) directly inhibit the H^+,K^+-ATPase pumps in the parietal cell and are more potent acid inhibitors than are H2 receptor antagonists (see Table III). Ulcer healing by PPIs occurs within 2 weeks in 70–80% of patients compared to 50–60% of those taking H2 receptor antagonists. The profound suppression of gastric acidity raises gastrin levels by two- to fourfold, but neuroendocrine tumors have not occurred in humans.

Histamine-2 Receptor Antagonists

H2 antagonists are particularly useful in the prevention of nocturnal acid production; a single nightly dose has been introduced for patients with nocturnal gastroesophageal reflux disease (GERD) and peptic ulcer healing.

TABLE III Acid Suppressive Therapy

Therapy	Dose (mg)[a]
Histamine-2 receptor antagonist	
Cimetidine	400
Ranitidine	150
Famotidine	20
Nizatidine	150
Proton pump inhibitor	
Omeprazole	20
Lansoprazole	30
Pantoprazole	40
Rabeprazole	20
Esomeprazole	40

[a] For histamine-2 receptor antagonists, the dose is twice/day or, alternatively, a double dose at night; for proton pump inhibitors, the dose is once/day.

TABLE IV High-Risk Ulcer Patients for Ulcer Recurrence and Ulcer Hemorrhage

Patient factors	Medication factors
Previous peptic ulcer or bleeding	High-dose, multiple NSAIDs
Elderly age	Concomitant corticosteroid usage
Comorbidities (e.g., heart or lung disease)	Concomitant anticoagulant usage

Long-Term Maintenance Therapy

Long-term maintenance therapy is advisable for high-risk patients and also for patients who have had ulcer complications while on NSAIDs or aspirin and who must remain on these medications (Table IV). Patients should be treated to eradicate *H. pylori* if it is present. The choice of maintenance therapy includes an H2 receptor antagonist, a PPI, or misoprostol (a prostaglandin E1 analogue). A high-dose H2 receptor antagonist such as famotidine (40 mg, twice daily) significantly reduces duodenal ulceration associated with NSAIDs. A PPI and misoprostol are also highly effective, but diarrhea with misoprostol limits its use, and 30% of patients cannot tolerate the drug at the required dose.

Eradication of *Helicobacter pylori*

Eradication of *H. pylori* substantially reduces the risk of ulcer recurrence and, for most patients, avoids the need for long-term maintenance therapy in preventing ulcer recurrence. The odds ratio of ulcer recurrence following eradication is 0.20 (95% CI, 0.13–0.31). At 6 months, ulcer treatment is successful in 1 out of 2.8 patients. Table V shows the ulcer recurrence rates according to successful or failed *H. pylori* eradication. Because the reinfection rate is low (0.6–2% per year), eradication therapy is both an effective and cost-effective means of treating ulcers.

Treatment

Eradication regimens (Table VI) that achieve a cure of >90% on a "per protocol" basis and >80% on "intention to treat" are recommended. First-line therapy consists of a PPI and two antibiotics—amoxicillin and clarithromycin or metronidazole given twice daily for 7 days. The original bismuth triple therapy, bismuth, metronidazole, and tetracycline (BMT) for 2 weeks, is less well tolerated and less convenient because it

TABLE V Meta-analyses on Duodenal Ulcer Relapse Rates Depending on *Helicobacter pylori* Status after Treatment

	Pooled ulcer recurrence rate	
Study	*H. pylori* eradicated	*H. pylori* not eradicated
Laine *et al.* (1998)	20	56
Hopkins *et al.* (1996)	6	67
Tytgat (1995)	3	58
Penston (1994)	7	67

involves a four-times-daily regimen. The combination of BMT with a PPI (so-called quadruple therapy) is an important option for patients who have had failed first-line therapy.

Confirmation of Eradication

No eradication regimen has a 100% cure rate. Patient compliance may not be satisfactory and antibiotic resistance is a growing concern. Therefore, confirmation of eradication is recommended for all patients who receive treatment, but is mandatory for those who present with ulcer complications. The urea breath test or stool antigen test are the preferred means of posteradication testing and should be preformed at least 4 weeks after treatment. Antisecretory drugs should be ceased at least 1 week prior to testing (Fig. 1).

Vaccination and Hygiene

Vaccination to control spread of *H. pylori* is under research and likely to be available in the future. Transmission of the organism because of poor hygiene and overcrowding is also a problem that needs to be addressed.

TABLE VI *Helicobacter pylori* Eradication Regimen[a]

Regimen[b]	Days of treatment	Eradication success rate (%)
PBMT	7	85
PCM	7	84
PCA	7	82
PMA	7	76
BMT	14	80
PC	14	65
BMA	14	62
PA	14	58

[a] Meta-analysis of *H. pylori* eradication regimens (119 studies, 6416 patients). Modified from Taylor *et al.* (1997).

[b] P, Proton pump inhibitor; B, bismuth; M, metronidazole; T, tetracycline; C, clarithromycin; A, amoxicillin.

Treating Other Causes of Peptic Ulcers

Aspirin and NSAIDs

Drugs less ulcerogenic than aspirin and NSAIDs should be used if patients develop duodenal ulcers from taking these medications. Clopidogrel, an antiplatelet agent that targets the adenosine diphosphate receptor on platelets, can be substituted for aspirin. This agent is safe, well tolerated, and clinically has a superior efficacy in cardiovascular secondary prophylaxis compared to aspirin, and has lower gastrointestinal toxicity. Higher cost, however, limits its widespread use. A COX-2 inhibitor may be substituted for NSAIDs. COX-2 inhibitors generally have a better gastroduodenal safety profile than non-selective NSAIDs, especially in short-term treatment of less than 6 months. COX-2 inhibitors also do not have the antithromboxane effects of NSAIDs, and do not protect against cardiovascular diseases. The addition of aspirin to COX-2 inhibitors negates the beneficial effects of COX-2 inhibitors in the upper gastrointestinal tract.

Smoking

Smoking increases the risk of peptic ulcers and therefore patients should be advised to stop smoking.

Recurrent Duodenal Ulcers and Idiopathic Ulcers

Despite the successful eradication of *H. pylori*, ulceration may still recur in some people. The causes may include unreported or inadvertent use of NSAIDs or aspirin, failed *H. pylori* eradication, or a reinfection with *H. pylori*. One report from Hong Kong shows that 17% of endoscopically diagnosed duodenal ulcers are not associated with *H. pylori* or NSAIDs, and 64% of these patients have concomitant diseases. The pathogenesis of these ulcers is unknown. Managing these patients includes treating the underlying disease and long-term administration of acid suppressants.

CONCLUSIONS

The incidence of duodenal ulcers is expected to decline as the prevalence of *H. pylori* infection is reduced worldwide and the use of the less toxic COX-2 inhibitors is increased. The growing availability and use of acid suppressant drugs and, in particular, proton pump inhibitors will also reduce the incidence ulcer disease. However, with the aging population and the increasing numbers of patients requiring aspirin for cardiovascular prophylaxis, more elderly patients are at risk of ulcer disease and complications. They may remain

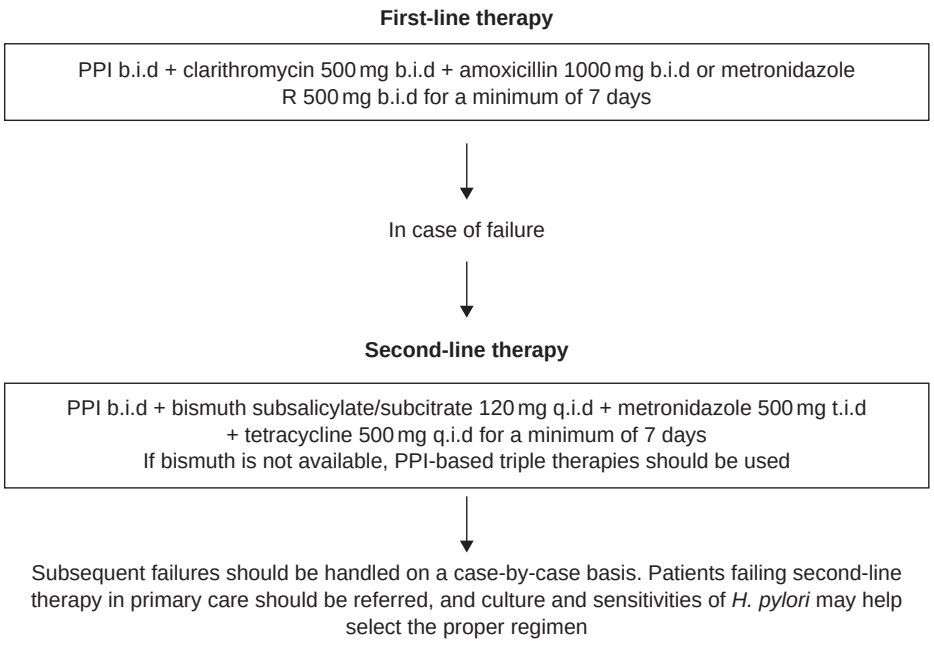


FIGURE 1 Summary of the recommended treatment strategy for eradication of *Helicobacter pylori*. In first-line therapy, the proton pump inhibitor (PPI)/clarithromycin/amoxicillin combination is preferred. Abbreviations: b.i.d., bis in die (twice daily); q.i.d, quater in die (four times daily); t.i.d, ter in die (three times daily). Adapted from Malfertheiner *et al*. (2002).

asymptomatic until the development of an ulcer complication. Due to their comorbidities, ulcer morbidity and mortality may remain substantial. There is also a worldwide trend of increasing incidence of non-*H. pylori*, non-NSAID ulcers, the pathogenic mechanism of which remains poorly understood. These ulcers tend to affect patients with concomitant diseases. Further study into this phenomenon is required.

See Also the Following Articles

Breath Tests • Functional (Non-Ulcer) Dyspepsia • Gastric Ulcer • Gastrointestinal Tract Malignancies, Radiation and Chemotherapy • H2-Receptor Antagonists • *Helicobacter pylori* • Mastocytosis, Gastrointestinal Manifestations of • Mucosa-Associated Lymphoid Tissue (MALT) • NSAID-Induced Injury • Proton Pump Inhibitors • Smoking, Implications of

Further Reading

Holcombe, C., Omotara, B. A., Eldridge, J., and Jones, D. M. (1992). *H. pylori*, the most common bacterial infection in Africa: A random serological study. *Am. J. Gastroenterol.* **87**, 28–30.

Hopkins, R. J., Girardi, L. S., and Turney, E. A. (1996). Relationship between *Helicobacter pylori* eradication and reduced duodenal and gastric ulcer recurrence: A review. *Gastroenterology* **110**(4), 1244–1252.

Laine, L., Hopkins, R. J., and Girardi, L. S. (1998). Has the impact of *Helicobacter pylori* therapy on ulcer recurrence in the United States been overstated? A meta-analysis of rigorously designed trials. *Am. J. Gastroenterol.* **93**(9), 1409–1415.

Malfertheiner, P., Mégraud, F., O'Morain, C., Hungin, A. P. S. Jones, R., Axon, A., Graham, D. Y., and Tytgat, G. (2002). Current concepts in the management of *Helicobacter pylori* infection. The Maastricht 2-2000 Consensus Report. *Aliment. Pharmacol. Ther.* **16**, 167–180.

Marshall, B. (ed.). (2002). "Helicobacter Pioneers." Blackwell Publ., Victoria, Australia.

Penston, J. G. (1994). Review article: *Helicobacter pylori* eradication—Understandable caution but no excuse for inertia. *Aliment. Pharmacol. Ther.* **8**(4), 369–839.

Taylor, J. L., Zagari, M., Murphy, K., and Freston, J. W. (1997). Pharmacoeconomic comparison of treatments for the eradication of *H. pylori*. *Arch. Intern. Med.* **157**, 87.

Tytgat, G. N. (1995). Peptic ulcer and *Helicobacter pylori*: Eradication and relapse. *Scand. J. Gastroenterol. Suppl.* **210**, 70–72.

Duodenitis

M. M. WALKER* AND N. J. TALLEY†

Imperial College School of Science, Technology and Medicine, London and †University of Sydney, Nepean Hospital

granuloma Focal nodular accumulation of histiocytes in tissue.

H2 blocker Compound that inhibits gastric acid secretion; also known as histamine-2 receptor antagonist.

Helicobacter pylori Bacterium that infects the stomach, induces gastritis, and increases the risk for peptic ulcers in the duodenum and stomach.

metaplasia Replacement of one type of epithelium with another type normally not present.

proton pump inhibitor Group of compounds of the substituted benzimidazole class; inhibit the H^+,K^+-ATPase of the parietal cell (proton pump), thus inhibiting gastric acid secretion.

Duodenitis is inflammation of the duodenal mucosa and may be defined by either characteristic endoscopy appearances (from erythema to erosions or ulceration) or by histopathology (reflecting a definitive increase in chronic inflammatory cells in the lamina propria with or without a neutrophil infiltrate).

PEPTIC DUODENITIS

Most gastroenterologists view duodenitis as part of the peptic diathesis due to gastric *Helicobacter pylori* infection or, more rarely, nonsteroidal antiinflammatory drugs (NSAIDs). Ischemia may be another etiologic factor. Idiopathic erosive duodenitis has been described. At endoscopy, duodenitis may be manifest as an erythematous or exudative, erosive, hemorrhagic, or nodular appearance, as described in the endoscopic arm of the Sydney classification of gastritis. Other endoscopic classifications include that of Joffe, in which the term "salami duodenum" is embodied, meaning erosions associated with or without petechial hemorrhages. Nodular duodenitis, characterized endoscopically by multiple erythematous nodules in the proximal duodenum, is a distinct entity. Although the association of duodenitis with dyspepsia remains unclear, erosive duodenitis probably causes symptoms. Erosive duodenitis signifies increased duodenal ulcer risk and should be treated as such with confirmation of *H. pylori* status and exclusion of NSAID use. Nonerosive duodenitis may also reflect ulcer risk but the link is controversial. Duodenitis may also be due to infection, particularly in immunosuppressed patients, as well as associated with inflammatory bowel disease (e.g., Crohn's disease), and rarely may be seen in celiac disease.

EPIDEMIOLOGY

In a large series of patients with dyspepsia at endoscopy, 9% had endoscopic duodenitis. By histology, duodenitis has been described in 32% of an asymptomatic population, but less than 10% of these patients had moderate to severe duodenal inflammation.

PATHOLOGY AND BIOPSY OF THE DUODENUM

The correlation between endoscopic duodenitis and histology is relatively good. Mild duodenitis is defined histologically by an increase in cellularity of the lamina propria. This can be difficult to assess because the duodenal mucosa has normally many mononuclear cells present—therefore, a definite increase in lymphocytes should be seen. The normal range of intraepithelial lymphocytes is 10−30/100 epithelial cells. If erosion of the surface epithelium is seen or neutrophils are present in the lamina propria or epithelium, this is regarded as active duodenitis. Gastric metaplasia (due to chronic acid insult) is restricted to the surface epithelium and should be confirmed by a periodic acid Schiff (PAS) stain to show the mucin pattern of the surface epithelial cells. Brunner gland hyperplasia may cause a nodular endoscopic appearance and is believed to be an adaptive response to acid. Hyperplastic polyps occur occasionally. Granulomas and excess macrophages may also be seen. Pathogens such as *Giardia lamblia* should be actively excluded, because these do not always provoke inflammation.

The duodenum should be biopsied by site according to clinical indication. The duodenum is divided into four parts, but the fourth part is rarely reached with the standard forward-viewing endoscope. Peptic

 Copyright 2004, Elsevier (USA). All rights reserved.

duodenitis and gastric metaplasia occur in the first and second parts, partial villous atrophy occurs in the second and third parts, and pathogens may infect any part. Antral mucosa may protrude into the duodenum in tongues, up to 10 mm distal to the pylorus. Duodenitis and gastric metaplasia can be patchy and at least two biopsies are recommended, one from the anterior wall and one from the roof. Partial villous atrophy is also patchy and it is preferable to take two biopsies from each of the second and third parts. Culture of biopsy or duodenal aspirates can aid identification of pathogens.

PATHOGENS

If there is immunosuppression, the common duodenal pathogens are microsporidia, cryptosporidia, mycobacteria, and cytomegalovirus. In mycobacterial infection, endoscopy may show a coarse granular mucosa or characteristic fine white nodules, which are particularly prominent in the duodenum. Infection with the protozoan *G. lamblia* also shows nonspecific endoscopic and pathological changes. Whipple's disease, due to *Tropheryma whippelii*, is a rare multisystem bacterial infection with characteristic duodenal pathology. Schistosomiasis can cause duodenal ulceration.

INFLAMMATORY BOWEL DISEASE

Crohn's disease can be present throughout the duodenum and has a high histological prevalence (up to 12%) in patients with this disease. This condition is strongly associated with active inflammation and, in the absence of *H. pylori* infection, elsewhere this finding should arouse suspicion of Crohn's disease. Ulcerative colitis is only rarely associated with duodenitis, although reports are increasing in frequency, particularly in children.

TREATMENT

Treatment of duodenitis involves establishing the underlying cause and instituting appropriate therapy. In peptic duodenitis, this involves eradication of *H. pylori* or cessation of NSAIDs. If NSAIDs are unavoidable, appropriate antiulcer cotherapy with a proton pump inhibitor (PPI) or H2-blocker (H2B) must be instituted. Idiopathic erosive duodenitis associated with dyspepsia usually responds to acid suppression with a PPI or H2B. Pathogens, once identified by biopsy or culture, can be suitably treated.

See Also the Following Articles

Crohn's Disease • Duodenal Ulcer • H2-Receptor Antagonists • Helicobacter pylori • NSAID-Induced Injury • Proton Pump Inhibitors

Further Reading

Caselli, M., Gaudio, M., Chiamenti, C. M., Trevisani, L., Sartori, S., Saragoni, L., Boldrini, P., Dentale, A., Ruina, M., and Alvisi, V. (1998). Histologic findings and *Helicobacter pylori* in duodenal biopsies. *J. Clin. Gastroenterol.* **26**, 74–80.

Joffe, S. N., Lee, F. D., and Blumgart, L. H. (1978). Duodenitis. *Clin. Gastroenterol.* **7**, 635–650.

Misiewicz, J. J., Price, A. B., and Tytgat, G. N. J. (1991). Working party report to the World Congress of Gastroenterology. Sydney 1990. The Sydney system: A new classification of gastritis. *J. Gastroenterol. Hepatol.* **6**, 207–234.

Talley, N. J., and Phillips, S. F. (1988). Non-ulcer dyspepsia: Potential causes and pathophysiology. *Ann. Intern. Med.* **108**, 865–879.

Walker, M. M., and Crabtree, J. E. (1998). *Helicobacter pylori* infection and the pathogenesis of duodenal ulceration. *N.Y. Acad. Sci.* **859**, 96–111.

Duodenum, Anatomy

ROBERT ANDERS
University of Chicago Hospitals

duodenum The first 25 cm of the small intestine.
mucosa An epithelial-lined surface with supporting tissues that forms the digestive and absorptive surface of the small intestine.

The duodenum forms the first part of the small intestine. Medieval anatomists measured the duodenum at 12 finger lengths and named it for the Latin word for twelve (*duodecim*). Starting at the gastric outlet and ending at the duodejejunal junction, the duodenum is C-shaped and has been arbitrarily divided into the superior, descending, horizontal, and ascending segments. The superior segment travels posteriorly at the level of the first lumbar vertebra. The remaining segments are located in the retroperitoneal space. The descending segment contains the ampulla of Vater, a sphincter-guarded orifice through which the biliary and pancreatic ducts drain. An additional minor duodenal papilla allows for drainage of the accessory pancreatic duct but is variably present. The horizontal segment, at the level of the third lumbar vertebra, crosses the body's midline. The duodenum ends as it emerges from behind the peritoneum where a fibrous band, the suspensory ligament of the duodenum (ligament of Treitz), marks the beginning of the jejunum.

IMPORTANT GROSS ANATOMY RELATIONS

The blood supply to the duodenum is from both the celiac artery via the superior pancreaticoduodenal branches of the common hepatic artery and the gastroduodenal artery and the superior mesenteric artery via the inferior pancreaticoduodenal artery. The gastroduodenal artery is closely applied to the posterior wall of the duodenum and perforation from deep ulceration can lead to life-threatening hemorrhage. The head of the pancreas is tucked in the curve of the C-shaped duodenum. Space-occupying lesions of the pancreas can deform the duodenum or affect the function of the major duodenal papilla. Since the duodenum is the only small intestinal segment located in the retroperitoneum, the duodenum is firmly attached to the body wall in the duodenal fossa. This feature allows for identification of the freely mobile jejunum during endoscopic procedures.

MICROSCOPIC ANATOMY

The wall of the duodenum is composed of several distinct layers. The innermost three layers constitute the mucosa. First, there is a layer of stratified ciliated columnar epithelial cells facing the lumen; second, there is the lamina propria, which contains mucus-secreting Brunner's glands, connective tissue, lymphoid tissue, vascular tissue, and neural tissue; and third, there is a layer of smooth muscle called the muscularis mucosa. The submucosa is an interface between the mucosa and the muscularis propria layers. The submucosa is of variable thickness and forms a distribution network of vascular and lymphatic channels as well as nerve fibers and ganglion cells forming the Meissner's plexus. The next layer is the muscularis propria, which is composed of an inner circular smooth muscle layer and an outer longitudinal smooth muscle layer, separated by the neural plexus of Auerbach. The outermost layer is the serosa, which contains connective, adipose, and vascular tissue with a single mesothelial covering.

IMPORTANT MICROSCOPIC CONSIDERATIONS

The epithelial cell functions to control the digestion and absorption of intestinal nutrients. The functional surface area of the duodenum (and the rest of the small intestine) is critical to intestinal function. The mucosal surface area is increased by dense circumferential mucosal folds referred to as plicae circularis. A further expansion of the surface is seen in the formation of villous projections and crypts or invaginations between villi. The brush border is a special ciliated epithelial interface with microvilli present at the apex of the columnar epithelial cells that project into the gastrointestinal lumen. Goblet, endocrine, and paneth cells are specialized neighboring epithelial cells that are also found in the surface epithelium. Nonepithelial cells, such as lymphocytes, eosinophils, mast cells, and macrophages, are scattered throughout the lamina propria and are also present in follicular aggregates. These cells play a role in the immune-mediated defense of the intestine.

 Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Biliary Tract, Anatomy • Duodenal Motility • Gastrointestinal Tract Anatomy, Overview • Small Intestine, Anatomy • Stomach, Anatomy

Further Reading

Crawford, J. (1999). The gastrointestinal tract. *In* "Robbins Pathologic Basis of Disease" (R. Cotran, V. Kumar, and T. Collins, eds.), 6th Ed., pp. 775–844. W. B. Saunders, Philadelphia, PA.

Donald, A., and Madara, J. (1998). Functional anatomy of the gastrointestinal tract. *In* "Pathology of the Gastrointestinal Tract" (S. Ming and H. Goldman, eds.), 2nd Ed., pp. 13–46. Williams & Wilkins, Baltimore, MD.

Phillips, A. (1995). The small intestinal mucosa. *In* "Gastrointestinal and Oesophageal Pathology" (R. Whitehead, ed.), 2nd Ed., pp. 33–43. Churchill Livingston Press, Edinburgh, UK.

Dysphagia

DEVANG N. PRAJAPATI AND REZA SHAKER
Medical College of Wisconsin, Milwaukee

achalasia An esophageal motor disorder characterized by failure of the lower esophageal sphincter to relax and loss of esophageal peristalsis.

fundoplication Surgical procedure for the treatment of gastroesophageal reflux disease that involves strengthening the lower esophageal sphincter by wrapping the fundus around it.

Zenker's diverticulum A pharyngeal out-pouching above the upper esophageal sphincter as the result of weakness in the posterior pharyngeal wall.

Dysphagia, from the Greek meaning "difficulty swallowing," is a symptom describing the sensation of food being held up from passage into the digestive tract after it is swallowed. Although any number of abnormalities involving the swallowing mechanism can result in dysphagia, it is most commonly associated with esophageal disorders. An accurate incidence of dysphagia is difficult to determine; however, population-based estimates have suggested a prevalence of 7% in the general population and up to 50% of elderly institutionalized patients.

PATHOPHYSIOLOGY

Swallowing involves a complex network of coordinated neuromuscular activities that serve to transport food and fluid from the oral cavity to the stomach, while at the same time protecting the airway. A defect in any part of the swallowing mechanism can lead to symptoms of dysphagia. A volitional component of the deglutitive process transfers the bolus to the pharynx and an automatic component transports this bolus into the esophagus and stomach. The deglutitive process has been broken down into various phases: the preparatory oral phase, pharyngeal phase, and esophageal phase. The oral phase involves the preparation of the food into a bolus that can be transported into the esophagus and involves the oral musculature as well as the salivary glands for digestion and lubrication. Once the bolus is ready, it is transferred by the tongue to the posterior pharynx. A series of reflexive responses then occurs, resulting in a conformational change of the pharynx from a respiratory to a digestive pathway. The proximal esophagus transports the bolus down to the stomach in an orderly fashion by peristalsis aided by gravity. The duration of the oropharyngeal phase is usually a couple of seconds and the esophageal phase is less than 10 s. Dysphagia results when there is a failure of the oropharyngeal or esophageal neuromusculature to propel a food bolus into the stomach in a normal pattern or when there is an obstruction to normal passage due to narrowing of the esophagus.

CLASSIFICATION

Dysphagia has traditionally been categorized into two groups on the basis of the location of the problem,

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Biliary Tract, Anatomy • Duodenal Motility • Gastrointestinal Tract Anatomy, Overview • Small Intestine, Anatomy • Stomach, Anatomy

Further Reading

Crawford, J. (1999). The gastrointestinal tract. *In* "Robbins Pathologic Basis of Disease" (R. Cotran, V. Kumar, and T. Collins, eds.), 6th Ed., pp. 775–844. W. B. Saunders, Philadelphia, PA.

Donald, A., and Madara, J. (1998). Functional anatomy of the gastrointestinal tract. *In* "Pathology of the Gastrointestinal Tract" (S. Ming and H. Goldman, eds.), 2nd Ed., pp. 13–46. Williams & Wilkins, Baltimore, MD.

Phillips, A. (1995). The small intestinal mucosa. *In* "Gastrointestinal and Oesophageal Pathology" (R. Whitehead, ed.), 2nd Ed., pp. 33–43. Churchill Livingston Press, Edinburgh, UK.

Dysphagia

DEVANG N. PRAJAPATI AND REZA SHAKER
Medical College of Wisconsin, Milwaukee

achalasia An esophageal motor disorder characterized by failure of the lower esophageal sphincter to relax and loss of esophageal peristalsis.

fundoplication Surgical procedure for the treatment of gastroesophageal reflux disease that involves strengthening the lower esophageal sphincter by wrapping the fundus around it.

Zenker's diverticulum A pharyngeal out-pouching above the upper esophageal sphincter as the result of weakness in the posterior pharyngeal wall.

Dysphagia, from the Greek meaning "difficulty swallowing," is a symptom describing the sensation of food being held up from passage into the digestive tract after it is swallowed. Although any number of abnormalities involving the swallowing mechanism can result in dysphagia, it is most commonly associated with esophageal disorders. An accurate incidence of dysphagia is difficult to determine; however, population-based estimates have suggested a prevalence of 7% in the general population and up to 50% of elderly institutionalized patients.

PATHOPHYSIOLOGY

Swallowing involves a complex network of coordinated neuromuscular activities that serve to transport food and fluid from the oral cavity to the stomach, while at the same time protecting the airway. A defect in any part of the swallowing mechanism can lead to symptoms of dysphagia. A volitional component of the deglutitive process transfers the bolus to the pharynx and an automatic component transports this bolus into the esophagus and stomach. The deglutitive process has been broken down into various phases: the preparatory oral phase, pharyngeal phase, and esophageal phase. The oral phase involves the preparation of the food into a bolus that can be transported into the esophagus and involves the oral musculature as well as the salivary glands for digestion and lubrication. Once the bolus is ready, it is transferred by the tongue to the posterior pharynx. A series of reflexive responses then occurs, resulting in a conformational change of the pharynx from a respiratory to a digestive pathway. The proximal esophagus transports the bolus down to the stomach in an orderly fashion by peristalsis aided by gravity. The duration of the oropharyngeal phase is usually a couple of seconds and the esophageal phase is less than 10 s. Dysphagia results when there is a failure of the oropharyngeal or esophageal neuromusculature to propel a food bolus into the stomach in a normal pattern or when there is an obstruction to normal passage due to narrowing of the esophagus.

CLASSIFICATION

Dysphagia has traditionally been categorized into two groups on the basis of the location of the problem,

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Causes of Oropharyngeal Dysphagia

Neurological diseases
Cerebrovascular disease
Parkinson's disease
Amyotrophic lateral sclerosis
Multiple sclerosis
Huntington's disease
Dementia
Brainstem neoplasms
Muscular diseases
Myasthenia gravis
Inflammatory and toxic myopathies (e.g., polymyositis, thyrotoxicosis)
Muscular dystrophies
Structural
Cricopharyngeal bar/achalasia
Cervical osteophytes
Zenker's diverticulum
Oropharyngeal neoplasms

either oropharyngeal or esophageal. Oropharyngeal dysphagia, as the name suggests, results from disorders of the oropharyngeal phase of swallowing and is commonly associated with neuromuscular disorders. A list of the common causes of oropharyngeal dysphagia is presented in Table I. Esophageal dysphagia is typically a result of mechanical obstruction or motor dysfunction of the esophagus, although both causes are not mutually exclusive. The degree of esophageal narrowing required before symptoms develop is variable but most patients will not experience dysphagia at an esophageal diameter above 20 mm. A list of the causes of esophageal dysphagia is given in Table II. Globus sensation is a common complaint, though not true dysphagia. Although the patient describes a sensation of a lump in the throat or neck, there is no clear structural or mechanical etiology, and it is usually unrelated to swallowing. In fact, globus is often most noticeable when the patient is not eating and may in fact improve when they eat.

TABLE II Causes of Esophageal Dysphagia

Mechanical (narrowing of the esophageal lumen)
Intrinsic mucosal stricture
Peptic (gastroesophageal reflux-related)
Caustic injury
Pill injury
Esophageal ring
Esophageal web
Malignancy
Extrinsic compression
Mediastinal mass
Vascular compression—Aortic arch or aberrant other vascular structures such as right brachial cephalic artery (dysphagia lusoria)
Motor (abnormality in esophageal smooth muscle function)
Achalasia
Esophageal spasm
Hypertensive LES
Scleroderma

EVALUATION

History and Physical Examination

An accurate history is vital in the diagnosis of the etiology of dysphagia and can facilitate further work-up. Asking the patient to describe what happens when they eat or asking them to describe a swallow can be very helpful in distinguishing oropharyngeal from esophageal dysphagia. Certain historical features that suggest an oropharyngeal cause of dysphagia include poor bolus transfer, cough during swallow, and nasal regurgitation. On the other hand, patients with esophageal dysphagia typically describe food "sticking" or getting "hung up" in the chest when they swallow. Unfortunately, the patient's localization of where the food sticks does not help differentiate between a proximal or a distal esophageal problem, although if the patient describes a more distal esophageal location, there is a greater correlation. Clinical points to elucidate from the history are whether the dysphagia is related to solids or liquids or both. Dysphagia to solids occurs prior to liquids in most causes of mechanical esophageal obstruction. On the other hand, motor disorders typically affect both liquids and solids. Progressive dysphagia is more suggestive of a progressive narrowing of the esophagus, often related to malignancy (rapid progression) or gastroesophageal reflux disease (slow progression), whereas intermittent dysphagia is most often related to esophageal rings. Odynophagia refers to pain with swallowing, typically in the chest, and is most commonly associated with pill-induced esophagitis. In immunosuppressed patients, it is typically seen with infectious esophagitis.

In addition to a description of the nature of the dysphagia itself, a thorough history will assist in determining the etiology of the dysphagia. Gastroesophageal reflux may lead to dysphagia by two mechanisms. First, chronic reflux can lead to esophageal strictures that can cause dysphagia. Second, peptic esophagitis can result in esophageal dysmotility that may lead to dysphagia. Conditions such as scleroderma are associated with esophageal motor disorders that can cause dysphagia.

Physical examination in most patients with dysphagia is normal. Occasionally, watching the patient swallow will demonstrate a neuromuscular abnormality. Listening to the side of the neck during swallow may reveal a "sloshing," which is related to the presence

of food or fluid in a Zenker's diverticulum. A brief neurologic examination may reveal an underlying neurologic cause of dysphagia.

Laboratory and Radiologic Evaluation

Laboratory studies are normal in most patients with dysphagia. Anemia may be seen from bleeding as a result of a number of esophageal mucosal abnormalities. Iron deficiency anemia of any cause may result in dysphagia as a result of a web in the proximal esophagus, a condition referred to as Plummer-Vinson syndrome. Thyroid testing is routinely performed, since subtle abnormalities in thyroid functions may result in muscular weakness leading to dysphagia.

Barium fluoroscopic studies play an important role in the evaluation of patients with dysphagia. Symptoms suggestive of an oropharyngeal cause of dysphagia are best evaluated by an oropharyngoesophagram. This test involves the administration of oral contrast medium while cineradiography of the oropharyngeal and esophageal portions of the swallow mechanism is performed, often in conjunction with a speech pathologist.

The barium esophagram has traditionally been performed to evaluate any esophageal motor or mechanical abnormality. If the patient complains of dysphagia to solids, a challenge with a barium-coated solid, such as a marshmallow, will help determine the site of obstruction if the patient's symptoms are reproduced. Generally, subtle mechanical obstruction and most motor disorders are better detected with a barium esophagram than endoscopy. Certain motor disorders, such as achalasia and scleroderma, have classic radiographic patterns that can be diagnostic.

Endoscopy and Esophageal Manometry

Endoscopy in patients with dysphagia is helpful in two respects: it allows direct visualization of the esophageal mucosa and the opportunity to perform esophageal dilation, with either balloons or bougies to help relieve symptoms of dysphagia. In patients with symptoms suggestive of a motor cause of their dysphagia, esophageal manometry can help determine the nature of the disorder. Manometry involves placement of a fine flexible tube through the nose into the esophagus to measure contractile pressures of the esophageal body and lower esophageal sphincter. Disorders such as achalasia and esophageal spasm can be detected with this technique. Manometry of the upper esophageal sphincter can also be performed; however, this is largely a clinical research tool at this time.

MANAGEMENT

Although diagnosing the cause for dysphagia is not typically difficult, treating it can be challenging. Particularly frustrating is oropharyngeal dysphagia, which can be very debilitating with an increased risk for aspiration and sometimes necessitating a feeding tube. Cerebrovascular accidents are the most common cause of oropharyngeal dysphagia and fortunately many patients recover their swallowing function with time. Certain oropharyngeal exercises can strengthen weak muscles and improve symptoms of dysphagia.

Esophageal dysphagia treatment is tailored to its cause. Gastroesophageal reflux-related dysphagia can be treated with dilation followed by acid suppressive medication or surgical fundoplication. Benign structural disorders, such as rings, webs, and strictures, are typically treated with endoscopic dilation with balloon or bougienage, although many patients require repeated dilations over time. Occasionally, patients are able to self-dilate their esophagus with bougies that they pass themselves as needed. Medical therapy for motor disorders is typically limited to smooth-muscle-relaxing agents, such as nitrates and calcium channel blockers. Although they can be effective, the durability of the effect is often short. A recent development has been the use of botulinum toxin in patients with achalasia. Injection of this agent at the time of endoscopy is helpful in the treatment of this disorder; however, the duration of benefit is limited. Finally, surgical resection of the esophagus is therapy for malignancy, if cure is considered to be possible by performing this procedure. Endoscopically placed esophageal stents have been used for palliation of dysphagia secondary to malignancy; however, newer types of stents are being developed for use in benign conditions.

SUMMARY

Dysphagia is a symptom associated with a myriad of disorders involving the esophagus and oropharynx. A detailed history can often determine the etiology of the symptom, which can be further evaluated with barium radiography and endoscopy. Although therapies for oropharyngeal dysphagia are limited, esophageal dysphagia can be treated effectively, largely by endoscopic means. In the future, newer methods of treatment will help patients with this symptom resume as normal deglutition as possible.

See Also the Following Articles

Achalasia • Esophageal Strictures • Swallowing • Zenker's Diverticulum

Further Reading

Barer, D. H. (1989). The natural history and functional consequence of dysphagia after hemisphere stroke. *J. Neurol. Neurosurg. Psychiatry* **52**, 236–241.

Castell, D. O., and Donner, M. W. (1987). Evaluation of dysphagia: a careful history is crucial. *Dysphagia* **2**, 65–71.

Cook, I. J, and Kahrilas, P. J. (1999). AGA technical review on management of oropharyngeal dysphagia. *Gastroenterology* **116**, 455–478.

Dodds, W. J., Logemann, J. A., and Stewart, E. T. (1990). Radiologic assessment of abnormal oral and pharyngeal phases of swallowing. *Am. J. Roentgenol.* **154**, 965–9741.

Kahrilas, P. J., Clouse, R. E., and Hogan, W. J. (1994). American Gastroenterological Association technical review on the clinical use of esophageal manometry. *Gastroenterology* **107**, 1865–1884.

Martin-Harris, B., Logemann, J. A., McMahon, S., and Schleicher, M. (2000). Clinical utility of the modified barium swallow. *Dysphagia* **15**, 136–141.

Miller, R. M., and Langmore, S. E. (1994). Treatment efficacy for adults with oropharngeal dysphagia. *Arch. Phys. Med. Rehabil.* **75**, 1256–1262.

Richter, J. E. (1998). Dysphagia, odynophagia, heartburn, and other esophageal symptoms. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., Vol. 1, pp. 97–105. W. B. Saunders, Philadelphia.

Shaker, R., Easterling, C., Kern, M., Nitschke, T., Massey, B., Daniels, S., Grande, B., Kazandjian, M., and Dikeman, K. (2002). Rehabilitation of swallowing by exercise in tube fed patients with pharyngeal dysphagia secondary to abnormal UES opening. *Gastroenterology* **122**, 1314–1321.

Shaker, R. (1992) Oropharyngeal dysphagia: Practical approach to diagnosis and management. *Semin. Gastrointest. Dis.* **3**, 115–128.

Spechler, S. J. (1999). American Gastroenterologic Association technical review on treatment of patients with dysphagia caused by benign disorders of the distal esophagus. *Gastroenterology* **117**, 233–254.

Wilcox, C. M., Alexander, L. N., and Clark, W. S. (1995). Localization of an obstructing esophageal lesion. Is the patient accurate? *Digest. Dis. Sci.* **40**, 2192–2196.

E

Ehlers–Danlos Syndrome

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

arthrochalasia Joint instability or tendency to dislocate.
collagen Strong, elastic, fibrous protein; the foundation of all connective tissues in the body.
dermatosparaxis Skin fragility.
kyphoscoliosis Backward and lateral curvature of the spine.

Ehlers–Danlos syndrome refers to a heterogeneous group of heritable collagen disorders characterized by joint hypermobility and increase in skin elasticity and tissue fragility. Originally described in 1892, the disorder was expanded to include 10 subtypes, but these groupings were simplified in 1997 to include six types. The most current classification scheme recognizes these six types.

ETIOLOGY AND PATHOGENESIS

A variety of mutations have been described in Ehlers–Danlos syndrome (EDS) (see Table I), and most involve defects in posttranslational modification of collagen. Mutations in the type V collagen genes account for up to 50% of cases of classic EDS. The most well-characterized genetic defect is found in the vascular type of EDS, which results from mutations in the gene for type III procollagen (*COL3A1*). The collagen is thus structurally abnormal and affords diminished support of cutaneous, skeletal, vascular, and other structures.

TABLE I Classification Scheme for Ehlers–Danlos Syndrome

Type	Description
Classic (former EDS I and II)	Autosomal dominant, accounts for 80% of all cases and characterized mainly by skin and joint involvement
Hypermobile (former EDS III)	Autosomal dominant, with predominantly joint involvement
Vascular (former EDS IV)	Variable inheritance, with common vascular and bowel complications
Arthrochalasia and dermatosparaxis (former EDS VIIA, B, and C)	Autosomal dominant (arthrochalasia) and autosomal recessive (dermatosparaxis), with short stature and round faces
Kyphoscoliosis or Ocular-Scoliotic (former EDS VI)	Autosomal recessive with kyphoscoliosis, marfanoid habitus, and ocular involvement
Other	Poorly differentiated EDS types; includes X-linked EDS

CLINICAL FEATURES

Cutaneous

In EDS, there is an increase in skin laxity and fragility. Traction on the ears or elbows will reveal this increase in elasticity. The skin may have a velvety feel. Bruising and varicose veins are common.

Musculoskeletal

Joint hypermobility with characteristic hyperextensibilty of the joints is a classic finding in EDS. Joint dislocation and kyphoscoliosis are particularly prominent in some subtypes of EDS. Facial dysmorphism and variations in stature are common in certain subtypes of EDS.

Gastrointestinal

Breakdown of gums and loss of teeth is common in EDS. Hiatal hernias are a common structural defect. Complications of peptic ulcer disease are more frequent, including perforation and bleeding. Erosion of duodenal ulcers posteriorly into the aorta can be catastrophic. Perforation of the small and especially the large intestine are among the most commonly reported gastrointestinal complications of EDS, especially in patients with the vascular type of EDS, because type III collagen is the major supportive constituent of the bowel wall.

Copyright 2004, Elsevier (USA). All rights reserved.

DIAGNOSIS

The diagnosis of EDS is based on clinical history, family history, physical examination, and analysis of genetic defects (when available).

THERAPY AND PROGNOSIS

Therapy is largely supportive and involves joint protection, avoidance of trauma, and care in treating wounds. Pregnancy is contraindicated in the vascular type of EDS because of the risk of uterine rupture.

See Also the Following Articles

Duodenal Ulcer • Hernias • Hiatal Hernia

Further Reading

Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P., and Wenstrup, R. J. (1998). Ehlers–Danlos syndromes: Revised nosology, Villefranche, 1997. Ehlers–Danlos National Foundation (USA) and Ehlers–Danlos Support Group (UK). *Am. J. Med. Genet.* **77**(1), 31–37.

Mao, J. R., and Bristow, J. (2001). The Ehlers–Danlos syndrome: On beyond collagens. *J. Clin. Invest.* **107**(9), 1063–1069.

Pepin, M., Schwarze, U., Superti-Furga, A., and Byers, P. H. (2000). Clinical and genetic features of Ehlers–Danlos syndrome type IV, the vascular type. *N. Engl. J. Med.* **342**(10), 673–680.

Solomon, J. A., Abrams, L., and Lichtenstein, G. R. (1996). GI manifestations of Ehlers–Danlos syndrome. *Am. J. Gastroenterol.* **91**(11), 2282–2288.

Electrogastrography

KENNETH L. KOCH

North Carolina Baptist Medical Center and Wake Forest University

bradygastria Abnormally slow gastric myoelectrical activity (1.0–2.5 cpm).

electrogastrogram The myoelectrical signal recorded with electrogastrography methods.

electrogastrography Methods for recording and analyzing gastric myoelectrical activity from electrodes positioned on the abdomen.

eugastria Normal gastric myoelectrical activity (2.5–3.7 cpm).

gastric dysrhythmia Abnormal gastric myoelectrical rhythm (e.g., tachygastria).

gastroparesis Delayed emptying of a test meal from the stomach.

mixed dysrhythmias (nonspecific) Gastric dysrhythmias characterized by intermittent bradygastria and tachygastria.

running spectral analysis Computerized analysis of the frequencies in the electrogastrogram signal over a period of time using Fourier transform.

tachygastria Abnormally rapid gastric myoelectrical activity (3.7–10.0 cpm).

Electrogastrography is the method for recording and analyzing electrogastrograms. Electrogastrograms reflect gastric myoelectrical activity, the electrical and muscular activities of the stomach. The stomach is a complex neuromuscular organ that receives, mixes, and empties ingested foodstuffs. The neuromuscular work of the stomach is coordinated by gastric pacesetter potentials, the pacemaker activity of the stomach that controls the frequency and propagation of gastric peristaltic waves. Normal electrogastrogram patterns have been described in response to provocative tests. Abnormal gastric electrical activity (gastric dysrhythmia) is recorded using electrogastrography methods. Gastric dysrhythmias (tachygastrias, bradygastrias, and mixed dysrhythmias) are associated with upper gastrointestinal symptoms such as nausea. Clinical and research uses for recording gastric myoelectrical activity with electrogastrography are described.

INTRODUCTION

Electrogastrography refers to the methods for recording and analyzing gastric myoelectrical activity from electrodes placed on the abdominal surface. The

Copyright 2004, Elsevier (USA). All rights reserved.

DIAGNOSIS

The diagnosis of EDS is based on clinical history, family history, physical examination, and analysis of genetic defects (when available).

THERAPY AND PROGNOSIS

Therapy is largely supportive and involves joint protection, avoidance of trauma, and care in treating wounds. Pregnancy is contraindicated in the vascular type of EDS because of the risk of uterine rupture.

See Also the Following Articles

Duodenal Ulcer • Hernias • Hiatal Hernia

Further Reading

Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P., and Wenstrup, R. J. (1998). Ehlers–Danlos syndromes: Revised nosology, Villefranche, 1997. Ehlers–Danlos National Foundation (USA) and Ehlers–Danlos Support Group (UK). *Am. J. Med. Genet.* **77**(1), 31–37.

Mao, J. R., and Bristow, J. (2001). The Ehlers–Danlos syndrome: On beyond collagens. *J. Clin. Invest.* **107**(9), 1063–1069.

Pepin, M., Schwarze, U., Superti-Furga, A., and Byers, P. H. (2000). Clinical and genetic features of Ehlers–Danlos syndrome type IV, the vascular type. *N. Engl. J. Med.* **342**(10), 673–680.

Solomon, J. A., Abrams, L., and Lichtenstein, G. R. (1996). GI manifestations of Ehlers–Danlos syndrome. *Am. J. Gastroenterol.* **91**(11), 2282–2288.

Electrogastrography

KENNETH L. KOCH

North Carolina Baptist Medical Center and Wake Forest University

bradygastria Abnormally slow gastric myoelectrical activity (1.0–2.5 cpm).

electrogastrogram The myoelectrical signal recorded with electrogastrography methods.

electrogastrography Methods for recording and analyzing gastric myoelectrical activity from electrodes positioned on the abdomen.

eugastria Normal gastric myoelectrical activity (2.5–3.7 cpm).

gastric dysrhythmia Abnormal gastric myoelectrical rhythm (e.g., tachygastria).

gastroparesis Delayed emptying of a test meal from the stomach.

mixed dysrhythmias (nonspecific) Gastric dysrhythmias characterized by intermittent bradygastria and tachygastria.

running spectral analysis Computerized analysis of the frequencies in the electrogastrogram signal over a period of time using Fourier transform.

tachygastria Abnormally rapid gastric myoelectrical activity (3.7–10.0 cpm).

Electrogastrography is the method for recording and analyzing electrogastrograms. Electrogastrograms reflect gastric myoelectrical activity, the electrical and muscular activities of the stomach. The stomach is a complex neuromuscular organ that receives, mixes, and empties ingested foodstuffs. The neuromuscular work of the stomach is coordinated by gastric pacesetter potentials, the pacemaker activity of the stomach that controls the frequency and propagation of gastric peristaltic waves. Normal electrogastrogram patterns have been described in response to provocative tests. Abnormal gastric electrical activity (gastric dysrhythmia) is recorded using electrogastrography methods. Gastric dysrhythmias (tachygastrias, bradygastrias, and mixed dysrhythmias) are associated with upper gastrointestinal symptoms such as nausea. Clinical and research uses for recording gastric myoelectrical activity with electrogastrography are described.

INTRODUCTION

Electrogastrography refers to the methods for recording and analyzing gastric myoelectrical activity from electrodes placed on the abdominal surface. The

Copyright 2004, Elsevier (USA). All rights reserved.

noninvasive techniques of electrogastrography are used to record electrogastrograms (EGGs). The frequencies in the EGG signal reflect the pacesetter potential activity of the stomach, the electrical events that control the frequency and propagation of gastric peristaltic contractions. The normal human gastric pacesetter potential activity and the normal EGG signal are approximately 3 cycles per minute (cpm). Gastric dysrhythmias are abnormal gastric myoelectrical activities termed tachygastria, bradygastria, and mixed dysrhythmias. Gastric dysrhythmias are found in a variety of conditions in which nausea and upper gastrointestinal symptoms are prominent. Gastric dysrhythmias are commonly present in patients with gastroparesis, a severe neuromuscular abnormality of the stomach. In patients with unexplained nausea and dyspepsia symptoms, EGG tests diagnose eugastria or gastric dysrhythmias and appropriate treatment can be planned. Electrogastrographic techniques are also used in clinical research studies where the effects of various therapies on symptoms and gastric electrical rhythms are assessed.

MYOELECTRICAL ACTIVITY OF THE STOMACH

Pacesetter Potentials

The stomach is a complex neuromuscular organ (Fig. 1). Contractions and relaxations of the gastric fundus, corpus, and antrum are neuromuscular events under myoelectrical, neurological, and hormonal control. Gastric peristaltic contractions are controlled by the pacesetter potential activity of the stomach. Gastric pacesetter potentials are also termed gastric slow waves. Gastric pacesetter potentials originate in a pacemaker region on the greater curvature of the stomach, between the fundus and the corpus (Fig. 1). These electrical depolarization and repolarization waves travel circumferentially and migrate distally at a rate of approximately 3 cpm in humans (Fig. 2). The gastric pacesetter potentials control the frequency and propagation of velocity of gastric peristaltic contractions. The fundus of the stomach does not have intrinsic pacesetter potential activity.

Rhythmicity of the gastric pacesetter potential activity originates in the interstitial cells of Cajal. These cells form networks within and between the muscular layers of the stomach wall. The interstitial cells of Cajal that lie near the circular muscle layer of the stomach wall control the frequency of depolarization of the circular muscle (Fig. 1). The interstitial cells of Cajal are also in close proximity to the enteric neurons of the intrinsic nervous

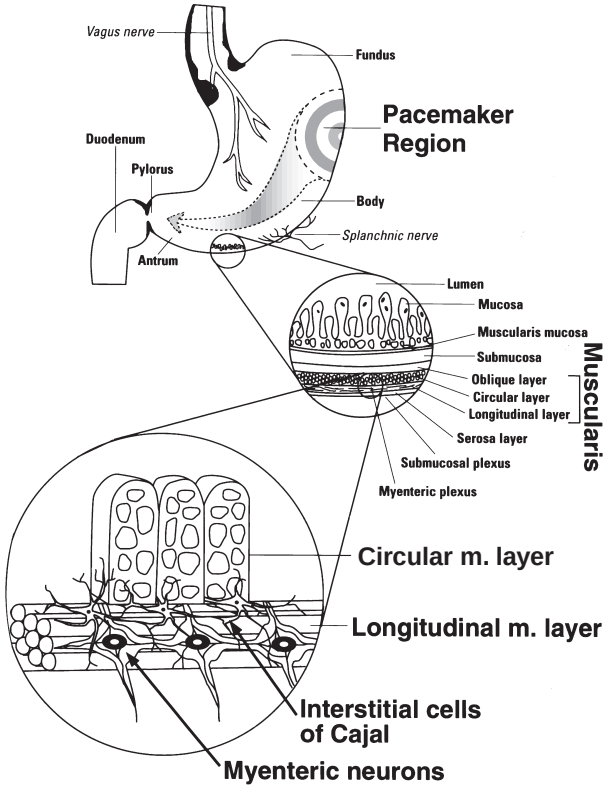


FIGURE 1 Diagram of major anatomical regions of the stomach (top), muscle layers of the gastric wall (middle circle), and detail of the relationships among longitudinal muscle, circular muscle, interstitial cells of Cajal, and myenteric neurons (bottom circle). See text for details.

system of the gut. Enteric nervous system and parasympathetic nervous system inputs to the stomach are transmitted to the pacemaker cells and thus to the circular muscle cells through these relationships.

Action Potentials and Plateau Potentials

Circular muscle contractions of the stomach wall occur during action potential or plateau potential activity. Action potential activity and plateau potential activities are linked to the pacesetter potentials that bring the circular smooth muscle membrane to the threshold of depolarization and contraction. Thus, a normal gastric peristaltic wave is composed of electrical components—the pacesetter potential linked with the action potential or plateau potential activity (Fig. 3). Consequently, increased electrical activity occurs at the pacesetter potential frequency of 3 cpm during recurrent peristaltic contractions.

These gastric neuromuscular events generate changes in electrical activity that are recorded from

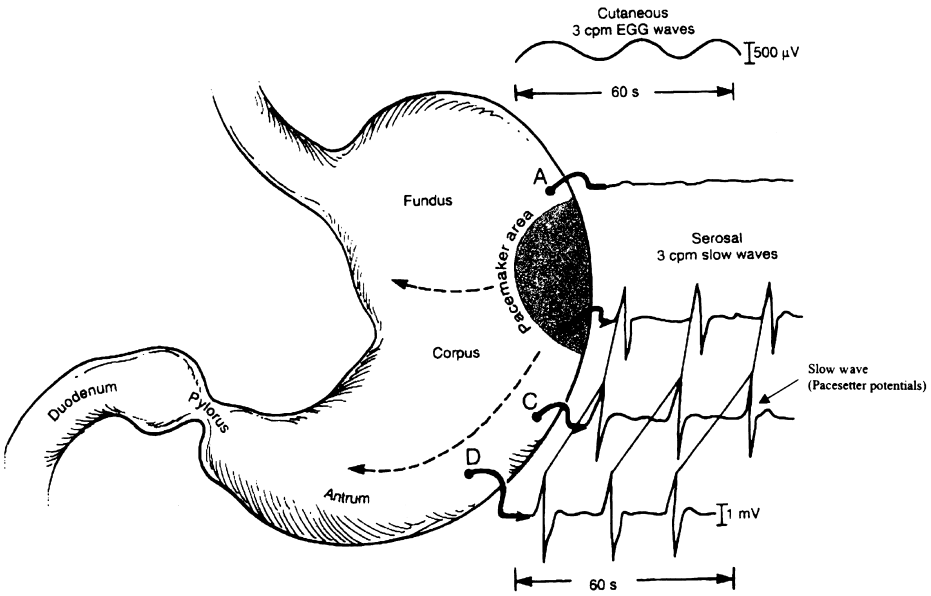


FIGURE 2 Gastric slow waves (pacesetter potentials) migrate circumferentially around the stomach and distally toward the pylorus at a rate of 3 cpm as recorded by the serosal electrodes (A–D). The fundus is electrically silent. The 3 cpm EGG waves recorded from electrodes placed on the surface of the epigastrium summate electrical events and yield a 3 cpm wave pattern.

electrocardiogram (EKG)-type electrodes placed on the epigastrium. These signals are termed electrogastrograms. During the fasted state, the electrical activity from the stomach reflects predominantly the gastric pacesetter potential activity. The EGG activity during fasting may be relatively unstable. In contrast, after a homogenous meal, such as water, barium, or yogurt, rhythmic gastric peristaltic contractions are evoked and the accompanying gastric electrical and EGG activity has greater amplitude and regularity than in the fasting condition.

RECORDING AND ANALYSIS OF ELECTROGASTROGRAMS

Recording EGGs

Electrogastrograms are recorded by placing EKG-type electrodes on the surface of the epigastrium as shown in Fig. 4. The first EGG was recorded in 1922 by Alvarez, who reported rhythmic 3 cpm electrical activity in a thin woman. The amplitude of the EGG signal ranges from 100 to 500 μV and the signal must be properly amplified and filtered. To reduce baseline drift and to filter out cardiac and respiratory signals, a 0.016 Hz high-pass filter and a 0.25 Hz low-pass filter are used. These filters create a window from approximately 1 to 15 cpm through which myoelectrical signals pass during the EGG recording. The normal EGG rhythm is approximately 3 cpm (2.5–3.7 cpm), whereas bradygastrias are signals from approximately 1.0 to 2.5 cpm and tachygastrias are from 3.7 to 10.0 cpm. The duodenal pacesetter potential frequency is approximately 12–14 cpm. Occasionally, the respiration rate is less than 15 breaths per minute and this respiratory frequency may be recorded in the EGG signal.

The EGG signal is digitized for computer analysis by using analog to digital technology. Digitization rates range from 1 to 4.267 Hz. The digital files are subjected to fast Fourier transform to extract the frequency information present in the EGG signal. Displayed over time the frequencies are often presented as running spectral analysis. Computer analysis allows for quantitative expression of the EGG signal as described below.

To avoid movement artifact, EGGs must be recorded in quiet rooms with the subject seated in a comfortable chair. A baseline EGG is recorded and is followed by a provocative test to assess the gastric myoelectrical activity responses and sensations or symptoms. Provocative tests using water loading stimulate the relaxation capacity (stretch) of the stomach, as well as the gastric myoelectrical response, without the confounding effects of a caloric meal. Caloric meals are also used to assess the effect of specific foods on EGG activity. Respiration rate must be recorded on a separate channel.

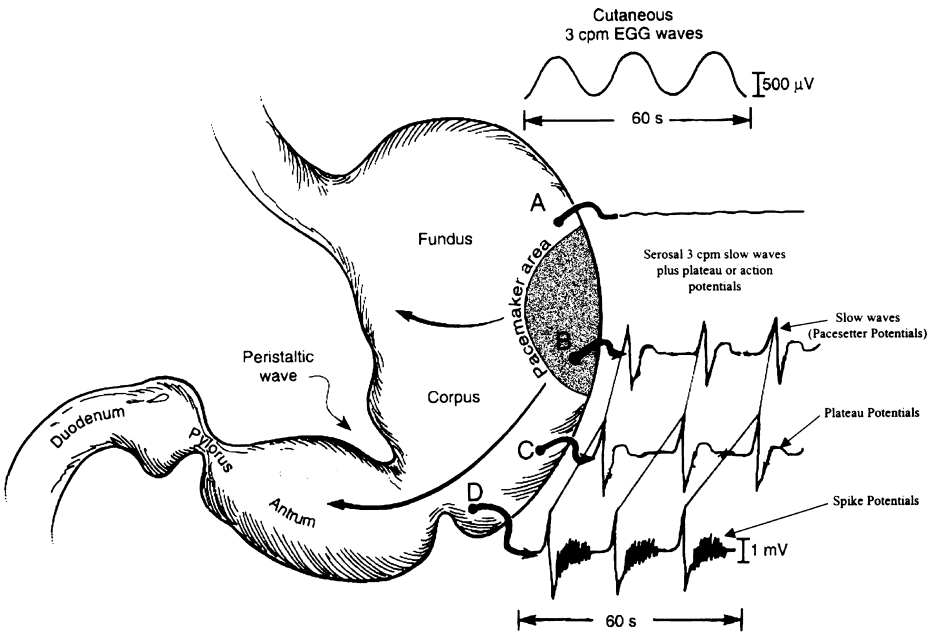


FIGURE 3 Gastric peristaltic waves during circular muscle contraction occur when plateau potentials or spike potentials occur and are linked with the propagating slow waves. Generally, there is an increased amplitude of the EGG wave during these electrical events associated with gastric peristaltic contractions.

Analysis of the EGG Signal

Visual Analysis

Proper analysis of the EGG requires visual inspection of the EGG signal. The EGG signal is reviewed for the presence of normal 3 cpm activity, tachygastrias, or bradygastrias (Fig. 5). It is extremely important to identify artifacts in the EGG signal, which may be created by movement of the limbs, deep breathing, coughing, or talking. An overall EGG diagnosis of normal versus abnormal EGG is made on the basis of the EGG rhythm strip.

Computer Analysis

Frequencies in EGG signal Computer analysis of the EGG signal emphasizes the frequency components. The frequency component is the most reliable and stable of the EGG parameters that can be quantified. The human gastric pacesetter potential activity is approximately 3 cpm and 2 standard deviations around the mean is the normal EGG frequency range of 2.5–3.7 cpm. Normative ranges for EGG activity have been defined in response to the water load test or test meals. The EGG response is reproducible in response to the water load test.

Percentage distribution of EGG power A certain percentage of the overall EGG activity occurs in four relevant frequency ranges: the normal range and the bradygastria, tachygastria, and duodenal frequency ranges. Thus, a useful method of expressing the EGG activity is to determine the percentage distribution of the total EGG power in these ranges. Power is the log of the microvolts squared (log μV^2) of the frequencies present in the EGG signal. Normal values in control subjects and patients with upper gastrointestinal symptoms have been reported.

Power ratio The postprandial to preprandial ratio of 3 cpm EGG power is generally >1 and indicates a normal increase in the 3 cpm activity after ingestion of a test meal. This parameter is also used to assess normal and abnormal neuromuscular activity of the stomach in response to test meals.

NORMAL ELECTROGASTROGRAM PATTERNS

Fasting

The neuromuscular activity of the stomach during the fasting state is unstable compared with that in the postprandial state. Contractile patterns in the stomach during fasting range from quiescence (phase I) to intermittent contractions (phase II) to sustained regular antral contractions (phase III). Thus, the electrical

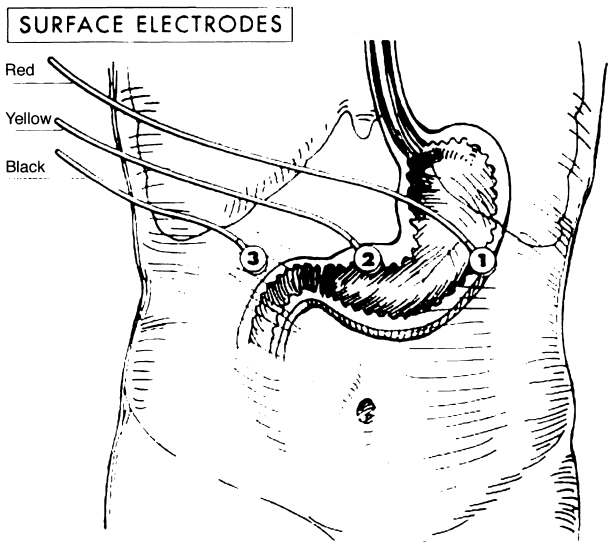


FIGURE 4 Position of electrodes on the surface of the epigastrium for recording electrogastrograms.

activity during fasting is also unstable in that the 3 cpm electrical activity is inconsistent. Phase III contractile activity also is reflected in increased 3 cpm EGG activity 50% of the time.

Postprandial EGG Patterns

Barium Meals and Water Load Test

The fasting EGG pattern, which may be unstable, is normally converted to regular 3 cpm activity when the stomach is loaded with a variety of nutrient or nonnutrient test meals. Figure 6 shows the effect of a barium meal on EGG activity. The barium meal stimulates regular 3 cpm EGG waves that correspond to the peristaltic gastric waves. Loading the stomach with water also stimulates regular 3 cpm waves after an initial period of suppression of 3 cpm waves (Fig. 7). This example also shows the EGG frequency response as a running spectral plot. (See legend to Fig. 7 for details.)

Yogurt and Caloric Meals

A variety of test meals ranging from yogurt to soup to milk to scrambled eggs also stimulate increased 3 cpm EGG activity and an increased power ratio in the normal frequency range. The exact pattern of onset of 3 cpm activity and other frequencies depends upon the caloric content of the meals and the proportions of fat and carbohydrate.

ABNORMAL ELECTROGASTROGRAM PATTERNS

Gastric Dysrhythmias

Tachygastrias

Tachygastrias are abnormally rapid gastric electrical events from 3.7 to 10.0 cpm. High-amplitude tachygastrias occur during the acute onset of motion sickness. Tachygastrias are frequently low-amplitude EGG signals recorded from patients with chronic nausea conditions such as morning sickness, functional dyspepsia, unexplained nausea, or idiopathic gastroparesis. The tachygastria frequency is often very unstable and may range from 4 to 7 to 9 cpm within a 30 min recording period. Other tachygastria patterns are fixed at a single frequency (e.g., 6 cpm) (Fig. 5). Tachygastrias have been recorded from the antrum with serosal electrodes. The presence of tachygastrias is associated with decreased rates of gastric emptying. The cause of tachygastrias is unknown but they have been induced by vagotomy, epinephrine, glucagon, and hyperglycemia.

Bradygastrias

Bradygastrias are abnormally slow EGG frequencies from approximately 1 to 2.5 cpm. Bradygastrias

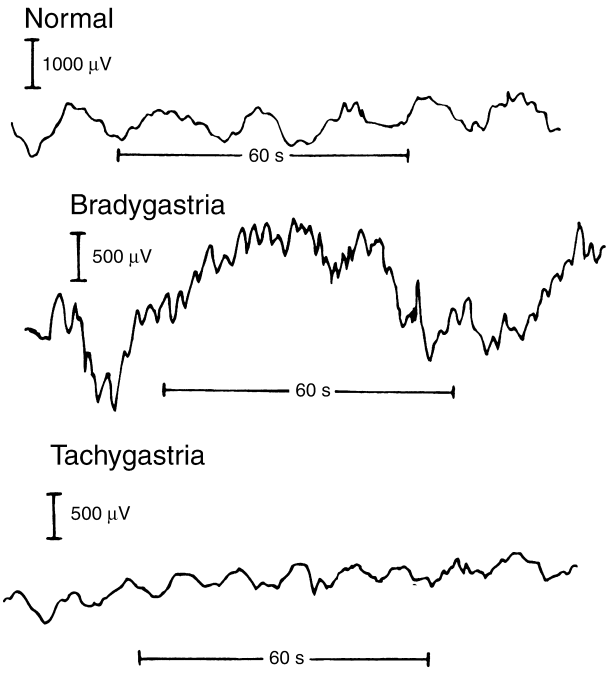


FIGURE 5 Normal 3 cpm EGG recording showing 3 cpm waves; bradygastria example shown as a 1 cpm EGG wave with superimposed respiratory rhythm and tachygastria example shown as 6 cpm regular, low-amplitude waves.

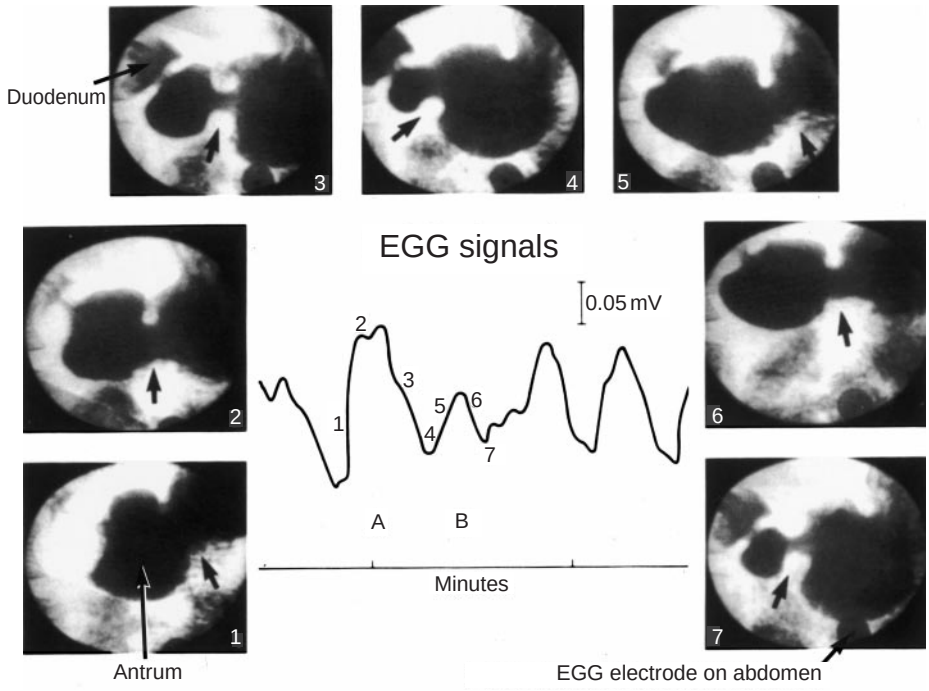


FIGURE 6 EGG waves (A and B) recorded during fluoroscopy of the barium-filled stomach. EGG wave A is numbered 1, 2, 3, and 4 to illustrate the relationship between this EGG wave and the gastric peristaltic wave shown in radiographic panels 1, 2, 3, and 4. In the radiographic panels, the arrow indicates a circular contraction that is migrating from the corpus (1) to the mid-antrum (2, 3) and finally to the distal antrum (4). EGG wave B is numbered 5, 6, and 7 and is similarly associated with the next gastric peristaltic wave shown in radiographic panels 5, 6, and 7. Reprinted from Koch *et al.* (1987), Effect of barium meals on gastric electromechanical activity in man. *Digest. Dis. Sci.* **32**, 1217–1222, with permission from Kluwer Academic.

are high-amplitude 1 to 2 cpm EGG waves or low-amplitude waves (Fig. 5). The bradygastrias have been recorded from the gastric corpus with several electrodes. Bradygastrias are present in a variety of patients with unexplained nausea, dyspepsia symptoms, and gastroparesis.

Mixed Dysrhythmia

EGG recordings that do not clearly fit into a tachygastria or bradygastria pattern are termed mixed (nonspecific) dysrhythmias. These patients have combinations of tachygastria and bradygastria, abnormal patterns that are brought out by provocative testing with the water load test or test meals.

Decreased Power Ratio

The power ratio is a calculation of the EGG power in the postprandial period compared with the preprandial EGG power. An abnormal power ratio is 1 or less and indicates failure of the expected increase in 3 cpm electrical activity after the provocative test. An abnormal power ratio of the signal at 3 cpm is associated with delayed gastric emptying.

GASTRIC DYSRHYTHMIAS AND CLINICAL DISORDERS

The measurement of EGG patterns in response to provocative testing is a clinical test to evaluate gastric myoelectrical activity in patients with suspected gastric motility disorders. The EGG pattern complements the solid-phase gastric emptying test, which tests the efficiency of the global muscular work of the stomach. The EGG reveals the mechanisms that underlie gastroparesis. Thus, the combination of EGG and gastric emptying studies reveals pathophysiological conditions of the stomach that may cause the patient's symptoms. In the sections below, the EGG patterns found in a variety of clinical conditions are reviewed.

Nausea of Pregnancy

Gastric dysrhythmias ranging from tachygastria to bradygastrias have been recorded in women with nausea

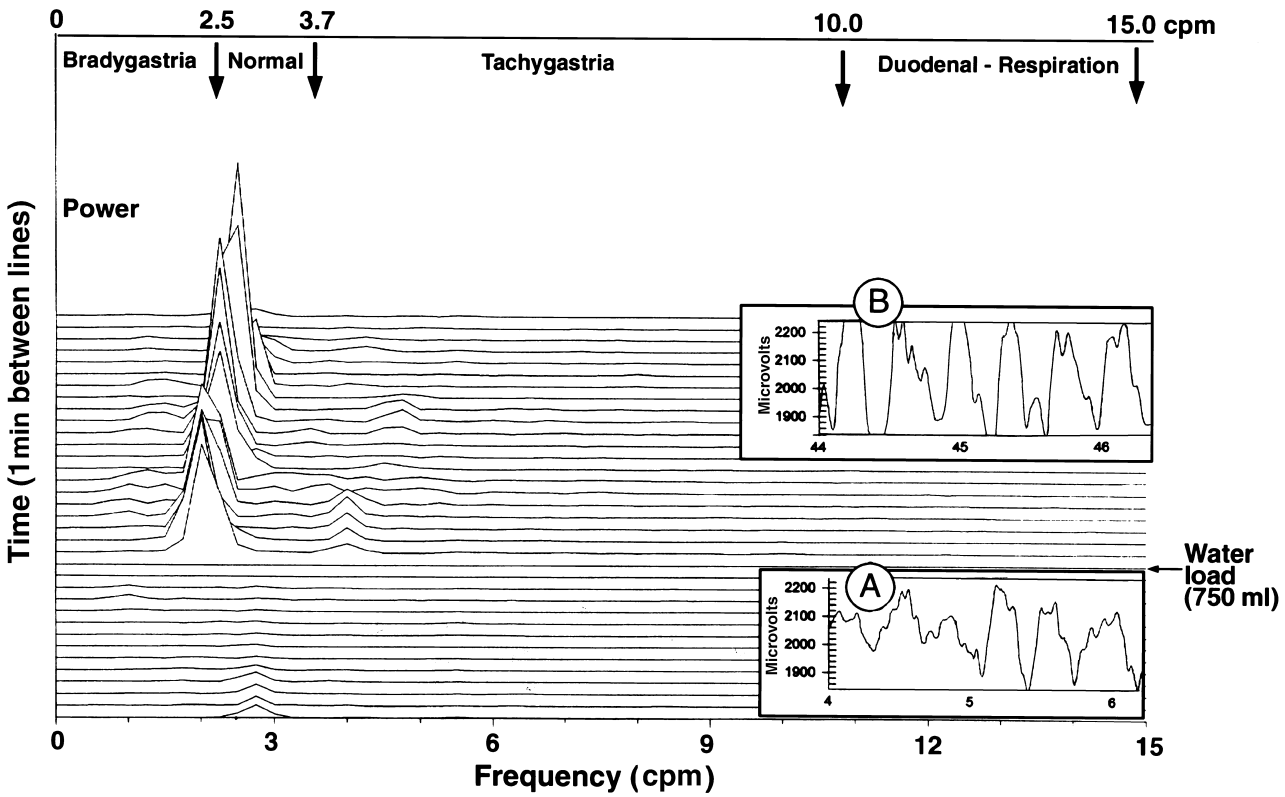


FIGURE 7 Running spectral analysis of the frequencies in the EGG signal recorded before and after a water load test in a healthy subject. The insets show the EGG signal. At baseline (A), some 3 cpm rhythm is seen in the EGG. After the water load was ingested, there is higher amplitude and more regular 3 cpm activity (B). The running spectral analysis is the frequency analysis of the EGG. The *X* axis displays the frequencies in the EGG in cycles per minute. The *Y* axis indicates time, with each line representing 4 min of EGG signal with 75% overlap. The *Z* axis represents the power (log of the μV^2) of any frequencies contained in the EGG signal from 0 to 15 cpm. The four relevant frequencies are indicated at the top of the graph. In this healthy subject, there are several low-power peaks near 3 cpm at baseline. After ingestion of water, there is a frequency shift to approximately 2 cpm after ingestion of 750 cc of water. Thereafter, there is a progressive increase in power in the peaks as they enter the normal range of 2.5 to 3.7 cpm.

of pregnancy. Pregnant women with no nausea had normal 3 cpm EGG patterns. Thus, the presence or absence of gastric dysrhythmias is associated with the presence or absence of nausea. Administration of estrogen and progesterone also produces gastric dysrhythmias.

Dysmotility-like Functional Dyspepsia

Gastric dysrhythmias have been recorded in approximately 60% of patients with dysmotility-like functional dyspepsia (unexplained nausea and vomiting, upper abdominal discomfort, early satiety, bloating, and fullness). Endoscopic examinations in these patients reveals no mucosal abnormalities and ultrasound or computerized tomography scans of the abdomen reveal no structural abnormalities to account for symptoms. Thus, gastric dysrhythmias are a pathophysiological mechanism for these symptoms. Correction of gastric dysrhythmias (which range from bradygastria to tachygastria) is associated with improved upper gastrointestinal (GI) symptoms. A minority of patients with gastric dysrhythmias have frank gastroparesis. These patients have a severe electrical and contractile abnormality of the stomach.

Gastroparesis

Gastroparesis indicates severely impaired emptying of meals from the stomach. The gastric emptying test does not reveal the cause of the gastroparesis. The EGG test diagnoses the presence of gastric dysrhythmias or normal myoelectrical signals in these patients.

Diabetic Gastroparesis

Patients with diabetic gastroparesis have a high incidence of gastric dysrhythmias. Patients with

predominantly bloating and distension symptoms in the upper abdomen after eating may also have delayed emptying, but normal 3 cpm electrical activity. Hyperglycemia also induces tachygastrias and bradygastrias.

Ischemic Gastroparesis

Patients with gastroparesis and those with histories of peripheral vascular disease, smoking, and weight loss may have ischemic gastroparesis. Abdominal bruits are present in approximately 50% of these patients. The median accurate ligament may obstruct the celiac artery and produce ischemic gastroparesis. Bypass of the obstructions in the mesenteric arteries results in resolution of the gastric dysrhythmias, gastroparesis, and other symptoms.

Obstructive Gastroparesis

Patients with gastric outlet obstruction due to pyloric stenosis or postbulbar obstructions have high-amplitude, consistently normal 3 cpm EGG activity despite the presence of gastroparesis. This discordance between normal EGG rhythm and delayed emptying suggests that mechanical or functional obstruction may account for the gastroparesis.

Idiopathic Gastroparesis

In patients with idiopathic gastroparesis, gastric dysrhythmias are commonly present. The dysrhythmias range from bradygastria to tachygastria. The origins of idiopathic gastroparesis are unknown but range from degenerative disorders of the enteric nervous system, damage to the interstitial cells of Cajal, and myopathies of various severity. Damage to the neuromuscular apparatus of the stomach may be a sequelae to a previous viral infection.

Drug-Induced Gastric Dysrhythmias

Gastric dysrhythmias and nausea are induced by hormones such as glucagon, epinephrine, estrogen, and progesterone. The narcotic drug morphine induces a variety of gastric dysrhythmias ranging from bradygastrias to tachygastrias. Nicotine induces gastric dysrhythmias and nausea.

Motion Sickness

Nausea is one of the most noxious symptoms of motion sickness. Nausea reported during vection-induced motion sickness is accompanied by tachygastrias and increased plasma vasopressin and epinephrine levels. Gastric dysrhythmias and motion sickness symptoms are reduced by carbohydrate meals, acustimulation, ginger, and a variety of other treatments.

CLINICAL APPLICATIONS OF ELECTROGASTROGRAPHY

EGG tests are obtained in patients who have persistent upper GI symptoms despite normal upper GI barium studies, endoscopy, and ultrasound examinations. In these patients, the presence of a gastric dysrhythmia provides a diagnosis of gastric neuromuscular dysfunction, whereas a normal EGG test is associated with normal gastric emptying and suggests the possibility of other pathophysiological mechanisms causing the symptoms.

The noninvasive diagnosis of gastric dysrhythmias with the EGG test has analogies to the noninvasive diagnosis of cardiac dysrhythmias with the EKG test. The EGG test provides a diagnosis of gastric dysrhythmia or eugastria in response to a provocative test and is an objective finding of neuromuscular abnormality of the stomach for patients with unexplained nausea and functional dyspepsia symptoms. The gastric dysrhythmias may reflect relatively mild gastric neuromuscular disorders that underlie symptoms. On the other hand, the gastric dysrhythmia may be associated with significant contractile abnormality of the stomach and frank gastroparesis. A normal or eugastric EGG pattern suggests that other pathophysiological mechanisms may account for the upper GI symptoms (e.g., gallbladder disease). Gastric visceral hypersensitivity, gastric electro-contractile abnormalities, or obstruction must also be considered.

RESEARCH APPLICATIONS OF ELECTROGASTROGRAPHY

Electrogastrography methods are used to identify patient populations with upper GI symptoms, such as nausea and gastric dysrhythmias (versus symptoms and normal EGGs). A variety of drugs, such as domperidone, metoclopramide, and cisapride, eradicate gastric dysrhythmias and the eradication of the dysrhythmia is associated with improvement in upper GI symptoms. Drug or nondrug therapies with potential anti-arrhythmic or prokinetic effects may be studied at baseline and with follow-up noninvasive EGG recordings. EGG testing may be performed multiple times without discomfort or adverse effects for the subject.

Acknowledgment

The author acknowledges the excellent secretarial assistance of Ms. Pamela Petito.

See Also the Following Articles

Basic Electrical Rhythm • Diabetic Gastroparesis • Gastric Motility

Further Reading

Brzana, R. J., Bingaman, S., and Koch, K. L. (1998). Gastric myoelectrical activity in patients with gastric outlet obstruction and idiopathic gastroparesis. *Am. J. Gastroenterol.* **93**, 1083–1089.

Chen, J., Schirmer, B. D., and McCallum, R. W. (1994). Serosal and cutaneous recordings of gastric myoelectrical activity in patients with gastroparesis. *Am. J. Physiol.* **266**, G90–G98.

Hasler, W. L., Souda, H. C., Dulai, G., and Owyang, C. (1995). Mediation of hyperglycemia-evoked gastric slow wave dysrhythmias by endogenous prostaglandins. *Gastroenterology* **108**, 727–736.

Koch, K. L. (1998). Review article: Clinical approaches to unexplained nausea and vomiting. *Adv. Gastroenterol. Hepatol. Clin. Nutr.* **3**, 163–168.

Koch, K. L. (2001). Electrogastrography: Physiological basis and clinical application in diabetic gastropathy. *Diab. Technol. Therapeut.* **3(1)**, 51–62.

Koch, K. L., Hong, S.-P., and Xu, L. (2000). Reproducibility of gastric myoelectrical activity and the water load test in patients with dysmotility-like dyspepsia symptoms and in control subjects. *J. Clin. Gastroenterol.* **31(2)**, 125–129.

Koch, K. L., and Stern, R. M. (1993). Electrogastrography. *In* "Illustrated Guide to Gastrointestinal Motility" (D. Kumar and D. Wingate, eds.), pp. 290–307. Churchill Livingstone, London.

Koch, K. L., Stern, R. M., Vasey, M., Botti, J. J., Creasy, G. W., and Dwyer, A. (1990). Gastric dysrhythmias and nausea of pregnancy. *Digest. Dis. Sci.* **35**, 961–968.

Koch, K. L., Stewart, W. R., and Stern, R. M. (1987). Effect of barium meals on gastric electrical mechanical activity in man: A fluoroscopic-electrogastrographic study. *Digest. Dis. Sci.* **32**, 1217–1222.

Lin, Z., Eaker, E. Y., Sarosiek, I., and McCallum, R. W. (1999). Gastric myoelectrical activity and gastric emptying in patients with functional dyspepsia. *Am. J. Gastroenterol.* **94**, 2384–2389.

Verhagen, M. A. M. T., Van Schelven, C. J., Samson, M., and Smout, A. J. P. M. (1999). Pitfalls in the analysis of electrogastrographic recordings. *Gastroenterology* **117**, 453–460.

Emesis

IVAN M. LANG, REZA SHAKER

Medical College of Wisconsin, Milwaukee

blood–brain barrier Physiological and anatomical barriers to the free diffusion of larger molecules from the vascular system to the neural circuits of the brain.

circumventricular organs Group of brain areas lining the cerebral ventricles; have no blood–brain barrier and may have chemosensory or secretory functions.

electrical control activity Ongoing spontaneous changes in membrane potential of the gastrointestinal smooth muscle that controls the timing of contractions.

electrical response activity Electrical activity of gastrointestinal smooth muscle associated with contractions.

enteric nervous system Set of integrative neural circuitry of the digestive tract that includes the myenteric and submucosal plexuses.

enterochromaffin cells Subset of hormone- or paracrine-secreting cells of the mucosa of the digestive tract that contain serotonin.

prodromata Set of autonomic responses that accompany and often precede emesis or nausea.

Emesis refers to the specific set of physiological responses associated with retching and vomiting that culminate in the expulsion of gastric contents. Emesis does not accurately refer to the act of expulsion of substances from the stomach, because this can occur during belching, regurgitation, rumination, coughing, and perhaps other situations. Nausea is a sensation that accompanies emesis or is activated by emetic stimuli.

INTRODUCTION

Emesis acts to protect an organism from ingested noxious substances. The protective function of emesis is signaled by two sets of receptors located at different levels of the absorptive pathway. A preabsorptive set of receptors is located in the digestive tract and a postabsorptive set is located within the vascular system.

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Basic Electrical Rhythm • Diabetic Gastroparesis • Gastric Motility

Further Reading

Brzana, R. J., Bingaman, S., and Koch, K. L. (1998). Gastric myoelectrical activity in patients with gastric outlet obstruction and idiopathic gastroparesis. *Am. J. Gastroenterol.* **93**, 1083–1089.

Chen, J., Schirmer, B. D., and McCallum, R. W. (1994). Serosal and cutaneous recordings of gastric myoelectrical activity in patients with gastroparesis. *Am. J. Physiol.* **266**, G90–G98.

Hasler, W. L., Souda, H. C., Dulai, G., and Owyang, C. (1995). Mediation of hyperglycemia-evoked gastric slow wave dysrhythmias by endogenous prostaglandins. *Gastroenterology* **108**, 727–736.

Koch, K. L. (1998). Review article: Clinical approaches to unexplained nausea and vomiting. *Adv. Gastroenterol. Hepatol. Clin. Nutr.* **3**, 163–168.

Koch, K. L. (2001). Electrogastrography: Physiological basis and clinical application in diabetic gastropathy. *Diab. Technol. Therapeut.* **3(1)**, 51–62.

Koch, K. L., Hong, S.-P., and Xu, L. (2000). Reproducibility of gastric myoelectrical activity and the water load test in patients with dysmotility-like dyspepsia symptoms and in control subjects. *J. Clin. Gastroenterol.* **31(2)**, 125–129.

Koch, K. L., and Stern, R. M. (1993). Electrogastrography. *In* "Illustrated Guide to Gastrointestinal Motility" (D. Kumar and D. Wingate, eds.), pp. 290–307. Churchill Livingstone, London.

Koch, K. L., Stern, R. M., Vasey, M., Botti, J. J., Creasy, G. W., and Dwyer, A. (1990). Gastric dysrhythmias and nausea of pregnancy. *Digest. Dis. Sci.* **35**, 961–968.

Koch, K. L., Stewart, W. R., and Stern, R. M. (1987). Effect of barium meals on gastric electrical mechanical activity in man: A fluoroscopic-electrogastrographic study. *Digest. Dis. Sci.* **32**, 1217–1222.

Lin, Z., Eaker, E. Y., Sarosiek, I., and McCallum, R. W. (1999). Gastric myoelectrical activity and gastric emptying in patients with functional dyspepsia. *Am. J. Gastroenterol.* **94**, 2384–2389.

Verhagen, M. A. M. T., Van Schelven, C. J., Samson, M., and Smout, A. J. P. M. (1999). Pitfalls in the analysis of electrogastrographic recordings. *Gastroenterology* **117**, 453–460.

Emesis

IVAN M. LANG, REZA SHAKER

Medical College of Wisconsin, Milwaukee

blood–brain barrier Physiological and anatomical barriers to the free diffusion of larger molecules from the vascular system to the neural circuits of the brain.

circumventricular organs Group of brain areas lining the cerebral ventricles; have no blood–brain barrier and may have chemosensory or secretory functions.

electrical control activity Ongoing spontaneous changes in membrane potential of the gastrointestinal smooth muscle that controls the timing of contractions.

electrical response activity Electrical activity of gastrointestinal smooth muscle associated with contractions.

enteric nervous system Set of integrative neural circuitry of the digestive tract that includes the myenteric and submucosal plexuses.

enterochromaffin cells Subset of hormone- or paracrine-secreting cells of the mucosa of the digestive tract that contain serotonin.

prodromata Set of autonomic responses that accompany and often precede emesis or nausea.

Emesis refers to the specific set of physiological responses associated with retching and vomiting that culminate in the expulsion of gastric contents. Emesis does not accurately refer to the act of expulsion of substances from the stomach, because this can occur during belching, regurgitation, rumination, coughing, and perhaps other situations. Nausea is a sensation that accompanies emesis or is activated by emetic stimuli.

INTRODUCTION

Emesis acts to protect an organism from ingested noxious substances. The protective function of emesis is signaled by two sets of receptors located at different levels of the absorptive pathway. A preabsorptive set of receptors is located in the digestive tract and a postabsorptive set is located within the vascular system.

Copyright 2004, Elsevier (USA). All rights reserved.

A number of motor and secretory events of the respiratory and digestive tracts occur before, during, and after emesis. The brain contains the vascular chemoreceptive area as well as the pattern generators for the motor and secretory responses, and pharmacological agents may block emesis at the sensory or integrative areas.

SENSORY MECHANISMS

Digestive Tract

Mechanoreceptors

Mechanical stimulation of the digestive tract from the pharynx to the small intestine by stroking the mucosa, distension, compression, or obstruction can activate nausea and vomiting. Slowly or rapidly adapting mechanoreceptors located in the mucosa, muscularis, or serosa may be involved in mechanical stimulation-induced emesis. Serosal slowly adapting mechanoreceptors, in particular, may be responsible for obstruction- or peritonitis-induced emesis.

Chemoreceptors

Noxious stimulation of the digestive tract using a variety of chemical substances, including acidic or alkaline solutions, $CuSO_4$, hypertonic saline, and hydrogen peroxide, can activate nausea and vomiting by stimulation of mucosal chemoreceptors of the stomach or duodenum.

Nonphysiological Stimulation of Receptors

Two clinically important treatments for cancer, cytotoxins and X-radiation, activate emesis probably by damage to physiological receptors or their afferent pathways. These emetic stimuli may activate vagal afferents directly or indirectly through the release of serotonin from enterochromaffin (EC) cells.

Chemoreceptor Trigger Zone

The chemoreceptor trigger zone (CTZ) is located in the area postrema (AP), one of the circumventricular organs of the brain. The CTZ is similar to other vascular chemoreceptors except that it functions as a postabsorptive receptor for the protection of the organism from ingested noxious substances. The CTZ has no blood–brain barrier and responds to numerous chemicals and hormones, including dopaminergic, α2-adrenergic, and opioid agonists, and cardiac glycosides, cytotoxins, and $CuSO_4$.

MOTOR RESPONSES OF THE DIGESTIVE TRACT

The digestive tract from pharynx to rectum is involved in the responses associated with emesis. The increase in salivation and swallowing prior to emesis probably functions to assist in the buffering of gastric juice. Salivary secretions protect the esophagus against swallowing-associated gastric reflux and probably do the same for emesis-associated gastric reflux. The subsequent digestive tract responses are organized into two sets of responses that occur in a hierarchical fashion. The gastrointestinal responses occur first and have a lower initiation threshold compared to the pharyngoesophageal responses, such that the gastrointestinal responses may occur without the pharyngoesophageal responses.

Gastrointestinal Responses

Motor Responses

The first gastrointestinal responses are relaxation of the gastric fundus and lower esophageal sphincter (LES), and these responses last the duration of the vomiting episode (Fig. 1A). A retrograde giant contraction (RGC) of the gastrointestinal tract (Fig. 1B) then empties the contents of the upper half of the small intestine into the stomach, thereby removing the offending noxious substance from the absorptive areas of the digestive tract. This RGC, in its progression through the duodenum, also empties the Brunner's glands and deposits its secretions into the stomach. The Brunner's gland secretions have significant buffering capacity, therefore, like salivary secretions, these Brunner's gland secretions may assist in protecting the esophagus during vomiting. Simultaneous with the RGC is the activation of a series of contractions of the lower half of the small intestine (Fig. 1B) that propagate distally. These contractions are of similar amplitude and duration as the usual spontaneous contractions, but they propagate distally a longer extent. These contractions act to milk the contents of the lower part of the small intestine into the colon. Defecation often follows emesis, therefore the contents of the lower half of the small intestine are eliminated from the organism just as vomitus expulsion eliminates noxious substances from the stomach and small intestine.

Myoelctric Responses

The electrical control activity (ECA), also known as electrical slow waves, of the entire gastrointestinal tract is altered during this phase of emesis (Fig. 2B). The ECA of the stomach becomes highly disorganized or highly regular at elevated (tachygastria) or reduced (bradygastria) frequencies before or during the

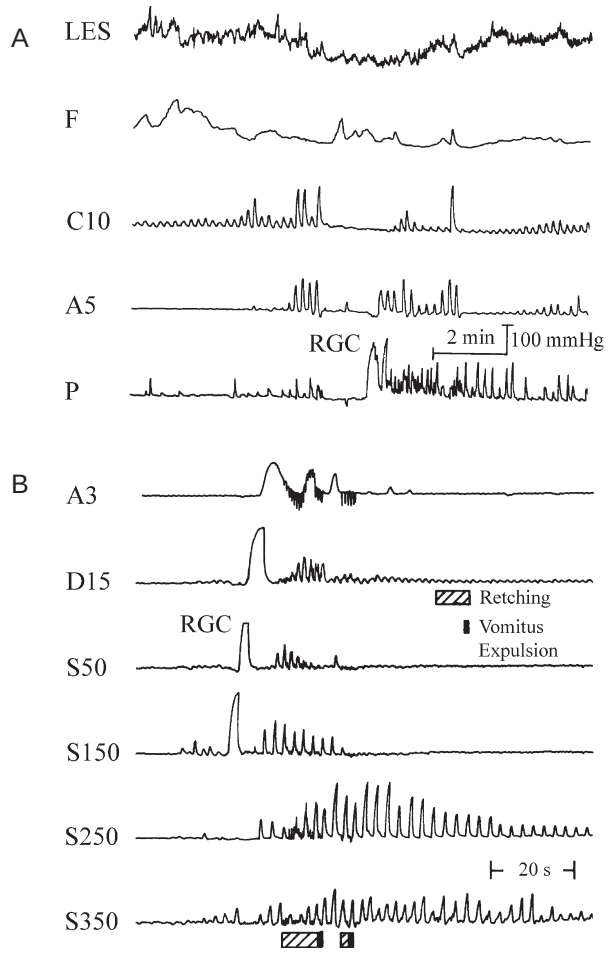


FIGURE 1 Gastrointestinal motor responses associated with emesis. (A) Gastric responses; (B) intestinal responses. LES, Lower esophageal sphincter; F, fundus; C, corpus; A, antrum; P, pylorus; D, duodenum; S, small intestine; RGC, retrograde giant contraction. The numbers next to abbreviated terms indicate the distance in centimeters of recording site from P. Note that LES and F relax before the occurrence of the RGC (A), the RGC occurs in upper small intestine only (B), and after the RGC a series of phasic contractions occur, especially in the distal small intestine (B).

vomiting episode. Contractions are not associated with tachygastria or bradygastria, and these electrical changes may be associated with active inhibition of the muscle. The RGC is associated with the highly irregular pattern of electrical activity, and this change may represent the electrical response activity (ERA) of the stomach. The ECA of the upper small intestine ceases (for up to 30 s) before and during the propagation of the RGC. This cessation of ECA is necessary for propagation of the RGC, because the ECA controls the normal orthograde propagation of contractions and this must be eliminated before this retrograde contraction can occur. The ECA of the lower small intestine slows significantly and becomes more entrained (time locked) in a distal direction, and this change in ECA is responsible for the increased propagation distance of the stripping contractions of the distal small intestine.

Pharyngoesophageal Responses

After intestinal contents have been refluxed into the stomach, the upper digestive tract responses begin. The pharyngoesophageal responses are coordinated closely with respiratory movements of retching and vomitus expulsion. The first response occurs just prior to

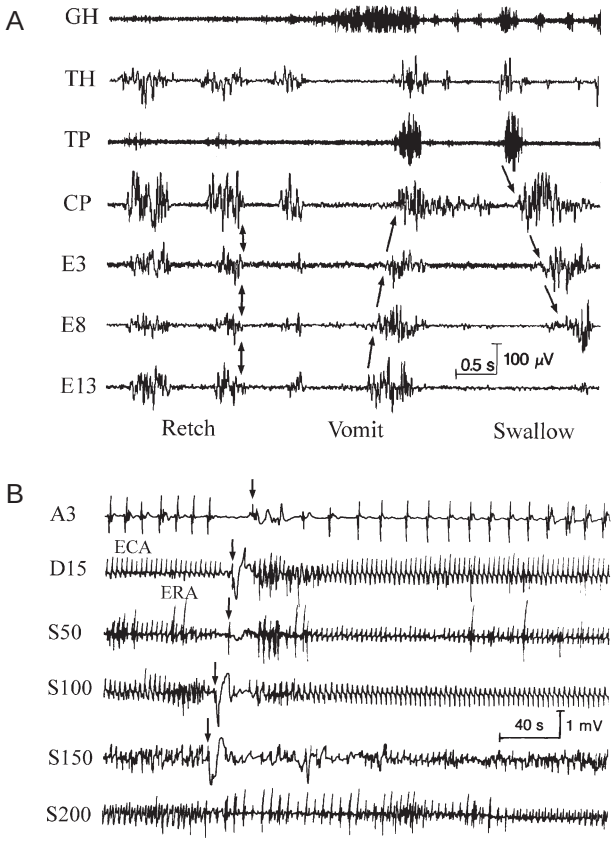


FIGURE 2 Gastrointestinal and esophagopharyngeal myoelectrical responses associated with emesis. (A) Esophagopharyngeal responses; (B) gastrointestinal responses. GH, Geniohyoideus; TH, thyrohyoideus; TP, thyropharyngeus; CP, cricopharyngeus; E, esophagus; A, antrum; D, duodenum; S, small intestine; ECA, electrical control activity; ERA, electrical response activity. The numbers next to abbreviated terms indicate the distance of recording site from CP (A) or A (B). Note that a retrograde propagated electrical event of E occurs at the end of the vomit (A), the retrograde myoelectric response of the gastrointestinal tract is associated with a loss of ECA (B), and the ECA frequency decreases after the retrograde myoelectrical response, especially in the distal small intestine (B).

retching and consists of longitudinal contraction of the striated muscle portion of the esophagus such that the LES rises into the chest. The effect of this action in conjunction with prior relaxation of the LES and fundus is to eliminate the gastroesophageal angle that forms a physical barrier between stomach and esophagus. Retching then occurs about once per second as a series of simultaneous contractions and relaxations of the diaphragm and abdominal muscles. The stomach rises and falls during retching, which mixes gastric contents with intestinal secretions, imparts a momentum to the gastric bolus to assist expulsion, and may stimulate gastric receptors that signal vomiting. Some gastric contents may be expelled during retching, but this is limited by contraction of the diaphragmatic hiatus that pinches off the gastroesophageal junction. Vomitus expulsion occurs when diaphragmatic dome and abdominal muscles contract strongly and simultaneously with relaxation of the diaphragmatic hiatus. These events always occur between retches when the stomach is rising and the gastric contents are already moving orad. As the bolus is expelled to the esophagus, a retrograde contraction of the striated muscle portion of the esophagus captures the bolus, assisting its transport to the pharynx and preventing the bolus from returning to the stomach (Fig. 2A). When the bolus reaches the upper esophageal sphincter (UES), the UES is maximally pulled open by anterior traction of the suprahyoid (Fig. 2A) and infrahyoid muscles and by posterior traction of the suprapharyngeal muscles.

NEURAL CONTROL

Afferent Pathways

The vagus nerves are the primary afferent nerves mediating emesis due to stimulation of the digestive tract. Splanchnic afferent nerves seem to serve little or no role in emesis, although they may mediate peritonitis-induced emesis and provide a significant afferent pathway for radiation-induced emesis.

Efferent Pathways

The gastrointestinal motor correlates of vomiting are controlled by vagal efferent fibers that are transmitted to the jejunum and ileum through the celiac branch of the vagus nerve and the mesenteric nerves. The innervation of the duodenum and stomach are through the duodenal and ventral branches of the vagus nerve, respectively. The neural pathway controlling initiation of the RGC descends about 20 cm within the intestinal wall, probably through the enteric nervous system, before innervating the muscle. The retrograde propagation of the RGC is controlled primarily by the central nervous system, but is modulated by the enteric nervous system. The upper digestive tract responses are controlled by branches of the vagus nerves, including the pharyngoesophageal, recurrent laryngeal, and superior laryngeal nerves.

Central Control

The central emetic pattern generator (CEPG), i.e., the area of the brain controlling the motor and secretory responses associated with emesis, may not be a single nucleus but may comprise a distributed network of brain nuclei. Some areas of the brain important for controlling emesis include the nucleus tractus solitarius, retrofacial nucleus, and dorsolateral reticular formation. However, the roles of specific areas of the brain in mediating and controlling emesis and the associated responses are unknown.

PHARMACOLOGY OF EMESIS

Emesis may be initiated from different sensory pathways, therefore specific types of emesis may be mediated by different neurotransmitters and blocked by different antagonists. Regardless of the emetic stimulus, all emesis is mediated by the CEPG, therefore some antagonists may block all forms of emesis.

Peripheral Control

Digestive tract-initiated emesis is mediated by multiple peripheral serotonergic (5-hydroxytryptamine) receptors. $CuSO_4$-induced and perhaps noxious chemical-induced emesis is mediated by 5-hydroxytryptamine isotype 4 (5-HT4) receptors, probably on chemosensory afferents, but radiation- or cytotoxin-induced emesis is mediated by 5-HT3 receptors, probably on vagal afferents supplying the EC cells. Presynaptic 5-HT3 receptors on vagal afferents supplying the AP may also mediate cytotoxin-induced emesis.

Chemoreceptor Trigger Zone

Numerous neurotransmitters activate emesis by stimulation of the CTZ, but emesis caused by these agonists is blocked only by the specific antagonists. Therefore, these neurotransmitters probably act at the level of the chemoreceptors rather than at the neurotransmitter level. Cholinergic or histaminergic antagonists block only motion sickness, therefore these agents may affect only the motion-sensing neural circuitry.

Central Pattern Generator

A few pharmacologic agents, because they block all forms of emesis, appear to block emesis at the level of the CEPG. The ability of agonists to block emesis suggests that there may be an endogenous emetic inhibitory mechanism. Perhaps the most powerful of the general antiemetic pharmacological agonists is opioid agonists, but cannabinoid and vanilloid receptor agonists may also be antiemetic. Neurokinin-1 receptor antagonists also block emesis due to a variety of emetic stimulants.

NAUSEA AND THE DIGESTIVE TRACT

Nausea is a sensation that is felt in the digestive tract, therefore investigators have sought a digestive tract source of this sensation for decades. Many digestive tract responses have been implicated, but all have subsequently been discarded. It has been recently suggested that changes in the gastric ECA, i.e., tachygastria, cause nausea. However, there is no one-to-one correspondence between tachygastria and nausea, changes in ECA occur at all levels of the gastrointestinal tract, and changes in ECA are just one of the many prodromata that accompany nausea. The prodromata do not cause the feeling of nausea, rather they are the result of the stimulus for nausea. It is probable that the unpleasant feeling in the digestive tract during nausea is a referred sensation similar to referred pain.

See Also the Following Articles

Enteric Nervous System • Hyperemesis Gravidarum • Nausea • Rumination Syndrome

Further Reading

Andrews, P. L. R., and Hawthorn, J. (1988). The neurophysiology of vomiting. *Bailliere's Clin. Gastroenterol.* **2**, 141–168.

Borison, H. L. (1989). Area postrema: Chemoreceptor circumventricular organ of the medulla oblongata. *Prog. Neurobiol.* **32**, 351–390.

Endo, T., Minami, M., Hirafuji, M., Ogawa, T., Akita, K., Nemoto, M., Saito, H., Yoshioka, M., and Parvez, S. H. (2000). Neurochemistry and neuropharmacology of emesis: Role of serotonin. *Toxicology* **153**, 189–201.

Grundy, D., and Scratcherd, T. (1989). Sensory afferents from the gastrointestinal tract. *In* "The Gastrointestinal System. Alimentary Canal" (J. D. Wood, ed.), Sect. 6, Vol. 1, pp. 593–620. American Physiological Society, Bethesda, Maryland.

Hornby, P. J. (2001). Central neurocircuitry associated with emesis. *Am. J. Med.* **111**(Suppl. 8A), 106S–112S.

Lang, I. M. (1990). Digestive tract motor correlates of vomiting and nausea. *Can. J. Physiol. Pharmacol.* **68**, 242–253.

Lang, I. M. (1999). Noxious stimulation of emesis. *Dig. Dis. Sci.* **44**, 58S–63S.

Lang, I. M., and Tansy, M. F. (1982). Minireview: Neural and hormonal control of Brunner's gland secretion. *Life Sci.* **30**, 409–417.

Lang, I. M., Dana, N., Medda, B. K., and Shaker, R. (2002). Mechanism of airway protection during retching, vomiting, and swallowing. *Am. J. Physiol.* **283**, G529–G536.

Miller, A. D., and Leslie, R. A. (1994). The area postrema and vomiting. *Front. Neuroendocrinol.* **15**, 301–320.

Endocrine Pancreas

SUSAN BONNER-WEIR
Joslin Diabetes Center, Boston

alpha cells Pancreatic islet cells that produce glucagon.
beta cells Pancreatic islet cells that produce insulin.
glucagon Hormone produced by alpha cells in the islets of Langerhans; acts to elevate blood glucose.
insulin Glucose-regulating hormone produced and secreted by beta cells in the islets of Langerhans; has broad anabolic effects on body metabolism.
islets of Langerhans Clusters of several hundred to several thousand endocrine cells embedded in the pancreas.
pancreatic duodenal homeobox factor-1 Patterning transcription factor; required for the differentiation of the pancreas.

The endocrine pancreas consists of the islets of Langerhans, which are scattered throughout the exocrine pancreas in all vertebrates evolutionarily higher than the bony fish (teleosts). Islets are complex structures of four primary endocrine cell types that function both separately as microorgans and in concert as the endocrine pancreas to regulate physiologic glucose homeostasis. Although the direct secretion of insulin and glucagon from islets into the portal vein has obvious advantages with respect to influence on hepatic function, it is not clear why the endocrine pancreas is dispersed throughout the exocrine pancreas. The most appealing suggestion is that the local insular–acinar portal system helps regulate the exocrine function of the pancreas.

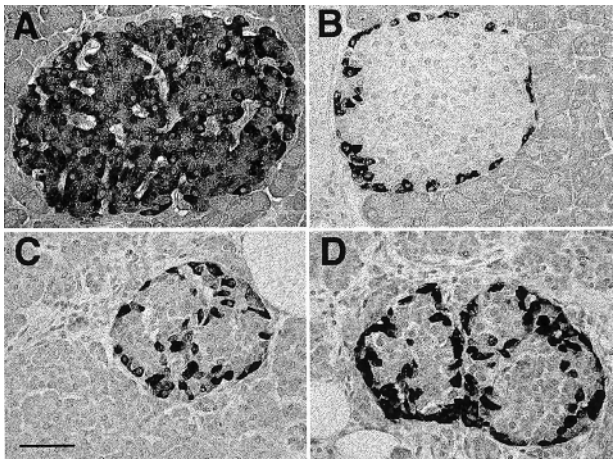


FIGURE 1 The islets of Langerhans of most species have a mantle of non-beta cells (consisting of glucagon-, somatostatin-, and pancreatic polypeptide-producing cells) and a core of insulin-producing beta cells. (A) Mouse islet immunostained for insulin, showing the extensive vasculature between the cords of endocrine tissue. (B) Mouse islet immunostained for glucagon, showing the incomplete mantle surrounding the core of beta cells. (C and D) Human islets immunostained for glucagon, showing a similar pattern of mantle/core but with greater complexity, suggestive of subunits similar to those in the rodent islet. Magnification bar = 50 μm.

INTRODUCTION

Four cell types of the endocrine pancreatic islets of Langerhans are common to all species: beta, or insulin producing; alpha, or glucagon producing; delta, or somatostatin producing; and pancreatic polypeptide (PP) producing. Each islet is a highly vas cularized cluster of these cells organized in a nonrandom manner, with the delta, alpha, and PP (the non-beta) cells occurring as a discontinuous mantle 1 to 3 cells thick around a central core of beta cells (Fig. 1). Islets of humans and other primates have a somewhat more complex arrangement. Sections of human pancreas show many different islet profiles, including oval and cloverleaf patterns, differences that have fueled controversy about whether they actually have a mantle–core arrangement (Fig. 1, C and D). Nonetheless, in three dimensions, human islets can be considered as composites of several mantle–core subunits or as lobulated with mantle–core lobules.

In the adult mammal, the islets make up 1–2% of the pancreatic mass. The relative volume of the islet cell mass varies with age, being much greater in fetuses and the young, presumably because the growth of the islet and exocrine tissues is discordant. The pancreas of an adult human has roughly 1 g of islet tissue, 500,000–1,000,000 islets, and contains about 200 units, or 8 mg, of insulin; the adult rat pancreas contains about 100 μg of insulin. The average rat islet is 150 μm in diameter and contains about 45 ng of insulin. Islet size can range from only a few cells and less than 40 μm in diameter to about 10,000 cells and 400 μm in diameter. Islets smaller than 160 μm in diameter represent 75% of the islets in number but only 15% of the islet

 Copyright 2004, Elsevier (USA). All rights reserved.

volume, whereas islets larger than 200 μm in diameter represent only 15% of the islets in number but 60% of the islet volume.

In adults, 70–80% of the islet consists of insulin-producing beta cells; 5% is somatostatin-producing delta cells, and 15–20% is either glucagon-producing alpha cells or pancreatic polypeptide-producing cells. However, at birth, the beta cells are usually only 50% of the islet, but with postnatal replication of the beta cells and an increase in cell volume, their proportion increases. The beta cell population is dynamic, undergoing compensatory changes in function and mass in order to maintain euglycemia.

PANCREATIC DEVELOPMENT

The pancreas is derived from outpocketings of the primitive gut. There are two or three pancreatic anlagen or primordia: a dorsal outpocketing of the gut across from the budding liver and, after a short temporal lag, one or two ventral buds (which fuse very early) from the base of the biliary floor of the gut. The dorsal anlage that gives rise to the splenic portion (tail and body) of the pancreas fuses to various degrees with the ventral anlage that gives rise to the head or duodenal portion of the pancreas. Islet cells are seen first as single cells along the terminal pancreatic tubules and then as clusters of cells within the epithelial basement membrane. These clusters become separated from the ductal epithelium to form islets.

The origin of the pancreas as separate primordia is thought to be the basis of the regional distribution of glucagon-producing and pancreatic polypeptide-producing cells. The dorsal pancreas, supplied with blood by the celiac trunk via the gastroduodenal and splenic arteries and drained by one main pancreatic duct, contains the glucagon-rich islets, with few pancreatic polypeptide-producing cells. The opposite distribution is found in the ventral pancreas, which is supplied with blood from the superior mesenteric artery via the inferior pancreaticoduodenal artery and is drained by a separate duct. Here the islets contain pancreatic polypeptide-producing cells and few, if any, glucagon-producing cells. The degree of fusion of these ducts differs among species.

The alpha cells appear before the others, followed by the beta cells and finally by the delta cells (PP cells have been reported several days before birth in rats). As they form, endocrine cells may express more than one hormone, with a progressive restriction to just one hormone. The finding that ablation of a homeobox transcription factor, pancreatic duodenal homeobox factor-1(pdx-1) [also known as insulin promoter factor-1 (IPF-1)], resulted in the failure to develop a pancreas brought great energy into defining the cascade of transcription factors involved in the differentiation of pancreas from endoderm and subsequently to the varied pancreatic cell types (exocrine and endocrine) (Fig. 2). Recent studies have shown that p48 [also known as pancreatic transcription factor-1 (PTF-1)], a transcription factor previously thought only to be the determinant for pancreatic acinar tissue, is also necessary for pancreatic formation.

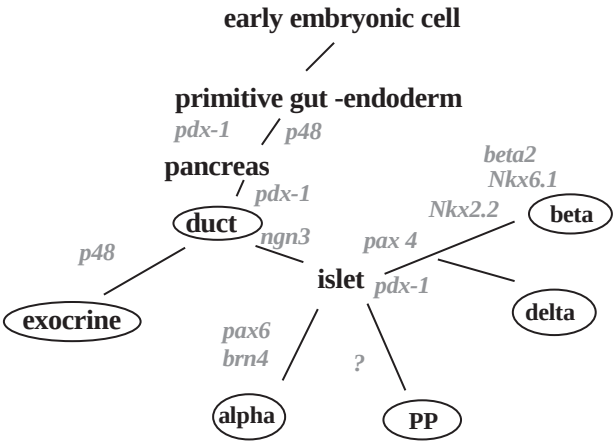


FIGURE 2 The cascade of transcription factors that are necessary for formation of the endocrine pancreas. Both pdx-1 and p48 are necessary for pancreas organogenesis, but each is limited to expression in certain cells in the mature pancreas, with pdx-1 expressed in insulin-producing beta cells and p48 expressed in acinar cells. Neurogenin 3 (ngn3) has been shown to be a marker of the endocrine progenitor cells. PP, Pancreatic polypeptide-producing cell. As indicated by the question mark, additional transcription factors remain to be placed in this schema.

COMPONENTS OF THE ISLETS OF LANGERHANS

Endocrine Cells

The four major endocrine cell types in mammalian islets are distinguished by ultrastructural and immunocytochemical techniques. The beta cells are polyhedral, being truncated pyramids, and are usually well granulated with secretory granules 250–300 nm in diameter. It has been estimated that each mouse beta cell is about 1000 μm^3 and contains about 10,000 insulin granules. Insulin granules can be found as mature granules that have an electron-dense core and a loosely fitting granule-limiting membrane with the appearance of a spacious halo, or as immature granules with little or no halo with moderately electron-dense contents (Fig. 3). Immature granules have been shown to be

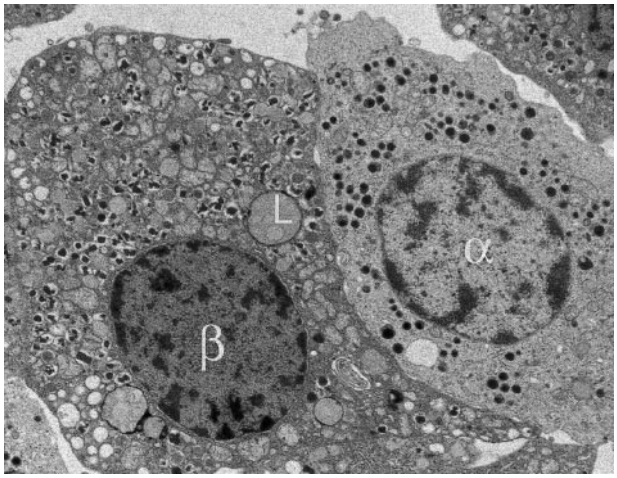


FIGURE 3 Insulin-producing beta cells and glucagon-producing alpha cells can be distinguished by secretory granule morphology. In a human islet, mature insulin granules in a beta cell have electron-dense crystalline cores and loosely fitting, granule-limiting membranes with the appearance of a spacious halo; immature granules have little or no halo and moderately electron-dense contents. Beta cells also contain lipid inclusions (L) that increase in numbers with cell age. In alpha cells, glucagon-containing granules are electron dense and have a narrow halo of less dense material and a tightly fitting, granule-limiting membrane.

the major, if not the only, site of conversion of proinsulin to insulin. In some species, the electron-dense core of the mature granule is visibly crystalline. Insulin is only one of at least 120 proteins in the granules.

The alpha cells are usually smaller and more columnar than the beta cells and well granulated with granules 200–250 nm in diameter. The granules are electron dense with a narrow halo of less dense material and a tightly fitting granule-limiting membrane; there is little species variation. The delta cells are usually smaller than either alpha or beta cells, are well granulated, and are often dendritic in shape. Within a delta cell, the electron density of granules varies greatly. Each granule, 200–250 nm in diameter, contains material of homogeneous moderate density that fills the granule-limiting membrane.

The PP cells are the most variable among species. In humans, the granules are elongated, very electron dense, and 120–160 nm in diameter, but in dogs and cats the granules are spherical, variable in electron density, and about 300 nm in diameter.

Microvasculature

The islets of Langerhans have a glomerular-like capillary network with a direct arteriolar blood supply. One to three arterioles penetrate each islet through discontinuities of the non-beta cell mantle and enter directly into the beta cell core, where each branches into a number of fenestrated capillaries. These capillaries follow a tortuous path, passing first through the beta cell core and then through the non-beta cell mantle. Often, a capillary will pass along the inside of the mantle before penetrating it to leave the islet. The pattern of microvasculature varies with islet size. In small islets ($<160\,\mu m$ in diameter), the efferent capillaries pass through exocrine tissue for 50 to 100 μm before coalescing into collecting venules. Large islets ($>200\,\mu m$ in diameter) are selectively located near the larger ducts and blood vessels. Their efferent vessels coalesce within the islet capsule, thus they probably have little effect on surrounding exocrine tissue. However, the vascular pattern of the small islets and their abundance would lead to an effective insuloacinar portal system.

The blood flow to the islets has been found to be disproportionately large (10–20% of the pancreatic blood flow) for the 1–2% of pancreatic volume. High concentrations of glucose have been shown to enhance pancreatic blood flow and to preferentially increase islet blood flow. Lymphatic vessels, although common in the pancreas, are not found within the islets.

Capsule

The islet capsule is only a single layer of fibroblasts and the collagen fibers laid down by these fibroblasts. Frequently, the capsule is discontinuous and is absent between islet and exocrine cells, such that the only separations between exocrine and islet cells are their respective basement membranes The capsule overlies the efferent blood vessels of the islet and thus defines a subcapsular interstitial space. On collagenase digestion for islet isolation, the capsule is disrupted and lost.

Nerves

The pancreas is innervated by sympathetic fibers via the celiac and superior mesenteric ganglia and by parasympathetic fibers from the vagus nerve. These parasympathetic fibers synapse in small ganglia dispersed in the pancreas; in some species, these ganglia are close to, or even within, the islets. They may act as pacemakers for the oscillations in hormone secretion that occur without extrinsic nervous connections, as in the isolated perfused canine pancreas. Within the islets, the nerves follow the blood vessels and terminate within the pericapillary space, within the capillary basement membrane, or closely apposed to the endocrine cells. Because no specialized synapses are found, it has been suggested that neurotransmitters released into

the interstitial space may affect a number of neighboring islet cells.

THE ISLET AS AN ORGAN

Functional Implications of Islet Organization

It is important to understand how the three-dimensional organization of the islet may determine islet function. Within the islet, endocrine cells are organized around the microvasculature. In the rat, 8 to 10 beta cells form a tube or rosette around a capillary, but each beta cell abuts a second capillary. In experimentally degranulated beta cells, the remaining granules are polarized toward the central "venous" capillary, suggesting an *in situ* polarity of the beta cells.

The interrelations of the islet cells and their relations with the microvasculature have functional implications. The microvascular pattern of the islet confers a directionality to the blood flow—from the point of arteriolar entry outward through the beta cell core to the peripheral non-beta cell mantle. Physiologic data from anterograde and retrograde passive neutralization perfusions of dog, rat, monkey, and human pancreases support a beta to alpha to delta (B—A—D) directional pattern of blood flow. This pattern of flow would favor insulin having an effect on the alpha and delta cells by its being transported in high concentrations from the core to the mantle. The reverse, that of the beta cells being influenced by local blood-borne somatostatin or glucagon, is not supported by the vascular pattern.

However, some paracrine effects may occur by simple diffusion through the interstitium. The dynamics and direction of flow of the interstitial fluid are not known, but we would expect it to be in the same direction as the blood flow. One could speculate that diffusion in the interstitial space allows somatostatin and glucagon to influence not only each other (adjacent cells) but also those adjacent beta cells. In large islets, there would be a central core of beta cells that would be isolated by their distance from the non-beta cells and thus from all but systemically circulating levels of the other islet hormones. In small islets, essentially all beta cells might be close enough to the non-beta cells to be influenced by their hormones diffusing into the interstitial fluid.

The intercellular contacts between the endocrine cells may also influence their function. Stimulated secretion is enhanced when beta cells have contact with at least one other beta cell. This enhancement is thought to be due to electrical coupling through gap junctions but may also involve cell adhesion molecules such as integrins.

Heterogeneity of Islets

We now realize that the islets within one pancreas are not all alike. Islets differ in cellular composition due to the regional distribution of the glucagon-producing alpha cell and the pancreatic polypeptide-producing cells. Increasing evidence suggests even that not all beta cells within an islet are identical. Studies on single islet cells have shown that individual beta cells may have different thresholds for glucose-induced insulin release; thus, with increasing concentrations of glucose, more beta cells are recruited in the response. This cellular heterogeneity has been extended to include the redox state and the threshold for glucose-induced biosynthesis of proinsulin and glucose-induced cytoplasmic free calcium concentration. It is unknown whether the heterogeneity is intrinsic and constant for a particular cell through its lifetime, is related to the age of the cell, or is imprinted by a factor from the environment. In addition, recent studies have questioned if such heterogeneity were artifactual.

Insular—Acinar Portal System

It has been hypothesized that one reason the endocrine pancreas is dispersed as islets throughout the exocrine pancreas is that the islet hormones regulate exocrine function and growth locally through an insuloacinar portal system. Microvascular corrosion casts show efferent vessels from islets passing through the exocrine tissue before coalescing into venules; these vessels comprise an insuloacinar portal system. Such vessels are mostly from smaller islets (160 μm in diameter and smaller) that are usually embedded in exocrine tissue. The efferent vessels of the larger islets are postcapillary and collecting venules that are considered the "leakiest" or most permeable in other tissues, thus allowing diffusion of the hormones into the surrounding acinar tissue. The presence of periinsular halos of enlarged and highly granulated exocrine cells adjacent to islets has been cited as histological evidence of tropic effects of islet hormones on the exocrine pancreas.

LIFETIME CHANGES IN THE ENDOCRINE PANCREAS

For many years, the accepted concept was that all the pancreatic beta cells present throughout life were present at birth. Many researchers thought that insulin resistance would lead to diabetes without change in the beta cells. However, evidence over the past decade has

been compelling; in most cases, the beta cells can, and do, compensate for added demand arising from pregnancy, obesity, or insulin resistance. It is important to remember that only 15–20% of obese or severely insulin-resistant humans become diabetic; the others maintain normoglycemia due to beta cell compensation. The new concept is that the beta cell mass is dynamic, with increases and decreases both in function and mass maintaining the glycemic level within a very narrow physiological range. The changes in mass can be both in number (hyperplasia) and in individual volume of beta cells (hypertrophy). When the mass cannot increase adequately, diabetes ensues.

During embryonic development, multipotent cells of the primary pancreatic buds differentiate into specific pancreatic phenotypes of ductal, acinar, and the four endocrine islet cells. Until late gestation, most beta cells result from the differentiation of ductal precursor cells to islet cells, a process called neogenesis. Shortly before birth, replication becomes the major mechanism for adding new beta cells, but both neogenesis (the differentiation process) and replication continue throughout adult life. In rodents, during the first days after birth, many new islets are formed, and a second wave of neogenesis occurs around the time of weaning. Beta cell replication is significantly higher during late gestation and the neonatal period compared to the period following weaning, with little change in replication rates occurring beyond 30 to 40 days of age.

From birth to adulthood, the volume of islet tissue with respect to that of the pancreas (relative volume) decreases, but the actual volume of islet tissue increases; the more slowly growing islet tissue is diluted by an exuberant postnatal growth of the acinar tissue. The beta cell mass has been shown to increase 12- to 15-fold from birth to adulthood. In fact, beta cell mass is linearly correlated with body weight in adult mice and rats and with body mass index (BMI) in humans. A longitudinal study shows that the addition of new cells accounts for the increase in beta cell mass throughout most of the life span of rats, but from 15 months onward hypertrophy of the beta cell is the main mechanism of for the expansion of beta cell mass.

The beta cell mass is dynamic in nature and increases or decreases in order to maintain euglycemia. Functional compensation, or changes in the amount of insulin each beta cell secretes, can be due to changes in the responsive threshold for glucose-induced insulin secretion, possibly caused by activation of glucokinase, or by recruitment of cells with heterogeneous response thresholds. The beta cell mass is the major factor in determining the amount of insulin that can be secreted. The mechanisms involved in this compensatory regulation are the determinants of beta cell mass: rate of replication, neogenesis and apoptosis, and cell size (hypertrophy vs. atrophy).

DETERMINANTS OF BETA CELL MASS

Neogenesis

Neogenesis is clearly the main pathway of beta cell increase during early to midgestation, but still occurs in the normal postnatal animal. In adults, islet cells of all types can be immunolocalized in the pancreatic ducts as occasional single cells or small budding islets. This process can be stimulated to form substantial numbers of new islets under numerous *in vivo* experimental conditions. Additionally, adult human ductal tissue can be differentiated into islets *in vitro*. In some experimental situations, such as the regeneration after partial pancreatectomy in the adult rodent, new lobes of pancreas (exocrine and endocrine) are added. There is some suggestion that new lobes of pancreas continue to form in adults.

Replication

Beta cell replication is significantly higher during late gestation and the neonatal period compared to the period following weaning, with little change in replication rates occurring beyond 30 to 40 days of age in rodents. The cell cycle has been determined using isolated rat islets synchronized in culture by exposure to hydroxyurea. A calculated cell cycle of 14.9 hours has been assumed to reflect that of the beta cell, the predominant cell of the islet. The length of the cell cycle does not change with glucose stimulation or with the age of the animals. Instead, the growth rate is regulated by the number of beta cells that can enter the division phase (G_1) from the resting (G_0) phase. Postnatal proliferation rate of beta cells is low, but the proliferation of differentiated beta cells must not be underestimated; it is adequate to maintain a slowly increasing beta cell mass. Replication can be further stimulated by prolactin, growth hormone, glucose, and glucagon-like peptide-1 (GLP-1). Replicating beta cells have been identified by bromodeoxyuridine and tritiated thymidine incorporation in cells immunolabeled for insulin, and numerous electron micrographs of mitotic figures in cells with insulin granules have been published. There may be functional differences in beta cells that can replicate and those that cannot, and as a cell enters the cell cycle, it may transiently lose function.

Hypertrophy

Hypertrophy, an increase in cell size without replication, can be an efficient and economical mechanism for rapid and transient compensation and has been recognized in many models, including late pregnancy.

Apoptosis

Apoptosis is a normal determinant of beta cell mass. The frequency of apoptotic beta cells in the adult rat has been found to be about 0.5%; this frequency cannot be equated with a rate because it is not known how long the apoptotic process takes nor how long apoptotic bodies are visible. The concept of beta cell turnover became evident with mathematical modeling of the beta cell mass. With beta cell replication of 2–3%/24 hours, as found in the adult rodent pancreas, the beta cell mass can double in about 1 month. Between 1 and 2, 2 and 3 months of age, the rat beta cell mass does almost double, but this monthly doubling does not continue, suggesting a loss of cells. The life span of a rat beta cell has been estimated as approximately 58 days. The slowly increasing beta cell mass results from the balance between cell formation and cell loss, such that the endocrine pancreas must be considered a slowly renewed tissue.

In the second week after birth, remodeling of the endocrine pancreas occurs simultaneous with a high rate of replication of beta cells and a high frequency of apoptosis. This remodeling coincides with marked changes in the message RNA levels of both insulin-like growth factor-1 and -2 (IGF-1 and IGF-2) as well as transient appearances of IGF binding protein-1 and -2 (IGFBP-1 and IGFBP-2), leading to the suggestion that an inadequate availability of survival factors, such as the IGFs, is the cause of the increased cell loss. This hypothesis is supported by evidence of the suppression of normal neonatal apoptosis by persistent IGF-2 in overexpressing transgenic mice. This remodeling may have significance in the priming of autoimmune diabetes, as seen in non-obese, diabetic (NOD) mice, BioBreeding (BB) rats, and humans, because apoptotic bodies are highly immunogenic.

Sources of Cells for Renewal: Progenitor and Stem Cells

With the acceptance of the concept of postnatal growth of the beta cell mass, interest in the adult "pancreatic stem cell" has increased. The characterization of such a stem/progenitor cell has been elusive. One likely source of progenitor cells is the pancreatic ducts, because adult duct epithelium retains the ability to give rise to all the differentiated cell types of the pancreas. Current research is focused on characterizing the cells capable of generating new islet cells in the adult pancreas.

GLUCOSE EFFECTS ON BETA CELLS

The effects of elevated glucose levels can be both beneficial, as in stimulation of beta cell proliferation, and detrimental, as in cell death; the term "glucose toxicity" has been used to describe the latter effect. There are a number of secondary effects of a hyperglycemic environment, including the loss of glucose-induced insulin secretion and even loss of specific beta cell differentiation. Additionally, beta cells in some strains or species undergo apoptosis when exposed to hyperglycemia, but in other strains, beta cells can have more adaptive responses and maintain a stable mass, even with chronic hyperglycemia.

See Also the Following Articles

Exocrine Pancreas • Pancreas, Anatomy • Pancreas, Development

Further Reading

Bonner-Weir, S. (2000). Islet growth and development in the adult. *J. Mol. Endocrinol.* **24**, 1–6.

Bonner-Weir, S., and Orci, L. (1982). New perspectives on the microvasculature of the islets of Langerhans in the rat. *Diabetes* **31**, 883–939.

Dean, P. M. (1973). Ultrastructural morphometry of the pancreatic β-cell. *Diabetologia* **9**, 115–119.

Edlund, H. (1999). Pancreas: How to get there from the gut? *Curr. Opin. Cell Biol.* **11**, 663–668.

Finegood, D. T., Scaglia, L., and Bonner-Weir, S. (1995). (Perspective) Dynamics of β-cell mass in the growing rat pancreas: Estimation with a simple mathematical model. *Diabetes* **44**, 249–256.

Hellerstrom, C., Swenne, I., and Andersson, A. (1988). Islet cell replication and diabetes "The Pathology of the Endocrine Pancreas in Diabetes" (P. J. Lefebvre, and D. G. Pipeleers, eds.), pp. 141–170. Springer-Verlag, Heidelberg.

Henderson, J. R. (1969). Why are the islets of Langerhans? *Lancet* **2**, 469–470.

Orci, L. (1985). The insulin factory: A tour of the plant surroundings and a visit to the assembly line. *Diabetologia* **28**, 528–546.

Pipeleers, D. G. (1992). Heterogeneity in pancreatic β-cell population. *Diabetes* **41**, 777–781.

Schwitzgebel, V. M., Scheel, D. W., Conners, J. R., Kalamaras, J., Lee, J. E., Anderson, D. J., Sussel, L., Johnson, J. D., and German, M. S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542.

Weir, G. C., and Bonner-Weir, S. (1990). Islets of Langerhans: The puzzle of intraislet interactions and their relevance to diabetes. *J. Clin. Invest.* **85**, 983–987.

Endomysial and Related Antibodies

KANTHI YALAMANCHILI AND KIMBERLY M.PERSLEY
Presbyterian Hospital of Dallas

antibody A Y-shaped immunoglobulin protein on the surface of B cells that is secreted into the blood or lymph in response to an antigenic stimulus.

celiac sprue (gluten-sensitive enteropathy) A malabsorptive disorder in which there is an atrophic and inflamed proximal small intestinal mucosa that usually improves morphologically and clinically on treatment with a gluten-free diet but relapses when gluten is reintroduced.

Antibodies to gliadin, reticulin, endomysium, and tissue transglutaminase have been detected in patients with celiac sprue. Studies have shown that endomysial antibodies, tissue transglutaminase antibodies, and anti-gliadin antibodies seem to be the most useful in screening patients for celiac sprue.

ENDOMYSIAL ANTIBODIES

Endomysial antibody (EMA) is directed against endomysium, a connective tissue protein found between myofibrils in the smooth muscle cells of the gastrointestinal tract of primates. Once EMAs were found to be associated with gluten-sensitive enteropathy, further tests were developed to detect them. An indirect immunofluorescence assay was implemented in which the binding of the serum EMAs to a tissue substrate produces a characteristic staining pattern. The initial tissue substrate used for the indirect immunofluorescence assay, primate esophagus, is limited in supply. However, now with more readily available human umbilical cord as the tissue substrate, wider application of serum EMA testing has been possible. Most patients with celiac sprue have a serum EMA of the immunoglobulin A (IgA) subtype. However, EMA of the IgG_1 subtype has also been employed for diagnosis (mostly in IgA-deficient patients). The sensitivity and specificity of IgA EMAs are reported to be 85–98% and 97–100%, respectively. Thus, serum EMA testing has a high positive predictive value.

TISSUE TRANSGLUTAMINASE

The target autoantigen for EMA in celiac disease has been identified as a tissue transglutaminase (tTG). Autoantibodies (IgA and IgG) to tissue transglutaminase have also been recently detected via more easy-to-perform enzyme-linked immunosorbent assays (ELISAs). Studies have shown that the sensitivity of IgA tissue transglutaminase antibodies (TTAs) is 90–98% and the specificity is 95–97%. These high positive and negative predictive values of IgA TTA have led more investigators to recommend the IgA TTA as the initial screening serologic test to diagnose celiac sprue, followed by EMA testing if IgA TTA is borderline, followed by a small intestinal biopsy if either EMA or TTA or both are positive. However, it is not yet clear how extensively TTA and EMA results overlap since both seem to bind to the same substrate, namely, tissue transglutaminase.

ANTI-GLIADIN ANTIBODIES

Finally, anti-gliadin antibodies (AGAs) are antibodies to wheat gliadin and have also been extensively studied in celiac sprue. Two anti-gliadin antibody subtypes exist, IgA and IgG. The sensitivity of IgG AGAs is 75–85% and the specificity is 75–90%. The sensitivity of IgA AGAs is 80–90% and the specificity is 85–95%. Therefore, IgA AGAs are more sensitive and specific than IgG AGAs and a negative IgA AGA result has a high negative predictive value. Interpretation of positive AGAs, however, must be scrutinized carefully because other conditions, such as cow's milk intolerance, IgA nephropathy, Crohn's disease, eosinophilic enteritis, and tropical sprue, may raise AGA levels.

In patients with celiac disease who are also IgA-deficient, all IgA antibody-based tests are negative. In these patients, elevated levels of IgG_1 EMAs and IgG TTAs support the diagnosis of celiac sprue.

The benefit of screening for asymptomatic celiac disease is debatable, but potential advantages include reducing the risk of enteropathy-associated T-cell lymphoma, reducing the risk of nutritional deficiency states, preventing mild intestinal symptoms, and improving general well-being, assuming adoption of and strict adherence to a gluten-free diet.

Although the IgA TTA assay has a high positive and negative predictive value and is an easy-to-perform ELISA-based test, it is relatively new. EMA testing has been the primary screening tool through the years and thus continues to be a comfortable initial assay with

 Copyright 2004, Elsevier (USA). All rights reserved.

which to screen populations with a high pretest probability of celiac sprue. EMAs in combination with TTAs or AGAs have positive and negative predictive values approaching 100%. Therefore, if either test is positive, the patient or the screened individual should be referred for diagnostic biopsy of the small intestine. If both tests are negative, it is highly unlikely that the individual has celiac disease. These serologic markers are useful not only in supporting the diagnosis but also in following the response to a gluten-free diet because it has been shown that these tests often but not always revert to negative when patients with celiac sprue comply with a gluten-free diet.

See Also the Following Articles

Celia Disease • Celiac Disease, Pediatric • Immunoglobulins

Further Reading

Biagi, F. (2001). Endomysial and tissue transglutaminase antibodies in celiac sera: A comparison not influenced by previous serological testing. *Scand. J. Gastroenterol.* **36**, 955–958.

Ciclitira, P. J. (2001). AGA technical review on celiac sprue. American Gastroenterology Association. *Gastroenterology* **120**, 1526–1540.

Farrell, R. J., and Kelly, C. P. (2002). Celiac sprue and refractory sprue. *In* "Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 1817–1838. W. B Saunders, Philadelphia, PA.

Lagerqvist, C. (2001). Screening for adult coeliac disease—Which serologic marker(s) to use? *J. Intern. Med.* **250**, 241–248.

Troncone, R. (1999). IgA antibodies to tissue transglutaminase: An effective diagnostic test for celiac disease. *J. Pediatr.* **134**, 166–171.

Endoscopic Ultrasonography

KEVIN M. MCGRATH
University of Pittsburgh Medical Center

adjuvant Postoperative cancer treatment.
anechoic Without echoes (black); generally correlates with a fluid-filled structure.
echo Reflected sound wave.
endoscopic ultrasonography Application of a high-frequency ultrasound probe to the tip of an endoscope; gastrointestinal tract pathology and lesions can be evaluated immediately adjacent to probe tip.
frequency Periodicity of sound waves per unit of time.
hertz (Hz) Unit of ultrasound frequency equal to one cycle per second.
high frequency More sound waves transmitted per unit of time, greater image resolution, less tissue penetration.
hyperechoic Brighter in appearance than the reference tissue.
hypoechoic Darker in appearance than the reference tissue.
linear array Ultrasound viewing field is in a narrow plane that is parallel to the shaft of the scope.
low frequency Less sound waves transmitted per unit of time, greater tissue penetration, less image resolution.
megahertz (MHz) One million hertz.
neoadjuvant Preoperative cancer treatment.
posterior enhancement Bright (hyperechoic) area on the far side of an anechoic (fluid-filled) structure; due to increased sound wave transmission through a fluid medium.
radial scanning Ultrasound field (360°) is in a plane that is perpendicular to the shaft of the scope.

Endoscopic ultrasound, a combination of both endoscopy and ultrasonography, allows for high-resolution imaging of lesions within and adjacent to the gastrointestinal tract. This system has been accomplished by mounting a high-frequency (5.0–20.0 MHz) ultrasound transducer on the tip of an endoscope. The endoscope can be introduced into the upper or lower digestive tract, providing detailed ultrasound imaging of the gastrointestinal wall or adjacent structures, such as the pancreas, bile duct, liver, mediastinum, and major blood vessels. Endoscopic ultrasound therefore plays a major role in staging cancers of the digestive system. With the advent of fine needle

Copyright 2004, Elsevier (USA). All rights reserved.

which to screen populations with a high pretest probability of celiac sprue. EMAs in combination with TTAs or AGAs have positive and negative predictive values approaching 100%. Therefore, if either test is positive, the patient or the screened individual should be referred for diagnostic biopsy of the small intestine. If both tests are negative, it is highly unlikely that the individual has celiac disease. These serologic markers are useful not only in supporting the diagnosis but also in following the response to a gluten-free diet because it has been shown that these tests often but not always revert to negative when patients with celiac sprue comply with a gluten-free diet.

See Also the Following Articles

Celia Disease • Celiac Disease, Pediatric • Immunoglobulins

Further Reading

Biagi, F. (2001). Endomysial and tissue transglutaminase antibodies in celiac sera: A comparison not influenced by previous serological testing. *Scand. J. Gastroenterol.* **36**, 955–958.

Ciclitira, P. J. (2001). AGA technical review on celiac sprue. American Gastroenterology Association. *Gastroenterology* **120**, 1526–1540.

Farrell, R. J., and Kelly, C. P. (2002). Celiac sprue and refractory sprue. *In* "Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 1817–1838. W. B Saunders, Philadelphia, PA.

Lagerqvist, C. (2001). Screening for adult coeliac disease—Which serologic marker(s) to use? *J. Intern. Med.* **250**, 241–248.

Troncone, R. (1999). IgA antibodies to tissue transglutaminase: An effective diagnostic test for celiac disease. *J. Pediatr.* **134**, 166–171.

Endoscopic Ultrasonography

KEVIN M. MCGRATH
University of Pittsburgh Medical Center

adjuvant Postoperative cancer treatment.
anechoic Without echoes (black); generally correlates with a fluid-filled structure.
echo Reflected sound wave.
endoscopic ultrasonography Application of a high-frequency ultrasound probe to the tip of an endoscope; gastrointestinal tract pathology and lesions can be evaluated immediately adjacent to probe tip.
frequency Periodicity of sound waves per unit of time.
hertz (Hz) Unit of ultrasound frequency equal to one cycle per second.
high frequency More sound waves transmitted per unit of time, greater image resolution, less tissue penetration.
hyperechoic Brighter in appearance than the reference tissue.
hypoechoic Darker in appearance than the reference tissue.
linear array Ultrasound viewing field is in a narrow plane that is parallel to the shaft of the scope.
low frequency Less sound waves transmitted per unit of time, greater tissue penetration, less image resolution.
megahertz (MHz) One million hertz.
neoadjuvant Preoperative cancer treatment.
posterior enhancement Bright (hyperechoic) area on the far side of an anechoic (fluid-filled) structure; due to increased sound wave transmission through a fluid medium.
radial scanning Ultrasound field (360°) is in a plane that is perpendicular to the shaft of the scope.

Endoscopic ultrasound, a combination of both endoscopy and ultrasonography, allows for high-resolution imaging of lesions within and adjacent to the gastrointestinal tract. This system has been accomplished by mounting a high-frequency (5.0–20.0 MHz) ultrasound transducer on the tip of an endoscope. The endoscope can be introduced into the upper or lower digestive tract, providing detailed ultrasound imaging of the gastrointestinal wall or adjacent structures, such as the pancreas, bile duct, liver, mediastinum, and major blood vessels. Endoscopic ultrasound therefore plays a major role in staging cancers of the digestive system. With the advent of fine needle

Copyright 2004, Elsevier (USA). All rights reserved.

aspiration, endoscopic ultrasound can guide needle biopsies of tumors, lymph nodes, and other lesions that are in close proximity to the gastrointestinal tract. This has greatly expanded diagnostic capability, and new indications continue to evolve.

INTRODUCTION

Endoscopic ultrasound (EUS) was originally designed in 1980 to better image the pancreas. Early echoendoscopes were big, bulky, and difficult to maneuver, and image quality was poor. With technologic advances over the past decade, these scopes are now smaller, more flexible, and provide high-quality endoscopic and ultrasound images.

Certain substances (fluid, blood) transmit or conduct ultrasound, whereas others reflect the sound waves (air, bone). Ultrasound images are created by sound waves that are reflected back to the transducer by objects or tissue components in the wave path. These reflected sound waves are termed "echoes." The detail of the image (resolution) and the depth of ultrasound (US) penetration into the tissue are a function of the ultrasound frequency. Low-frequency sound waves have an increased tissue penetration capability, but image resolution is less defined. High-frequency sound waves increase the image resolution, but the penetration depth is much less. EUS employs frequencies between 5 and 20 MHz., with penetration depths ranging from 1 to 7 cm. In contrast, transabdominal ultrasound utilizes lower frequency sound waves (2.5–5.0 MHz) to visualize more distant internal organs.

EQUIPMENT

EUS utilizes specialized endoscopes, called echoendoscopes, that incorporate ultrasound transducers on the distal tip. There are two types of echoendoscopes, defined by the orientation of the ultrasound transmission with respect to the shaft of the endoscope. The optical or endoscopic image is generated from fiber-optic or computerized video technology.

Radial Scanning Echoendoscopes

The radial scanning echoendoscope uses a mechanical rotating transducer that creates a 360° image in a circular plane perpendicular to the shaft of the scope (Fig. 1). A small balloon is placed over the transducer, so that a water bath can be created to enhance US transmission, or acoustic coupling. The viewing optics are located just proximal to the transducer and are oriented at a 60° oblique angle. Current radial echoendoscopes

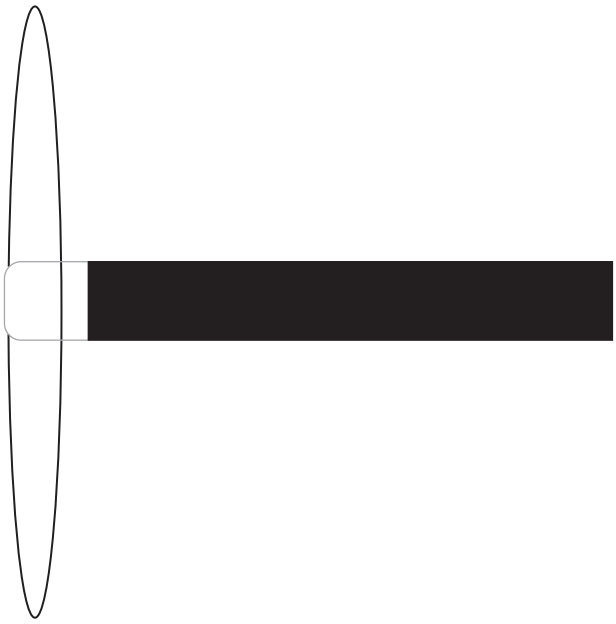


FIGURE 1 Diagram of radial scanning echoendoscope showing the US scanning plane perpendicular to the shaft of the scope. Reproduced with permission from Rosch and Classen (1992).

(Fig. 2) have switchable ultrasound frequencies of 5, 7.5, 12 and 20 MHz. The 360° ultrasound image resembles a computed tomography (CT) scan image.

A transverse array echoendoscope also exists; in this system, an electronic transducer creates a 270° image in a plane that is perpendicular to the shaft of the scope. This instrument has forward-viewing optics and switchable frequencies (5, 7.5, and 10 MHz). Although current

FIGURE 2 The tip of the radial scanning echoendoscope. Courtesy of Olympus America, Inc.

experience with this instrument is limited, it appears to be a promising addition to the EUS armamentarium.

Curved Linear Array Echoendoscopes

The curved linear array (linear) echoendoscope uses an electronic transducer that scans sagitally in a narrow viewing plane that is parallel to the shaft of the scope (Fig. 3). The major advantage of the linear scanner is the ability to perform fine needle aspiration (FNA). The needle exits the scope in the ultrasound plane, therefore it can be sonographically visualized as it is advanced into a target lesion. Additionally, the linear scanner has Doppler capability so that vascular structures can be identified and avoided during FNA. Two linear echoendoscope models are available; one with switchable frequencies between 5 and 7.5 MHz and a 120° scanning angle (Fig. 4), and one with switchable frequencies between 5, 7.5, and 10 MHz and a 180° scanning angle (Fig. 5).

Miniprobes

Specialized high-frequency radial "miniprobes" are available; these can be passed through the operating channel of a standard endoscope to provide detailed imaging of mucosal and other superficial lesions. Available frequencies range from 12 to 30 MHz.

EUS PROCEDURE

In the vast majority of cases, EUS is performed as an outpatient procedure using conscious sedation. If a patient cannot be adequately sedated, then general anesthesia is required to perform the exam. Depending on the indication, the duration of the exam can range from 10 to 90 minutes. The echoendoscope is introduced through the mouth or anus and is maneuvered to the area of interest. By positioning the ultrasound transducer in certain locations within the digestive tract, particular organs or structures may be viewed in great detail. As sound waves are reflected by air, excess air within the gut lumen is continuously suctioned through the scope. As previously mentioned, a small balloon surrounds the transducer to allow for creation of a water bath to enhance acoustic coupling. Additionally, water can be instilled into the gut lumen to decrease air artifact and enhance imaging.

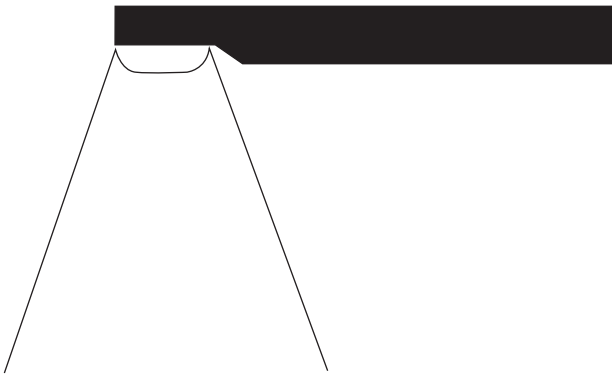


FIGURE 3 Diagram of the linear array echoendoscope showing the US scanning plane parallel to the shaft of the scope. Reproduced with permission from Rosch and Classen (1992).

FIGURE 4 Tip of a linear array echoendoscope with FNA needle. Courtesy of Pentax Precision Instruments, Inc.

EUS-GUIDED FINE NEEDLE ASPIRATION

Utilizing linear array echoendoscopes, FNA can be performed under real-time ultrasound guidance. Needle devices (19, 22, or 25 gauge) can be passed down the operating channel of the echoendoscope, and when an abnormality has been targeted, the needle can be advanced through the gut wall and into the lesion under constant ultrasound guidance. Doppler analysis is employed to ensure that the path of the needle is clear of intervening blood vessels. Once the needle is in position, suction is applied and the needle is moved to and fro within the lesion to obtain a cytologic sample. Slides are prepared within the endoscopy room, and a cytologist provides immediate feedback on the sample. In clinical practice, one to seven biopsies may be performed to obtain the diagnosis. Both solid and cystic lesions can be aspirated; cyst fluid is generally sent for tumor marker analysis in addition to cytology.

FIGURE 5 Tip of a linear array echoendoscope with a stent protruding from the operating channel. Courtesy of Olympus America, Inc.

The overall diagnostic yield for EUS-guided FNA is 77%. The complication rate (significant bleeding) associated with this technique is 1–2%.

INDICATIONS FOR EUS

EUS is most commonly performed to diagnose and stage malignancies of the gastrointestinal (GI) tract. However, indications for this sophisticated procedure continue to expand with improvement in technology. Listed in Table I are the more common indications for EUS.

Staging of Gastrointestinal Malignancies

The gastrointestinal wall is viewed in great detail as a five- layer structure (Fig. 6), with the EUS layers corresponding to the histopathologic layers of the gut wall. The ability to obtain such detailed wall layer imaging has made EUS the most accurate staging modality for several GI malignancies. Cancer survival correlates with the stage of disease, and accurate staging directly guides treatment, be it minimally invasive endoscopic therapy, surgery with/without chemoradiation, or palliation alone. The information obtained by EUS serves as a cornerstone for this decision process.

TABLE I Indications for Endoscopic Ultrasound[a]

Staging of esophageal, gastric, and rectal cancer
Evaluation of abnormalities of the gastrointestinal wall or adjacent structures (submucosal masses, extrinsic compression)
Evaluation of thickened gastric folds
Diagnosis (FNA) and staging of pancreatic cancer
Evaluation of pancreatic abnormalities (suspected masses, cystic lesions including pseudocysts, suspected chronic pancreatitis)
Staging of ampullary neoplasms
Diagnosis and staging of cholangiocarcinoma
Evaluation of suspected choledocholithiasis
Celiac plexus neurolysis for chronic pain due to intraabdominal malignancy or chronic pancreatitis
Staging of non-small-cell lung cancer (subcarinal and posterior mediastinal lymph nodes)
Evaluation/diagnosis of posterior mediastinal masses

[a] Courtesy of the *North Carolina Medical Journal*.

The tumor/node/metastasis (TNM) staging system is universally employed for cancer staging; tumor staging (T stage) for luminal malignancies generally consists of four different stages: T1, tumor extension limited to the mucosa or submucosa; T2, tumor invades the muscularis propria, but does not extend into the adventitia or serosa; T3, tumor invades through the muscularis propria into the adventitia or serosa; and T4, tumor invades adjacent organs or major blood vessels.

Nodal (N) staging is represented by the presence (N1) or absence (N0) of lymph node metastases. EUS criteria have been developed to suggest malignant lymphadenopathy: (1) size greater than 1 cm, (2) round or oval shape, (3) hypoechoic echo pattern, and (4) well-demarcated borders. If all four criteria are met, accuracy rates for predicting lymph node metastasis range from 70 to 80%. CT scanning is capable only of evaluating lymph node size. EUS-guided FNA of suspicious nodes further increases the accuracy to approximately 90%.

Esophageal Cancer

EUS is the most accurate modality to determine the locoregional stage of esophageal cancer. Because it is the only modality that can actually delineate gut wall layers, it can determine local tumor extension (T stage) with an accuracy ranging from 76 to 89%. Node stage accuracy is 72–80% for EUS using the established criteria. In comparison, the accuracy of computed tomography for locoregional staging of esophageal cancer ranges from 50 to 60% (T stage) and 45 to 60% (N stage). Computed tomography still remains invaluable as a staging modality, and should be the first study performed after diagnosis to rule out distant metastases, the presence of which significantly changes management. Because EUS has only limited depth penetration, it frequently cannot determine the presence of distant metastases. Many large centers are now using neoadjuvant chemoradiation for the treatment of locally advanced esophageal cancer. This decision is directly influenced by EUS evaluation.

Gastric Cancer

Surgical resection is the main treatment for gastric cancer. Disease stage still reflects survival, and

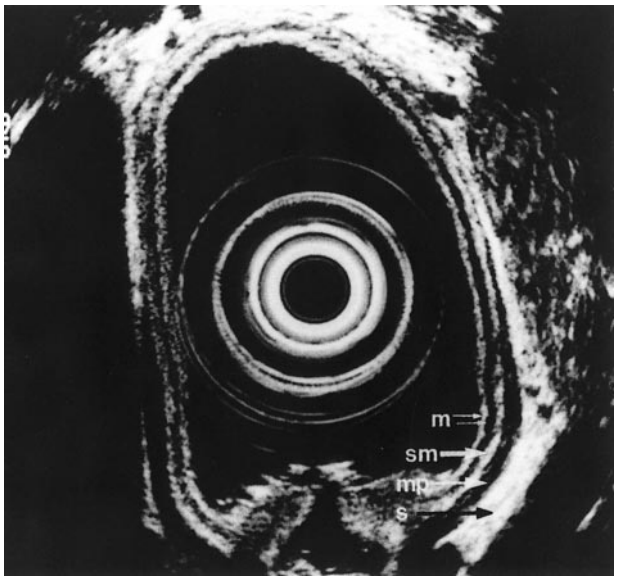


FIGURE 6 Radial image of the normal five-layer structure of the GI tract wall. Layer 1, superficial mucosa (m); layer 2, deep mucosa (m); layer 3, submucosa (sm); layer 4, muscularis propria (mp); and layer 5, serosa/adventitia (s). Reprinted courtesy of the *North Carolina Medical Journal*.

historically, adjuvant chemotherapy has not improved survival in gastric cancer. A new study has suggested a modest survival benefit using chemotherapy in a neoadjuvant fashion. EUS again remains a valuable tool in guiding treatment decisions, specifically in the setting of neoadjuvant protocols. Additionally, EUS can guide endoscopic mucosal resection (EMR) of superficial cancers. The accuracy of EUS-determined T (92%) and N (78%) stages is superior to that of CT scanning (42 and 48%, respectively).

Rectal Cancer

Patients with locally advanced rectal cancer ($\geq$T3, or N1 disease) are best treated with neoadjuvant chemoradiation prior to surgical resection. Given the detailed wall layer imaging achieved, EUS is well suited for the staging of rectal cancer. The accuracy of T and N staging is 84 and 84%, respectively. For superficial tumors (T1) without evidence of nodal disease, EUS can direct transanal excision or endoscopic resection. EUS does not play a role in colon cancer staging. There is little impact on therapy, because treatment of colon cancer is surgical resection with possible adjuvant chemotherapy pending the pathologic stage.

Evaluation of the Pancreaticobiliary System

Originally developed more than 20 years ago to better image the pancreas, EUS has had a tremendous impact on the diagnosis and evaluation of pancreatic pathology. Given the anatomic relationships, placement of the transducer within the duodenum and stomach allows for detailed imaging of the pancreas and bile duct.

Pancreatic Cancer

EUS provides highly accurate staging information in the evaluation of pancreatic cancer. In the absence of distant metastases, EUS is used to assess resectability based on local vascular invasion. EUS has an accuracy rate on the order of 80–95% and 65–85% for T and N staging of pancreatic cancer, respectively. New dual-phase helical CT scanning is also quite accurate for tumor and nodal staging (86 and 77%, respectively). A major advantage of EUS is the ability to accurately stage and diagnose via FNA a pancreatic mass in a single procedure. Accuracy rates for EUS-guided FNA of pancreatic masses range between 85 and 95%. EUS is also more sensitive in detecting small pancreatic tumors $<$2 cm in size, which can be missed by CT.

EUS-guided FNA offers several advantages over percutaneous CT-guided biopsies. These include shorter needle tracks (with less concern for peritoneal seeding), smaller diameter needles, use of real-time color Doppler analysis to avoid blood vessels, and better access to

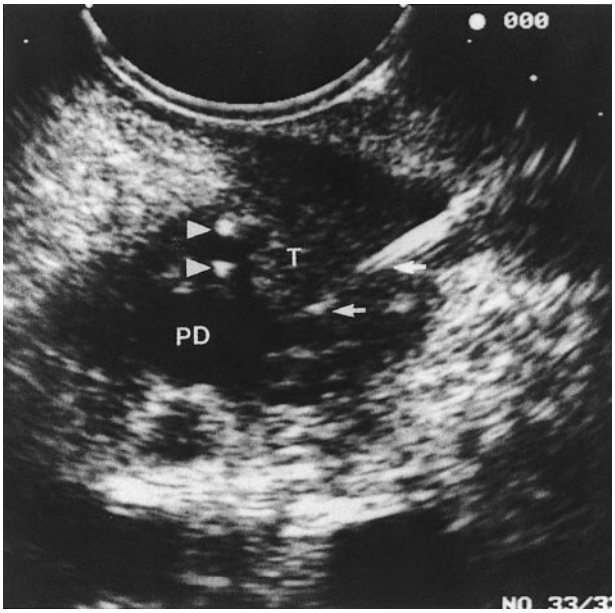


FIGURE 7 Linear image of a small 2-cm hypoechoic tumor (T) in the head of the pancreas. A biliary stent (white arrowheads) is in the common bile duct, above a dilated pancreatic duct (PD). Fine-needle aspiration reveals adenocarcinoma (white arrows, the needle tip). Reprinted courtesy of the *North Carolina Medical Journal*.

smaller lesions (Fig 7). Additionally, small hepatic metastases that may have been missed by CT can be detected by EUS.

Neuroendocrine Tumors

Given its ability to detect small pancreatic lesions, EUS is perhaps the most sensitive modality to localize neuroendocrine tumors of the pancreas. Sensitivity for detection of islet cell tumors ranges between 79 and 93%. Additionally, simultaneous FNA can be performed for definitive diagnosis.

Cystic Lesions of the Pancreas

Pancreatic cysts are rare, but they are being discovered more frequently due to increased utilization of CT. On discovery, anxiety generally mounts due to the possibility of a premalignant cystic neoplasm. Previously, the patient could be offered surgical resection or radiologic surveillance. Now, EUS is very helpful in characterizing pancreatic cystic lesions, because morphologic detail can be assessed and cyst fluid can be aspirated for cytology and tumor markers. Cyst infection is a potential risk, and prophylactic antibiotics are warranted.

EUS evaluation of pancreatic pseudocysts is helpful in guiding endoscopic drainage. The gastric wall can be surveyed and marked for cyst gastrostomy. EUS guidance can lower the risk of transgastric stent placement, because blood vessels can be avoided.

Chronic Pancreatitis

The diagnosis of chronic pancreatitis is obvious in the setting of parenchymal calcifications or intraductal stones as seen on abdominal radiograph, CT, or endoscopic retrograde cholangiopancreatography (ERCP). These findings are also easily seen on EUS. However, there are patients with symptomatology suggestive of chronic pancreatitis, yet imaging modalities are normal or nondiagnostic. This cohort of patients may have early or minimal-change chronic pancreatitis. EUS can accurately assess for subtle parenchymal and ductal changes in these patients. Defined EUS criteria, which have been correlated to excised pathologic specimens, exist for the diagnosis of chronic pancreatitis. Histologic confirmation is the gold standard for the diagnosis of chronic pancreatitis, and new needle systems are being developed that will allow for a core pancreatic biopsy via EUS guidance.

Choledocholithiasis

EUS is an excellent modality to visualize the biliary tract, given the proximity to the duodenal bulb. EUS is superior in the detection of common bile duct stones (choledocholithiasis) as compared to CT and transabdominal US. This is particularly true in patients with small stones and normal-caliber bile ducts. The sensitivity for the detection of bile duct stones is equivalent to that of endoscopic retrograde cholangiopancreatography (95%). Given the risk of pancreatitis associated with ERCP, EUS can be helpful in evaluating patients in whom there is a low suspicion for choledocholithiasis. It is useful in evaluating pregnant patients with suspected bile duct stones, because unnecessary fetal exposure to radiation can be avoided. EUS cannot replace ERCP as the test of choice in patients with a high suspicion of choledocholithiasis, because ERCP can offer therapeutic removal of the stone.

Celiac Plexus Neurolysis

Celiac plexus neurolysis is a therapeutic application of EUS. The celiac plexus cannot actually be visualized by EUS, but its ganglia straddle the celiac trunk. Given the ability to visualize the takeoff of the celiac trunk from the aorta, anesthetic or neurolytic agents can be injected with excellent localization and, therefore, minimal morbidity (Fig. 8). Injection of absolute alcohol has proved effect in relieving the pain associated with pancreatic cancer. Temporary blocks using anesthetics and steroids have been performed to decrease pain caused by chronic pancreatitis. Results have been less promising,

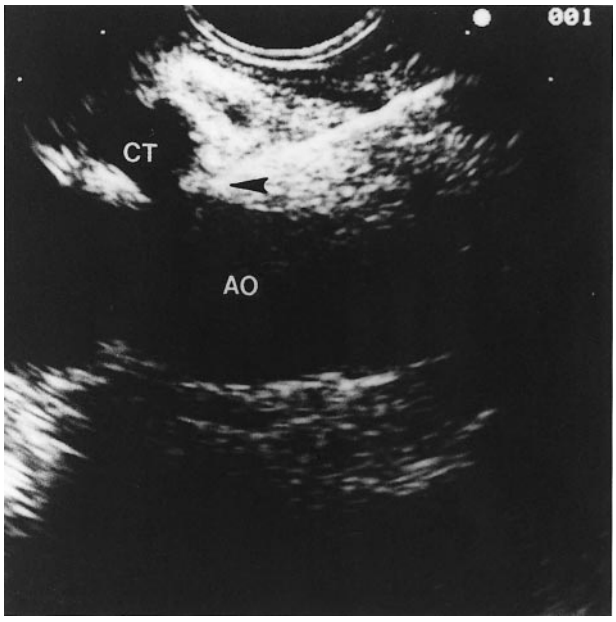


FIGURE 8 Linear image representing celiac plexus neurolysis in a patient with intractable pain from pancreatic cancer. The needle tip (arrowhead) is seen at the takeoff of the celiac trunk (CT) from the aorta (AO). Reprinted courtesy of the *North Carolina Medical Journal*.

but approximately 50% of patients do experience transient improvement in their pain.

Intramural Tumors and Extrinsic Compression

EUS can accurately distinguish between intramural masses and extrinsic compression in the GI tract. Intramural masses such as cysts, lipomas, and stromal cell tumors have defining echo characteristics and arise from characteristic gut wall layers. Foregut duplication cysts appear as anechoic submucosal cysts, whereas lipomas image as hyperechoic submucosal masses. These are both benign entities. Gastrointestinal stromal cell tumors (GISTs) are hypoechoic and arise from the muscularis propria (Fig. 9). Defined EUS criteria that raise suspicion for malignancy can guide recommendations for surgical resection. Extrinsic compression of the GI lumen can result from normal organs (liver, spleen), adjacent tumors, or cysts.

Lung Cancer Staging

The staging of non-small-cell lung cancer is an indication for which EUS and FNA have proved very useful. EUS is an excellent modality for imaging the posterior mediastinum, therefore enlarged lymph nodes can easily be sampled. The accuracy of diagnosing malignant mediastinal lymph node involvement for CT- and EUS-guided FNA ranges from 49 to 74% and 96 to 100%, respectively. EUS is able preferentially to access the aortopulmonary (AP) window and subcarinal space (American Thoracic Society stations 5 and 7). These are the two most difficult stations to access with mediastinoscopy. EUS-guided FNA is most helpful in the preoperative evaluation of patients with suspicious contralateral mediastinal or bulky subcarinal lymph nodes, because malignant involvement drastically alters management (Fig. 10). Some centers now consider EUS-guided FNA as the technique of choice for sampling the subcarinal space and AP window. Routinely, the left adrenal gland is sought when staging lung cancer, and adrenal metastases have been diagnosed via EUS-guided FNA.

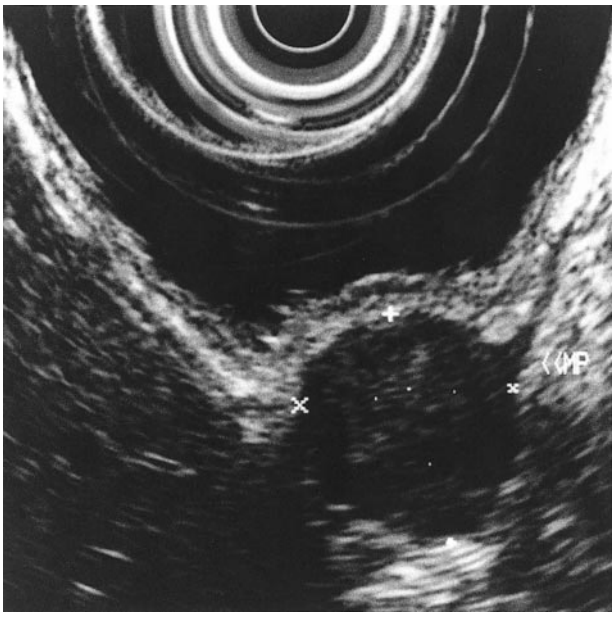


FIGURE 9 Radial image of a 15- × 14-mm round, well-demarcated hypoechoic mass arising from the muscularis propria (MP). This finding is consistent with a gastrointestinal stromal tumor. Reprinted courtesy of the *North Carolina Medical Journal*.

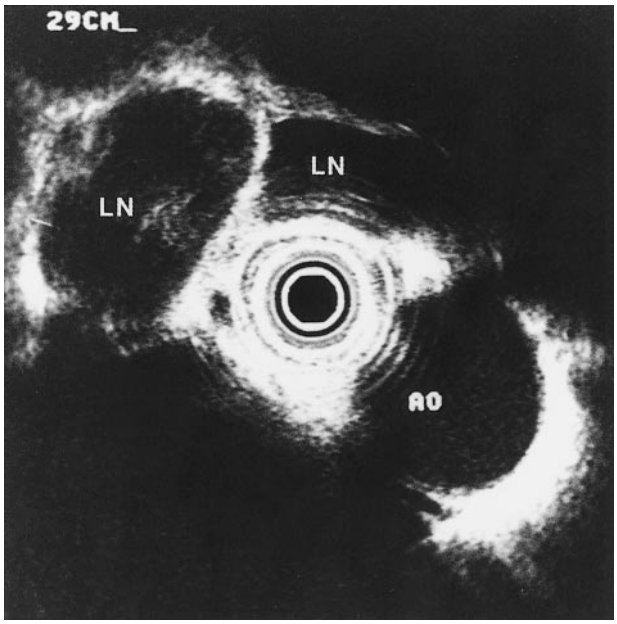


FIGURE 10 Radial image of two bulky subcarinal lymph nodes (LN) in a patient with non-small-cell lung cancer (AO, aorta). Reprinted courtesy of the *North Carolina Medical Journal*.

Evaluation of Mediastinal Masses

Mediastinal masses can be evaluated with EUS, given the fact that the esophagus traverses the mediastinum. Posterior mediastinal masses are most easily seen, because tracheal air artifact can obscure imaging of superior and anterior mediastinal tumors. Posterior mediastinal masses are amenable to EUS-guided FNA, with an accuracy rate for a cytologic diagnosis of 83–89%. As an outpatient procedure, this takes 20 to 30 minutes and requires conscious sedation only. It is a cost-effective and minimally invasive alternative to mediastinoscopy for posterior mediastinal mass evaluation.

LIMITATIONS OF EUS

Endoscopic ultrasound is not without its limitations. Understaging of malignancies can occur due to microscopic invasion that is not apparent on EUS examination. Perhaps the biggest drawback of EUS is the inability, based on imaging alone, to distinguish between benign and malignant tissues, therefore overstaging due to peritumor inflammation can occur.

EUS and EUS-guided FNA are technically challenging procedures that are operator dependent. Advanced training in EUS is a requirement for proficiency in performing this procedure. Additionally, the equipment is quite costly, therefore, at this juncture, EUS tends to be limited to academic medical centers.

SUMMARY

EUS has rapidly emerged as a valuable asset to the practice of gastroenterology. EUS accurately stages GI malignancies, evaluates intramural tumors, and diagnoses pancreatic cancer via FNA. Essentially, any mass lesion in close proximity to the reach of the echoendoscope can be biopsied for diagnostic purposes. Interventional EUS currently incorporates celiac plexus neurolysis and cyst gastrostomy, but this field will likely expand in the future to include EUS-guided radiofrequency ablation and injection therapy (chemotherapy, attenuated viruses) for the treatment of gastrointestinal malignancies.

See Also the Following Articles

Cancer, Overview • Endoscopy, Complications of • Ultrasonography • Upper Gastrointestinal Endoscopy

Further Reading

Bhutani, M. S., Hawes, R. H., and Hoffman, B. J. (1997). A comparison of the accuracy of echo features during endoscopic ultrasound (EUS) and EUS-guided fine-needle aspiration for diagnosis of malignant lymph node invasion. *Gastrointest. Endosc.* **45**, 474–479.

Brugge, W. R. (1998). Endoscopic ultrasonography: The current status. *Gastroenterology* **115**, 1577–1583.

Caletti, F., and Fusaroli, P. (2001). Endoscopic ultrasonography. *Endoscopy* **33**, 158–166.

Canto, M. I., Chak, A., Stellato, T., *et al.* (1998). Endoscopic ultrasonography versus cholangiography for the diagnosis of choledocholithiasis. *Gastrointest. Endosc.* **47**, 439–448.

Chak, A., Canto, M. I., Rosch, T., *et al.* (1997). Endoscopic differentiation of benign and malignant stromal cell tumors. *Gastrointest. Endosc.* **45**, 468–473.

Gress, F., and Bhattacharya, I. (2001). "Endoscopic Ultrasonography." Blackwell Science, Malden, MA.

Gress, F. G., Hawes, R. H., Savides, T. J., *et al.* (1997). Endoscopic ultrasound-guided fine needle aspiration biopsy using linear array and radial scanning endosonography. *Gastrointest. Endosc.* **45**, 243–250.

Gress, F. G., Savides, T. J., Sandler, A., *et al.* (1997). Endoscopic ultrasonography, fine-needle aspiration biopsy guided by endoscopic ultrasonography, and computed tomography in the preoperative staging on non-small cell lung cancer: A comparison study. *Ann. Intern. Med.* **127**, 604–612.

Mallery, S., and Van Dam, J. (2001). Current status of diagnostic and therapeutic endoscopic ultrasonography. *Radiol. Clin. North Am.* **39**, 449–463.

McGrath, K. M., and Jowell, P. S. (2000). Endoscopic ultrasound: State of the art. *North Carolina Med. J.* **61**, 408–412.

Rosch, T. (1995). Staging of pancreatic cancer: Analysis of literature results. *Gastrointest. Endosc. Clin. North Am.* **5**, 735–739.

Rosch, T., and Classen, M. (eds.). (1992). *In* "Gastroenterologic Endosonography," pp. 1 and 6. Thieme Med. Publ., New York.

Wiersema, M. J., and Wiersema, L. M. (1996). Endosonography-guided celiac plexus neurolysis. *Gastrointest. Endosc.* **44**, 656–662.

Endoscopy, Complications of

MICHAEL F. BYRNE, ROBERT M. MITCHELL, AND JOHN BAILLIE
Duke University Medical Center

colonoscopy A flexible video (or fiber-optic) endoscopy that allows direct visualization of the lining of the rectum and colon.

endoscopic retrograde cholangiopancreatography (ERCP) An endoscopic procedure that permits detailed radiological imaging of the pancreaticobiliary ductal system. One can also perform interventional procedures during ERCP such as removal of stones, placement of stents, and tissue sampling.

endoscopic ultrasound An endoscopic technique that utilizes an endoscope with an ultrasound probe at its tip. This allows the physician to see beyond the endoscopic image provided by standard endoscopes. It is particularly useful for cancer staging.

esophagogastroduodenoscopy A flexible video (or fiber-optic) endoscopy that allows direct visualization of the lining of the esophagus, stomach, and proximal duodenum.

pancreatitis An inflammatory process in which pancreatic enzymes autodigest the gland. It is a well-described complication of endoscopic retrograde cholangiopancreatography.

percutaneous endoscopic gastrostomy (PEG) An endoscopic procedure during which a feeding tube is placed between the stomach and the anterior abdominal wall to allow for direct feeding into the stomach. The most common indications for PEG tube placement are neurologic conditions associated with impaired swallowing and neoplasms of the oropharynx, larynx, and esophagus.

polypectomy A procedure used to remove a polyp(s). If polyps are detected during endoscopy, these will be removed by passing a snare or biopsy forceps through the endoscope and severing the attachment of the polyp from the intestinal wall, sometimes with the use of cautery.

There is a risk of complication with any endoscopic procedure, many of which are inherent to the techniques used. It is critical that endoscopists be aware of the potential complications with each endoscopic procedure performed, that they know how to limit the chance of such complications occurring, and that they recognize and treat complications when they arise. In this article, some general complications applicable to all gastrointestinal endoscopy are described. This is followed by a more detailed overview of the risks associated with diagnostic and therapeutic upper gastrointestinal endoscopy, colonoscopy, endoscopic retrograde cholangiopancreatography, and endoscopic ultrasound.

GENERAL COMPLICATIONS

Complications of Conscious Sedation

Gastrointestinal (GI) endoscopy is performed under conscious sedation in the United States and many parts of Europe. The most common practice is to use intravenous doses of benzodiazepines, usually midazolam, and opiates such as meperidine (demerol) or fentanyl. However, morbidity rates of 1:200 to 1:2000, and occasional mortality, usually due to cardiorespiratory complications, have been reported using these agents. Thus, the American Society of Anesthesiologists (ASA) and the American Society of Gastrointestinal Endoscopy (ASGE) devised more formal practice guidelines for conscious sedation that included the use of supplemental oxygen and revised dosing guidelines recommending titration of sedative medications rather than bolus doses.

Midazolam is a good agent to use for endoscopy as it has both anxiolytic and amnestic effects, but its use also results in residual sedative effects and anterograde amnesia. Propofol (2,6-diisopropyl phenol), an intravenous anesthetic agent often used with other agents for delivery of general anesthesia, can be used in lower doses to induce conscious sedation. For this latter indication, propofol has some advantages over benzodiazepines and opiates such as shorter half-life and quicker hepatic clearance, resulting in quicker recovery from sedation and less hangover effects. However, there is significant debate regarding the safety of propofol use by personnel who are not anesthesiologists. For example, significant respiratory depression can occur during propofol-induced conscious sedation and one can inhibit the gag and cough reflexes at higher doses. It is also important to be aware that, unlike the benzodiazepines, propofol has no specific antagonist that can be employed to reverse its effects.

In general, whichever conscious sedation protocol is used, one needs to be aware of the risk of respiratory depression, aspiration, hypotension, paradoxical

 Copyright 2004, Elsevier (USA). All rights reserved.

agitation, and anaphylaxis. Special care should be taken in the very young and elderly.

Cardiopulmonary Complications during Endoscopy

Most of the cardiopulmonary complications during endoscopy are medication related, as described above. However, some of these complications can be attributed to vagal- or sympathetic-mediated changes associated with endoscopy. Simple arrhythmias, such as sinus tachycardia, and transient hypoxia are common during endoscopy. Current guidelines recommend that supplemental oxygen and pulse oximetry are used throughout endoscopy. It has also been suggested by some that capnography be used to detect early respiratory depression, especially if propofol is used, but the place of capnography in endoscopic practice remains to be determined. Cardiac ischemia is rare after endoscopy and recent myocardial infarction is not a contraindication to endoscopy. However, one should consider each such case individually. Tachydysrhythmias and heart block should be treated before endoscopy, whenever possible.

Complications of Topical Oropharyngeal Medication

Topical oropharyngeal anesthesia is used routinely by many endoscopists for upper GI endoscopy. Topical lignocaine and benzocaine are used most commonly and usually are safe. However, topical anesthesia can increase the risk of aspiration as the gag reflex is suppressed. One needs to be aware of this, particularly if there are significant amounts of blood or food in the stomach. In addition, there have been a limited number of reports of methemoglobinemia induced by topical anesthesia, a condition treated by use of supplemental oxygen and intravenous methylene blue.

Consent

Commonly occurring complications of endoscopy should be discussed with the patient prior to endoscopy when gaining informed consent.

Infection—General Comments

Infectious complications of endoscopy can result from the procedure itself or from contaminated equipment. Risk of infection is higher with particular endoscopic techniques, particularly therapeutic procedures, and this is discussed in the relevant sections below. Transient bacteremia may occur after endoscopic procedures but the risk of infective endocarditis or other infectious complications is low. ASGE guidelines for the use of antibiotic prophylaxis for endoscopy include high-risk patients (such as those with prosthetic heart valves, synthetic vascular grafts, history of endocarditis) undergoing high-risk procedures [such as sclerotherapy, stricture dilation, or endoscopic retrograde cholangiopancreatography (ERCP) in the setting of biliary obstruction]. It also appears that transmission of infection from the endoscope to the patient is rare. This is a concern, especially in the era of human immunodeficiency virus, hepatitis C, and other viruses. However, although iatrogenic infection transmitted by endoscope has been shown for several bacteria, including *Pseudomonas, Salmonella,* and *Helicobacter pylori*, there is little convincing evidence of viral transmission. If published guidelines on cleaning and disinfection of endoscopic equipment are followed, the risk of infection is low.

UPPER GASTROINTESTINAL ENDOSCOPY

Diagnostic upper GI endoscopy is very safe, with an overall complication rate of less than 0.1%. However, although rare, aspiration and perforation during upper GI endoscopy can be fatal. Most of the complications associated with upper GI endoscopy result from intervention.

Perforation

This is a rare but potentially serious complication of diagnostic upper GI endoscopy. Perforation rates are as low as 1 in 5000. Perforation is particularly concerning in the debilitated and elderly. It is more common in the presence of certain predisposing factors such as Zencker's diverticulum, esophageal stricture and malignancy, and anterior cervical osteophytes. Perforation of the esophagus has a particularly high mortality, at approximately 25% overall, with mortality higher for perforation of the intrathoracic esophagus. Perforation in the cervical esophagus is more common with rigid esophagoscopy. Upper gastrointestinal perforations present with pain and one or more of the following symptoms: pleuritic chest pain, leukocytosis, fever, and surgical emphysema. Any perforation related to endoscopy is best managed in cooperation with surgeons.

Bleeding

Significant bleeding is a rare complication of diagnostic upper GI endoscopy, even in the setting of

coagulopathy or thrombocytopenia. However, significant coagulopathy and severe thrombocytopenia should be corrected before biopsies are taken. Mallory–Weiss tears occur rarely (<0.01%) during routine upper endoscopy and are usually related to struggling and retching during the procedure.

Complications of Hemostatic Procedures

Injection Therapy

Epinephrine, at a dilution of 1:10,000, is the most commonly used injection agent for the treatment of bleeding peptic ulcers. This is a safe practice overall, with no reports of perforation but tissue necrosis and ulceration at the site of injection may occur. Epinephrine injection directly into a blood vessel can cause transient tachycardia, which usually resolves within 30–60 s. The risk of complications is higher if sclerosants are used but this practice is becoming less common.

Thermal Hemostasis

Hemostasis of bleeding peptic lesions can be achieved using thermal methods such as bipolar or multipolar electrocoagulation or the heater probe. These are fairly safe hemostatic methods as the low-powered thermal energy causes tissue desiccation rather than vaporization. For example, the perforation rate for multipolar electrocoagulation is reported to be between 0 and 2%. Bleeding may also occur during delivery of therapy, but this is rarely significant. Another problem with these probes is tissue sticking as coagulum builds up. Repeated cleaning by removal of the probe or with the water jet may be necessary; this probably lessens the chance of posttreatment bleeding as the probe does not adhere to the treatment site. Repeat thermocoagulation within 24–48 h may be associated with an increased perforation rate and is best avoided.

Laser Hemostasis

Delivery of heat to a site of bleeding or potential bleeding by the Nd:YAG laser causes tissue erosion. Theoretically, this should result in a higher rate of perforation than for the other thermal methods described above, but this does not appear to translate into significant risk in clinical practice.

Treatment of Varices

The two main techniques for treatment of esophageal varices are endoscopic sclerotherapy and banding. Sclerotherapy has an overall significant complication rate between 20 and 40%, with a mortality of up to 2%. Risks of sclerotherapy include esophageal perforation, ulceration, stricture, and motility disturbances. Less common complications include spinal cord paralysis, pneumothorax, bacteremia, and pleural effusion. If sclerotherapy is performed, a second session should not be performed earlier than 7 days after initial treatment, as this appears to lessen the risk of ulceration. The recent literature favors the use of band ligation on grounds of clinical efficacy and safety. The overall complication rate for banding is approximately half that of sclerotherapy. In particular, the rate of stricture formation and ulceration is much lower with banding.

Complications of Dilation Procedures

Esophageal dilation is relatively safe but can be complicated by perforation. Overall, the risk of perforation is less than 1%, but the rate depends upon whether dilation is performed for benign strictures, malignant strictures, or achalasia. For dilation of benign esophageal strictures, there appears to be no difference in the rate of perforation whether balloon dilators or wire-guided polyvinyl dilators (such as Savary-Gilliard) are used; blind use of Maloney dilators does increase the risk of perforation and this practice should be discouraged. For achalasia, the risk of perforation after pneumatic balloon dilation is approximately 5%, although there is evidence that a graded-dilation technique may limit this risk. Dilation of malignant strictures results in perforation in perhaps up to 10% of cases.

For patients with pyloric outlet obstruction caused by benign pyloric stenosis, balloon dilation carries a perforation rate between 0 and 6% and is comparable in safety to wire-guided push-type dilators. However, a high stenosis recurrence rate suggests that this procedure should be reserved for patients of high surgical risk.

Esophageal Stenting

This procedure is performed for palliation of malignant strictures. Originally, use of rigid Silastic plastic stents resulted in a complication rate of approximately 20%, with risk of bleeding, perforation, tumor ingrowth, and stent migration. These risks are significantly less with the more recent practice of using expandable metal stents.

Treatment of Esophageal Malignancies

There are several endoscopic options for debulking esophageal tumors. Injection with sclerosants such as 95% ethanol using a sclerotherapy needle is effective but there is a risk of extensive tissue necrosis, predisposing to perforation and mediastinitis. This is more

likely with injections over 10 ml in volume. All of the thermal energy delivery methods described above for treatment of hemostasis (heater probe, bipolar/multipolar coagulation, and laser therapy) have been used for tumor ablation; it is reported that perforation and fistula development may occur in up to 10% of cases. Photodynamic therapy (PDT) is a technique that is being used in both tumor destruction and ablation of Barrett's mucosa. In theory, targeted tumor therapy is achieved using PDT as malignant tissue preferentially concentrates the photosensitizer given prior to treatment. However, perforation, bleeding, fistula, and stricture development have all been reported with this technique. Patients should be warned about the risk of cutaneous photosensitivity after ingesting photosensitizers.

Foreign Body Retrieval

Food boluses that become impacted in the esophagus can often be gently pushed into the stomach using the tip of the endoscope, but sometimes they must be removed. A particular concern is loss of the bolus in the region of the larynx. Use of an overtube limits the risk of aspiration, but carries a risk of perforation and bleeding. Papain-based enzyme preparations to tenderize meat boluses can result in severe pulmonary complications and should never be used. Sharp objects should be removed using an overtube or a sleeve over the end of the endoscope. Irregular, hard, or sharp foreign bodies (e.g., dental plate) should be removed by rigid esophagoscopy, usually under general anesthesia, otherwise there is a significant risk of perforation and bleeding on foreign body removal. One should also resist the temptation to attempt endoscopic removal of condoms or plastic bags containing narcotics, as these are easily ruptured with disastrous consequences. Surgical removal is the only safe option.

Percutaneous Endoscopic Gastrostomy

Percutaneous endoscopic gastrostomy (PEG) is a technique used increasingly for the delivery of nutritional support to a range of patients. It is an excellent feeding option in many neurological conditions and also allows bowel decompression proximal to any malignant intestinal obstruction. However, major complications are reported in 2.7% of cases and minor complications in up to 7% of cases. The mortality rate for this procedure is approximately 0.7%.

The major complications of PEG tube placement include perforation, aspiration, hemorrhage, peritonitis, and gastrocolic fistula. Aspiration is the most commonly reported PEG complication, probably due to the underlying neurological defect with poor or absent oropharyngeal reflexes, delayed gastric emptying, and reflux of feeds. Taking certain feeding precautions, such as elevation of the head during feeding, will limit the risk of aspiration. Peritonitis can occur if the gastrostomy tube becomes dislodged, if there is an unrecognized perforation, or if there is leakage of stomach contents into the peritoneal cavity around the stoma site. A through-the-tube water-soluble contrast study will aid in the diagnosis and management of PEG tube leaks. Potential infectious complications of PEG tube placement include wound infection and necrotizing fasciitis. Use of prophylactic antibiotics significantly reduces the risk of stomal infections. Necrotizing fasciitis is a concern in patients with diabetes mellitus, alcoholism, malnutrition, and those who are immunosuppressed. Occasionally, one may puncture the colon during PEG tube placement, if there is a loop of colon anterior to the stomach, resulting in a gastrocolocutaneous fistula. Migration of the feeding tube and impaction in the abdominal wall ("buried bumper syndrome") can occur if there is significant traction on the internal bumper, resulting in mucosal ulceration and eventual erosion through the mucosa. Once the fistulous tract is mature, loosening of the external bumper will help prevent this complication. There have been reports of development of metastases at the PEG insertion site, raising concern that placement of a PEG tube using the oropharyngeal route in patients with oropharyngeal cancers may facilitate metastasis. A direct introducer method may be preferable in these cases, if the primary tumor cannot be removed.

ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY

Endoscopic retrograde cholangiopancreatography carries many of the risks associated with all endoscopic techniques as well as having complications particular to this technique. Risks specific to instrumentation of the pancreaticobiliary system include pancreatitis, cholangitis, bleeding, retroduodenal perforation, and stent dysfunction.

Pancreatitis

The serum amylase level is elevated in up to 50% or more of asymptomatic patients after ERCP, but the clinical syndrome of post-ERCP pancreatitis occurs in only 3–10% of cases on average. In the majority of cases, post-ERCP pancreatitis is mild and self-limiting but it can be severe (Fig. 1) and occasionally fatal. Factors that appear to increase the risk of pancreatitis include

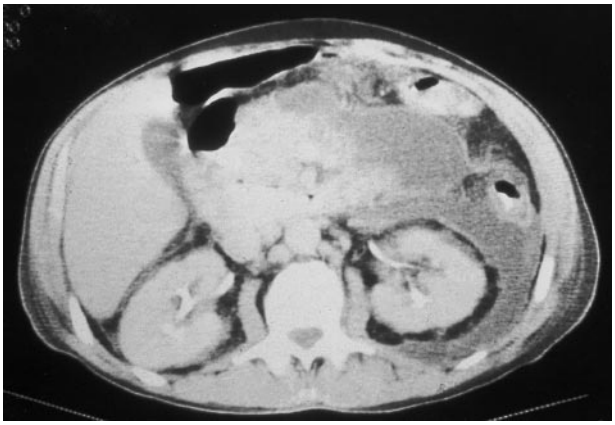


FIGURE 1 CT scan of the abdomen showing severe post-ERCP pancreatitis.

overfilling or acinarization of the pancreatic parenchyma, multiple cannulations of the pancreatic duct, endoscopic sphincterotomy, and manometry (risk of pancreatitis as high as 20%). Measures that show promise in preventing post-ERCP pancreatitis are still under investigation but there are some encouraging early data regarding the prophylactic use of the protease inhibitor, gabexate (not currently available in the United States), intravenous somatostatin (or its synthetic analogue, octreotide), and interleukin-10, and the placement of pancreatic stents.

Bleeding

Sphincterotomy can cause significant bleeding (Fig. 2). The incidence of complications after sphincterotomy ranges from 5 to 10% (9.8% in the Freeman multicenter study). In particular, uncontrolled (or "zipper") cuts predispose to bleeding. Fortunately, bleeding from a sphincterotomy site is more often venous than arterial. It is rare for a patient to require a blood transfusion after a sphincterotomy bleed, possibly because of the widespread use of blended rather than pure-cut current. If bleeding occurs when the cannulotome is still in the duct, one can often achieve hemostasis with a brief application of coagulation current. If oozing continues, flushing the site with dilute epinephrine solution may arrest bleeding, but injection of epinephrine into the sphincterotomy site may be required. Using an inflated stone-extraction balloon to tamponade the site may also be effective. If bleeding persists and it proves impossible to deliver local therapy, celiac angiography with subsequent embolization of the offending vessel should be performed but, in some cases, surgery is necessary. Delayed bleeding is a rare but potentially serious complication of sphincterotomy, particularly as many patients are discharged home within hours of ERCP. Rarely, ERCP may provoke hemobilia from trauma to friable hilar tumors or a guide-wire penetrating the bile duct wall, creating a biliovenous fistula.

Perforation

Perforation can occur with any endoscopic procedure as described earlier. During ERCP, perforation is associated with sphincterotomy. However, the risk is still low, at less than 1% of sphincterotomies. In this setting, perforations are usually retroduodenal (Fig. 3) but can occasionally communicate with the peritoneal cavity. The risk of perforation is greater the larger the sphincterotomy (Fig. 4) and the more the cut deviates from the axis of the bile duct. The risk of perforation may be even greater if a precut papillotomy is performed.

Retroduodenal perforation may be recognized during the procedure as a diffuse leak of contrast behind the duodenum, but it is not uncommon to miss this at the time of ERCP as they are often small. The radiologic appearances are often subtle. Patients may take several hours or even days to develop abdominal pain, elevated white count, and fever, suggestive of occult sepsis. One should be especially suspicious of perforation if a patient, behaving as if he or she has post-ERCP pancreatitis, has a normal serum amylase. Most retroduodenal

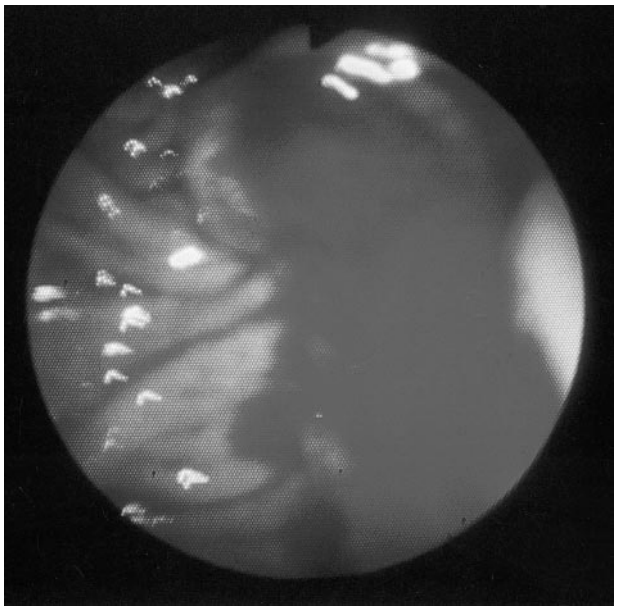


FIGURE 2 A case of postsphincterotomy bleeding. Reprinted from Baillie, J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK, with permission.

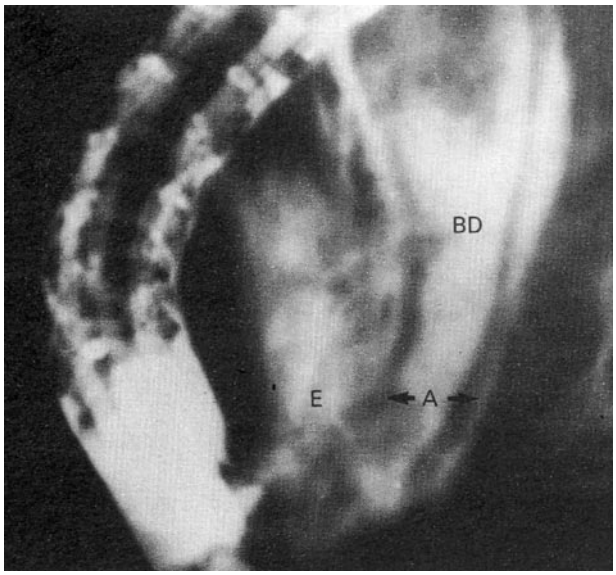


FIGURE 3 A retroperitoneal perforation after endoscopic sphincterotomy. A, air outside the bile duct wall (free air); E, retroduodenal extravasation of contrast; BD, common bile duct containing contrast. Reprinted from Baillie, J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK, with permission.

perforations can be managed conservatively, although percutaneous or surgical drainage is required if a retroperitoneal abscess develops.

Sepsis

Cholangitis is a particular concern if ERCP is performed in patients with biliary obstruction. Sepsis after ERCP can be fatal. Patients with biliary obstruction should be given broad-spectrum antibiotics prior to instrumentation. The risk of significant post-ERCP sepsis can be decreased by avoiding injection of large volumes of contrast into an obstructed bile duct. Aspirating bile before injecting contrast should be encouraged in all cases. It is critical to relieve biliary obstruction if contrast has been injected, whether that is by removal of stones, stenting in cases of tumor or retained stones, or placement of a nasobiliary drain. If none of these measures is successful, percutaneous or surgical drainage should be performed as soon as practically possible, and certainly within 48 h, or earlier if possible. The patient should be kept on intravenous antibiotics until adequate biliary drainage is achieved.

Pathogenic organisms may be introduced by the unsuspected use of contaminated duodenoscopes and catheters. One should be alerted to this possibility of iatrogenic infection when *Serratia* or *Pseudomonas* species are cultured from bile or blood after ERCP. Thorough disinfection of all instruments used at ERCP and a regular review of infection control practices is indicated. A series of cases involving the same organism in particular raises concern regarding the possibility of endoscope contamination.

Occasionally, patients may develop acute cholecystitis after ERCP. This may occur within the first 10 days or so, but more commonly it is delayed months or even years. Inability of the gallbladder to contract and clear bacteria that reflux into the biliary system after sphincterotomy or stent placement is a possible cause. Standard management of acute cholecystitis should be employed.

Stenting

Polyethylene stents placed during ERCP commonly occlude or kink, causing stent failure. Stent occlusion is felt to be due to clogging with sludge and bacterial infection may play a part in this. In addition, plastic stents occasionally migrate from their original position. Complications such as these result in the need for repeated procedures. Placement of self-expanding metal stents appears to result in fewer stent failures but careful patient selection is critical. Metal stents should only be used to palliate malignant strictures. They are

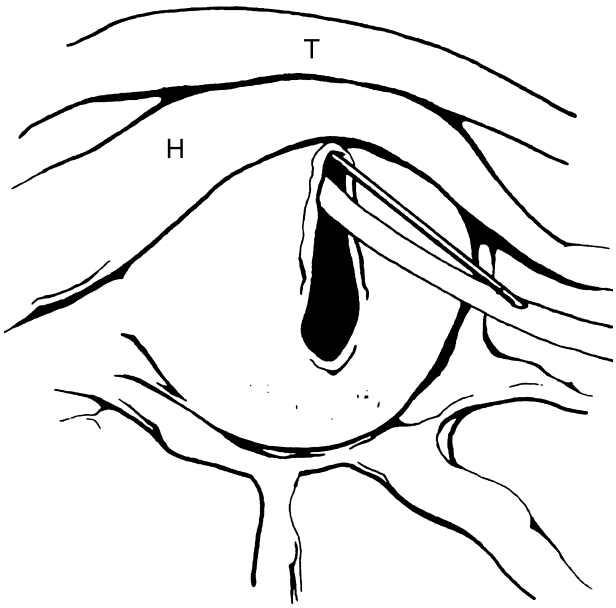


FIGURE 4 Papillotomy in progress. The cut may proceed across the hooding fold (H) but the transverse fold (T) is often the upper limit for a safe papillotomy. Reprinted from Baillie, J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK, with permission.

cost-effective only if the patient's projected survival is greater than 2 months.

COLONOSCOPY

Overall, colonoscopy carries a complication rate of approximately 0.1%, with a mortality rate between 0.008 and 0.02%. The risk of a complication is increased in the setting of interventional procedures such as polypectomy. The main risks of colonoscopy are bleeding and perforation.

Perforation

This is the most feared of colonoscopic complications. It occurs in approximately 0.05% of examinations. Perforation can be either a mechanical or a pneumatic injury. Significant forces are often exerted on the bowel wall by a colonoscope and occasionally these forces can result in a transmural tear. Advancing the colonoscope forcibly through a loop can overcome the compliance of the colon (Fig. 5). During diagnostic procedures, this most commonly occurs in the rectosigmoid area. The practice of "sliding by" in the sigmoid colon has also been reported to increase the risk of perforation. Mechanical perforation can also occur with use of biopsy forceps (particularly in the cecum) and polypectomy or dilation (see below). Pneumatic perforation is less common than mechanical perforation but can occur (usually on the anti-mesenteric aspect of the colon), especially with excessive insufflation of the colon proximal to a stricture (Fig. 6).

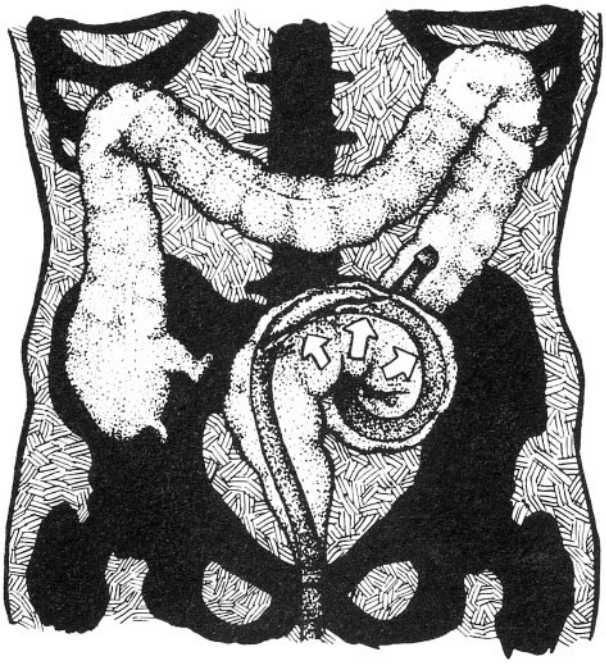


FIGURE 5 Linear tear (perforation) of the sigmoid colon caused by excessive force used to advance through a loop. Reprinted from Baillie, J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK, with permission.

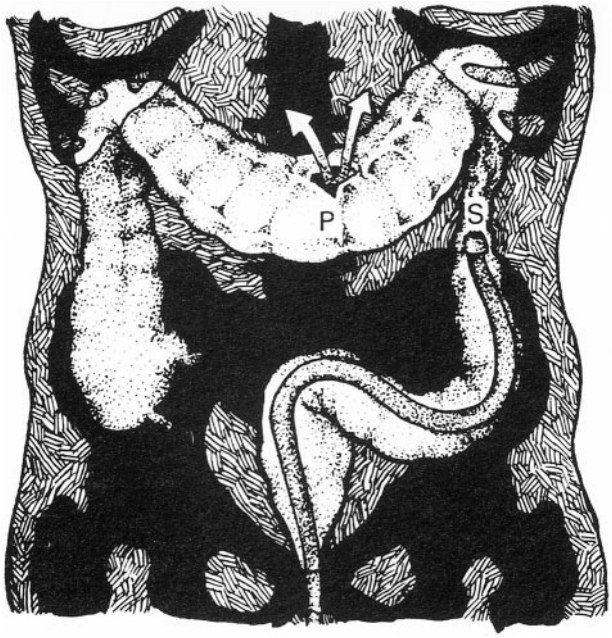


FIGURE 6 Pneumatic perforation of the colon. The site of perforation (P) is proximal to a stricture (S) that cannot be negotiated by the colonoscope. Reprinted from Baillie, J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK.

Bleeding

Hemorrhage is the most common serious complication of colonoscopy. Hemorrhage occurs in approximately 0.02% of diagnostic colonoscopies, with this figure rising to 1.6% if polypectomy is performed. Bleeding may be immediate or delayed. Immediate bleeding after biopsy or polypectomy usually ceases spontaneously but occasionally, especially after polypectomy has been performed with suboptimal coagulation of the polyp stalk, local measures to arrest bleeding are necessary. These measures include injection of epinephrine, electrocoagulation, or placement of endoscopic clips. More rarely, angiography with embolization or surgery is required to achieve hemostasis. Delayed bleeding (up to 1 month after the procedure) occurs either after sloughing of the eschar formed by coagulation or when the initial mucosal "injury" erodes over time into a deeper blood vessel. Reversal of anticoagulation or correction of a known coagulopathy prior to the procedure will

obviously decrease the incidence of significant bleeding, as will discontinuation of aspirin (ASA) and NSAIDs that impair platelet function. Patients should be advised to avoid ASA and NSAIDs for 7–10 days after colonoscopic polypectomy.

Complications of Bowel Preparation

Poor bowel preparation often leads to inability to reach the cecum or, even if the cecum is reached, poor visualization of the mucosa. Thus, pathology may be missed, repeat examinations may be scheduled at earlier intervals than otherwise planned, and the procedure is more difficult and often more time-consuming to perform. In addition, poor bowel preparation can increase the risk of significant complications. The endoscope may inadvertently be introduced into diverticula obscured by feces. Perforation resulting from this or from other maneuvers has graver consequences in the setting of inadequate bowel preparation, as there is increased risk of peritoneal contamination. A rare complication that should never occur these days is combustion of bowel gases when electrocautery is used in the setting of inadequate bowel preparation (i.e., residual feces).

The two predominant bowel cleansing regimens currently employed involve the use of polyethylene glycol-electrolyte lavage solutions or oral sodium phosphate. Sodium phosphate is the more easily completed preparation and may be preferable in patients without comorbidities such as renal insufficiency or congestive cardiac failure that preclude the safe use of sodium phosphate.

Postpolypectomy Syndromes

Postpolypectomy coagulation syndrome (otherwise known as transmural burn or serositis) presents with abdominal pain, fever, and leukocytosis but no evidence of free perforation. It is more common after removal of sessile rather than pedunculated polyps. It is a reaction to thermal damage to the colonic wall deep to the submucosa; it usually settles spontaneously but keeping the patient nil by mouth and using broad-spectrum antibiotics are recommended measures.

Microperforation of the colon is also recognized as a complication of polypectomy. This is essentially an intermediate between the deep thermal damage as described above and "free" perforation in which there is escape of bowel contents and frank peritonitis. Most cases of microperforation do well with conservative management. If the patient does not improve significantly within 24 h, the possibility of more extensive perforation and the need for surgery should be considered.

Tattooing

There have been some concerns about the safety of injection of India ink to label colonic lesions or polypectomy sites, as it appears to promote histological inflammation. There have been a few reports of abscesses and perforation but, overall, it appears that as long as small volumes of sterile ink are used, colonic tattooing is relatively safe.

ENDOSCOPIC ULTRASOUND

Endoscopic ultrasound (EUS) carries the same general risks of all endoscopic procedures such as perforation, bleeding, and sedation reactions. One procedure increasingly used during EUS is that of EUS-guided fine-needle aspiration (FNA). EUS–FNA is usually performed using linear-array echoendoscopes as the whole length of the needle can be seen in real time. EUS-guided FNA using linear array has a complication rate less than 1%. FNA is possible using radial array systems but is technically more difficult, carries a higher complication rate for bleeding and perforation, and is not recommended. A recent study described a low complication rate (0.5%) for FNA of solid lesions but a higher rate of 14% for FNA of cystic lesions, predominantly due to infectious or hemorrhagic events after puncturing cystic lesions. Prophylactic antibiotics should be given for FNA of cystic lesions. The theoretical risk of tumor seeding has not yet been reported.

Another problem with EUS and staging of esophageal cancers arises with advanced and stenotic lesions. It may prove impossible to advance the EUS probe through the stenosis and in this situation some authors have suggested dilating the stricture to allow for EUS evaluation. Although an increased incidence of perforation has been reported, study results vary and it now appears to be safe if general principles of esophageal dilation are applied.

CONCLUSION

There have been a great number of advances in the fields of diagnostic and therapeutic gastrointestinal endoscopy over the last number of years. Although these techniques have contributed greatly to diagnosis and disease treatment, there are attendant complications, some of which are common to all endoscopic techniques and others are associated with particular endoscopic interventions. Undoubtedly, with the advances in interventional EUS, endoscopic fundoplication, and widespread colorectal cancer screening, there will be an increasing number of endoscopies performed.

Being aware of the types of complications that may arise, and their early recognition, is a standard to which all endoscopists should adhere.

See Also the Following Articles

Colonoscopy • Endoscopic Ultrasonography • Fistula • Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Occult Gastrointestinal Bleeding • Percutaneous Endoscopic Gastronomy (PEG) • Perforation • Upper Gastrointestinal Bleeding • Upper Gastrointestinal Endoscopy

Further Reading

Baillie J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK.

Chan, M. (1996). Complications of upper gastrointestinal endoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 287–303.

Eisen, G., Baron, T., Dominitz, J., Faigel, D., Goldstein, J., Johanson, J., Mallery, J., Raddawi, H., Vargo II, J., Waring, J., Fanelli, R., and Wheeler-Harbough, J. (2002). Complications of upper GI endoscopy. *Gastrointest. Endosc.* **55**, 784–793.

Froehlich, F., Gonvers, J., Vader, J., Dubois, R., and Burnand, B. (1999). Appropriateness of gastrointestinal endoscopy: Risk of complications. *Endoscopy* **31**, 684–686.

Lee, J., and Lieberman, D. (1996). Complications related to endoscopic hemostasis techniques. *Gastrointest. Endosc. Clin. North Am.* **6**, 305–321.

Mergener, K., and Baillie, J. (1998). Complications of endoscopy. *Endoscopy* **30**, 230–243.

Minocha, A., and Srinivasan, R. (1998). Conscious sedation: Pearls and perils. *Digest. Dis. Sci.* **431**, 835–844.

Waye, J., Kahn, O., and Auerbach, M. (1996). Complications of colonoscopy and flexible sigmoidoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 343–377.

Enteral Nutrition

TIMOTHY O. LIPMAN

Veterans Affairs Medical Center and Georgetown University

disease-specific formula An enteral formula designed to treat a specific disease.

enteral formula The nutrition product delivered via tube into the stomach or intestines. Usually, the formula is synthesized commercially from fixed ingredients. Rarely, the formula may consist of blenderized food.

enteral nutrition The provision of nutrition and nutrients directly by tube into the gastrointestinal tract, bypassing normal eating mechanisms.

immune-enhancing formula An enteral formula with added substrates (often arginine, ω-3 fatty acids, ribonucleic acids, and glutamine). Such formulas are postulated to enhance intestinal immune function.

parenteral nutrition The provision of nutrients directly into the venous side of the vascular system.

predigested formula An enteral formula consisting of amino acids and/or short peptides and/or hydrolyzed protein plus simple sugars plus minimal fat. These products are fiber-free and require less digestion by the gastrointestinal system, but still require intact intestinal absorptive function.

Enteral nutrition is the provision of nutrients directly into the gastrointestinal tract via a tube. As such, it is nonvoluntary, or nonvolitional, and is regarded as a substitute for eating in individuals with functional gastrointestinal tracts, but who cannot eat for a variety of reasons. Many commercial nutritional products are available for gastrointestinal infusion. However, enteral nutrition is an invasive technology associated with a degree of risk.

INTRODUCTION

Enteral nutrition involves direct access to the gastrointestinal tract with a tube—that is, there is bodily

 Copyright 2004, Elsevier (USA). All rights reserved.

Being aware of the types of complications that may arise, and their early recognition, is a standard to which all endoscopists should adhere.

See Also the Following Articles

Colonoscopy • Endoscopic Ultrasonography • Fistula • Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Occult Gastrointestinal Bleeding • Percutaneous Endoscopic Gastronomy (PEG) • Perforation • Upper Gastrointestinal Bleeding • Upper Gastrointestinal Endoscopy

Further Reading

Baillie J. (1992). "Gastrointestinal Endoscopy: Basic Principles and Practice." Butterworth-Heinemann, Oxford, UK.

Chan, M. (1996). Complications of upper gastrointestinal endoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 287–303.

Eisen, G., Baron, T., Dominitz, J., Faigel, D., Goldstein, J., Johanson, J., Mallery, J., Raddawi, H., Vargo II, J., Waring, J., Fanelli, R., and Wheeler-Harbough, J. (2002). Complications of upper GI endoscopy. *Gastrointest. Endosc.* **55**, 784–793.

Froehlich, F., Gonvers, J., Vader, J., Dubois, R., and Burnand, B. (1999). Appropriateness of gastrointestinal endoscopy: Risk of complications. *Endoscopy* **31**, 684–686.

Lee, J., and Lieberman, D. (1996). Complications related to endoscopic hemostasis techniques. *Gastrointest. Endosc. Clin. North Am.* **6**, 305–321.

Mergener, K., and Baillie, J. (1998). Complications of endoscopy. *Endoscopy* **30**, 230–243.

Minocha, A., and Srinivasan, R. (1998). Conscious sedation: Pearls and perils. *Digest. Dis. Sci.* **431**, 835–844.

Waye, J., Kahn, O., and Auerbach, M. (1996). Complications of colonoscopy and flexible sigmoidoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 343–377.

Enteral Nutrition

TIMOTHY O. LIPMAN

Veterans Affairs Medical Center and Georgetown University

disease-specific formula An enteral formula designed to treat a specific disease.

enteral formula The nutrition product delivered via tube into the stomach or intestines. Usually, the formula is synthesized commercially from fixed ingredients. Rarely, the formula may consist of blenderized food.

enteral nutrition The provision of nutrition and nutrients directly by tube into the gastrointestinal tract, bypassing normal eating mechanisms.

immune-enhancing formula An enteral formula with added substrates (often arginine, ω-3 fatty acids, ribonucleic acids, and glutamine). Such formulas are postulated to enhance intestinal immune function.

parenteral nutrition The provision of nutrients directly into the venous side of the vascular system.

predigested formula An enteral formula consisting of amino acids and/or short peptides and/or hydrolyzed protein plus simple sugars plus minimal fat. These products are fiber-free and require less digestion by the gastrointestinal system, but still require intact intestinal absorptive function.

Enteral nutrition is the provision of nutrients directly into the gastrointestinal tract via a tube. As such, it is nonvoluntary, or nonvolitional, and is regarded as a substitute for eating in individuals with functional gastrointestinal tracts, but who cannot eat for a variety of reasons. Many commercial nutritional products are available for gastrointestinal infusion. However, enteral nutrition is an invasive technology associated with a degree of risk.

INTRODUCTION

Enteral nutrition involves direct access to the gastrointestinal tract with a tube—that is, there is bodily

 Copyright 2004, Elsevier (USA). All rights reserved.

invasion. Although usually considered equivalent to eating, enteral nutrition differs from simple eating:

- Enteral nutrition bypasses normal eating mechanisms of smell, taste, chewing, and swallowing as well as the cephalic phase of digestion.
- Enteral formulas are usually fixed, defined, and unvaried, as opposed to the wide variety of nutrients in food.
- Enteral feeding is involuntary, or nonvolitional, and often delivered continuously, as opposed to the intermittent and voluntary intake of oral food.
- Enteral nutrition is associated with risks and complications.

This article will focus on tubes and access to the gastrointestinal tract, types of enteral formulas, indications, complications, and drug–formula interactions.

ACCESS AND TUBES

Access into the gastrointestinal tract is generally through the nose or through the abdominal wall. Tubes through the nose may be placed into the stomach (nasogastric) or beyond the pylorus into the proximal small intestine (nasointestinal or nasoenteric). Passage of tubes into the stomach is usually done "blindly." Although intestinal passage may be done blindly at the bedside, more often nasointestinal tubes are passed with the assistance of gastrointestinal endoscopy or fluoroscopy. The use of nasoenteric tubes is somewhat controversial, as it is not clear that the clinical outcome is always enhanced or that complications are prevented. Additionally, nasoenteric passage is more difficult and costly and the tubes often are displaced into the stomach. As opposed to the stiff polyvinyl chloride composition of traditional large-bore (5–6 mm) nasogastric tubes, nasogastric and nasoenteric feeding tubes are usually manufactured from polyurethane and are narrower (2–4 mm), are softer, and have a larger proportional internal diameter. Tubes placed through the abdominal wall may go into the stomach (gastrostomy), through the stomach into the jejunum (jejuno-gastrostomy), or directly into the jejunum (jejunostomy). Currently, the most common technique for direct abdominal wall insertion is via a percutaneous approach using gastroendoscopic guidance. Percutaneous insertion may be achieved also with radiologic guidance. Open or laparoscopic surgical insertion of tubes also occurs. Gastrostomy and jejunostomy tubes are usually of larger diameter (6–10 mm) than nasogastric or nasointestinal tubes, are composed of silicone, and are more flexible. These tubes are designed to be more permanent than those used through the nose, but may deteriorate over time.

ENTERAL FORMULAS

A host of enteral formulas exist. Table I lists important differentiating features that can be used for the selection and administration of enteral feeding products.

INDICATIONS

It is generally assumed that enteral nutrition is the same as eating, but this is not necessarily true, as enteral formulas are not the same as food and delivery may result in complications. Additionally, there are some data indicating that enteral formulas may not have the same effect on gastrointestinal function and structure as real food. Because of the difference between food and enteral formulas and because of the risk involved from inserted tubes, enteral nutrition should be thought of as an invasive medical intervention with potential benefits and risks.

Since prolonged starvation is not considered beneficial, enteral nutrition may be initiated in clinical situations in which normal eating is not possible for at least a week, in which there is possible and safe access to a functional gastrointestinal tract, and in which there is a clinical condition for which prolonged nonvolitional feeding is thought to provide a benefit.

Prospective clinical trials have suggested benefit from enteral nutrition after hip fractures. Enteral nutrition appears to be better than placebo for primary treatment of Crohn's disease (regional enteritis), but is not as good as corticosteroids. Enteral nutrition appears to be associated with less infectious morbidity after acute abdominal trauma and in acute pancreatitis when compared to parenteral intravenous feeding. Nonvolitional enteral feeding with "immune-enhancing" formulas may also result in less infectious morbidity in certain clinical situations. Increasing evidence indicates that gastrostomy feeding provides no benefit for patients with dementia of the Alzheimer's type. Some patients recovering from strokes have enhanced rehabilitation when gastrostomy feeding is added to total care. Gastrostomy or jejunostomy feedings may improve response to therapy in patients with head and neck or esophageal carcinomas and may provide palliation for selected individuals with advanced, but not terminal obstructing cancers.

TABLE I Considerations in Enteral Nutrition Formula Composition

Water	Origin	Macronutrient content	Added substrates	Disease specific[a]
Sufficient free water (1 kcal/cc formulas) vs concentrated formulas with free water removed (1½ −2 kcal/cc)[b]	Blenderized whole food[c] vs defined nutrient substrates	Intact protein vs amino acids and/or peptides[d] Complex carbohydrates vs simple sugars[d] "Normal" fat content vs reduced fat +/− medium-chain triglycerides[d] Lactose-free[e]	Soluble fiber "Immune-enhancing" nutrients: ω-3 fatty acids, RNA, arginine, +/− glu- tamine Modules of protein, carbohydrate, or fat can be added prior to patient administration to increase protein or energy content	Liver Pulmonary Renal Diabetes Critical care

[a] Formula substrates are modified to "treat" various disease states; there is little documentation that these modifications provide clinical benefit.

[b] Free-water-removed formulas may cause dehydration.

[c] Commercial blenderized formulas are rarely utilized.

[d] Formulas with amino acids and/or peptides, simple sugars, and low-fat +/− medium-chain triglycerides are "predigested," in that they do not need normal pancreatic and intestinal digestive processing; however, they still require intact intestinal absorptive function.

[e] Virtually all commercial "nutritionally complete" formulas are lactose-free.

COMPLICATIONS

Complications from nasogastric or nasoenteric feeding tube insertion include the following: inadvertent passage through the tracheobronchial tree into the lungs (resulting in pneumonia, pneumothorax, or hydropneumothorax), esophageal perforation, laryngeal injury, and even cranial insertion. Nasogastric and nasoenteric tubes often "fall" out or are pulled out by patients, resulting in a "nuisance" need to reinsert the tube with the repeated potential for insertion-related complications and the need for additional chest X rays to assess proper placement. Complications from indwelling nasoenteric feeding tubes include nasal alar pressure necrosis, sinusitis and otitis, and esophageal stricture. Major complications from percutaneous gastrostomy or jejunostomy insertion may occur in 5−10% of insertions and include the following: death, bleeding, esophageal laceration or perforation, gastric and colonic perforation, liver laceration, cellulitis, peritonitis, necrotizing fasciitis, exit site leakage, migration of tubes out of the stomach into the peritoneal cavity, and migration of balloon gastrostomy tubes into the pylorus causing gastric outlet obstruction. Enteral formula delivery-related complications include the following: pulmonary aspiration and aspiration pneumonia (gastrostomy and jejunostomy do not reduce the risk of aspiration), intravenous administration of formula, tube occlusion, glucose and electrolyte abnormalities, and dehydration from "concentrated, free-water-removed" formulas. When the intestine is normal, diarrhea is generally not a complication related to enteral formulas, but more likely results from concomitant medication administration, antibiotic-associated diarrhea, *Clostridium difficile* colitis, liquid medications formulated with sorbitol, intestinal "failure" in the intensive care unit as part of multiorgan failure, and rarely, bacterial contamination of the enteral product consequent to a prolonged "hang-time" at room temperature or inappropriate handling.

Nutrition support teams, composed of multidisciplinary health care professionals with expertise in the delivery of enteral and parenteral nutrition, reduce complications associated with enteral nutrition.

DRUG−NUTRIENT INTERACTIONS

Solid pills generally do not go through narrow feeding tubes and medications often are not formulated to be delivered directly into the jejunum. For these and other reasons, it is necessary to be cognizant of some basic issues surrounding the administration of medications via feeding tubes. Ideally, only medications available in liquid form should be administered through a feeding tube. Even if in liquid form, if the tube is in the small intestines, it should be ascertained that the liquid formulation is safe to infuse directly into the small bowel. Often, solid medications are "crushed" for delivery through a feeding tube. Protocols should be followed to safely liquefy solid medications. Medications formulated in delayed-release form should not be crushed for feeding delivery, as crushing will nullify the sustained release mechanisms and may result in

gastrointestinal tract toxicity or excessive delivery of medication.

With respect to several specific medications, Dilantin (phenytoin) is inactivated by enteral nutrition formulas and must be administered with tube feeding "turned off." The vitamin K content in enteral products may alter anti-coagulation responses to Coumadin. Sucralfate- and fiber-containing formulas can produce esophageal, gastric, and intestinal bezoars with complete obstruction.

See Also the Following Articles

Gastrostomy • Nutritional Assessment • Parenteral Nutrition • Percutaneous Endoscopic Gastronomy (PEG)

Further Reading

American Society for Parenteral and Enteral Nutrition web site. Available at http://www.nutritioncare.org/.

Finucane, T. E., Christmas, C., and Travis, K. (1999). Tube feeding in patients with advanced dementia. A review of the evidence. *J. Am. Med. Assoc.* **282**, 1365–1370.

Heyland, D. K., Novak, F., Drover, J. W., Jain, M., Su, X., and Suchner, U. (2001). Should immunonutrition become routine in critically ill patients? A systematic review of the evidence. *J. Am. Med. Assoc.* **286**, 944–953.

Lipman, T. O. (1998). Grains or veins: Is enteral nutrition really better than parenteral nutrition? A look at the evidence. *J. Parenteral Enteral Nutr.* **22**, 167–182.

Rombeau, J. L., and Rolandelli, R. H. (eds.) (1997). "Clinical Nutrition: Enteral and Tube Feeding," 3rd Ed. W. B. Saunders Co., Philadelphia.

Enteric Nervous System

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

AH-type enteric neuron Identified by specialized electrical behavior that includes long-lasting after-hyperpolarizing potentials associated with its action potentials and Dogiel Type II neuron morphology.

Auerbach's plexus Another name for the myenteric plexus of the enteric nervous system.

Dogiel Type I neurons Enteric neurons with multiple short dendrites and a single long axon.

Dogiel Type II neurons Multipolar enteric neurons with smooth cell bodies and multiple long and short processes in a variety of configurations.

Meissner's plexus Another name for the submucosal plexus of the enteric nervous system.

myenteric plexus Ganglionated plexus of the enteric nervous system, positioned between longitudinal and circular muscle coats of the intestine.

S-type enteric neuron Identified by specialized electrical behavior that includes elevated excitability and the universal presence of fast excitatory postsynaptic potentials and Dogiel Type I neuron morphology.

submucosal plexus Ganglionated plexus of the enteric nervous system positioned between the circular muscle coat and mucosa of the intestine.

The sympathetic, parasympathetic, and enteric divisions of the autonomic nervous system make-up the autonomic innervation of the digestive tract. The enteric division, i.e., the enteric nervous system, functions as a "brain" in the gut. It can be construed as a minibrain placed close to the effector systems it controls. Rather than packing the 2×10^8 neurons required for control of gut functions into the skull as part of the brain, and relying on signal transmission over long, unreliable pathways to the gut, natural selection has distributed the integrative neural networks along the gut at the site of the effectors.

INTRODUCTION

The enteric nervous system (ENS) controls effector systems of the digestive tract, consisting of the musculature, secretory glands, and blood vessels. As in the central nervous system, circuits at the effector sites have evolved as an organized array of different kinds of neurons interconnected by chemical synapses. Function in the circuits is determined by generation of action

 Copyright 2004, Elsevier (USA). All rights reserved.

gastrointestinal tract toxicity or excessive delivery of medication.

With respect to several specific medications, Dilantin (phenytoin) is inactivated by enteral nutrition formulas and must be administered with tube feeding "turned off." The vitamin K content in enteral products may alter anti-coagulation responses to Coumadin. Sucralfate- and fiber-containing formulas can produce esophageal, gastric, and intestinal bezoars with complete obstruction.

See Also the Following Articles

Gastrostomy • Nutritional Assessment • Parenteral Nutrition • Percutaneous Endoscopic Gastronomy (PEG)

Further Reading

American Society for Parenteral and Enteral Nutrition web site. Available at http://www.nutritioncare.org/.

Finucane, T. E., Christmas, C., and Travis, K. (1999). Tube feeding in patients with advanced dementia. A review of the evidence. *J. Am. Med. Assoc.* **282**, 1365–1370.

Heyland, D. K., Novak, F., Drover, J. W., Jain, M., Su, X., and Suchner, U. (2001). Should immunonutrition become routine in critically ill patients? A systematic review of the evidence. *J. Am. Med. Assoc.* **286**, 944–953.

Lipman, T. O. (1998). Grains or veins: Is enteral nutrition really better than parenteral nutrition? A look at the evidence. *J. Parenteral Enteral Nutr.* **22**, 167–182.

Rombeau, J. L., and Rolandelli, R. H. (eds.) (1997). "Clinical Nutrition: Enteral and Tube Feeding," 3rd Ed. W. B. Saunders Co., Philadelphia.

Enteric Nervous System

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

AH-type enteric neuron Identified by specialized electrical behavior that includes long-lasting after-hyperpolarizing potentials associated with its action potentials and Dogiel Type II neuron morphology.

Auerbach's plexus Another name for the myenteric plexus of the enteric nervous system.

Dogiel Type I neurons Enteric neurons with multiple short dendrites and a single long axon.

Dogiel Type II neurons Multipolar enteric neurons with smooth cell bodies and multiple long and short processes in a variety of configurations.

Meissner's plexus Another name for the submucosal plexus of the enteric nervous system.

myenteric plexus Ganglionated plexus of the enteric nervous system, positioned between longitudinal and circular muscle coats of the intestine.

S-type enteric neuron Identified by specialized electrical behavior that includes elevated excitability and the universal presence of fast excitatory postsynaptic potentials and Dogiel Type I neuron morphology.

submucosal plexus Ganglionated plexus of the enteric nervous system positioned between the circular muscle coat and mucosa of the intestine.

The sympathetic, parasympathetic, and enteric divisions of the autonomic nervous system make-up the autonomic innervation of the digestive tract. The enteric division, i.e., the enteric nervous system, functions as a "brain" in the gut. It can be construed as a minibrain placed close to the effector systems it controls. Rather than packing the 2×10^8 neurons required for control of gut functions into the skull as part of the brain, and relying on signal transmission over long, unreliable pathways to the gut, natural selection has distributed the integrative neural networks along the gut at the site of the effectors.

INTRODUCTION

The enteric nervous system (ENS) controls effector systems of the digestive tract, consisting of the musculature, secretory glands, and blood vessels. As in the central nervous system, circuits at the effector sites have evolved as an organized array of different kinds of neurons interconnected by chemical synapses. Function in the circuits is determined by generation of action

 Copyright 2004, Elsevier (USA). All rights reserved.

potentials within single neurons and chemical transmission of information at the points of contact (i.e., synapses) between neurons. Action potentials transmit coded information from one region of the neuron to another. Synapses at the boundaries transform the action potential codes into chemical codes. Action potential codes arriving in the presynaptic neuronal terminals are transformed to chemical signals for transmission across the synapse and are then retransformed as action potentials in the postsynaptic neurons. Postsynaptic neurons integrate large numbers of synaptic inputs and represent the fundamental component in the computation and processing of neural information.

The enteric microcircuits in the various specialized regions of the digestive tract are wired with large numbers of neurons and synaptic sites where information processing occurs. Multisite computation generates output behavior from the integrated circuits that could not be predicted from properties of their individual neurons and synapses. Emergence of complex behaviors is a fundamental property of the neural networks of the ENS, in the same way as in the brain and spinal cord.

As in the brain and spinal cord, the enteric circuitry processes information and generates patterns of signals that determine the timing and sequence of behavior of the muscles and other effector systems. Processing of information in the circuits is a function whereby computation of input signals results in output to the effectors that is a meaningful transform of the input. Like electronic calculators, which receive numbers as input and calculate outputs according to specific equations preprogrammed into the circuits, the neural circuits receive sensory signals and compute outputs that control behavior of the effectors in ways adapted for overall performance of the specialized regions of gut.

Processing of sensory signals is one of the major functions of the ENS neural networks. The sensory signals are generated by sensory nerve endings and coded in the form of action potentials. The code may represent the status of an effector system (e.g., tension in a muscle) or may signal a change in a parameter in the environment, such as luminal acidity. Sensory signals are deciphered by the neural networks to generate output signals that initiate homeostatic adjustments in the behavior of the effector system.

MYENTERIC AND SUBMUCOSAL PLEXUSES

The ENS consists of ganglia, primary interganglionic fiber tracts, and secondary and tertiary fiber projections to the effector systems (i.e., musculature, glands, and blood vessels). Two ganglionated plexuses are the most obvious structures of the ENS (Fig. 1). The myenteric plexus, also known as Auerbach's plexus, is located between the longitudinal and circular muscle layers of most of the digestive tract. The submucosal plexus, also known as Meissner's plexus, is situated in the submucosal region between the circular muscle and mucosa. The submucosal plexus is most prominent as a ganglionated network in the small and large intestine. It does not exist as a ganglionated plexus in the esophagus and is sparse in the submucosal space of the stomach. In larger mammals (e.g., humans), the intestinal submucosal plexus consists of an inner submucosal network (Meissner's plexus) located at the serosal side of the muscularis mucosae and an outer plexus (Schabadasch's plexus) adjacent to the luminal side of the circular muscle coat. In human small and large intestine, a third intermediate plexus lies between Meissner's and Schabadasch's plexus.

Most motor innervation of the intestinal circular and longitudinal muscle coats originates from motor neurons that have their cell bodies in the myenteric plexus. Motor innervation of the intestinal secretory glands (i.e., crypts of Lieberkühn) and villi originates in the submucosal plexus. Neurons in submucosal ganglia send fibers to the myenteric plexus and also receive synaptic input from axons projecting from the myenteric plexus. The interconnections link the two networks into a functionally integrated nervous system.

Structure, function, and neurochemistry of enteric ganglia differ from those of other autonomic ganglia. Unlike most sympathetic or parasympathetic autonomic ganglia, for which function is mainly relay-distribution centers for signals transmitted from the brain

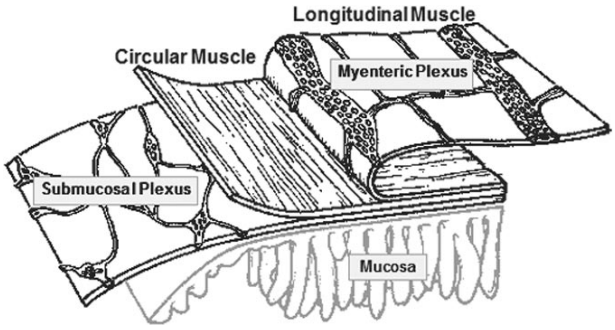


FIGURE 1 Structural relations of the intestinal musculature and the enteric nervous system. Longitudinal and circular muscle coats are the main components of the musculature involved in propulsive motility. Ganglia and interganglionic fiber tracts of the myenteric and submucosal plexuses are the main constituents of the enteric nervous system. The myenteric plexus is situated between the circular and longitudinal muscle coats; the submucous plexus is between the mucosa and circular muscle coat.

and spinal cord, enteric ganglia are interconnected to form a nervous system with mechanisms for integration and processing of information like those found in the brain and spinal cord. In view of this, the ENS is sometimes referred to as "the brain-in-the-gut."

HEURISTIC MODEL OF THE ENS

The heuristic model for the ENS is the same as for the brain and spinal cord (Fig. 2). In fact, the ENS has as many neurons as the spinal cord. Like the central nervous system, sensory neurons, interneurons, and motor neurons are connected synaptically for flow of information from sensory neurons, to interneuronal integrative networks, to motor neurons, to effector systems. The ENS organizes and coordinates the activity of each effector system into meaningful behavior of the integrated organ. It is integrated with the central nervous system and bidirectional communication occurs between the centers in the central nervous system and the ENS.

NEURONAL TYPES

Dogiel Types I and II are the principal morphological classes of neurons in the ENS. The German neuroanatomist, A. S. Dogiel, described two morphological types of enteric ganglion cells that have been named for him. These two types are each distributed in both myenteric and submucosal plexuses. Both types of neurons are distributed in a two-dimensional plane (i.e., the cell bodies are not stacked one on the other). This may adapt the ENS for the changing forces within the gut wall experienced during contraction of the musculature or stretching of the wall as the hollow organs fill.

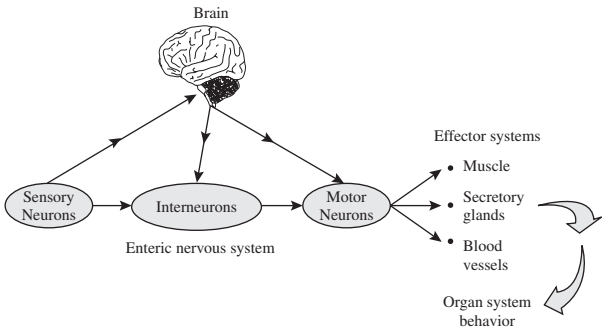


FIGURE 2 Sensory neurons, interneurons, and motor neurons are synaptically interconnected to form the integrated microcircuits of the enteric nervous system. Like the central nervous system, sensory neurons, interneurons, and motor neurons are connected synaptically for flow of information from sensory neurons to interneuronal integrative networks, to motor neurons, to effector systems. The enteric nervous system organizes and coordinates the activity of each effector system into meaningful behavior of the whole organ. Bidirectional communication occurs between the central and the enteric nervous system.

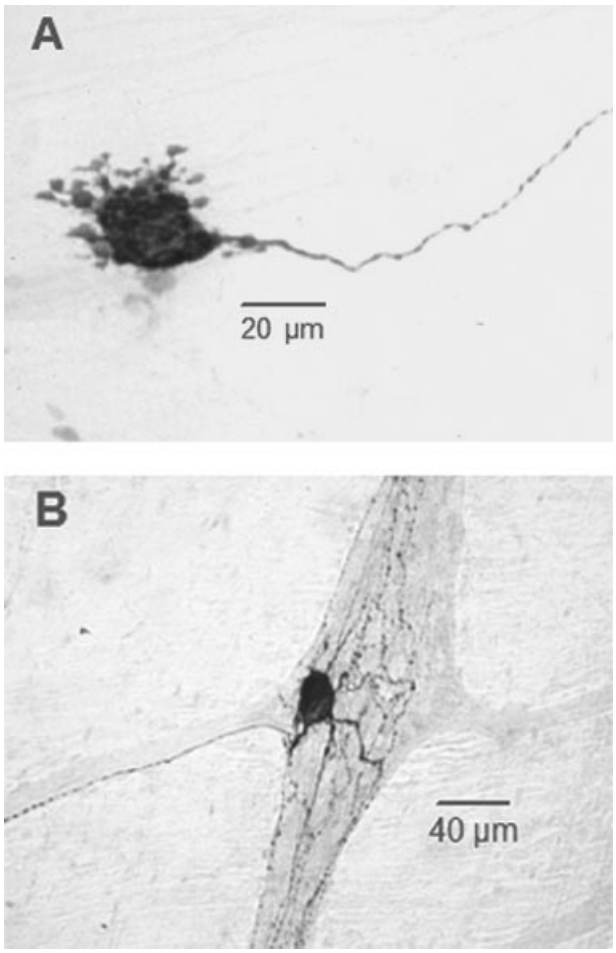


FIGURE 3 Dogiel Type I and Type II morphologies represent the major neurons in the enteric nervous system. (A) Dogiel morphologic Type I neuron in the submucosal plexus of the small intestine of the guinea pig. Dogiel Type I neurons have multiple short dendrites at the edges of the cell body and a single axon. (B) Dogiel morphologic Type II neuron in the myenteric plexus of the guinea pig small intestine. Dogiel Type II neurons have multiple long processes of variable length that ramify throughout the ganglion and into interganglionic fiber tracts. These neurons were marked by the injection of biocytin from micropipettes that also served as recording microelectrodes.

Dogiel Type I neurons have cell bodies with many short processes and a single long process (Fig. 3). These are flat neurons with the processes extending from the cell body in the circumferential and longitudinal planes of the wall. The short processes are dendrites, which receive synaptic input; the long process is an axon. The axon of Dogiel Type I neurons projects for relatively long distances through interganglionic fiber tracts and many rows of ganglia. Still, the projections are

not long in absolute length; the longest known projections of any enteric ganglion cell are only about 2–3 cm. Dogiel I neurons projecting in a specific direction may express a specific neurotransmitter. Aborally projecting Dogiel I axons release nitric oxide and vasoactive intestinal polypeptide as neurotransmitters; those projecting in the oral direction release substance P and acetylcholine. Some of the Dogiel Type I neurons are motor neurons to the musculature and secretory epithelium; others are interneurons.

The cell bodies of ENS neurons with Dogiel Type II morphology have smooth surfaces with long and short processes arising in a variety of configurations (Fig. 3). The long processes may extend through interganglionic fiber tracts across several rows of ganglia in either the circumferential direction, oral direction or aborally. Shorter processes may only project within the home ganglion. Almost all Dogiel Type II neurons in the myenteric plexus project a process to the submucosa/mucosa. Dogiel Type II processes in the mucosal regions are fired by mechanical and chemical stimulation of the mucosa. Dogiel Type II neurons are involved at an early stage in the handling of sensory information, as well as interneuronal functions in the microcircuits of the ENS. They express the neurotransmitters, substance P and acetylcholine.

ELECTRICAL BEHAVIOR OF NEURONS

Intracellular microelectrodes are used to record the electrical and synaptic behavior of enteric neurons in preparations removed from the various organs of the digestive tract. Differences in electrical and synaptic behavior are a basis for classification of enteric neurons into two main subgroups, AH type and S type. The S-type neurons were first identified on the basis of nearly universal occurrence of fast excitatory postsynaptic potentials and the AH-type neurons were identified based on the occurrence of long-lasting after-hyperpolarization associated with the action potential (see Figs. 4 and 5). Electrophysiological behavior of the AH type is found mainly in enteric neurons with Dogiel Type II morphology, and S-type behavior is found mainly in Dogiel Type I neurons.

Repetitive action potential discharge always can occur in S type; repetitive discharge generally does not occur in AH-type enteric neurons (Fig. 4). Membrane excitability is higher in S-type as compared to AH-type enteric neurons. The S-type neurons fire repetitively during the intraneuronal injection of long-duration depolarizing current pulses through the recording microelectrode. The frequency of the repetitive spike discharge increases in proportion to the amplitude of the depolarization produced by the injected pulses. The AH-type neurons do not fire repetitively during sustained membrane depolarization when they are in a resting state. Irrespective of the strength of a depolarizing pulse, these neurons fire only one or two spikes at the onset of the pulse. During slow synaptic excitation (see Fig. 5), AH neurons become hyperexcitable and fire repetitively like S-type neurons.

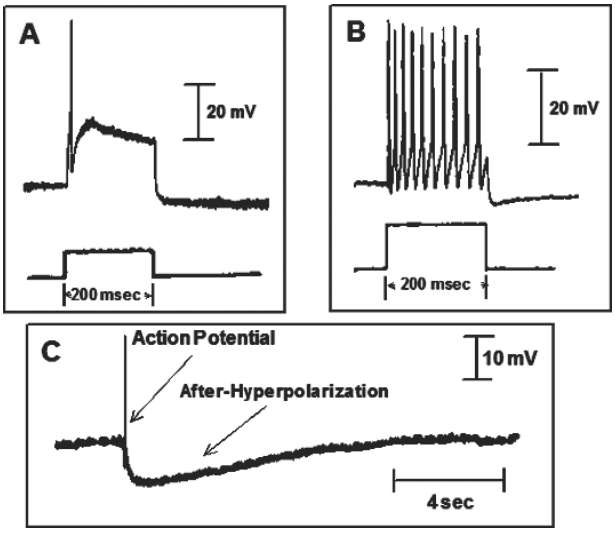


FIGURE 4 Enteric neurons are classified as S-type 1 or AH-type 2 based on their electrophysiologic behavior. (A) An AH-type 2 neuron that had low excitability, as reflected by discharge of only a single action potential during intraneuronal injection of a 200-msec depolarizing current pulse. (B) An S-type 1 neuron that had relatively high excitability, as reflected by repetitive discharge of action potentials during intraneuronal injection of a 200-msec depolarizing current pulse. The bottom traces in A and B show the current pulses and the top traces show the neuronal responses to the depolarizing current pulses. (C) The AH-type 2 neurons are characterized by long-lasting membrane hyperpolarization (i.e., after-hyperpolarization) following the discharge of an action potential.

In the guinea pig, which is the most commonly studied model, the occurrence of AH-type neurons is limited to the gastric antrum and small and large intestine. The S-type neurons are found throughout the digestive tract, including gastric corpus and antrum, gall bladder, esophagus, and pancreas.

NEUROTRANSMISSION

Synaptic events in the ENS are basically the same as in the brain and spinal cord. Excitatory postsynaptic potentials, inhibitory postsynaptic potentials, and presynaptic inhibition and facilitation are the principal synaptic events in the enteric minibrain. Both slow

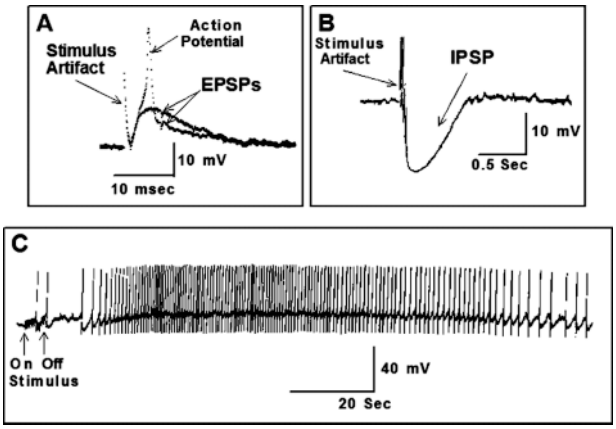


FIGURE 5 Fast and slow excitatory postsynaptic potentials (EPSPs) and slow inhibitory postsynaptic potentials (IPSPs) are the main kinds of synaptic potentials in enteric neurons. (A) Two fast EPSPs were evoked by successive stimuli and are shown as superimposed records. Only one of the EPSPs reached threshold for discharge of an action potential. The time course of the EPSPs is in the millisecond range. The fast EPSPs were evoked by single electrical shocks applied to the axon that formed the synapse with the recorded neuron. (B) Slow IPSP evoked by stimulation of an inhibitory input to the neuron. This is a hyperpolarizing synaptic potential that will suppress excitability (i.e., decrease probability of action potential discharge), as compared with enhanced excitability during slow EPSPs. (C) Slow EPSP evoked by repetitive electrical stimulation of the synaptic input to the neuron. Slowly activating depolarization of the membrane potential continues for more than 2 min after termination of the stimulus. Repetitive discharge of action potentials reflects enhanced neuronal excitability during the slow EPSP.

and fast synaptic mechanisms are operational. Fast synaptic potentials have durations in the millisecond range; slow synaptic potentials last for several seconds or minutes. Fast synaptic potentials are usually excitatory postsynaptic potentials. The slow synaptic events may be either excitatory postsynaptic potentials or inhibitory postsynaptic potentials.

Fast Excitatory Postsynaptic Potentials

Fast excitatory postsynaptic potentials are membrane depolarizations with durations less than 50 msec (Fig. 5). They are found at synapses in ENS microcircuits in myenteric and submucosal plexuses, where they are most prominent in neurons with S-type electrical behavior in all of the specialized organs of the digestive tract. Most of the fast excitatory postsynaptic potentials are mediated by acetylcholine at nicotinic postsynaptic receptors. The actions of serotonin at the serotonergic receptor subtype 3 and of purine nucleotides at P_{2X} purinergic receptors behave much like fast excitatory postsynaptic potentials.

Slow Excitatory Postsynaptic Potentials

Slow excitatory postsynaptic potentials are involved in neurotransmission in both the myenteric and submucosal plexuses and in both AH- and S-type neurons. They are most dramatic in Dogiel morphologic Type II neurons with AH-type electrophysiologic behavior when conversion from hypo- to hyperexcitability is involved. Neurons with slow excitatory synaptic inputs are found in the small and large intestine and gastric antrum, but not the gastric corpus or gallbladder. They seem to be associated with specialized regions in which propulsive motility is a significant function.

Slowly activating membrane depolarization continuing for several seconds to minutes after termination of release of the neurotransmitter from the presynaptic terminal identifies slow excitatory postsynaptic potentials (EPSPs) (Fig. 5). Enhanced excitability, reflected by long-lasting trains of impulses, is the hallmark of slow excitatory postsynaptic neurotransmission. Enhanced excitability is apparent experimentally as repetitive spike discharge during depolarizing current pulses. The AH-type neurons, which fire only a single spike at the beginning of a depolarizing current pulse in the inactivated state, will fire repetitively in response to depolarizing pulses when the slow EPSP is in effect. When activated by slow synaptic inputs, behavior of AH-type neurons is much like that of S-type neurons and may be confused as such, if the AH-type neurons happen to be in an activated state due to ongoing release of the transmitter. Postspike hyperpolarization in AH-type neurons is suppressed during slow excitatory neurotransmission. Suppression of the after-hyperpolarization is part of the mechanism that permits repetitive spike discharge at increased frequencies during the enhanced state of excitability.

Slow excitatory postsynaptic potentials are a mechanism for long-lasting activation or inhibition of gastrointestinal effector systems. The prolonged neuronal firing during a slow excitatory postsynaptic potential drives the release of neurotransmitter from the neuron's axon for the duration of the spike discharge. Prolonged inhibition, or excitation at neuronal synapses in the processing circuits and at neuroeffector junctions, is the functional outcome of slow synaptic excitation. This governs the functional behavior of the effector systems. Compared to twitches of skeletal muscles, contractile responses of the gut musculature are sluggish events that last for several seconds from start to completion. The prolonged trainlike firing during

slow excitatory postsynaptic potentials is the neural correlate of long-lasting excitatory or inhibitory responses of the muscle groups in the functioning gastrointestinal tract. Prolonged secretory responses in the intestinal crypts are also related to the sustained firing during slow synaptic excitation.

Several messenger substances found in neurons, endocrine cells, or immune cells of the brain and gut mimic slow EPSPs when applied experimentally to enteric neurons. Receptors for more than one of the messenger substances may be present on the same neuron. The current list of substances (and their receptor subtypes, in parentheses) includes the following compounds: (1) acetylcholine (muscarinic M_1), (2) cholecystokinin (A), (3) bombesin, (4) calcitonin gene-related peptide, (5) thyrotropin releasing hormone, (6) 5-hydroxytryptamine (5-HT1P), (7) norepinephrine, (8) pituitary adenylate cyclase activating peptide, (9) interleukin-1β, (10) tumor necrosis factor, (11) adenosine (A_2), (12) histamine (H2), (13) vasoactive intestinal peptide, (14) cerulein, (15) gastrin-releasing peptide, (16) tachykinins (NK_3 and NK_1), (17) corticotropin releasing hormone, (18) γ-aminobutyric acid, (19) motilin, (20) glutamate (group I metabotropic), (21) interleukin-6, (22) platelet-activating factor, and (23) bradykinin (B_2).

Slow Inhibitory Postsynaptic Potentials

Slow inhibitory postsynaptic potentials are hyperpolarizing synaptic potentials found in neurons in both myenteric and submucosal plexuses of the small and large intestine and in myenteric neurons of the gastric antrum. Slow inhibitory postsynaptic potentials activate relatively slowly and continue for several seconds after termination of the stimulation (Fig. 5). The characteristics of slow synaptic inhibition are inverse to those of slow synaptic excitation in that hyperpolarization instead of depolarization of the membrane potential occurs (e.g., a change from −50 to −75 mV). Hyperpolarization of the membrane potential decreases excitability and the probability that the neuron will fire impulses. Inhibitory postsynaptic potentials are more readily demonstrated in the submucous than in the myenteric plexus of animal preparations.

Several putative neurotransmitters evoke responses similar to slow inhibitory postsynaptic potentials when experimentally applied to enteric neurons. Some of these substances are peptides, others are purine compounds, and another is norepinephrine. Receptors for two or more of these substances may be localized to the cell body of the same neuron. Substances (and their receptor subtypes, in parentheses) that may be found in enteroendocrine cells, as well as in enteric neurons, include (1) acetylcholine, (2) 5-hydroxytryptamine (5-HT1A), (3) neurotensin, (4) somatostatin, (5) adenosine (A_1), (6) nociceptin, (7) opioid peptides, (8) norepinephrine (α2), (9) cholecystokinin, (10) ATP, and (11) galanin. These substances, which are present both in the brain and in the gastrointestinal tract, simulate slow synaptic inhibition when applied experimentally to enteric neurons.

See Also the Following Articles

Autonomic Innervation • Glial Cells (Enteric) • Parasympathetic Innervation • Sympathetic Innervation

Further Reading

Gershon, M. (1998). "The Second Brain." HarperCollins Publ., New York.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Wood, J. (2002). Enteric nervous system. *In* "Atlas of Gastrointestinal Motility" (M. Schuster, M. Crowell, and K. Koch, eds.), 2nd Ed., pp. 19–42. B. C. Decker, Inc., Hamilton, Ontario.

Wood, J., Alpers, D., and Andrews, P. (1999). Fundamentals of neurogastroenterology. *Gut* **45**(suppl. II), II6–II16.

Enterochromaffin-like (ECL) Cells

CHRISTIAN PRINZ
Technische Universität, Munich

chromogranin A protein present in cellular storage granules or vesicles containing bioactive amines such as histamine and norepinephrine.

enterochromaffin-like (ECL) cells Neuroendocrine cells in the gastric epithelium that control the peripheral regulation of acid secretion by releasing histamine as a paracrine stimulant.

enterochromaffin-like (ECL) cell carcinoid A tumor of gastric ECL cells that usually does not metastasize, but can infiltrate into deeper layers of the gastric wall. Elevated plasma gastrin levels (hypergastrinemia) play a crucial role in the development of these tumors.

gastrin Gastrointestinal hormone released from G cells in the gastrin antrum in response to food.

histamine Paracrine regulator derived from the amino acid histidine.

pituitary adenylate cyclase-activating peptide A neuropeptide present in the vagal nerve endings innervating the gastric mucosa.

Enterochromaffin-like cells are neuroendocrine cells in the gastric mucosa that control acid secretion by releasing histamine as a paracrine stimulant. Their name derives from their resemblance to chromaffin cells, which stain with chromium salts. They also synthesize and secrete other biogenic amines.

PHYSIOLOGY

During food intake, endocrine and neural signals such as gastrin and PACAP (pituitary adenylate cyclase-activating peptide), which directly stimulate enterochromaffin-like (ECL) function, are released. Gastrin is released from antral G cells in response to meal stimulation, reaches the ECL cell via the systemic circulation, and binds to cholecystokinin-B (CCK-B; gastrin) receptors on ECL cells. PACAP is a neural mediator and is released from neural endings of the vagus nerve that are activated during the cephalic as well as gastric phases of digestion. Stimulation of ECL cells induces both secretion of histamine and its synthesis, the latter through transcriptional regulation of the histidine decarboxylase gene, which codes for the enzyme that converts histamine to histidine. Prolonged stimulation of ECL cells can lead to cell growth and division.

Secretion of histamine by ECL cell stimulation is mediated by a rise in intracellular calcium resulting from the release of intracellular stored calcium and entry of calcium across the plasma membrane from extracellular fluid. L-type voltage-gated calcium channels are activated upon stimulation by gastrin. Blockage of L-type channels can decrease histamine release. K^+ and Cl^- channels have also been identified in gastric ECL cells and act to maintain the intracellular negative membrane potential at -60 mV. Histamine available for secretion in ECL cells is stored in membrane-bound vesicles. V-type ATPases transport protons into these vesicles to maintain a low pH. The proton gradient serves to energize vesicular monoamine transporters of subtype 2, which mediate the uptake of histamine into the vesicles. Secretion occurs by exocytosis and involves interacting proteins on the vesicle and the plasma membrane. In addition to histamine, the secretory vesicles contain peptides of the chromogranin family, including pancreostatin, that have been used as histological markers for the presence of ECL cells and to assess the number of ECL cells. Furthermore, circulating pancreostatin has been used as a functional marker for ECL cell secretion.

PATHOPHYSIOLOGY

The gastric mucosa is frequently infected with *Helicobacter pylori*, leading to elevated mucosal levels of pro-inflammatory cytokines such as tumor necrosis factor α (TNFα) or interleukin 1 β (IL-1β). These pro-inflammatory cytokines inhibit the secretory response of ECL cells and prevent *de novo* histamine synthesis. Furthermore, IL-1β and TNFα can induce ECL cell apoptosis. These findings may explain the clinical observations that *H. pylori*-positive patients with or without duodenal ulceration have significantly lower gastric histamine concentrations and histidine decarboxylase activity than *H. pylori*-negative subjects. Part of the

 Copyright 2004, Elsevier (USA). All rights reserved.

inhibitory response may be mediated by the release of prostaglandin E2 from ECL cells, which in turn prevents histamine secretion from this cell type.

TUMORIGENESIS: ECL CELL CARCINOIDS

The distribution of the vesicular monoamintransporters in embryonal and adult cells suggests that ECL cells arise from cells within the neural crest. ECL cell growth is differentially regulated from other gastric epithelial or parietal cells that derive from gastric mucosal stem cells. Most importantly, the life cycle and the proliferative response are directly controlled by the antral hormone gastrin, which stimulates the proliferation of ECL cells. Hypergastrinemia is observed during conditions of gastric mucosal atrophy, such as autoimmune gastritis (AIG). The destruction of mucosal cells during AIG can therefore be associated with tumors of ECL cells (type 1 ECL cell carcinoids). In contrast, type 2 carcinoids are associated with multiple endocrine neoplasia and type 3 carcinoids occur sporadically. The development of carcinoids during AIG has not been explored; however, it appears that hypergastrinemia may increase the production of growth factors such as epidermal growth factor or basic fibroblast growth factor in ECL cells, which in turn prevents ECL cell apoptosis and favors the persistence and proliferation of ECL cells in the chronically inflamed gastric mucosa. Furthermore, mutations of the CCK-B receptor with continuous intrinsic activity may play a role in this tumorigenesis.

TREATMENT OF CARCINOIDS

Endoscopic resection can be performed to dissect carcinoid tumors; however, gastric antrectomy may be the treatment of choice to remove the gastrin-producing cells and to close this vicious cycle of achlorhydria, hypergastrinemia, and the resulting proliferation and carcinogenesis of ECL cells. Future research may provide specific, long-acting CCK-B receptor antagonists in order to prevent gastrin activity and to enable medical treatment of these tumors.

See Also the Following Articles

Exocytosis • Gastric Acid Secretion • Gastrin • Pituitary Adenylate Cyclase Activating Peptide (PACAP)

Further Reading

Bordi, C., D'Adda, T., Azzoni, C., Pilato, F. P., and Caruana, P. (1995). Hypergastrinemia and gastric enterochromaffin-like cells. *Am. J. Surg. Pathol.* **19**(Suppl. 1), S8S19.

Calender, A., Cadiot, G., and Mignon, M. (2001). Multiple endocrine neoplasia type 1: Genetic and clinical aspects. *Gastroenterol. Clin. Biol.* **25**, B38–B48.

Hakanson, R., Boettcher, G., Ekblad, F., Panula, P., Simonsson, M., Dohlsten, M., Hallberg, T., and Sundler, F. (1994). The ECL cells. *In* "Physiology of the Gastrointestinal Tract" (D. H. Alpers, J. Christensen, E. D. Jacobsen, and J. H. Walsh (eds.), 3rd Ed., pp. 1171–1184. Raven Press, New York.

Mahr, S., Neumayer, N., Gerhard, M., Classen, M., and Prinz, C. (2000). IL-1β-induced apoptosis in rat gastric enterochromaffin-like cells is mediated by iNOS, NF-κB, and Bax protein. *Gastroenterology* **118**, 515–524.

Modlin, I. M., and Tang, L. H. (1996). The gastric enterochromaffin-like cell: An enigmatic cellular link. *Gastroenterology* **111**, 783–810.

Prinz, C., Neumayer, N., Mahr, S., Classen, M., and Schepp, W. (1997). Functional impairment of rat enterochromaffin-like cells by interleukin 1 beta. *Gastroenterology* **112**, 364–375.

Prinz, C., Scott, D. R., Hurwitz, D., Helander, H. F., and Sachs, G. (1994). Gastrin effects on isolated rat enterochromaffin-like cells in primary culture. *Am. J. Physiol* **267**, G663–G675.

Prinz, C., Zanner, R., Gerhard, M., Mahr, S., Neumayer, N., Höhne-Zell, B., and Gratzl, M. (1999). The mechanism of histamine secretion from gastric enterochromaffin-like cells. *Am. J. Physiol* **277**, C845–C855.

Enteroglucagon

JENS JUUL HOLST
University of Copenhagen

differential processing Formation of different processing products from the same prohormone in different tissues.

dipeptidyl peptidase IV A proteolytic enzyme occurring in the plasma membranes of many cells, including vascular endothelial cells, as well as in a soluble form in plasma. It is responsible for rapid inactivation of the glucagon-like peptides 1 and 2.

enterogastrones Gastrointestinal hormones that inhibit gastric motility and/or secretion.

ileal brake An endocrine mechanism, elicited by the presence of nutrients in the ileum, that causes inhibition of upper gastrointestinal secretion and/or motility.

incretin hormones Insulinotropic intestinal hormones responsible for enhanced insulin secretion after oral as opposed to intravenous glucose administration.

L cells The endocrine cell type in the intestinal mucosa that expresses the glucagon gene and secretes the proglucagon-derived peptides.

proglucagon-derived peptides Secreted products of proglucagon processing, including those from the pancreas (glicentin-related pancreatic polypeptide, glucagon, and major proglucagon fragment) and those from the gut (glicentin, oxyntomodulin, and glucagon-like peptides 1 and 2).

prohormone The biosynthetic precursor for a peptide/protein hormone. By cleavage of the prohormone, the mature hormone is produced.

prohormone convertases Enzymes responsible for the proteolytic processing (cleavage) of prohormones.

Enteroglucagon derives its name from the fact that the gene encoding the hormone glucagon is expressed not only in the pancreas but also in the gut. However, here the gene products (peptides) differ from those produced in the pancreas and the main products are the glucagon-like peptides 1 and 2.

HISTORY AND CHEMISTRY

The hormone glucagon was discovered in 1923 as a hyperglycemic substance derived from the pancreatic islets. However, in 1948 it was found that a similar hyperglycemic substance could be extracted from the gastric mucosa. Subsequent immunochemical studies confirmed the existence of "glucagon-like immunoreactivity" in gut extracts and detailed studies revealed that a specific, open (i.e., with apical cytoplasmic processes reaching the gut lumen) endocrine cell type, designated the L cell, was responsible for the immunoreactivity. A large number of reports from the 1970s and early 1980s contained descriptions of the secretion of "enteroglucagon," particularly after the introduction of radioimmunoassays that could distinguish pancreatic glucagon from the "total" glucagon immunoreactivity in plasma. By subtraction, one obtained enteroglucagon [also called "gut glucagon" or "gut GLI" (gut glucagon-like immunoreactivity)]. It was soon established that enteroglucagon was itself heterogeneous and composed of at least two components, complicating the interpretation of the results. Eventually, the two main components were purified from intestinal extracts and sequenced. The larger component was named glicentin, "gli-" for "glucagon-like imunoreactivity" and "-cent-" because it was originally thought to be composed of 100 amino acid residues. Eventually, it was discovered to consist of only 69 amino acids. Residues 33 to 61 corresponded exactly to the sequence of pancreatic glucagon. The smaller form was called "oxyntomodulin" and corresponded to glucagon plus the same 8 additional amino acids at the C-terminus as in glicentin; it was therefore considered to be a fragment of glicentin (see Fig. 1). Because of the presence of the entire glucagon sequence, it was assumed that glicentin might actually represent, if not the entire biosynthetic precursor of glucagon (known as "proglucagon"; glicentin was too small), then at least a fragment thereof. This notion was supported by the discovery in pancreatic extracts of a peptide that corresponded exactly to residues 1–30 of glicentin and was secreted in parallel with glucagon. This could also suggest that the intestinal and pancreatic glucagons could be products of the same gene. This turned out to be the case. In 1983, the sequence of mammalian proglucagon was deduced from a hamster islet cDNA library and subsequently the human gene was cloned and sequenced (see Fig. 1). The human proglucagon sequence included from the N-terminus the full glicentin sequence (in which the glucagon sequence occupies positions 33–61) and in addition contained two

 Copyright 2004, Elsevier (USA). All rights reserved.

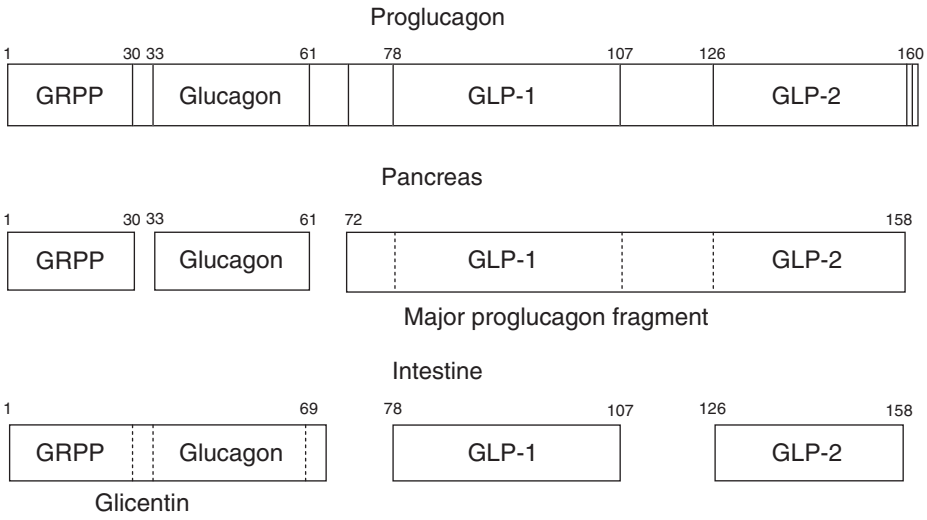


FIGURE 1 Schematic representation of the structure of proglucagon and its processing products in the pancreas and in the intestine. The numbers refer to the positions of the amino acids in the proglucagon sequence. The vertical lines indicate the major sites of proteolytic cleavage. GRPP, glicentin-related pancreatic polypeptide; GLP, glucagon-like peptide. See text for details.

glucagon-like sequences (designated glucagon-like peptides 1 and 2), flanked by pairs of basic amino acids (typical cleavage sites for prohormone processing enzymes, the so-called convertases) and occupying positions 72–108 and 126–160. Further studies confirmed that in mammals only a single gene could be identified and, furthermore, that only a single mRNA was produced in gut and pancreatic tissues (this is quite unlike the situation in birds and fish). The different proteins generated in the alpha cells of the pancreas and the L cells of the gut therefore had to result from differential processing of the prohormone in the two tissues, with the formation of glucagon in the pancreas and of glicentin in the gut. But what about the glucagon-like peptides? Detailed studies of the processing and secretion of the proglucagon-derived peptides from the pancreas and the gut revealed that the pancreatic products include proglucagon (PG) 1–30 (designated glicentin-related pancreatic polypeptide, GRPP), glucagon, and the so-called major proglucagon fragment comprising PG residues 72–158 (Fig. 1). The intestinal products include glicentin (PG1–69), oxyntomodulin (PG33–69; small amounts of GRPP are therefore also formed in the gut), and the glucagon-like peptides 1 and 2, which were apparently cleaved from the major proglucagon fragment. For some time, both of the new glucagon-like peptides seemed biologically inert, showing no glucagon-like biological activities, but natural GLP-1, isolated from intestinal extracts, was found to powerfully stimulate insulin secretion in a perfused pancreas preparation. Sequence analysis of the naturally occurring GLP-1 revealed that the peptide corresponded to PG78–107 (and was C-terminally amidated), indicating that it was cleaved out of proglucagon at the single basic amino acid residue at position 77 of PG. Natural GLP-2 was found to correspond to PG126–158. The peptide positioned between GLP-1 and GLP-2, the so-called intervening peptide 2, was found to correspond to residues 111–123 (and also to be C-terminally amidated); this peptide is also secreted, but thus far nothing is known about its physiology. The differential processing of proglucagon in the two tissues was thought to be due to expression of different processing enzymes (prohormone convertases; PCs) in the two cell types, and subsequent studies, including studies in mice with targeted deletions of the genes encoding the most prominent convertases, PC2 and PC1/3, indicated that PC2 is essential and apparently sufficient for the cleavage of proglucagon to release glucagon and that PC1/3 is essential for processing of PG to GLP-1, whereas it is still unclear whether PC1/3 is also sufficient to explain the full processing of PG into glicentin, GLP-1, and GLP-2. In agreement with these findings, PC2 is expressed in the pancreatic alpha cells and PC1/3 in the L cells.

With the complete knowledge of the intestinal processing of proglucagon and the structure of the products, it became possible to analyze their physiology and pathophysiology. Whereas glicentin (and oxyntomodulin) is currently thought to play a limited role in metabolism, GLP-1 turned out to be an important regulator of gastrointestinal functions and insulin secretion

and GLP-2 acts as a potent regulator of intestinal growth and repair. In the remainder of this article, emphasis will be placed on the glucagon-like peptides.

SECRETION OF THE PROGLUCAGON-DERIVED PEPTIDES

In principle, it should be possible to monitor L-cell secretory activity by measuring any of its products. Apparently, the three main products, glicentin, GLP-1, and GLP-2, are formed in equimolar amounts and their secretion is synchronous. The GLPs are eliminated rather rapidly, whereas glicentin seems to metabolized more slowly and, in theory, glicentin assays should be superior for monitoring of L-cell secretion. Specific assays for glicentin that are suitable for measurements of endogenous glicentin in plasma do not exist. Instead, assays directed against a midsequence of the glucagon constituent have been employed, but these require, as mentioned above, corrections for the contribution of glucagon, derived from the pancreas, and are therefore somewhat inaccurate and cumbersome. Nevertheless, since the secretion of gut-derived proglucagon-derived peptides (PGDPs) is often opposite to the secretion of pancreatic glucagon, many of the early reports of enteroglucagon/gut glucagon/gut GLI secretion are probably perfectly valid. In general, secretion is low in the fasting state (although not absent, as indicated by inhibitory effects of exogenous somatostatin administration). Ingestion of mixed meals elicits a robust secretion and it has been established that in particular, carbohydrates and lipids stimulate secretion. However, proteins are not without effect. Secretion seems to be determined mostly by the exposure of the gut lumen with nutrients, as reflected by a tight correlation between the gastric emptying rate and PGDP secretion. Furthermore, administration of α-glycosidase inhibitors (e.g., acarbose) increases the secretion of the PGDPs and reduces the secretion of proximal gastrointestinal hormones [such as gastric inhibitory polypeptide (GIP)], presumably by reducing the proximal concentration of stimulatory monosaccharides while increasing the load of nutrients to the distal small intestine, where the density of the L cells is higher. It has been suspected that only absorbable and metabolizable carbohydrates stimulate secretion, but currently it is not known whether the absorption and metabolism should take place in the L cells or in neighboring enterocytes, which could then signal to the L cells. Despite the higher density of L cells in the distal gut (in fact, there are numerous L cells even in the colon, but nothing is known about their function), the response to meal ingestion may be rapid (noticeable within 5–10 min), which has raised suspicion that secretion might be regulated by neural or hormonal mechanisms. It has been established that activation of the sympathetic nerve supply to the gut inhibits secretion, but the parasympathetic division seems to play a minor role, if any, and hormones with a clear influence on GLP-1 secretion have not been identified (these statements are based mainly on findings in humans and pigs; findings in rodents may be different).

METABOLISM OF INTESTINAL PROGLUCAGON-DERIVED PEPTIDES

The actual plasma concentrations of glicentin and the GLPs are influenced not only by their rate of secretion but also by their individual elimination rates. GLP-1 is metabolized with unique rapidity. Its primary metabolism is due to the ubiquitous enzyme dipeptidyl peptidase-IV (DPP-IV), which is located in the plasma membranes of many cell types, including vascular endothelial cells, and which also exists in a soluble form in plasma. The enzyme cleaves off the two N-terminal amino acids, leaving behind a truncated, inactive metabolite. Interestingly, this metabolite may even act as an antagonist at the GLP-1 receptor. Subsequently, both GLP-1 and the metabolite are cleared from the plasma, mainly in the kidneys, by a combination of filtration and peritubular uptake. The initial clearance of GLP-1 exceeds cardiac output and hence GLP-1 is significantly degraded while circulating, in agreement with the expression of the enzyme on vascular endothelial cell membranes. The apparent elimination half-life is on the order of 1–2 min (but there is no steady state, so the figure has little meaning). The disappearance half-life of the metabolite is approximately 4–5 min. The concentrations of intact GLP-1 are therefore very low (a few picomoles per liter), whereas those of the primary metabolite three to five times higher. GLP-2 is also cleaved by DPP-IV, but much more slowly than GLP-1. The elimination half-life of the intact GLP-2 (1–33) in humans is 7 min and that of the metabolite, GLP-2 (3–33), is 27 min. In humans, therefore, the plasma concentration of glicentin is usually the highest, that of intact GLP-2 is somewhat lower, and that of intact GLP-1 is very low. Because of the intravascular degradation of GLP-1, its rate of secretion cannot be ascertained from measurements of plasma concentrations of the intact hormone. Instead, one can use the sum of the metabolite and the intact hormone concentrations as a measure of L-cell secretion. In humans, where most of the GLP-1 molecules carry a C-terminal carboxyamidation, assays against the amidated carboxy-terminus will

provide a measure of this sum. Alternatively, L-cell secretion can be estimated by measurements of enteroglucagon or GLP-2 levels. Clearly, however, if the purpose is to estimate the biological potential of a certain concentration of GLP-1 in plasma, it is necessary to employ assays for the intact molecule.

BIOLOGICAL EFFECTS OF THE PROGLUCAGON-DERIVED PEPTIDES

There have been some controversies with respect to the biological effects of glicentin. The intact molecule seems to have little effect in humans on glucagon-responsive systems, such as gastric acid secretion, intestinal motility, and insulin secretion. Oxyntomodulin has clear glucagon-like actions, such as inhibition of acid secretion, stimulation of insulin secretion, and interaction with the hepatic glucagon receptors. However, its potency is generally much lower (approximately 1/50 of that of glucagon), and given that it also circulates in very low concentrations, it seems unlikely that it exerts major metabolic functions.

Both of the GLPs interact with specific receptors. The receptors belong to a subgroup of highly related G-protein-coupled receptors to which the glucagon, VIP, and secretin receptors, and also the PTH and calcitonin receptors, belong. The fact that both of the GLPs interact with a specific receptor also explains the finding that their actions diverge despite their close sequence homology.

Actions of Glucagon-like Peptide 1 on the Beta Cell

One of the prominent actions of GLP-1 is to potentiate glucose-induced insulin secretion. The intracellular machinery activated by GLP-1 has not been fully elucidated, but accumulation of cyclic AMP and elevation of intracellular Ca^{2+} are essential components. All steps of insulin biosynthesis are stimulated by GLP-1, including expression of the insulin genes and the genes required for the beta cell to react to glucose with increased insulin secretion, such as the glucose transporter (GLUT-2) and glucokinase genes. Furthermore, in numerous rodent studies, GLP-1 has been found to stimulate beta-cell replication and differentiation of progenitor cells into new beta cells in the pancreatic ducts. Also, beta-cell apoptosis rates are inhibited by GLP-1. GLP-1's function as an incretin hormone (the gastrointestinal hormones that enhance insulin secretion above that elicited by elevated postprandial substrate concentrations) has been proven in experiments with receptor antagonists and in mice with a targeted deletion of the GLP-1 receptor gene. These mice become glucose intolerant and the male mice may develop fasting hyperglycemia, i.e., diabetes. GLP-1 also inhibits glucagon secretion (the mechanism is thus far unknown) and the combined action on the islets therefore is to enhance insulin secretion and inhibit glucagon secretion. This in turn inhibits hepatic glucose production, which is regulated by the ratio of these hormones, and this inhibition, in combination with the stimulatory actions of elevated insulin concentrations on peripheral glucose uptake, ensures effective disposition of the absorbed carbohydrates.

Gastrointestinal Actions of Glucagon-like Peptide 1

GLP-1 powerfully inhibits upper gastrointestinal secretion and motility, inhibiting gastric acid and pancreatic enzyme secretion, gastric emptying of both liquids and solids, and intestinal transit (see Fig. 2). Because of these actions, GLP-1 is usually regarded as an important member of the enterogastrone hormones of the so-called ileal brake mechanism, an endocrine mechanism that inhibits upper gastrointestinal secretion and motility when unabsorbed nutrients are present in the ileum. Thus, on infusion into the ileum of lipids or carbohydrates in amounts that correspond to "physiological malabsorption," GLP-1 secretion is inversely correlated

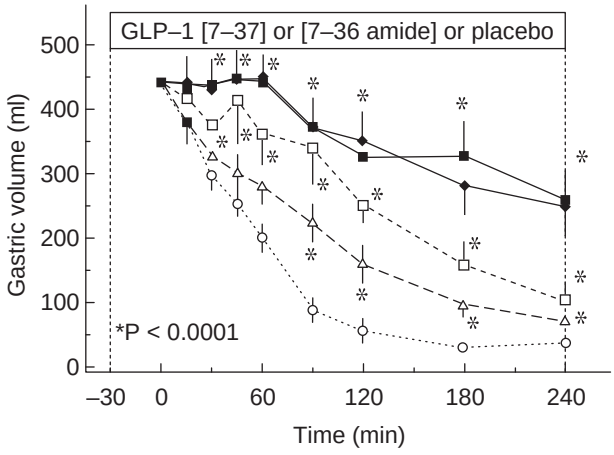


FIGURE 2 Effect of different rates of intravenous infusion of GLP-1 on gastric emptying of a liquid meal (measured as volume remaining in stomach) in healthy volunteers. Both the amidated and the nonamidated forms of GLP-1 were infused at the highest rate. (◯) Placebo; (■) GLP-1 [7–36 amide] 1.2 pmol · kg^{-1} · min^{-1}; (□) GLP-1 [7–36 amide] 0.8 pmol · kg^{-1} · min^{-1}; (△) GLP-1 [7–36 amide] 0.4 pmol · kg^{-1} · min^{-1}; (◆) GLP-1 [7–37] 1.2 pmol · kg^{-1} · min^{-1}. Data from Nauck *et al.* (1997). Glucagon-like peptide 1 inhibition of gastric emptying outweighs its insulinotropic effects in healthy humans. *Am J. Physiol. Endocrinol. Metab.* **273**, E981–E988. Reprinted with permission of the American Physiological Society.

with an inhibition of gastropancreatic secretion and motility. The inhibitory actions of GLP-1 seem to be mediated via interaction with parasympathetic outflow from the central nervous system. Thus, the inhibition of acid secretion by GLP-1 is lost in vagotomized subjects, whereas GLP-1 strongly inhibits purely neurally stimulated secretion, as elicited by "sham feeding." How GLP-1 interacts with the cerebral regulation of parasympathetic activity is not known. Circulating GLP-1 can gain access to the brain via leaks in the blood–brain barrier in the subfornical organ and the area postrema, where GLP-1 receptors are expressed, but there is also evidence that GLP-1 may affect sensory afferent neurons traveling to the brain stem. This may provide a clue to understanding its efficiency despite the surprisingly extensive degradation by DPP-IV as discussed above. Investigations have shown that a substantial percentage of the GLP-1 molecules, between 50 and 75%, are already truncated by DPP-IV and thereby inactivated before they leave the gut. Apparently, newly secreted GLP-1 molecules traverse the basal lamina of the gut epithelium and enter the lamina propria, where they may interact with nerve endings of sensory afferent neurons (see Fig. 3). Subsequently, GLP-1 enters the capillary system of the villi, where it is immediately degraded by DPP-IV, which is expressed on the endothelial surface of the vessels. According to this view, GLP-1 may physiologically act more as a paracrine regulator than as a true hormone. If, on the other hand, the exposure of the gut to unabsorbed nutrients is extensive, as seen after large meals or accelerated gastric emptying, GLP-1 secretion is enhanced and more intact hormone escapes to the systemic circulation. This allows the hormone to interact with the pancreatic islets in an endocrine fashion, thereby facilitating deposition of the substrates absorbed from this very large meal. According to this view, GLP-1 may be regarded as a second-in-line incretin hormone (GIP is the first), which is mainly activated during nutritional abundance and acts to diminish further intake (see below) and processing of nutrients and to promote disposition of already absorbed nutrients.

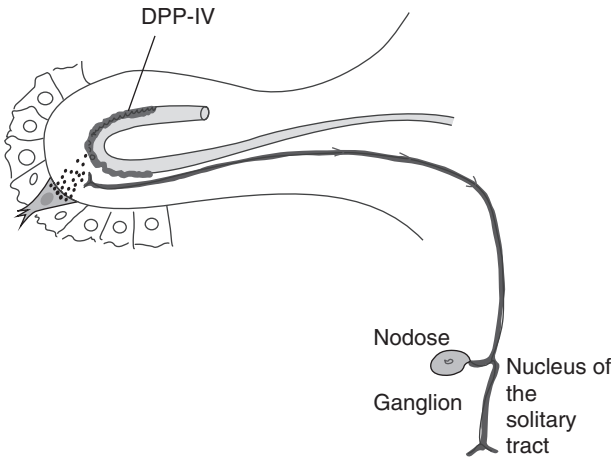


FIGURE 3 Secretion of GLP-1 from an endocrine cell in a villus of the intestinal mucosa and possible mechanism of activation of afferent sensory nerve fibers as well as inactivation of the hormone on entry into the capillaries. The peptide is secreted by exocytosis from the basal aspect of the open-type endocrine cell on lumenal stimulation. The peptide crosses the basal lamina and diffuses into the lamina propria, where it may interact with nerve endings of sensory nerves. Subsequently, it reaches the capillaries and is degraded by the enzyme dipeptidyl peptidase IV, localized on the endothelial surface of the capillaries.

Effects of Glucagon-like Peptide 1 on Appetite and Food Intake

Interestingly, GLP-1 infused intravenously also inhibits appetite and food intake. It is not known whether this is due to its inhibitory effect on gastrointestinal motility or whether it principally involves some of the brain centers regulating food intake. GLP-1 administered into the cerebral ventricles also inhibits food intake and investigations have shown that the GLP-1 receptor is expressed at numerous locations of the brain, including hypothalamic nuclei involved in appetite regulation. As mentioned above, these receptors are not immediately accessible to peripheral GLP-1, but are likely targets for GLP-1, produced in neurons of the nucleus of the solitary tract, the fibers of which project to the hypothalamus and other brain regions. According to studies of early gene expression (cFos), these neurons are not activated by extensive food intake but rather by interoceptive stress, harmful stimulation of the inner organs, e.g., elicited by administration of lipopolysaccharides or lithium chloride. Thus, intracerebrovascular (icv) administration of the GLP-1 receptor antagonist exendin 9-39 abolished the inhibition of food intake elicited by lithium chloride. The antagonist, however, also enhances spontaneous food intake as well as neuropeptide Y-induced food intake, suggesting that GLP-1 may nevertheless be tonically involved in the regulation of food intake. The mechanisms whereby peripheral GLP-1 and central GLP-1 inhibit food intake are apparently not related. Thus, the inhibitory effect of central GLP-1 on food intake is lost in animals neonatally treated with monosodium glutamate, a model of hypothalamic obesity in which the arcuate nucleus is destroyed, whereas peripheral GLP-1 is still effective. However, these findings do not exclude the possibility that neurons activated by GLP-1 on

food ingestion may transmit satiety signals, only that they do not transmit to GLP-1-producing neurons of the brain stem. Alternatively, brain GLP-1 receptors may differ with respect to transmission of mainly satiety signals and enteroceptive stress signals. Mice in which the GLP-1 receptor has been knocked out do not become obese, suggesting either that GLP-1 is not essential for the regulation of body weight or that other regulatory mechanisms compensate for the loss of the inhibitory peptide.

Biological Effects of Glucagon-like Peptide 2

The exact structure of human GLP-2 was only recently determined and synthetic GLP-2 with the correct structure has not been available until recently. The early assays for GLP-2 were unable to discriminate between GLP-2 and other GLP-2-containing products of proglucagon. In addition, using the available C-terminally extended synthetic GLP-2, most workers were unable to find biological activities for the peptide, although an early report claimed that there was activation of adenylate cyclase in sections of the hypothalamus. It has been hypothesized that GLP-2 could have a supportive role in the ileal brake mechanism and GLP-2 has been demonstrated to inhibit gastric motility and vagally induced acid secretion, but GLP-2 appears less potent than GLP-1. The GLP-2 receptor is expressed in the intestinal epithelium and also in the ventromedial nucleus of the hypothalamus and icv administration of GLP-2 was recently demonstrated to cause inhibition of food intake. It is possible therefore that GLP-1 and GLP-2 interact with respect to this cephalic action. Evidence for what appears to be the major biological activity of GLP-2 was provided by Drucker *et al*. in 1996, when they discovered that mice harboring subcutaneous glucagonomas expressing proglucagon-derived peptides in a gut-like pattern developed intestinal villous hyperplasia. Among the various proglucagon-derived peptides subsequently tested (GLP-1, glicentin, oxyntomodulin, and GLP-2), GLP-2 was the only one with significant intestinotropic activity.

The expansion of the villous epithelium was attributable both to increased crypt cell proliferation and to decreased enterocyte apoptosis. The major effect was seen in the small bowel epithelium, in which a marked increase in villus height and small bowel mass was observed, following treatment with GLP-2 for 7–10 days. Thus, it seems that GLP-2 is in fact the intestinal growth factor of the L cell predicted to exist from numerous animal experiments in which an association between adaptive growth and increased L-cell secretion was observed.

THERAPEUTIC APPLICATIONS OF GLUCAGON-LIKE PEPTIDE 1 AND GLUCAGON-LIKE PEPTIDE 2

Taken together, the actions of GLP-1 should make it ideally suited to therapy of type 2 diabetes mellitus and indeed, in such patients, infusions of GLP-1 may completely normalize blood glucose, presumably by a combined action of stimulated insulin secretion, inhibited glucagon secretion (thereby decreasing hepatic glucose production), and flattened meal-induced glucose excursions (due to the inhibition of gastric emptying). It could also be expected to inhibit food intake and thus promote a weight loss, which per se would be expected to enhance insulin sensitivity and improve glucose tolerance. Finally, and most importantly, its effect on beta-cell replication and differentiation raises the hope that with GLP-1, the effect on blood glucose regulation will be more long-lasting than that observed with conventional therapy, which generally fails after some years of treatment. The major problem associated with therapeutic application of GLP-1 is its rapid degradation and inactivation by DPP-IV. Several approaches are currently being investigated to circumvent this problem, including the use of resistant analogues of GLP-1 and administration of inhibitors of DPP-IV. Both approaches are currently undergoing clinical testing, inspired by successful experiments in animal models of diabetes where such approaches have resulted in lasting improvements in glucose tolerance, near normalization of glycated hemoglobin concentrations, and reductions of body weight. The fact that age-induced deterioration of glucose tolerance in rats can be prevented by GLP-1 suggests that its tropic effects can also be exploited clinically. Recently, in 2002, the efficacy of GLP-1 in diabetes treatment was documented by Zander *et al*. in a study in which GLP-1 was administered for 6 weeks by continuous subcutaneous infusion to patients with severe type 2 diabetes. In these patients, average plasma glucose values were lowered by approximately 5 mmol/liter, the concentrations of free fatty acids were lowered, and the concentration of glycated hemoglobin fell by 1.3%. In addition, the patients lost 2 kg of body weight and both their insulin sensitivity and beta-cell function were markedly improved. The treatment was without side effects. These observations have greatly increased interest in developing therapies based on GLP-1 receptor activation. Because GLP-1 does not cause hypoglycemia, it can also be given to euglycemic subjects and it is therefore hoped that GLP-1 can also be used to treat obesity.

Because of its intestinotropic effects, GLP-2 is currently being considered as treatment for a number of

gastrointestinal diseases associated with insufficient intestinal mucosal function. Thus, it has been shown that the mucosal atrophy associated with total parenteral nutrition could be prevented by administration of GLP-2 and that the adaptive growth responses to major intestinal resections can be augmented by GLP-2 treatment. Furthermore, in mice with dextran sulfate- or indomethacin-induced enteritis, their injuries, weight loss, and mortality may be significantly attenuated by GLP-2 treatment. Similarly, GLP-2 significantly enhanced mucosal repair after experimental intestinal ischemia. In a 2001 study in patients with short bowel syndrome, Jeppesen and co-workers demonstrated that GLP-2 significantly reduced malabsorption and promoted weight gain. In agreement with the finding that GLP-2 is less extensively degraded by DPP-IV than GLP-1, native GLP-2 can be used for these therapeutic applications. Thus, GLP-2 also seems to have an interesting spectrum of clinical applications.

See Also the Following Articles

Epithelium, Proliferation of • Gastric Motility • Glucose-Dependent Insulinotropic Polypeptide (GIP) • Pancreatic Enzyme Secretion (Physiology)

Further Reading

Bell, G. I., Sanchez-Pescador, R., Laybourn, P. J., and Najarian, R. C. (1983). Exon duplication and divergence in the human preproglucagon gene. *Nature* **304**, 368–371.

Drucker, D. J. (2003). Glucagon-like peptides: Regulators of cell proliferation, differentiation, and apoptosis. *Mol. Endocrinol.* **17**, 161–171.

Drucker, D. J., Erlich, P., Asa, S. L., and Brubaker, P. L. (1996). Induction of intestinal epithelial proliferation by glucagon-like peptide 2. *Proc. Natl. Acad. Sci. USA* **93**, 7911–7916.

Holst, J. J. (2002). Therapy of type 2 diabetes mellitus based on the actions of glucagon-like peptide-1. *Diabetes Metab. Res. Rev.* **18**, 430–441.

Holst, J. J., Bersani, M., Johnsen, A. H., Kofod, H., Hartmann, B., and Orskov, C. (1994). Proglucagon processing in porcine and human pancreas. *J. Biol. Chem.* **269**, 18827–18833.

Holst, J. J., and Deacon, C. F. (1998). Inhibition of the activity of dipeptidyl-peptidase IV as a treatment for type 2 diabetes. *Diabetes* **47**, 1663–1670.

Jeppesen, P. B., Hartmann, B., Thulesen, J., Graff, J., Lohmann, J., Hansen, B. S., Tofteng, F., Poulsen, S. S., Madsen, J. L., Holst, J. J., and Mortensen, P. B. (2001). Glucagon-like peptide 2 improves nutrient absorption and nutritional status in short-bowel patients with no colon. *Gastroenterology* **120**, 806–815.

Kinzig, K. P., D'Alessio, D. A., and Seeley, R. J. (2002). The diverse roles of specific GLP-1 receptors in the control of food intake and the response to visceral illness. *J. Neurosci.* **22**, 10470–10476.

Lovshin, J., and Drucker, D. J. (2000). New frontiers in the biology of GLP-2. *Regul. Pept.* **90**, 27–32.

Orskov, C., Holst, J. J., Knuhtsen, S., Baldissera, F. G., Poulsen, S. S., and Nielsen, O. V. (1986). Glucagon-like peptides GLP-1 and GLP-2, predicted products of the glucagon gene, are secreted separately from pig small intestine but not pancreas. *Endocrinology* **119**, 1467–1475.

Scrocchi, L. A., Brown, T. J., MaClusky, N., Brubaker, P. L., Auerbach, A. B., Joyner, A. L., and Drucker, D. J. (1996). Glucose intolerance but normal satiety in mice with a null mutation in the glucagon-like peptide 1 receptor gene. *Nat. Med.* **2**, 1254–1258.

Tang-Christensen, M., Larsen, P. J., Thulesen, J., Romer, J., and Vrang, N. (2000). The proglucagon-derived peptide, glucagon-like peptide-2, is a neurotransmitter involved in the regulation of food intake. *Nat. Med.* **6**, 802–807.

Turton, M. D., O'Shea, D., Gunn, I., Beak, S. A., Edwards, C. M., Meeran, K., Choi, S. J., Taylor, G. M., Heath, M. M., Lambert, P. D., Wilding, J. P., Smith, D. M., Ghatei, M. A., Herbert, J., and Bloom, S. R. (1996). A role for glucagon-like peptide-1 in the central regulation of feeding. [see comments]. *Nature* **379**, 69–72.

Wettergren, A., Wojdemann, M., and Holst, J. J. (1998). Glucagon-like peptide-1 inhibits gastropancreatic function by inhibiting central parasympathetic outflow. *Am. J. Physiol.* **275**, G984–G992.

Zander, M., Madsbad, S., Madsen, J. L., and Holst, J. J. (2002). Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and beta-cell function in type 2 diabetes: A parallel-group study. *Lancet* **359**, 824–830.

Eosinophilic Gastroenteritis

JON A. VANDERHOOF AND ROSEMARY J. YOUNG
University of Nebraska Medical Center, Omaha

eotaxin Specific eosinophil chemoattractant that is important in the process of eosinophil production.
mucosal eosinophilic gastroenteritis Eosinophilic infiltration primarily in the mucosal layer, resulting in vomiting, diarrhea, and abdominal pain.
muscular eosinophilic gastroenteritis Eosinophilic infiltration extending into the muscular layer of the gastrointestinal tissues, resulting in obstructive symptoms and stricture formation.
serosal eosinophilic gastroenteritis Eosinophilic infiltration extending into the serosal layer of the gastrointestinal tract, resulting in symptoms such as abdominal distension and pain.

Eosinophilic gastroenteritis is a rare disorder characterized by the infiltration of eosinophils in the gastrointestinal tissues. The location of the eosinophils, number of eosinophils, and depth of eosinophilic infiltration all are factors in the varied manifestations of this disorder. Classification of this condition can be related to the depth of the eosinophilic infiltration, such as mucosal, muscular, or serosal infiltration, and as either primary (immune-mediated) enteropathy, which is the subject of this article, or secondary enteropathy, which may occur in parasitic infections.

INTRODUCTION

Visual endoscopic findings in eosinophilic gastroenteritis may reveal prominent mucosal folds, hyperemia, and nodularity. Small numbers of eosinophils are normally present in the lamina propria of the gastrointestinal tract, thus there is no general agreement on what increased level constitutes a pathological condition. The presence of eosinophils in mucosal/submucosal, muscular, or serosal layers is considered abnormal. Attempts to grade the severity of eosinophilic gastroenteritis have been based on density of eosinophils, and it is generally agreed that greater than 20 eosinophils per high-power microscopic field is needed for histological diagnosis. It has also been noted that eosinophilic gastroenteritis can be patchy in nature and requires multiple biopsy specimens for identification. Superficial biopsy specimens lacking the muscular and serosal layers may also obscure findings and, therefore, full-thickness gastrointestinal biopsies may be desirable.

PATHOGENESIS

Epidemiologic studies to identify the frequency of eosinophilic gastroenteritis have not been done. It has been noted that this condition affects all age groups and races. The most frequent form affects the mucosal layer of the gastrointestinal tract. There is a definite association between patients having a personal or family history of other allergic disorders, such as asthma or eczema, and there is a slight male preponderance.

Eosinophils are thought to be recruited in excess numbers by eotaxin, a chemoattractant released from mast cells after stimulation by irritants or certain food antigens. Eosinophils accumulate and initiate mast cell degranulation and further release of eotaxin and other cytokines, such as histamine, platelet-activating factor, and leukotriene B4. The cytotoxic products cause tissue injury and subsequent development of symptoms. The up-regulation of eotaxin mRNA also occurs in eosinophilic gastroenteritis, but the mechanism that stimulates the eotaxin is not well understood.

Diagnosis of eosinophilic gastroenteritis generally relies on a combination of gastrointestinal symptoms and the demonstration of eosinophilic infiltration on biopsies. Many cases exhibit peripheral eosinophilia and Charcot–Leyden crystals in the stools. Absence of parasitic infestation and the lack of involvement of organs outside the gastrointestinal tract are also cited as diagnostic factors. The differential diagnosis of eosinophilic gastroenteritis includes parasitic infections, particularly with helminths, early inflammatory bowel disease, celiac disease, connective tissue diseases, immunological disorders (including malignancies), and adverse effects of pharmacological agents such as enalapril.

CLINICAL PRESENTATION

Signs and symptoms of eosinophilic gastroenteritis are dependent on the portion of the gastrointestinal tract involved and are highly variable. Some patients exhibit

 Copyright 2004, Elsevier (USA). All rights reserved.

symptoms only in response to certain trigger factors, such as food, whereas seasonal variations are noted in other patients. Oftentimes, the course of the disease starts out mild and variable, making the diagnosis difficult. Many times, severity of the symptoms increases over time and may become debilitating, regardless of treatment.

Typically, patients with eosinophilic esophagitis present with dysphagia, anorexia, and/or heartburn. Patients with eosinophilic duodenitis or enteritis have abdominal pain with or without diarrhea. Eosinophilic colitis may present with abdominal pain and/or diarrhea. Mucosal/submucosal infiltration often results in pain, nausea, vomiting, diarrhea, and weight loss. Eosinophilic infiltration in the rectum, recently described in children, may present as constipation. Muscular layer eosinophilia results in partial obstruction in the pylorus or intestine. The esophagus may appear with multiple contractions or rings, often called trachealization of the esophagus. Abdominal computer tomography (CT) and X-ray findings in this disorder may reveal areas of a thickened gastrointestinal bowel wall. Serosal layer disease involving the entire bowel wall is extremely rare, and these patients typically present with eosinophilic ascites.

TREATMENT

Because the etiology of eosinophilic gastroenteritis is not known, treatment of the condition primarily relies on management of symptoms. Because of the highly variable presentation, management often becomes a trial-and error-process. Younger children may outgrow the symptoms. Dietary and/or medication therapy are generally utilized.

Avoidance of certain foods is an initial mode of therapy. In younger children, cow's milk protein is most often an offending agent; however, restrictive elimination diets may be required for symptom control. A completely elemental diet based on free amino acids may also be required if symptoms are severe and triggered by multiple dietary components. Direct skin testing is not helpful due to a high false positive rate and the failure of a negative skin test to exclude a delayed hypersensitivity reaction. In children with identified allergic constipation, restriction of all dairy products has been shown to be of significant value.

Acid-reducing medications may be used for patients with mucosal/submucosal eosinophilic infiltration of the upper gastrointestinal tract. Medications such mesalamine may provide topical antiinflammatory effects for colonic eosinophilia. Other medications specifically targeting control of eosinophils have demonstrated some success. Case reports of sodium cromoglycate (Gastrocrom), montelukast (Singulair), and systemic steroids (Prednisone) have indicated positive benefits. If symptoms are severe, initiation of therapy with systemic steroid therapy may be required, with later weaning, and maintenance of remission with less potent antiinflammatories and diet therapy. Systemic steroids may intermittently be required if symptoms return. Most recently, the benefits of fluticasone (Flovent), an inhalable corticosteroid that is swallowed, has been shown to be of particular benefit for eosinophilic esophagitis. Additional medications for other symptoms, such as nausea, may include medications such as ondansteron hydrochloride (Zofran) or prokinetic agents such as metaclopromide (Reglan). Ketotifen (which inhibits secretion of mast cell mediators such as histamine) and suplatast tosilate (a selective T helper-2 cytokine interleukin-4 and interleukin-5 inhibitor) have been used with success in some cases of eosinophilic gastroenteritis.

CONCLUSION

Although previously considered to be rare, eosinophilic gastroenteritis is increasingly being diagnosed due to enhanced awareness of characteristics and better diagnostic methods. Pain control and chronic malnutrition are the major issues to be dealt with and fatality is uncommon unless obstructive symptoms occur. Placebo-controlled studies of various treatment modalities have been lacking due to small patient numbers and highly variable symptoms. Further research is needed to identify the exact etiology and the most appropriate therapy.

See Also the Following Articles

Cow Milk Protein Allergy • Gastritis • Gastroenteritis

Further Reading

Caldwell, J. H. (2002). Eosinophilic gastroenteritis. *Curr. Treat. Options Gastroenterol.* **5**(1), 9–16.

Daneshjoo, R., and Talley, N. (2002). Eosinophilic gastroenteritis. *Curr. Gastroenterol. Rep.* **4**(5), 366–372.

Faubion, W. A., Perrault, J., Bergart, L. J., *et al.* (1998). Treatment of eosinophilic esophagitis with inhaled corticosteroids. *J. Pediatr. Gastroenterol. Nutr.* **27**, 90–93.

Khan, S., and Orenstein, S. R. (2002). Eosinophilic gastroenteritis: Epidemiology, diagnosis and management. *Pediatr. Drugs* **4**(9), 563–570.

Melamed, I., Feanny, S. J., Sherman, P. M., *et al.* (1991). Benefit of ketotifen in patients with eosinophilic gastroenteritis. *Am. J. Med.* **90**(3), 310–314.

Perez-Millan, A., Martin-Lorente, J. L., Lopez-Morante, A., *et al.* (1997). Subserosal eosinophilic gastroenteritis treated

efficaciously with sodium cromoglycate. *Dig. Dis. Sci.* **42**(2), 342–344.
Sampson, H. A., and Anderson, J. A. (2000). Summary and recommendations: Classification of gastrointestinal manifestations due to immunologic reactions to foods in infants and young children. *J. Pediatr. Gastroenterol. Nutr.* **30**, S87–S94.
Schwartz, D. A., Parki, D. S., and Murray, J. A. (2001). Use of montelukast as steroid sparing agent for recurrent eosinophilic gastroenteritis. *Dig. Dis. Sci.* **46**(8), 1787–1790.
Shirai, T., Hashimoto, D., Suzuki, K., *et al.* (2001). Successful treatment of eosinophilic gastroenteritis with suplatast tosilate (letter). *J. Allerg. Clin. Immunol.* **107**(5), 924–925.
Talley, N. J. (1993). Eosinophilic gastroenteritis. *In* "Gastrointestinal Disease," (M. H. Sleisinger and J. S. Fordtran, eds.), 5th Ed., pp. 1224–1232. Saunders, Philadelphia.
Vanderhoof, J. A., and Young, R. J.. (2001). Allergic disorders of the gastrointestinal tract. *Curr. Opin. Clin. Nutr. Metab. Care* **4**(6), 553–556.

Epithelial Barrier Function

BRIAN A. BABBIN AND ASMA NUSRAT
Emory University

cytokines Regulatory proteins released by inflammatory and noninflammatory cells that act as mediators of immune responses.
ions Elements that have a positive (cationic) or negative (anionic) charge due to the loss or gain of one or more electrons.
kinase Enzyme that catalyzes the transfer of phosphate from ATP to substrates.
phosphatase Enzyme that catalyzes the hydrolysis of phosphate groups from substrates.
plasma membrane Semipermeable lipid bilayer that encloses the cytoplasm of a cell.

The gastrointestinal epithelium forms a selective and regulated barrier, allowing the uptake of nutrients while restricting pathogen access to underlying tissue compartments. Epithelial barrier function relies on the formation of cell–cell junctions, namely, the apical tight junction. The tight junction is the major structure that defines barrier function and selective paracellular transport. The tight junction and the underlying adherens junction form what is referred to as the apical junctional complex, which is linked to cytoskeletal elements via scaffolding and linker proteins. These complexes act to regulate tight junction structure and function. Alterations in epithelial barrier function occur in both physiologic and pathologic conditions.

GASTROINTESTINAL TRACT EPITHELIAL BARRIER

The epithelial lining of the gastrointestinal tract is an important barrier that separates two very distinct environments with different molecular compositions. Thus, the epithelium interfaces with and separates luminal contents from underlying tissue compartments. The epithelial barrier is not a static structure but is a dynamic barrier that is tightly regulated to control movement of fluid and solutes between the two environments. The epithelium permits uptake of luminal nutrients, ions, and water while restricting access of pathogens to the underlying tissue compartments. Transport across the epithelium occurs through both transcellular and paracellular pathways. Transcellular transport is an energy-dependent, directional transport mechanism that relies on the cell-specific transporters and channels positioned on the apical and basolateral cell membranes. Paracellular transport is also a regulated process that complements the transcellular transport pathway. The most apical intercellular junction in epithelial cells, referred to as the tight junction (TJ), is a major player in regulating paracellular permeability. In addition to regulating paracellular permeability, the TJ separates apical and basolateral plasma membrane domains, thereby

Copyright 2004, Elsevier (USA). All rights reserved.

efficaciously with sodium cromoglycate. *Dig. Dis. Sci.* **42**(2), 342–344.
Sampson, H. A., and Anderson, J. A. (2000). Summary and recommendations: Classification of gastrointestinal manifestations due to immunologic reactions to foods in infants and young children. *J. Pediatr. Gastroenterol. Nutr.* **30**, S87–S94.
Schwartz, D. A., Parki, D. S., and Murray, J. A. (2001). Use of montelukast as steroid sparing agent for recurrent eosinophilic gastroenteritis. *Dig. Dis. Sci.* **46**(8), 1787–1790.
Shirai, T., Hashimoto, D., Suzuki, K., *et al.* (2001). Successful treatment of eosinophilic gastroenteritis with suplatast tosilate (letter). *J. Allerg. Clin. Immunol.* **107**(5), 924–925.
Talley, N. J. (1993). Eosinophilic gastroenteritis. *In* "Gastrointestinal Disease," (M. H. Sleisinger and J. S. Fordtran, eds.), 5th Ed., pp. 1224–1232. Saunders, Philadelphia.
Vanderhoof, J. A., and Young, R. J.. (2001). Allergic disorders of the gastrointestinal tract. *Curr. Opin. Clin. Nutr. Metab. Care* **4**(6), 553–556.

Epithelial Barrier Function

BRIAN A. BABBIN AND ASMA NUSRAT
Emory University

cytokines Regulatory proteins released by inflammatory and noninflammatory cells that act as mediators of immune responses.
ions Elements that have a positive (cationic) or negative (anionic) charge due to the loss or gain of one or more electrons.
kinase Enzyme that catalyzes the transfer of phosphate from ATP to substrates.
phosphatase Enzyme that catalyzes the hydrolysis of phosphate groups from substrates.
plasma membrane Semipermeable lipid bilayer that encloses the cytoplasm of a cell.

The gastrointestinal epithelium forms a selective and regulated barrier, allowing the uptake of nutrients while restricting pathogen access to underlying tissue compartments. Epithelial barrier function relies on the formation of cell–cell junctions, namely, the apical tight junction. The tight junction is the major structure that defines barrier function and selective paracellular transport. The tight junction and the underlying adherens junction form what is referred to as the apical junctional complex, which is linked to cytoskeletal elements via scaffolding and linker proteins. These complexes act to regulate tight junction structure and function. Alterations in epithelial barrier function occur in both physiologic and pathologic conditions.

GASTROINTESTINAL TRACT EPITHELIAL BARRIER

The epithelial lining of the gastrointestinal tract is an important barrier that separates two very distinct environments with different molecular compositions. Thus, the epithelium interfaces with and separates luminal contents from underlying tissue compartments. The epithelial barrier is not a static structure but is a dynamic barrier that is tightly regulated to control movement of fluid and solutes between the two environments. The epithelium permits uptake of luminal nutrients, ions, and water while restricting access of pathogens to the underlying tissue compartments. Transport across the epithelium occurs through both transcellular and paracellular pathways. Transcellular transport is an energy-dependent, directional transport mechanism that relies on the cell-specific transporters and channels positioned on the apical and basolateral cell membranes. Paracellular transport is also a regulated process that complements the transcellular transport pathway. The most apical intercellular junction in epithelial cells, referred to as the tight junction (TJ), is a major player in regulating paracellular permeability. In addition to regulating paracellular permeability, the TJ separates apical and basolateral plasma membrane domains, thereby

Copyright 2004, Elsevier (USA). All rights reserved.

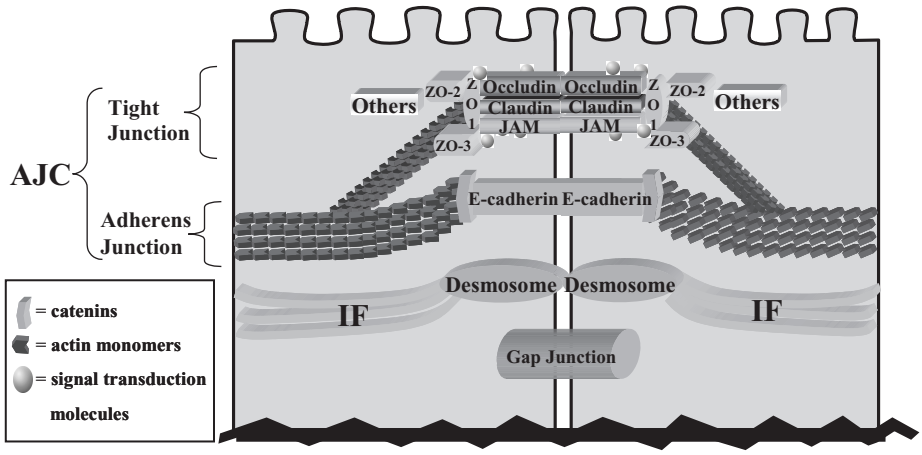


FIGURE 1 Epithelial intercellular junctions. The apical junctional complex (AJC) consists of the tight junction and underlying adherens junction. Both junctions consist of transmembrane proteins linked to the perijunctional actin cytoskeleton via linker or scaffolding proteins. Occludin, claudin(s), and junction adhesion molecules (JAMs) of the tight junction are linked to the cytoskeleton via the zonula occludens (ZO) proteins. E-Cadherin affiliates with the cytoskeleton via the catenins. Desmosomes anchor cell–cell junctions that bind intermediate filaments (IFs). Gap junctions form conduits in which molecules can pass from the cytosol of one cell to the cytosol of another cell. Figure courtesy of Thomas Brown.

contributing to epithelial cell polarity. Immediately subjacent to the TJ is the adherens junction (AJ), and these two intercellular junctions constitute the apical junctional complex. The development, stabilization, and regulation of the TJ appear to be dependent on interaction with the underlying actin cytoskeleton that encircles the apical perijunctional region of epithelial cells. Other intercellular junctional complexes in epithelial cells include the desmosomes and gap junctions (see Fig. 1). Because TJs are key structural elements that regulate epithelial paracellular permeability/barrier function, the focus of the remainder of this discussion is on these intercellular junctions.

TIGHT JUNCTION STRUCTURE

TJs were first identified by ultrathin-section electron microscopy as a series of discrete sites of apparent membrane fusion ("membrane kisses") involving the outer leaflet of adjacent plasma membranes (see Fig. 2). On freeze-fracture electron microscopy, these junctions appear as anastomosing strands or fibrils in the plasma membrane, with complementary grooves in the extracytoplasmic membrane leaflet. At sites of membrane kisses, the TJ strands in one cell are in close apposition with strands in the adjoining cell. The intervening grooves create aqueous permeation routes, or pores, in which certain solutes can diffuse. This netlike meshwork demonstrates ion and size selectivities of solutes that are dependent on the biochemical composition of the TJ and the aqueous pore size, which are regulated under both physiologic and pathologic conditions.

In recent years, the molecular composition of the TJ has begun to be elucidated. The basic biochemical structure of TJs consists of integral membrane proteins that are linked to the underlying actin cytoskeleton via linker or scaffolding proteins (see Fig. 1). The TJ multiprotein complex appears to be affiliated with distinct lipid-rich membrane microdomains, "membrane rafts," which cluster signal transduction proteins to allow for efficient regulation of TJ function. The TJ multiprotein complex consists of three major transmembrane proteins. These include occludin, claudin(s), and the junction adhesion molecule (JAM). Occludin, a 65-kDa protein, is predicted to consist of four transmembrane domains, two extracellular loops, a short cytosolic N terminus, and a long cytoplasmic C-terminal tail of approximately 255 amino acids. Both the N- and C-terminal domains are high in serine and threonine residues that are hyperphosphorylated when the protein is in its active form. The C-terminal domain interacts with the postsynaptic protein-95/discs-large/zonula occludin-1 (PDZ) domains of TJ cytoplasmic plaque proteins in the zonula occludens (ZO) family, namely, ZO-1 and ZO-3.

The claudin family of transmembrane proteins consists of over 20 members. Evidence supports that claudins are the major structural component or "backbone" of TJs. They localize to TJs of epithelial and

FIGURE 2 Electron micrograph of epithelial intercellular junctions. The arrows indicate the site of the tight junction where membranes of adjoining cells are in close apposition. The brackets outline the adherens junction with subjunctional actin filament condensation. The desmosomes are apparent beneath the adherens junction.

endothelial cells and their expression patterns vary among different tissues. Claudin-1, -2, and -4 are expressed in the gastrointestinal tract. In the rat gut, claudin-4 has been identified to be exclusively expressed on the colonic surface epithelium, whereas claudin-2 demonstrates a crypt-to-villous decrease in expression. Claudins are also predicted to have four transmembrane domains with a cytoplasmic N and C terminus, and two extracellular loops. They show no sequence homology with occludin; however, like occludin, they interact with the ZO proteins. Distinct species of claudins can polymerize to form "heteropolymers," and between adjacent cell strands, claudins adhere to each other in either a homotypic or heterotypic manner. The extracellular loops of claudins vary in the amount and nature of charged residues, suggesting that these loops and their positions in the TJ influence ion passage. It has recently been demonstrated that claudins create a charge-selective barrier. Claudin-4 has been shown to selectively alter paracellular flux of Na^+ in Madin–Darby canine kidney (MDCK) epithelial cells. Because the extracellular loops of occludin lack charged amino acid side chains and have a high tyrosine and glycine content, it is unlikely that this protein directly determines the ion selectivity of TJs. Rather, occludin is likely to have a regulatory role in assembly and maintenance of TJs.

The last major transmembrane component of TJs is JAM.Three JAM-related protein belong to the immunoglobulin superfamily. They have a single transmembrane domain. The extracellular domain of JAM is thought to be folded into two immunoglobulin-like domains. JAM has been demonstrated to be involved in cell adhesion, TJ assembly, and extravasation of monocytes across the endothelium. *In vitro* experiments have suggested that JAM also associates with TJ cytoplasmic plaque proteins such as ZO-1, acute lymphocytic leukemia fusion-6 (AF-6), and partitioning-defective protein 3 (PAR 3), implicating its role in the establishment of cell polarity.

The scaffolding proteins linking the transmembrane proteins to the underlying cytoskeleton also couple tight junctions to cytoplasmic regulation, intracellular signal transduction, cell polarity, and vesicle targeting. ZO-1 was the first TJ protein to be identified, and related ZO-2 and ZO-3 have subsequently been identified. In addition to their localization in the cytoplasmic plaque of TJs, the ZO proteins have also been identified in the nucleus of growing cells, suggesting a role in transcriptional events. They are members of the membrane-associated guanylate kinase-like homologues (MAGUKs). These proteins are characterized by one or more PDZ domains, a src homology (SH3) domain, and a guanylate kinase-like (GUK) domain. The have been demonstrated to link the TJ protein complex to the apical perijunctional actin–myosin ring. This affiliation is important in regulating paracellular permeability. Numerous additional proteins have been discovered in TJs. These include atypical protein kinase C (PKC) isotype-specific interacting protein (ASIP), symplexin, the phosphoprotein cingulin, Sec 6/8, vesicle-associated membrane protein-associated protein (VAP), and AF-6, as well as PAR. Details of mechanisms by which these proteins regulate TJ structure and function are still incompletely understood. What appears to be clear, however, is that affiliation of the TJ protein complex with the underlying actin cytoskeleton is important in regulation of TJ function. Numerous signal transduction pathways have been implicated in regulation of TJ function/paracellular permeability, as discussed in the following sections.

REGULATION OF TJ FUNCTION IN HEALTH AND DISEASE

Signal Transduction Pathways Regulating TJ Function

The tight junction is a dynamic intercellular junction that regulates paracellular permeability. It is

therefore not surprising that TJ function is regulated by a broad spectrum of signal transduction proteins that serve to maintain and regulate the association of the TJ multiprotein complex with the underlying actin cytoskeleton, which in turn controls movement of fluid and solutes in the paracellular pathway. There are key signaling mechanisms involved in regulation of paracellular permeability.

The G proteins encompass a growing family of signal transduction molecules with inducible activities that are dependent on reversible interactions with guanine nucleotides. Heterotrimeric and the small Ras superfamily of G proteins have been implicated in the regulation of TJ function. Of the many Ras family members, Rho, Rab 3B, and Rab 13 have been shown important in TJ regulation. It has been shown that the *Clostridium botulinum* toxin C3 transferase, which selectively blocks Rho effector coupling, results in reorganization of the perijunctional F-actin ring, disassembly of TJs, and increased paracellular permeability. Furthermore, studies using transfected epithelial cells expressing mutant Rho GTPases have documented the role for these proteins in TJ regulation. One group of downstream effectors mediating Rho signaling events includes the family of serine/threonine kinase isoenzymes, termed p160ROCK (ROCK I) and Rho kinase (ROCK II). These kinases are important in many cellular processes that involve actin cytoskeletal rearrangements, such as stress fiber formation, axonal growth, tumor cell invasion, and platelet activation. ROCK activity has been attributed to inactivation of myosin phosphatase and phosphorylation of myosin light chain kinase, leading to actin–myosin contraction. It has been previously shown that modification of ROCK function is associated with enhanced paracellular permeability without changes in organization of TJ proteins, occludin, JAM, and ZO-1. It is therefore likely that ROCK mediates its influence on TJs via modifications in the actin cytoskeleton and that the assembly of the TJ multiprotein complex requires other downstream effectors of Rho. The Rab proteins (Rab 11, Rab 3B) are implicated in the maintenance of intercellular junctional structures.

Affiliation of TJ proteins with each other, the plasma membrane, and the underlying actin cytoskeleton has previously been shown to require phosphorylation/dephosphorylation events. Although details of kinases and phosphatases involved in such events are still lacking, numerous studies have focused on proteins in the protein kinase C family. PKCs are phospholipid-dependent serine/threonine protein kinases that can be classified into three groups based on the structure of their regulatory domains [classical (cPKC), novel (nPKC), and atypical (aPKC)]. Numerous PKCs have been implicated in regulation of TJ structure/function. These include PKCs in the novel and atypical category. Physical association of PKC-ζ, an aPKC, with the cytoplasmic tail of the TJ transmembrane protein, occludin, has been documented. Studies in LLC-PK1 and MDCK cell lines have demonstrated that PKC activation regulates the phosphorylation and localization of occludin. In addition to TJs, PKCs have been implicated in endocytosis and recycling of the subjacent AJ transmembrane protein E-cadherin, implicating a more diverse regulatory role of these proteins.

Physiologic Regulation of Paracellular Permeability by Luminal Glucose

The first example of physiologic regulation of TJs by an extracellular stimulus was described in 1987. In these experiments, the addition of intraluminal glucose segments of rodent small intestine caused a significant increase in permeability to small molecules and a simultaneous decrease in transepithelial resistance (TER). Na^+–glucose cotransport-dependent TJ regulation has subsequently been demonstrated *in vivo* in healthy human subjects. Intrajunctional tight junction dilations and dissociation of ZO-1 have been observed during Na^+–glucose transport. Studies to define mechanisms by which luminal glucose regulates paracellular permeability have suggested that a Na^+–glucose cotransport-dependent activation of the Na^+/H^+ exchanger (NHE3) leads to alkalinization of the cytoplasm and subsequent myosin light chain phosphorylation, causing contraction of the perijunctional actomyosin ring, leading to increased paracellular permeability.

Modifications in Epithelial Paracellular Permeability by Leukocytes

Active inflammation in the gastrointestinal tract and patient symptoms are histologically correlated with transmigration of polymorphonuclear neutrophils (PMNs) across the epithelium. PMNs migrate across the epithelium from the basolateral to the apical/luminal compartments. Interactions of PMNs with epithelial cells have been demonstrated to alter barrier function in a manner that is dependent in part on the number of PMNs actively crossing the epithelium. With large-scale PMN migration, the resultant epithelial discontinuities may be the precursors for erosions and ulcers seen in many inflammatory conditions. However, low-density PMN transmigration is associated with little loss of barrier function, which recovers rapidly. The TJ serves as the rate-limiting step in PMN transmigration, and therefore mechanisms for PMNs to facilitate their migration

across this junction are likely to be important in minimizing deleterious effects of epithelial barrier disruption. The cross-talk between epithelial cells and PMNs is an area of intense investigation. A number of different epithelial intercellular junction proteins and lateral membrane proteins appear to play an essential role in this process. CD11b/CD18 (PMN β2-integrin) and CD47 (integrin-associated protein) are some key proteins in PMNs and epithelial cells that mediate interaction(s) between these two cell types. Interestingly, *in vitro* studies using model intestinal epithelial cells (T84 cells) and PMNs have shown that epithelial−PMN contact prior to PMN transmigration is sufficient to induce epithelial cell phosphorylation events that enhance paracellular permeability. This provides evidence that regulation of transepithelial migration is, at least in part, mediated by sequential signaling events between migrating PMNs and the epithelium.

Intraepithelial lymphocytes (IELs) comprise one of the largest lymphocytic populations in the body. IELs comprise a distinct T lymphocyte population that resides in close proximity with epithelial cells. Increased IELs are observed in a number of chronic inflammatory conditions, such as celiac sprue and lymphocytic colitis. E-Cadherin has been identified as the counterreceptor for the IEL-expressed integrin, αEβ7-integrin, which is important in the interactions between these two cell types. Affiliation with the epithelium, as well as elaboration of cytokines by IELs, may play a role in modifying epithelial barrier function.

Cytokine Regulation of Epithelial Barrier Function

The extracellular signals of many different cytokines and growth factors in a variety of physiologic and pathologic conditions influence TJ structure and function. Key proinflammatory cytokines such as interferonγ (IFNγ)and tumor necrosis factor α (TNFα) enhance paracellular permeability by influencing TJ structure. Studies have demonstrated that IFNγ induces displacement of occludin, ZO-1, and ZO-2 away from the apical lateral membrane with alterations of the perijunctional F-actin ring. TNFα has been shown to reduce the number of tight junction strands and induce loss of submembranous cortical F-actin with an increase in G-actin content.

Numerous other cytokines such as interleukin (IL)-1, IL-4, IL-13, insulin-like growth factor-1 (IGF-1), and vascular endothelial growth factor (VEGF) have been shown to decrease barrier function of endothelial and epithelial cells. Transforming growth factor-β (TGF-β)produced by platelets, lymphocytes, macrophages, and endothelial cells, in contrast, has been shown to increase the barrier function of epithelial cells and promote intestinal epithelial wound closure. In addition to inducing increased epithelial barrier function, TGF-β has additional protective effects by curtailing the influence of barrier-reducing cytokines such as IFN-γIL-4, and IL-10.

Disease Entities Influencing Epithelial Barrier Function

Idiopathic Inflammatory States

Impaired epithelial barrier function has been observed in diverse idiopathic pathologic conditions of the gastrointestinal tract. These entities range from microscopic colitis to inflammatory bowel disease (IBD). In IBD, numerous mechanisms have been described to explain the decrease in epithelial barrier function. These include release of proinflammatory cytokines, PMN transepithelial migration, and innate increased paracellular permeability not completely accounted for by the former mechanisms. Altered barrier function in microscopic colitis may be due to IEL−epithelial cell interactions.

IBD encompasses ulcerative colitis (UC) and Crohn's disease (CD), which display a waxing and waning course, with flares of disease activity correlating with active PMN transepithelial migration. Inflammatory mediators such as TNFα and IFN-are known to be elevated in these disorders and contribute to impairment of barrier function. This impairment is most pronounced in areas of active inflammation, and less so in quiescent areas. Innate differences in epithelial barrier function have been described in patients with IBD. Redistribution of TJ proteins in this disorder has been observed both in areas of active inflammation and in quiescent regions. Application of targeted gene deletion and transgenic approaches has revealed potential genetic factors that could contribute to the development of IBD. Models that include deletions of IL-2, IL-10, and TGF-β, as well as T cell receptor rearrangements, have resulted in chronic inflammatory bowel disease. However, these manipulations of single genetic loci are modulated by other genetic factors as well as environmental factors. Thus, it appears that IBD results from an environmental stimulus in a genetically predisposed individual.

Collagenous and lymphocytic colitis, also known as microscopic colitis, are characterized by watery diarrhea and a normal-appearing colonic mucosa during endoscopic examination. Histologically, these entities are associated with an increase in intraepithelial lymphocytes or thickening of epithelial basement membrane and epithelial damage. Diarrhea in these

patients may be attributed to reduced net ion absorption secondary to alterations in barrier function. These alterations could result from direct IEL–epithelial interaction as well as elaboration of cytokines by IELs. In collagenous colitis, the thickened subepithelial collagen plate may act as a diffusion barrier, decreasing ion uptake.

Pathogen-Induced Epithelial Damage and Decreased Epithelial Barrier Function

Many infectious agents, such as *Salmonella*, *Shigella, Clostridium difficile, Escherichia coli* (enterohemorrhagic strains), and *Campylobacter*, are capable of breaching the epithelial barrier and inducing disease. Some pathogens secrete toxins that specifically interact with TJ proteins or regulatory molecules in TJs, thereby enhancing paracellular permeability (see Table I). As an example, *C. difficile* elaborates two toxins (A and B). These toxins specifically glucosylate Rho GTPases and inhibit Rho effector coupling, resulting in modifications in TJ structure and enhanced paracellular permeability. The latter in turn exposes underlying tissues to luminal antigens/pathogens, resulting in further epithelial damage and influx of PMNs. Monocytes exposed to *C. difficile* toxins have been shown to secrete the PMN chemoattractant interleukin-8. In addition, specific TJ transmembrane proteins have been documented to function as receptors for viral pathogens. These include claudin-4, which is a receptor for *Clostridium perfringens* enterotoxin (CPE). This single polypeptide induces displacement of claudin-4 from the TJ, with subsequent degradation in MDCK cells. Viral receptors in TJs include JAM, which is the reovirus receptor, and the coxsackievirus and adenovirus receptor (CAR) protein. Thus, both bacterial and viral pathogens not only target TJ proteins and influence paracellular permeability, but also utilize TJ proteins as receptors or conduits for internalization and epithelial damage.

SUMMARY

TJs are dynamic structures that are regulated to control epithelial barrier function and selective paracellular transport of solutes. They also serve to segregate apical and basolateral membrane domains that are important in cell polarity. A variety of factors regulate TJs in physiologic and pathologic conditions. In inflammatory states, alterations in these structures are a major factor in the development of diarrhea. Although our understanding of these structures has greatly advanced over recent years, questions regarding the regulation of the TJ barrier, incorporation of TJ proteins into the tight junctions, and apical polarization of TJs have yet to be answered.

TABLE I Enteric Pathogens that Alter Barrier Function/Tight Junctions[a]

Enteric pathogen	Virulence factor	Cell lines/tissue studied	Proposed mechanism	TJ structural modification and pathophysiology
Clostridium difficile	Toxin A, toxin B	T84, human colon, rabbit ileum, Caco-2	Monoglucosylation of Rho at threonine 37	Alteration of apical F-actin with decreased TER in 2–4 hours
Clostridium perfringens	Enterotoxin	MDCK I	Cleavage of claudin-3 and claudin-4	Loss of claudins at TJs with decreased TER starting at 4 hours
Vibrio cholerae	Zonula occludens toxin	Rabbit ileum	Protein kinase C activation	Decreased numbers of TJ strands with decreased TER by 1 hour
	Hemagglutinin	MDCK I	Cleavage of occludin	Occludin degradation; ZO-1 and F-actin reorganization with decreased TER after 1 hour
Bacteroides fragilis	Metalloprotease toxin	T84, Caco-2, MDCK, human colon	Cleavage of E-cadherin	Dissociation of occludin and ZO-1 from TJ, absent E-cadherin with decreased TER starting in 15 minutes
Escherichia coli	Cytotoxic necrotizing factor 1	Caco-2	Deamination of Rho at glutamine 63	Alteration of actin filament formation with decreased TER in 1 hour

[a] Abbreviations: MDCK, Madin–Darby canine kidney; TER, transepithelial resistance.

See Also the Following Articles

Colonic Absorption and Secretion • Epithelium, Proliferation of • Epithelium, Repair of • Small Intestine, Absorption and Secretion

Further Reading

Anderson, J. M. (2001). Molecular structure of tight junctions and their role in epithelial transport. *News Physiol. Sci.* **16**, 126–130.

Clarke, H., Marano, C. W., Soler, A. P., and Mullin, J. M. (2000). Modulation of tight junction function by protein kinase C isoforms. *Adv. Drug Deliv. Rev.* **41**, 283–301.

Colegio, O. R., Van Itallie, C. M., McCrea, H. J., Rahner, C., and Anderson, J. M. (2002). Claudins create charge-selective channels in the paracellular pathway between epithelial cells. *Am. J. Physiol. Cell Physiol.* **283**, C142–C147.

Edens, H. A., Levi, B. P., Jaye, D. L., Walsh, S., Reaves, T. A., Turner, J. R., Nusrat, A., and Parkos, C. A. (2002). Neutrophil transepithelial migration: Evidence for sequential, contact-dependent signaling events and enhanced paracellular permeability independent of transjunctional migration. *J. Immunol.* **169**, 476–486.

Hopkins, A. M., Randall, D. L., Mrsny, R. J., Walsh, S. V., and Nusrat, A. (2000). Modulation of tight junction function by G protein-coupled events. *Adv. Drug Deliv. Rev.* **41**, 329–340.

Kucharzik, T., Walsh, S. V., Chen, J., Parkos, C. A., and Nusrat, A. (2001). Neutrophil transmigration in inflammatory bowel disease is associated with differential expression of epithelial intercellular junction proteins. *Am. J. Pathol.* **159**, 2001–2009.

Mitic, L. L., Van Itallie, C. M., and Anderson, J. M. (2000). Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: Lessons from mutant animals and proteins. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G250–G254.

Nusrat, A., Giry, M., Turner, J. R., Colgan, S. P., Parkos, C. A., Carnes, D., Lemichez, E., Bouquet, P., and Madara, J. L. (1995). Rho protein regulates tight junctions and perijunctional actin organization in polarized epithelia. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10629–10633.

Nusrat, A., Turner, J. R., and Madara, J. L. (2000). Molecular physiology and pathophysiology of tight junctions IV. Regulation of tight junctions by extracellular stimuli: Nutrients, cytokines, and immune cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G851–G857.

Parkos, C. A. (1997). Cell adhesion and migration I. Neutrophil adhesive interactions with intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **36**(4), G763–G768.

Schmitz, H., Barmeyer, C., Fromm, M., Runkel, N., Foss, H., Bentzel, C. J., Riecken, E., and Schulzke, J. (1999). Altered tight junction structure contributes to the impaired epithelial barrier function in ulcerative colitis. *Gastroenterology* **116**, 301–309.

Sears, C. L. (2000). Molecular physiology and pathophysiology of tight junctions V. Assault of the tight junction by enteric pathogens. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G1129–G1134.

Tsukita, S., Furuse, M., and Itoh, M. (2001). Multifunctional strands in tight junctions. *Nat. Rev.* 2, 285–293.

Turner, J. R., Cohen, D. E., Mrsny, R. J., and Madara, J. L. (2000). Noninvasive *in vivo* analysis of human small intestinal paracellular absorption: Regulation by Na^+-glucose cotransport. *Dig. Dis. Sci.* **45**(11), 2122–2126.

Walsh, S. V., Hopkins, A. M., Chen, J., Narumiya, S., Parkos, C. A., and Nusrat, A. (2001). Rho kinase regulates tight junction function and is necessary for tight junction assembly in polarized intestinal epithelia. *Gastroenterology* **121**, 566–579.

Epithelium, Proliferation of

PAUL E. STROMBERG AND CRAIG M. COOPERSMITH
Washington University School of Medicine, St. Louis, Missouri

apoptosis Evolutionarily conserved method of programmed cell death; also referred to as "cell suicide."
base One of the two functional zones of the gastric unit, found at the bottom of the gastric gland; site of cells producing the acid and enzymes necessary for breakdown of food in the gastric lumen.
cell cycle Series of four phases that a dividing or proliferating cell goes through. G_1 is the first of two growth phases and is the longest portion of the cell cycle, S phase is the period during which a cell's genetic material is replicated, G_2 is the second growth phase, and M phase is when a cell divides into two genetically identical cells.
columnar epithelium Classification of epithelium in which cell height is larger than cell width.
crypt of Lieberkühn Zone of proliferation in the small intestine, located underneath the villi. Crypts provide the cells that replace the epithelium of the villi of the small intestine.
deoxyribonucleic acid Genetic material of a cell.
differentiation Process of a cell maturing and acquiring a particular, specific function within the body, usually accompanied by a loss of the ability to proliferate.
foveolae Opening of the gastric unit into the lumen of the stomach.
goblet cell Responsible for the production and secretion of mucus into the lumen of the small intestine and colon.
homeostasis Ability of an organism to maintain an internal balance or equilibrium (i.e., a balance between cell loss and cell production).
hyperplasia Increase of proliferation to a level that is higher than normal.
hypoplasia Decrease of proliferation to a level that is lower than normal.
metaplasia Process of converting one tissue type to another.
Paneth cell Normally found only at the base of the crypt of Lieberkühn in the small intestine; responsible for secreting lysozyme and antimicrobial factors.
parietal cell One of the cell types of the gastric epithelium; responsible for acid production and secretion into the gastric lumen.
pit One of the two functional zones of the gastric unit, located superior to the gastric gland; site of pit cells that produce the protective mucus of the gastric epithelium.
restitution Ability of epithelium to rapidly replace or repair sections of denuded epithelial tissue.
squamous epithelium Flat type of epithelium classified by cells with a width much larger than their height.
stem cell Has the ability to self-replicate and ultimately produce multiple distinct types of differentiated cells.
villus Functional zone of the small intestinal epithelium; composed of columnar epithelium that projects into the lumen of the small intestine.
zymogenic cell One of the cell types of the gastric epithelium, found within the base; responsible for secreting enzymes needed in the chemical breakdown of food.

The gut epithelium is one of the most rapidly dividing tissues within the body. Rapid cellular renewal in the gastrointestinal tract is protective, serving to prevent bacterial infection or other diseases of the epithelial layer. However, alterations in gut proliferation may lead to a variety of diseases, including cancer and inflammatory bowel disease.

INTRODUCTION

The gastrointestinal epithelium serves several important roles within the body. It provides a physical barrier against infections and protects the body from the harsh environment of the gut lumen, exemplified by the stomach's highly acidic environment. It also serves as the major barrier against luminal carcinogens, toxins, or pathogens. In addition, the gastrointestinal epithelium provides a surface layer for selective absorption of nutrients and water from ingested foods.

Although specific details vary along the length of the gastrointestinal tract, the gut epithelium can generally be subdivided into two zones: a proliferative zone and a functional zone. In the proliferative zone, a small number of multipotent stem cells create transit daughter cells, which further divide, expanding the number of cells available to the epithelium. These cells then migrate to the functional zone, where they differentiate into their final form to perform a specialized function (such as producing mucus or absorbing nutrients) and lose their ability to divide.

CELL CYCLE AND METHODS FOR STUDYING PROLIFERATION

Proliferating cells progress through four stereotypical phases, called the cell cycle, as they undergo the journey from being a single cell to having identical cellular offspring (Fig. 1). The presynthetic gap phase (G_1) is the

 Copyright 2004, Elsevier (USA). All rights reserved.

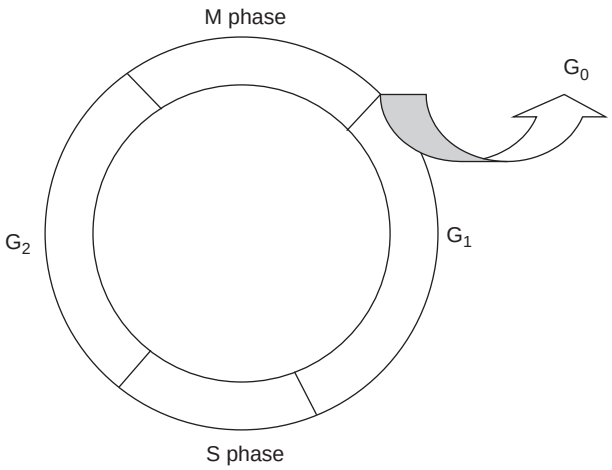


FIGURE 1 The cell cycle of proliferating cells. Once cells enter the functional zone, they lose the ability to proliferate and are in G_0.

first and longest portion of this cycle and is a period of cell growth. This is followed by a synthesis (S) phase in which the cell's DNA is replicated, leading to a doubling of the cell's genetic material. The postsynthetic gap phase (G_2) follows next and is another growth phase, during which the cell is preparing to undergo division. The final step of the cell cycle is the mitosis (M) phase. During M phase, the cell divides into two daughter cells that are genetically identical to the original parental cell. When a cell migrates from the proliferative zone of the gastrointestinal epithelium to the functional zone, where it can differentiate to perform specialized processes, it loses the ability to proliferate. Differentiated cells that lack the ability to divide under normal circumstances are said to be in the G_0 phase.

Of the four parts of the cell cycle, mitosis is the easiest to detect; it is the only phase detectable on conventional light microscopy without specialized staining. Mitotic cells have a characteristic morphology, including the presence of mitotic spindles. This is readily apparent on hematoxylin and eosin staining, the simplest histologic technique for evaluating tissue microscopically.

Although there is no clear way to detect cells in G_1 or G_2 by means of conventional microscopy, several methods exist for identifying proliferating cells in the S phase of the cell cycle. These techniques are more specialized than the visual identification of M phase cells on simple histologic sections and are typically utilized in animal or cell culture research. Bromodeoxyuridine (BrdU) is a material similar to the DNA building block thymidine. Subcutaneous injection of BrdU leads all cells in S phase to mistakenly incorporate this molecule in replicating DNA. The incorporated BrdU can be visualized using standard staining techniques with antibodies to the molecule. A similar approach involves proliferating cell nuclear agent (PCNA), a small protein that is needed for the cell's DNA replication machinery to function. Cellular staining allows for the identification of this protein when levels are highest to indicate cells that are in S phase.

The S phase cells can also be identified as cells that incorporate tritiated thymidine, a radiolabeled building block of DNA, into their genetic material. Identification involves either using autoradiography, a photographic technique similar to X rays wherein radioactive emission from labeled cells causes a darkening of a silver-coated film that is laid over a section of tissue, or using a quantitation technique that counts radioactive emissions using a scintillation counter.

ORAL CAVITY AND ESOPHAGUS

The epithelium of the oral mucosa and esophagus provides protection to the salivary glands and muscles within the esophagus. It also serves as a barrier against cytotoxic agents and possible carcinogens in ingested substances. The hard palate and gingiva are protected by the presence of a keratinized epithelia in areas where mechanical breakdown of food occurs. Keratin, a family of scleroproteins also seen in nails, hair, and horn tissue, adds additional strength to this portion of the gastrointestinal epithelium. All other mucosa of the oral cavity and esophagus consists of a softer, nonkeratinized epithelium. The epithelium also serves as a flexible sheath to help transfer food to the stomach via peristaltic contractions of smooth muscles in the esophagus.

Epithelium of the oral mucosa and the esophagus is characterized as stratified squamous epithelium. Stem cells are located in the basal layer of the epithelium. Cells migrate upward from the basal layer to the superficial layer, the surface of the epithelium. As cells migrate toward the epithelial surface they begin to differentiate and lose the ability to divide. To ensure that the epithelial layer does not get too thick, cells are sloughed off into the oral cavity or into the lumen of the esophagus. Homeostasis is therefore maintained, and cell loss in the superficial layer is balanced by proliferation in the basal layer.

STOMACH

The stomach contains thousands of invaginations called gastric units; these have the greatest spatial and geographic complexity of any epithelia in the gastrointestinal tract. Like all epithelium in the gastrointestinal tract except for oral mucosa and esophagus, the gastric epithelium is a columnar epithelium. Its gastric units

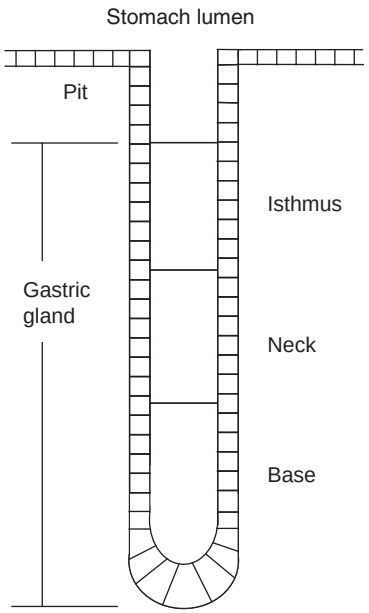


FIGURE 2 Gastric unit of the stomach epithelium. Multipotent stem cells that give rise to all the cell types of the gastric epithelium proliferate in the isthmus.

open into the lumen of the stomach via the foveolae. Each gastric unit can be divided into two subunits, the pit and the gastric gland, with the pit located superior to the gastric gland (Fig. 2). The gastric gland is further subdivided into three subunits (in descending order from the junction with the pit)—the isthmus, the neck, and the base. The proliferative zone of the gastric unit is located in the isthmus of the gastric gland. Multipotent stem cells located in the isthmus are capable of producing the three cell lineages found within the gastric epithelium—mucus-producing pit cells, acid-producing parietal cells, and enzyme-producing zymogenic cells.

Although gastric units exist throughout the stomach, differences exist in the fundus, corpus (body), and antrum, the stomach's three main anatomical regions (from proximal to distal). The two functional zones of the stomach's epithelium are located at the opening of the gastric unit to the lumen and at the base of the gastric gland. The functional zone located near the opening of the gastric lumen is called the pit, which is composed primarily of pit cells, responsible for secreting a protective layer of mucus, and a few acid-producing parietal cells. The base of the gastric gland is the location of the second functional zone of the gastric unit and contains parietal cells and/or zymogenic cells, depending on its anatomic location.

The neck region of the gastric gland in the corpus serves as a transitional zone for zymogenic cells. As prezymogenic cells travel through the neck, they undergo a process of maturation and lose the ability to divide. This process of maturation and migration downward from the isthmus to the base takes prezymogenic cells approximately 14 days. Migration for pit cells is much more rapid; they travel from the isthmus to the pit in 12 hours. Parietal cells exhibit a bidirectional migratory pattern, with mature parietal cells being found in all compartments of the gastric unit.

Although stem cells lead to daughter cells and eventually differentiated cells throughout the stomach, local geographic influences lead to differences in gastric unit size and type throughout the stomach. For instance, the gastric unit of the corpus is smaller than the gastric unit of the antrum. In mice, gastric units in the corpus contain 200 cells, compared to 250 cells in the antrum. The length of the pit also varies between gastric units within the corpus and the antrum, with the pit being much larger in the antrum. The base also differs between regions; zymogenic cells are present in the corpus but are absent in the antrum. Therefore, even though stem cells can give rise to each of the differentiated cell types of the gastric epithelium, local environmental factors can influence what type and how many cells are produced.

Turnover rates of cells in the gastric epithelium vary depending on the cell type. Parietal cells have an average life span of 54 days, whereas zymogenic cells live an average of 194 days, after which they are eliminated in an unclear fashion, potentially by being engulfed by macrophages in the extracellular matrix. In contrast, pit cells are rapidly renewed and have a turnover rate of 3 days, after which they are shed into the gastric lumen. Despite the obvious disparity in life span of differentiated epithelial cells, stem cells and their resultant daughter cells are able to match cell loss with cell production to maintain gastric homeostasis.

SMALL INTESTINE

The primary function of the small intestine is absorption of water and nutrients from ingested foods. Enzymatic digestion occurs within the small intestine whereas mechanical (and to a lesser degree chemical) breakdown occurs in the stomach. In order to increase the surface area of the small intestine for greater absorption, the intestinal epithelium is divided into hundreds of thousands of fingerlike projections, villi, into the lumen. Each villus is supplied with cells from the small intestine's proliferative units, flask-shaped invaginations called the crypts of Lieberkühn (crypts) (Fig. 3). Proliferation has been studied extensively in the small intestine; the mechanisms underlying this process are well delineated in mice and are more completely understood in humans than they are for other locations in the

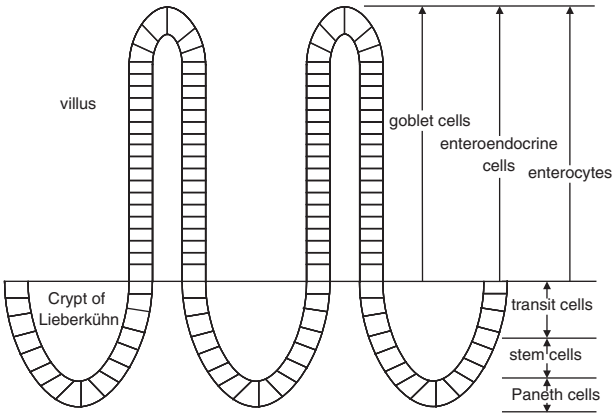


FIGURE 3 Crypt–villus unit of the small intestinal epithelium. Multipotent stem cells are located three to five positions above the crypt base. These give rise to a transit cell population and ultimately to all four cell lineages of the small bowel.

gastrointestinal tract. All cells within the small intestinal epithelium are produced in the crypts. Each crypt has an average cell population of 450, and although the average number of crypts per intestine is unknown in humans, it is approximately 1×10^6 in mice. An individual crypt supplies two to three villi. In contrast, each villus is supplied by approximately six crypts.

Multipotent stem cells that supply the entire small intestinal epithelium are located near the base of the crypt at approximately three to five cell positions above the base of the crypt. The number of stem cells in each crypt is unclear, with estimates ranging from 4 to 16, accounting for less than 5% of crypt cells. Stem cells divide approximately once every 36 hours or greater and give rise to a zone of rapidly dividing transit or daughter cells in the region immediately superior to the stem cell niche. There are six or greater generations of transit cells per crypt, and the cell cycle takes 20–73 hours to complete in this population. Cumulatively, a human's small intestinal crypts produce approximately 50×10^6 cells/minute, a total of 0.25 kg/day.

Although human studies examining how a multipotent stem cell gives rise to multiple differentiated villus cells are obviously difficult to perform, experiments performed in mice demonstrate that this is an extremely complex process. The daughter transit cells resulting from the stem cell vary in both their life span as well as in the type of cells they can produce. Daughter cells can be short-lived, with a life span of less than 10 days, or long-lived, with a life span of over 100 days. In addition, some transit cells can parent both columnar-shaped enterocytes and mucus-producing goblet cells, whereas other types can produce only a single cell type.

As animals age, their intestines become larger. As such, intestinal homeostasis is a process in continual evolution, with production of both new crypts and new villi as well as continued division of existing crypt cells and migration into existing villi. The crypt/villus ratio increases over time, with a disproportionate increase in new crypts. Up to 10% of crypts are estimated to be dividing at any given time, after the number of cells contained inside reaches a certain, as yet unknown, threshold.

Stem cells inside a crypt are capable of producing all four cell lineages of the small intestinal epithelium. Three cell types (absorptive enterocytes, mucus-producing goblet cells, and enteroendocrine cells) migrate toward the villus apex. As cells travel up the crypt, they begin to differentiate into their final form. Once an epithelial cell passes the crypt–villus junction, it is fully committed into its mature state and can no longer divide. It then takes between 3 and 5 days for a differentiated cell to migrate to the villus tip. Epithelial cells are removed from the villus primarily by extrusion into the intestinal lumen, although a small subset undergoes apoptosis before reaching the villus apex. Proliferation studies of both mouse and human intestines suggest that the entire epithelium of the small intestine renews every 3 to 5 days.

The fourth small intestinal cell lineage, Paneth cells, differentiate as they migrate toward the base of the crypt, where they reside for approximately 3 weeks. Paneth cells produce lysozyme and other antimicrobial factors. The downward migration of this cell lineage takes approximately 5 to 8 days, and cells are ultimately eliminated by either apoptosis or by being phagocytosed by macrophages within the extracellular matrix or basement membrane.

COLON AND RECTUM

The function of the large intestine is absorption of the water in what remains of the ingested nutrients after their passage through the small intestine. The large intestine is also responsible for helping to maintain the balance of sodium and other minerals within the body. Like the small intestinal and gastric epithelium, the colonic epithelium is columnar in organization. Unlike the small intestine, the colon has no formal villi to reach into the lumen—each unit ends in a hexagonal cuff of cells ringing the opening of the crypt.

The epithelium of the large intestine has three cell types—mucus-producing goblet cells, enteroendocrine cells, and columnar cells, which are also referred to as colonocytes. Because no Paneth cells exist in the large intestine, stem cells are located at the base of the colonic crypt, with transit cells found above. Like the small intestine, multipotent stem cells can ultimately produce

all three cell types found in the colonic epithelium. However, colonic crypts are much larger than their small intestinal counterparts, with an average of 2250 cells per crypt. Stem cells divide every 36 hours or greater, and daughter transit cells take 36–96 hours to divide. The proliferative region occupies the basal two-thirds of a colonic crypt, and there are greater than five to nine daughter transit cell generations, although the exact number is not known. The top third of the crypt is the zone of maturation, with the luminal surface being the functional zone of the crypt. The epithelial surface of the colon is replaced in mice every 8 days. Nearly all cells exit the colonic epithelium through extrusion into the lumen.

Organization of the rectal epithelium is identical to that seen in the colon. The epithelial layer is flat, without villi, and the proliferative zone comprises the basal two-thirds of the crypt, with stem cells located at the extreme base leading to production of columnar, goblet, and enteroendocrine cells. Cells at the top of the rectal crypt are eliminated by being sloughed into the gut lumen. Murine studies indicate that the rectal epithelium is replaced every 4–5 days.

FACTORS INFLUENCING EPITHELIAL PROLIFERATION

The presence and passage of food through the gastrointestinal tract are important stimuli for epithelial proliferation and thickness. Animal data suggest that mechanical passage of food through the esophagus stimulates cell production in the basal layers of the esophageal epithelium. Direct feeding into the stomach also causes an increase of esophageal epithelial proliferation over baseline via the hormone gastrin, suggesting that factors produced in the stomach can stimulate epithelial proliferation in distant anatomic locations. However, direct gastric feeding results in a less robust proliferative response compared to normal ingestion and illustrates that mechanical passage of food plays an important role in esophageal proliferation. Experiments in mice demonstrate that bypassed sections of the small intestine undergo substantial hypoplasia compared to those sections that receive luminal nutrition. This point is further supported by animal models that show general hypoplasia throughout the gastrointestinal tract during intravenous feeding. Regrowth of an atrophied epithelium occurs rapidly when luminal nutrition is reintroduced.

Injuries to the gastrointestinal epithelium may also affect proliferation. Brief treatment of rat gastric mucosa with 100% ethanol denudes the epithelium. Within moments after stopping the ethanol treatment, new epithelial cells begin to migrate toward the gastric lumen. An intact epithelial layer is reestablished within 1 hour of injury; this phenomenon is called restitution. Restitution has also been described in the colon of rabbits and humans.

Critical illness can affect gastrointestinal epithelial proliferation in disparate ways. Both *Pseudomonas aeruginosa* pneumonia and burn injury in mice cause decreases in small intestinal proliferation as measured by BrdU. These animals also have simultaneously elevated epithelial apoptosis resulting from their injuries. In contrast, a mouse model of ruptured appendicitis causes increased small intestinal epithelial proliferation in rats measured with tritiated thymidine.

A number of molecular mediators are also involved in the control of epithelial proliferation. Epithelial growth factor (EGF) is a strong inducer of epithelial proliferation and maturation in the small and large intestines, and also plays a role in intestinal development in both fetuses and infants. Transforming growth factor (TGF) also increases epithelial proliferation, although in a less potent fashion than EGF. Prostaglandins increase proliferation and influence cell migration in the gastric epithelium. Prostaglandins also impart a protective benefit on gastric epithelium and appear to play a role in gastric mucosal wound healing. Prostaglandins appear to stimulate proliferation in the small intestine but have no effect on proliferation in the colon. Gastrin, a hormone produced by the gastric epithelium, stimulates proliferation in the esophageal epithelium as previously outlined. Cells neighboring wounds within the gastrointestinal epithelium secrete trefoil peptides, small stable signal proteins with a distinctive three-dimensional configuration. Trefoil peptides are thought to be important in the process of restitution by influencing cell migration and possibly stimulating proliferation. Corticosteroids reduce both the rate of proliferation and cell migration within the gastrointestinal epithelium and may play a role in the development of ulcers in this tissue.

DISEASES OF THE GASTROINTESTINAL EPITHELIUM INVOLVING ABNORMAL PROLIFERATION

Although the proliferation rate is high in both the small intestine and the colon, colorectal cancer is the third most common type of neoplasia in the United States but small intestinal cancer is rare. The differential sensitivities of the two organs to crypt apoptosis are one potential explanation for this paradox. Theoretically, decreased programmed cell death coupled with

increased or abnormal proliferation could yield an increase in cancer risk. Evidence that supports this involves increased expression of the antiapoptotic protein, Bcl-2, in the stem cell compartments of the colonic and rectal epithelium, compared to the stem cell compartments in the small intestine, and the fact that it is much more common to see apoptotic crypt cells in the small intestine than in the colon. Some carcinogenic agents can also extend the proliferative zone into the functional zone of the colonic crypt or induce the production of an increased proliferative cell population.

Ulcerative colitis is primarily a disease of the rectal epithelium that may spread proximally into the colon. Symptoms of ulcerative colitis include bloody diarrhea and a thinning of the rectal mucosa. The rectal mucosa has an increased proliferative rate in ulcerative colitis, leading to a reduction in turnover time for the rectal epithelium from the typical 4–5 days down to 2–3 days. In addition, the proliferative zone of the affected rectal or colonic crypt extends to include the entire crypt, in contrast to the basal two-thirds seen in normal epithelium.

Gastric metaplasia involves both an increase in proliferation and an abnormal pattern of gastric epithelium. In this disorder, the gastric epithelium undergoes a shift in appearance to one resembling the small intestine, and the proliferation zone extends toward the base of the gastric gland. This enlargement of the zone of proliferation is accompanied by a marked increase in proliferation to two to three times the normal rate. Small intestine epithelial cell lines also begin to appear, with both Paneth and goblet cells present in the gastric epithelium. This can ultimately lead to gastric dysplasia and invasive cancer.

SUMMARY

The gastrointestinal epithelium is one of the most rapidly dividing organs in the body. In keeping with the gut's multiple functions, its epithelium is rapidly renewable and replaces itself on a routine basis. Multipotent stem cells located in the proliferative region of the gastrointestinal tract produce daughter or transit cells that eventually mature into each of the specialized cell types that make up the functional portion of the gastrointestinal epithelium. Because gut stem cells can differentiate into a number of cell types, they have attracted substantial research interest both for treatment of disorders of the gastrointestinal tract and for answering basic scientific questions about how epithelial cells proliferate. Further understanding of stem cell function and signaling may allow future manipulations of gut epithelial proliferation for therapeutic gain.

Acknowledgments

This work was supported by National Institutes of Health grants GM66202 and GM00709 (to C. M. C.).

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Epithelial Barrier Function • Epithelium, Repair of • Gastrin • Gastrointestinal Matrix, Organization and Significance • Growth Factors • Parietal Cells • Transforming Growth Factor-β (TGF-β)

Further Reading

Biasco, G., Paganelli, G. M., Miglioli, M., and Barbara, L. (1992). Cell proliferation biomarkers in the gastrointestinal tract. *J. Cell Biochem.* **16G**, 73–78.

Bjerkenes, M., and Cheng, H. (1999). Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7–14.

Brittan, M., and Wright, N. A. (2002). Gastrointestinal stem cells. *J. Pathol.* **197**, 492–509.

Cheng, H., and Bjerknes, M. (1982). Whole population cell kinetics of mouse duodenal, jejunal, ileal, and colonic epithelia as determined by radioautography and flow cytometry. *Anat. Rec.* **203**, 251–264.

Cheng, H., and Leblond, C. P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian theory of the origin of the four epithelial cell types. *Am. J. Anat.* **141**, 537–561.

Clatworthy, J. P., and Subramanian, V. (2001). Stem cells and the regulation of proliferation, differentiation and patterning in the intestinal epithelium: Emerging insights from gene expression patterns, transgenic and gene ablation studies. *Mech. Dev.* **101**, 3–9.

Eastwood, G. L. (1995). A review of gastrointestinal epithelial renewal and its relevance to the development of adenocarcinomas of the gastrointestinal tract. *J. Clin. Gastroenterol. Suppl.* **1**, S1–S11.

Karam, S. M. (1999). Lineage commitment and maturation of epithelial cells in the gut. *Front. Biosci.* **4**, D286–D298.

Mills, J. C., and Gordon, J. I. (2001). The intestinal stem cell niche: There grows the neighborhood. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12334–12336.

Potten, C. S. (1998). Stem cells in gastrointestinal epithelium: Numbers, characteristics and death. *Philos. Trans. R. Soc. Lond., B., Biol. Sci.* **353**, 821–830.

Potten, C. S., Kellet, M., Rew, D. A., and Roberts, S. A. (1992). Proliferation in human gastrointestinal epithelium using bromodeoxyuridine *in vivo*: Data for different sites, proximity to a tumour, and polyposis coli. *Gut* **33**, 524–529.

Squier, C. A., and Kremer, M. J. (2001). Biology of oral mucosa and esophagus. *J. Natl. Cancer Inst. Monogr.* **29**, 7–15.

Van Nieuwenhove, Y., Chen, D., and Willems, G. (2001). Postprandial cell proliferation in the esophageal epithelium of rates. *Regul. Pept.* **97**, 131–137.

Wong, W. M., and Wright, N. A. (1999). Cell proliferation in gastrointestinal mucosa. *J. Clin. Pathol.* **52**, 321–333.

Epithelium, Repair of

SHAUN V. WALSH
Ninewells Hospital, Dundee, United Kingdom

columnar epithelium A classification of epithelium composed of a type of cell whose height is larger than its width.
restitution The ability of epithelium to rapidly replace or repair sections of denuded epithelial tissue.
squamous epithelium A flat type of epithelium classified by a type of cell whose width is much larger than its height.

Epithelial repair or restitution is the process by which small superficial defects in a mucosal surface are sealed by coordinated epithelial cell migration. Restitution occurs throughout the gastrointestinal tract as a necessary response to noxious injury in both simple columnar and stratified squamous epithelia. Adequate restitution prevents bacteria and other foreign antigens from invading deeper mural tissues. Inadequate restitution, induced either by persistent noxious stimuli or by failure of epithelial responses, contributes to the production of persistent and larger breaches in the mucosal barrier and clinically significant ulcerative diseases. Therefore, restitution is central to the pathophysiology of many common ulcerative disorders including gastroesophageal reflux disease, duodenal ulceration, enteric infections, idiopathic inflammatory bowel disease, and ischemia. For the purposes of this article, epithelial restitution in each compartment of the gastrointestinal tract will be considered separately. In each section, existing knowledge concerning coordinated epithelial cell migration will be considered, underlying molecular mechanisms will be discussed, migration-enhancing and -inhibiting factors will be summarized, and the contribution of stromal factors will be assessed where appropriate.

EPITHELIAL RESTITUTION IN THE ESOPHAGUS

The human esophagus is lined by a stratified squamous epithelium. Numerous exogenous and endogenous stimuli may damage this barrier, most commonly reflux of low pH stomach contents in gastroesophageal reflux disease (GERD). The presence of such gastric contents is directly toxic to the esophageal epithelium, producing small defects or erosions in the mucosal surface and exciting an inflammatory response including intraepithelial eosinophils. These erosions heal by coordinated squamous epithelial cell migration into the eroded areas and subsequent basal cell proliferation to regenerate the normal barrier.

Most studies of epithelial restitution, including studies of esophageal epithelium, have relied on simple *in vitro* culture models of rabbit esophageal cells. Wounds in epithelial monolayers are made mechanically, thereby stimulating neighboring epithelial cells to migrate in coordinated sheets with preserved intercellular junctions into the wound space. Experiments using such a model have demonstrated that epithelial restitution decreases markedly in direct response to a falling pH and increased length of time of acid exposure. These conditions closely model the conditions of the lower esophageal epithelial surface in patients with GERD. This profound inhibition of restitution might explain why prolonged acid suppression is necessary for healing patients with esophagitis and ulcer disease.

At the molecular level, the role of growth factors such as hepatocyte growth factor (HGF) and epidermal growth factor (EGF) is controversial. Initial *in vitro* studies had demonstrated restitution-enhancing effects of HGF and EGF in a rabbit esophageal model, noting a greater sensitivity to these stimuli in esophageal epithelia than in gastric epithelia. However, a subsequent study, also of cultured rabbit esophageal epithelial monolayers, failed to detect a restitution-enhancing effect from HGF or EGF but instead detected enhanced cell proliferation. Of interest, tumor growth factor-β_1 (TGF-β_1) inhibited both restitution and cell proliferation. Overall, given evidence in other epithelial systems, it seems highly likely that several cytokines and growth factors play a critical role in modulating restitution. The likely complex interplay between restitution-enhancing and restitution-inhibiting growth factors has not been unraveled. More critically, the identity of central regulatory factors, likely targets for therapeutic intervention, is not known.

In contrast, the role of epithelial–stromal interactions has not been studied in the esophagus and specific proteins crucial in cell migration, such as the focal adhesion complex proteins, have not yet been intensively studied. Such investigations will be necessary

 Copyright 2004, Elsevier (USA). All rights reserved.

to further explore the role of restitution in the pathophysiology of GERD.

EPITHELIAL RESTITUTION IN THE STOMACH

The human stomach is lined by a simple mucus-producing columnar epithelium. Although well adapted to the low pH environment of the stomach lumen, the gastric mucosa nevertheless is commonly affected by toxic stimuli that produce mucosal defects or gastric erosions. Such noxious stimuli may be locally derived, such as chronic active gastritis driven by *Helicobacter pylori* or by exogenous chemical injury such as that caused by alcohol, nonsteroidal anti-inflammatory drugs (NSAIDs), and bile reflux. Failure to achieve restitution can result in the development of large or confluent erosions and eventually deep ulcer formation.

Studies performed in rats demonstrated that superficial erosions heal by migration of flat to low cuboidal epithelial cells from within gastric pits to the gastric surface. This process is rapid with migrating cells from pits extending lamellipodia over denuded basal lamina and moving at approximately 1–2 μm/min. These studies correlate well with *in vitro* studies of frog mucosa that demonstrate flattened cells migrating from gastric pits during restitution. Typical junctional complexes were present between adjacent cells. Subsequent studies demonstrated that lumenal ammonia, potentially derived from *H. pylori*, retarded restitution of guinea pig gastric mucosa *in vitro*.

Growth factors, as in other compartments of the gastrointestinal tract, have been associated with enhancing gastric epithelial cell migration and thus restitution. Acceleration of gastric epithelial restitution by platelet-derived growth factor (PDGF) has been demonstrated *in vitro*. HGF has also been shown to be a potent stimulant of rabbit gastric epithelial cell migration and proliferation, possibly via cyclooxygenase 2. In addition, EGF has been found to contribute to gastric mucosal restitution via stimulation of basolateral EGF receptors. EGF effects are at least in part produced by the stimulation of basolateral Na^+/H^+ exchangers, by formation of actin stress fibers, and via effects on focal adhesion complex proteins, notably focal adhesion kinase (FAK) and tensin. In culture models of canine oxyntic mucosa EGF, transforming growth factor-α (TGF-α), insulin-like growth factor-I, and interleukin-1 all enhanced cell migration and restitution. Finally, cytokines, such as PDGF derived from subepithelial fibroblasts, may not only influence epithelial function but also may have crucial autocrine effects on stromal cells during larger wound repair. It is likely that further growth factors and cytokines will be shown to regulate gastric restitution.

Although all of the early studies noted the coordinated, cohesive nature of gastric epithelial cell migration, few studies have concentrated on cell–cell adhesion and cell–cell communication during restitution. Increased expression of E-cadherin during cell migration and inhibition of cell migration by anti-E-cadherin antibodies demonstrated the critical role of this intercellular adhesion molecule in a culture model of rat gastric epithelial restitution. However, simple up-regulated linkage between migrating gastric epithelial cells is not the whole story. Inhibition of intercellular communication via gap junctions significantly inhibits gastric epithelial restitution in rats. This correlated with abnormal patterns of gap junction expression as evidenced by connexin 32 expression. Intercellular communication possibly via gap junctions may be necessary to coordinate the sheet-like movement of cells observed during migration.

The extracellular matrix also plays an important role in gastric epithelial restitution. This is adequately demonstrated by elegant studies of the differing effects of matrix components on gastric epithelial cell migration. At a molecular level, focal adhesion complexes serve to anchor the actin cytoskeleton of epithelial cells to the extracellular matrix. Focal adhesion complexes also serve as potent signal transducers between the extracellular matrix and the intracellular milieu. As noted above, the restitution-enhancing effects of EGF are associated with effects on focal adhesion complexes. Furthermore, the NSAID indomethacin delayed gastric restitution in association with inhibition of FAK phosphorylation. Clearly, extensive cross talk between gastric epithelial cells and their underlying matrix is vital in gastric epithelial restitution whether it occurs via direct stromal focal adhesion complex interaction or via other paracrine effectors, such as growth factors.

One of the most exciting fields of study in gastrointestinal epithelial restitution is the function of trefoil peptides. Their roles are well characterized in intestinal restitution but their role in the stomach is less certain. Trefoil peptides have been shown to stimulate canine gastric epithelial restitution *in vitro*. However, trefoil factor family 2 (TFF2) does not appear to affect restitution in studies of knockout mice. Instead, TFF2 influences gastric proliferation, acid secretion, and susceptibility to NSAID injury. Finally but importantly, a recent study has analyzed the energy dependence of gastric epithelial restitution and determined that the glycolytic pathway is essential for restitution after injury in the bullfrog gastric mucosa and that almost all repair

occurs in the absence of mitochondrial respiration, serving to highlight the rapidity of restitution. Pulling these various strands of information together, it is clear that cytokine stimulation, epithelial–stromal interactions, and efficient cell–cell communication are all likely to be required for gastric epithelial restitution. It is hoped that some of these recent advances in the understanding of gastric epithelial restitution and in particular the precise molecular mechanisms by which noxious stimuli affect restitution may well stimulate the discovery of novel therapies for gastric mucosal erosions and ulcers.

EPITHELIAL RESTITUTION IN THE INTESTINE

In direct contrast to the esophagus and stomach, an enormous volume of research has been performed on mechanisms of epithelial restitution in the intestine. This is no doubt due to the numerous diseases that produce mucosal erosions in the intestine, including enteric infection, ischemia, and idiopathic inflammatory bowel disease. It is not possible in this article to deal with the many and varied avenues of research into the intestinal restitution process and therefore certain general areas of promising research will be highlighted with reference to existing excellent reviews of the subject matter.

The mucosa of the intestines serves a vital function as a barrier between the highly noxious potentially pathogenic environment of the intestinal lumen and the vulnerable deeper tissues. Breaches in this epithelial barrier must be rapidly resealed by coordinated epithelial cell migration from surrounding uninjured mucosa in order to prevent the development of serious disease. Due to the wide variety of potential noxious stimuli, it is not surprising that the human small bowel and intestines have evolved a variety of mechanisms to protect against mucosal breaches and affect their rapid resealing.

Morphology of Epithelial Restitution in the Colon

No matter what the cause of the epithelial injury, areas of lamina propria denuded of epithelial cells provoke a rapid response from surrounding colonic epithelial cells. These epithelial cells flatten out in a coordinated fashion and migrate from all sides in a sheet to re-cover the exposed areas of lamina propria. Histopathologically, snapshots of this process may be identified in routine sections of erosions detected in biopsy material from patients with idiopathic inflammatory bowel disease or ischemia. In damaged colonic crypts, epithelial cells assume a flattened-out phenotype and extend long thin processes into the denuded areas. Importantly, they do not separate from their adjacent viable epithelium. The use of experimental models to analyze the molecular mechanism underlying these histopathological appearances will now be examined.

Intestinal Epithelial Restitution: Models of Repair

Most models of intestinal epithelial restitution are culture-based models composed of a monolayer of intestinal epithelial cells grown over a collagen-coated porous membrane. The existing commonly utilized models, both *in vivo* and *ex vivo*, are eloquently summarized by Wilson and Gibson in their review. The authors extensively discuss the benefits of simple *in vivo* model systems of intestinal epithelial cell migration. Such models allow the dissection of complex molecular pathways without the confounding and potentially confusing issues of host–pathogen or epithelial–stromal interactions. Then again, the limitations of using transformed cell lines that, however well differentiated, cannot precisely mimic normal epithelium is recognized. In addition, the particular difficulties surrounding the growth of model epithelia on matrix components are noted.

Intestinal Epithelial Restitution: Role of Peptides

A massive body of work now exists on the role of peptides in the modulation of intestinal epithelial restitution. These are excellently summarized in the review by Wilson and Gibson already mentioned but also in the excellent reviews by Podolski and by Dignass. Briefly, the role of peptide growth factors in intestinal epithelial restitution is now thought to be crucial. Numerous studies have recognized the central roles of TGF-β in this process. Importantly, numerous other restitution-enhancing growth factors, such as TGF-α, EGF, HGF, fibroblast growth factor, interleukin-1, interleukin-2, and interferon-γ, have been demonstrated to enhance restitution through a TGF-β-dependent pathway. However, the relative importance of each individual growth factor in enhancing restitution in a particular disease setting, such as idiopathic inflammatory bowel disease, remains to be determined.

Of equal importance in intestinal epithelial restitution is the role of trefoil factors or peptides. Members of this peptide family are differentially expressed throughout the gastrointestinal tract. Numerous investigators have determined the critical role of these peptides in recovery for mucosal injury and it appears that their effects are entirely independent of the TGF-β pathway.

Of great clinical interest is the fact that trefoil production is enhanced in sites as diverse as the stomach, small bowel, and colon. Dysregulation of trefoil peptide production has been demonstrated in both peptic ulceration and Crohn's disease. It is likely therefore that these peptides may provide useful therapeutic targets in healing such diseases.

Intestinal Epithelial Restitution: Role of Nonpeptide Factors

Again the reader is referred to the excellent reviews noted above. A wide variety of nonpeptide factors have been shown to modulate intestinal epithelial restitution. These include short-chain fatty acids, glutamine, nucleotides, bile salts, lysophosphatidic acid, trace elements, and numerous pharmacological agents. The extensive list of nonpeptide regulators serves as a reminder that, in the *in vivo* setting, each individual factor is a small part of a very complex puzzle. It has not been possible using *in vitro* models to assess in detail the complex interactions between numerous peptide and nonpeptide regulatory factors in governing epithelial restitution. Such studies are keenly awaited.

Intestinal Epithelial Restitution: Role of the Extracellular Matrix

Another fertile area of study has been the interaction between the intestinal epithelial cell and its supportive matrix. This has been excellently summarized in a recent review by Basson. This critical review notes that the extracellular matrix may influence restitution not only as a physical substrate but also by modulating relevant intracellular proteins, such as focal adhesion complex proteins, and by modulating the receptors for soluble factors derived from the extracellular environment. In this process, it appears that FAK plays a central role in integrating signals derived from the extracellular matrix. These signals are modulated in order to produce changes necessary to affect a phenotypic shift from a tall stationary columnar phenotype to a migratory flattened phenotype. Indeed, it is the rapid assembly and disassembly of focal adhesion complexes to integrins and thus to the extracellular matrix on one hand, and their linkage to the actin cytoskeleton providing contractile force on the other, that may serve as a trigger point for the influences of the many modulators of intestinal epithelial cell migration whether matrix or cytokine derived. Further studies have emphasized the role of subepithelial myofibroblasts enhancing epithelial migration. On a cautionary note, due to the difficulty of manipulating the extracellular matrix of the human intestinal lamina propria *in vivo*, the vast majority of studies of epithelial–stromal interaction are performed *in vitro*.

Intestinal Epithelial Restitution: Molecular Mechanisms

Due to the enormous variety of factors, both peptide and nonpeptide, that have been shown to modulate intestinal epithelial cell restitution, it is not surprising that the downstream molecular mechanisms by which these factors affect their responses are of great interest. Influencing a final common pathway governing cell migration would be of great therapeutic value. The role of focal adhesion complexes as signal integration sites is discussed above. Not surprisingly, the role of actin, critical in the formation of lamellipodia and intimately related to focal adhesion complexes, has been emphasized by several authors.

Similarly, the extensive junctional complexes between intestinal epithelial cells including the adherence junction complex may also serve as critical sites for the integration of numerous signals during restitution. Interestingly, recent studies have suggested that TFF3-modulated epithelial cell adhesion and migration occur via perturbing complexes between E-cadherin and β-catenin. The role of tight junctions and gap junctions in this regard remains to be studied. However, as epithelial migration occurs in a coordinated fashion, some intercellular signaling via gap junctions is to be expected. This has already been discussed with regard to gastric epithelium restitution. The role of the tight junction is less clear as preservation of mucosal integrity depends on the preservation of an intact tight junction during coordinated epithelial cell migration. However, this does not exclude the possibility of modulation of such junctions during migration. It could certainly be hypothesized that greater tractional force is exerted on tight junctions during monolayer migration than in the stationary setting. The expanding list of tight junction component proteins may yet contain critical agents necessary for reinforcing or otherwise altering tight junction function during migration. Clearly, much further study of junctions during migration is required.

CONCLUSION

Epithelial restitution throughout the gastrointestinal tract displays some recurring themes. First, restitution is a rapid, energy-dependent process. It occurs in a coordinated fashion and therefore inherently involves cell–cell adhesion complexes and molecules. Furthermore, the nature of epithelial–stromal interactions is crucial in modulating the restitution response. Finally,

peptide growth factors, trefoil peptides, and numerous other exogenous factors provide a busy information highway to the cell in regulating migration and thus restitution. Critical junctions on this highway at which integration of information occurs include focal adhesion complexes and intercellular junctions. A better understanding of these complex pathways may lead to the development of novel therapeutic strategies to enhance epithelial restitution and thus alleviate patient diseases.

See Also the Following Articles

Colitis, Ulcerative • Crohn's Disease • Duodenal Ulcer • Epithelial Barrier Function • Epithelium, Proliferation of • Gastric Ulcer • Gastroesophageal Reflux Disease (GERD) • Gastrointestinal Matrix, Organization and Significance

Further Reading

Basson, M. D. (2001). *In vitro* evidence for matrix regulation of intestinal epithelial biology during mucosal healing. *Life Sci.* **69**, 3005–3018.

Dignass, A. (2001). Mechanisms and modulation of intestinal epithelial repair. *Inflamm. Bowel Dis.* **7**, 68–77.

Farrell, J. J., Taupin, D., Koh, T. J., *et al.* (2002). TFF2/SP deficient mice show decreased gastric proliferation, increased acid secretion and increased susceptibility to NSAID injury. *J. Clin. Invest.* **109**, 193–204.

Jimenez, P., Lanas, A., Piazuelo, E., and Esteva, F. (1999). Effects of extracellular pH on restitution and proliferation of rabbit oesophageal epithelial cells. *Aliment. Pharmacol. Ther.* **13**, 545–552.

Jimenez, P., Lanas, A., Piazuelo, E., and Esteva, F. (1998). Effects of growth factors and prostaglandin E2 on restitution and proliferation of rabbit esophageal epithelial cells. *Digest. Dis. Sci.* **43**, 2309–2316.

Piazuelo, E., Lanas, A., Jimenez, P., *et al.* (1998). *In vitro* wound repair by human gastric fibroblasts: Implications for ulcer healing. *Digest. Dis. Sci.* **43**, 1230–1240.

Podolsky, D. K. (1999). Mucosal immunity and inflammation. V. Innate mechanisms of mucosal defense and repair: The best offense is a good defense. *Am. J. Physiol.* **277**, G495–G499.

Rhodes, D., Revis, D., and Lacy, E. R. (1994). Extracellular matrix constituents affect superficial gastric epithelial cell adhesion. *J. Gastroenterol. Hepatol.* **9**, S72–S77.

Suzuki, H., Yanaka, A., and Muto, H. (2000). Luminal ammonia retards restitution of guinea pig injured gastric mucosa. *Am. J. Physiol.* **279**, G107–G117.

Szabo, I. L., Pai, R., Jones, M. K., *et al.* (2002). Indomethacin delays gastric restitution: Association with the inhibition of focal adhesion kinase and tensin phosphorylation and reduced actin stress fibres. *Exp. Biol. Med.* **227**, 412–424.

Takahashi, M., Ota, S., Ogura, K., Nakamura, T., and Omata, M. (1995). Hepatocyte growth factor stimulates wound repair of rabbit esophageal epithelial cells in primary culture. *Biochem. Biophys. Res. Commun.* **2**, 298–305.

Terano, A., Sakata Horie, K., Shimada, T., *et al.* (2001). The role of cellular migration in the repair process of gastric epithelial cells. *Life Sci.* **69**, 3083–3089.

Watanabe, S., Wang, X. E., Hirose, M., *et al.* (1996). Platelet derived growth factor accelerates gastric epithelial restoration in a rabbit cultured model. *Gastroenterology* **110**, 775–779.

Wilson, A. J., and Gibson, P. R. (1997). Epithelial migration in the colon: Filling in the gaps. *Clin. Sci.* **93**, 97–108.

Yanaka, A., Suzuki, H., Shibahara, T., *et al.* (2002). EGF promotes gastric mucosal restitution by activating Na(+)/H(+) exchange of epithelial cells. *Am. J. Physiol.* **282**, G866–G876.

Erosive and Hemorrhagic Gastritis (Gastropathy)

RAJEEV JAIN
Presbyterian Hospital of Dallas

gastritis Gastric mucosal disorder associated with an inflammatory infiltrate.
gastropathy Gastric mucosal disorder without significant microscopic inflammatory infiltrate.

The classification of erosive and hemorrhagic gastritis (gastropathy) is based on endoscopic findings related to acute chemical or irritant injury to the stomach without associated microscopic inflammation. Depending on the etiologies of erosive and hemorrhagic gastritis, it is recommended that diagnoses be further defined (e.g., reactive gastritis, chemically induced gastritis).

HISTOLOGY

Although the term "gastritis" is often used to both describe clinical symptoms and endoscopic findings, the histologic definition of acute gastritis is mucosal injury associated with an inflammatory neutrophilic infiltrate. Chronic gastritis is characterized by an inflammatory infiltrate of mononuclear cells such as lymphocytes, plasma cells, and macrophages. Causes of gastritis include *Helicobacter pylori* and other infectious agents, allergic and hypersensitivity reactions, or drugs. In comparison, erosive and hemorrhagic gastritis (gastropathy) denotes endoscopic findings in which mucosal injury has occurred but no significant inflammatory infiltrate is seen microscopically. In response to mucosal injury, the gastric mucosa can display cellular regeneration with reactive changes of foveolar hyperplasia. Erosions are superficial mucosal breaks involving necrosis that does not extend beyond the muscularis mucosa, whereas ulcers penetrate through the muscularis mucosa.

ENDOSCOPY

The endoscopic findings of the gastropathy of erosive and hemorrhagic gastritis vary from petechial hemorrhages to erosions and ulcers. The hemorrhagic lesions can vary in intensity and distribution based on the causative agent. The erosive changes are typically manifested by multiple erosions that can progress to ulcer formation.

The endoscopic and histologic findings of gastritis are frequently discordant. For example, the endoscopic findings suggestive of inflammation are frequently incorrect when gastric biopsies are examined. Also, the mucosa can appear normal at endoscopy but biopsies reveal inflammation. Therefore, when describing endoscopic findings, the term "gastropathy" is preferred over "gastritis."

ETIOLOGIES

Alcohol

Acute ingestion of alcohol can cause subepithelial hemorrhages in the stomach. The histological characteristics of acute alcohol injury include superficial hemorrhage and edema. The correlation between chronic alcoholism and gastric mucosal injury is unclear because of factors such as *H. pylori* infection and concomitant aspirin and/or nonsteroidal antiinflammatory drug use.

Antineoplastic Therapy

Radiation injury to the stomach affects the parietal and chief cells by causing edema and necrosis. Of patients receiving 5500 cGy of radiotherapy to the stomach, 50% will develop ulceration, which typically is solitary and located in the antrum. Chemotherapy agents infused through the hepatic artery can inadvertently injure the stomach and cause gastric ulceration with marked epithelial atypia.

Bile Reflux

Bile salts can reflux from the proximal small intestine into the stomach through an operative stoma or because of an incompetent pylorus or abnormal duodenal motility. Bile acts as a topical irritant, leading to gastric mucosal breakdown. The gastric histologic

 Copyright 2004, Elsevier (USA). All rights reserved.

findings documented in a group of patients with gastroenteric anastamosis-associated bile reflux include both hyperplasia and fibrosis of the lamina propria with foveolar hyperplasia. Endoscopically, the mucosal changes are most prominent in the distal stomach or proximal to the gastroenteric anastamosis. The clinical presentation of bile reflux gastropathy can vary from no symptoms to nausea, vomiting, abdominal pain, and weight loss. Medical therapy with ursodeoxycholic acid, sucralfate, misoprostol, and cholestyramine has shown limited benefit. In some reports, patients with bile reflux related to a gastroenteric anastamosis have improved after undergoing surgery with a Roux-en-Y gastrojejunostomy designed to divert bile from the stomach.

Cocaine

Crack cocaine use can cause exudative erosions throughout the stomach, sometimes with complications of bleeding and perforation. It is believed that these lesions are caused by vasoconstriction and resultant mucosal ischemia.

Ischemia

Ischemic gastropathy has been described in patients with chronic mesenteric ischemia, most typically related to atherosclerosis, atheromatous embolization, and extreme athletic activity, as seen in long-distance runners, who can present with gastrointestinal hemorrhage and anemia.

Mechanical Prolapse

Mechanical injury to the gastric cardia occurs in some patients, with retching and vomiting when the gastric cardia prolapses into the esophageal lumen. At endoscopy, prolapse can be observed, with gastric cardia injury manifested by erosions and ulceration.

Medications

Aspirin and nonsteroidal antiinflammatory drugs (NSAIDs) are the leading etiology of gastropathy. These agents injure the gastric mucosa by both topical and systemic routes. Topical injury occurs because aspirin and many NSAIDs are weak acids and can pass through the gastric mucus barrier into the surface epithelium. Also, some NSAIDs are prodrugs that the liver converts to acidic active drugs, which are secreted in bile. Topical injury from biliary excretion occurs when the active acidic metabolite refluxes into the stomach from the duodenum. Systemic injury is related to prostaglandin inhibition in the gastric mucosa, which causes decreased mucus and bicarbonate secretion and reduces mucosal blood flow, with subsequent impairment of mucosal protection. Within minutes of ingestion of aspirin and NSAIDs, histologic changes of mucosal cell injury can be noted. Endoscopic findings of subepithelial hemorrhages and erosions can be seen within a few hours of ingestion. The entire stomach is vulnerable to injury, with most erosions and ulcers found in the antrum. The primary treatment and prevention of gastropathy is discontinuation and avoidance of the injurious agent. If aspirin or NSAIDs cannot be discontinued, active treatment requires antisecretory therapy with histamine-2 receptor antagonists or proton pump inhibitors. Prophylactic strategies include misoprostol and proton pump inhibitors.

Oral potassium and iron therapy have also been shown to cause gastric petechial hemorrhage and erosions.

Portal Hypertension

The findings of congestive gastropathy in the setting of portal hypertension are recognized at endoscopy as mucosal red spots in a mosaic pattern. This endoscopic appearance is attributed to mucosal congestion and edema, with the grooves surrounding the area gastrica displayed as pale lines, producing the typical mosaic pattern. Microscopically, superficial venular and capillary ectasia with or without organizing microvascular thrombi is seen. Clinically, patients can be asymptomatic or present with occult bleeding and anemia. Nonselective beta-blockers and decompressive shunts have been used to treat refractory cases of bleeding.

Stress

Stress gastropathy is seen in critically ill patients who require mechanical ventilation and/or have a coagulopathy. Other risk factors for stress injury include previous gastrointestinal bleeding, hypotension, trauma, burn injury (>35% of body surface), central nervous system injury, renal failure, hepatic failure, organ transplantation, and sepsis. Erosions and ulcerations develop from a multifactorial process, with the primary insult being gastric ischemia. In endoscopic studies, mucosal injury is found in greater than 75% of critically ill patients within 24 hours of admission. The typical mucosal injury seen is multiple, diffuse, superficial erosions in the gastric fundus and body; however, focal, deep ulceration can also be seen in both the stomach and duodenum. Patients in the intensive care unit can experience significant bleeding, which

is associated with increased mortality. Preventive strategies have focused on acid suppression with histamine-2 receptor antagonists and mucosal protection with sucralfate. Based on the available data, stress ulcer prophylaxis is warranted only in high-risk critically ill patients such as those requiring mechanical ventilation or those with significant coagulopathy. Intravenous histamine-2 receptor antagonists or proton pump inhibitors are the preferred preventive strategy in these high-risk patients.

See Also the Following Articles

Gastritis • H2-Receptor Antagonists • *Helicobacter pylori* • NSAID-Induced Injury • Proton Pump Inhibitors

Further Reading

Carpenter, H. A., and Talley, N. J. (1995). Gastroscopy is incomplete without biopsy: Clinical relevance of distinguishing gastropathy from gastritis. *Gastroenterology* **108**, 917–924.

Chamberlain, C. E. (1993). Acute hemorrhagic gastritis. *Gastrointest. Clin. North Am.* **22**(4), 843–873.

Dixon, M. F., Genta, R. M., Yardley, J. H., *et al.* (1996). Classification and grading of gastritis: the updated Sydney system. *Am. J. Surg. Pathol.* **20**, 1161–1181.

Laine, L., and Weinstein, W. L. (1988). Subepithelial hemorrhages and erosions of human stomach. *Dig. Dis. Sci.* **4**, 490–503.

Lee, E. L., and Feldman, M. (2002). Gastritis and other gastropathies. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 810–821. Saunders, Philadelphia.

Wolfe, M. M., Lichtenstein, D. R., and Singh, G. (1999). Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. *N. Engl. J. Med.* **340**, 1888–1899.

Esophageal Cancer

MARK LEMERT
University of California, San Diego

achalasia Absence of lower esophageal sphincter relaxation, resulting in functional obstruction at the distal esophagus.

Barrett's esophagus Replacement of normal stratified squamous mucosa by columnar-lined epithelium; extends upward from the gastroesophageal junction.

dysphagia Difficulty with swallowing.

gastric esophageal reflux disease Passage of stomach contents into the esophagus, resulting in esophageal mucosal injury. Gastric and possibly bile acids are the important mediators of this injury.

metaplasia Reversible change in which one adult cell type is replaced by another adult cell type; may represent an adaptive substitution of cells in response to environmentally mediated injury; if persistent, cancer transformation may occur.

neoadjuvant Before surgery.

odynophagia Pain with swallowing.

Squamous cell carcinoma and adenocarcinoma represent the two important histologic forms of esophageal cancer. These two tumor types differ significantly in terms of associated risk factors. Tobacco use and alcohol consumption are correlated with squamous cell carcinoma. The critical risk factor for adenocarcinoma is Barrett's esophagus, in which the normal squamous lining of the esophagus has been replaced by intestinal-type columnar cells in response to chronic gastric reflux disease. Overall, the survival for esophageal cancer is poor, because a majority of patients at diagnosis are found to have advanced disease.

INTRODUCTION

Approximately 12,000 people in the United States are diagnosed annually with esophageal cancer. Although considered a comparatively uncommon cancer in Western countries, esophageal cancer represents a major cause of cancer-related mortality worldwide. In

Copyright 2004, Elsevier (USA). All rights reserved.

is associated with increased mortality. Preventive strategies have focused on acid suppression with histamine-2 receptor antagonists and mucosal protection with sucralfate. Based on the available data, stress ulcer prophylaxis is warranted only in high-risk critically ill patients such as those requiring mechanical ventilation or those with significant coagulopathy. Intravenous histamine-2 receptor antagonists or proton pump inhibitors are the preferred preventive strategy in these high-risk patients.

See Also the Following Articles

Gastritis • H2-Receptor Antagonists • *Helicobacter pylori* • NSAID-Induced Injury • Proton Pump Inhibitors

Further Reading

Carpenter, H. A., and Talley, N. J. (1995). Gastroscopy is incomplete without biopsy: Clinical relevance of distinguishing gastropathy from gastritis. *Gastroenterology* **108**, 917–924.

Chamberlain, C. E. (1993). Acute hemorrhagic gastritis. *Gastrointest. Clin. North Am.* **22**(4), 843–873.

Dixon, M. F., Genta, R. M., Yardley, J. H., *et al.* (1996). Classification and grading of gastritis: the updated Sydney system. *Am. J. Surg. Pathol.* **20**, 1161–1181.

Laine, L., and Weinstein, W. L. (1988). Subepithelial hemorrhages and erosions of human stomach. *Dig. Dis. Sci.* **4**, 490–503.

Lee, E. L., and Feldman, M. (2002). Gastritis and other gastropathies. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 810–821. Saunders, Philadelphia.

Wolfe, M. M., Lichtenstein, D. R., and Singh, G. (1999). Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. *N. Engl. J. Med.* **340**, 1888–1899.

Esophageal Cancer

MARK LEMERT
University of California, San Diego

achalasia Absence of lower esophageal sphincter relaxation, resulting in functional obstruction at the distal esophagus.

Barrett's esophagus Replacement of normal stratified squamous mucosa by columnar-lined epithelium; extends upward from the gastroesophageal junction.

dysphagia Difficulty with swallowing.

gastric esophageal reflux disease Passage of stomach contents into the esophagus, resulting in esophageal mucosal injury. Gastric and possibly bile acids are the important mediators of this injury.

metaplasia Reversible change in which one adult cell type is replaced by another adult cell type; may represent an adaptive substitution of cells in response to environmentally mediated injury; if persistent, cancer transformation may occur.

neoadjuvant Before surgery.

odynophagia Pain with swallowing.

Squamous cell carcinoma and adenocarcinoma represent the two important histologic forms of esophageal cancer. These two tumor types differ significantly in terms of associated risk factors. Tobacco use and alcohol consumption are correlated with squamous cell carcinoma. The critical risk factor for adenocarcinoma is Barrett's esophagus, in which the normal squamous lining of the esophagus has been replaced by intestinal-type columnar cells in response to chronic gastric reflux disease. Overall, the survival for esophageal cancer is poor, because a majority of patients at diagnosis are found to have advanced disease.

INTRODUCTION

Approximately 12,000 people in the United States are diagnosed annually with esophageal cancer. Although considered a comparatively uncommon cancer in Western countries, esophageal cancer represents a major cause of cancer-related mortality worldwide. In

Copyright 2004, Elsevier (USA). All rights reserved.

developed countries over the past three decades, a substantial shift in relative incidence of the two major histologic types has been observed. Adenocarcinoma is now diagnosed more frequently than squamous cell carcinoma in the United States and western Europe. The reason for this change remains unclear, but is likely associated with the parallel increase in incidence of gastroesophageal reflux disease and, in turn, Barrett's metaplasia.

Solid food dysphagia is the most common clinical manifestation of esophageal carcinoma. Most patients present with advanced disease, because esophageal cancer appears to metastasize early in disease progression. Overall survival is directly dependent on the stage of disease at diagnosis, regardless of the histologic type of cancer. Five-year survival rates range from approximately 60% in patients with carcinoma *in situ* to less than 4% in patients with distant metastases. Because survival correlates closely to extent of disease, precise clinical staging is a critical component of determining appropriate treatment options. Endoscopic ultrasound has emerged as the most accurate diagnostic modality for locoregional staging. In select patients with limited disease, curative surgical resection represents a potential therapeutic approach.

SQUAMOUS CELL CARCINOMA

Epidemiology and Risk Factors

The incidence of esophageal squamous cell carcinoma varies greatly according to geographic location. The region extending from northern Iran across central Asia to northern China exhibits annual incidence rates exceeding 100 per 100,000 new cases. Additional high-incidence regions included Puerto Rico, South Africa, and India. Men and women appear to be equally effected in these high-incidence regions, whereas, in low-incidence regions, such as the United States and parts of Europe, men are more likely to be affected. In the United States, the annual incidence is approximately 3 to 4 per 100,000. African Americans have a fourfold increased risk as compared to European Americans. Further, most persons diagnosed with squamous cell carcinoma are over the age of 50 years. Such geographic differences in epidemiology suggest that dietary and environmental factors may play a central role in pathogenesis. A majority of cases in the United States and Europe are attributed to alcohol and tobacco consumption, but in endemic regions of Asia, exposure to foods containing carcinogenic nitrosamines and fungal contaminants has been implicated.

The presence of underlying esophageal disease also increases the risk for squamous cell carcinoma. Such disease processes include achalasia, chronic esophageal strictures, and previous caustic injuries to the esophagus. In addition, an association exists with rare diseases such as Plummer–Vinson syndrome, characterized by spoon-shaped concave fingernails, angular stomatitis, and esophageal webs, and tylosis, an autosomal-dominant disorder marked by hyperkeratosis of the soles and palms. As well, the infection of squamous cells with DNA viruses, including Epstein–Barr virus and human papilloma virus, has been linked to malignant transformation.

Pathology

Most squamous cell cancers arise from the middle esophagus. Early lesions may begin as small, focal areas of thickened mucosa. Over a period of months to years, these areas may progress to clinically overt disease (see Fig. 1). Gross features of advanced disease are generally characterized as polypoid, ulcerative, or infiltrating. In terms of spread, squamous cell cancers typically invade to the submucosa early in the disease process. As a consequence of an abundant network of submucosal lymphatics in the esophagus, dissemination to regional lymph nodes occurs rapidly. Common sites of such dissemination include the cervical, paratracheal, mediastinal, and celiac groups. Direct invasion of surrounding structures may also occur with further disease progression. Reported consequences of this process include gastrointestinal bleeding due to erosion into the aorta and aspiration as a result of tracheoesophageal fistula formation.

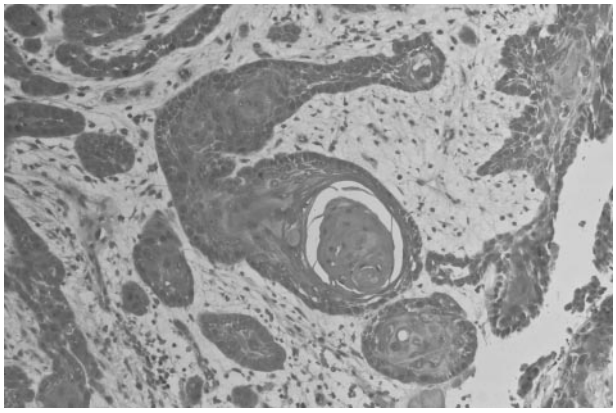


FIGURE 1 High-power microscopic view of a well-differentiated squamous cell carcinoma with a characteristic keratin pearl and invading nests of malignant squamous cells.

ADENOCARCINOMA

Epidemiology and Risk Factors

Several studies using population-based cancer registry data have suggested that incidence rates of adenocarcinomas of the esophagus have been increasing since 1970. This upward trend has been observed particularly in Caucasian men of developed countries such as the United States, Denmark, Sweden, and the United Kingdom. Among Caucasian males in the United States, the incidence of adenocarcinoma surpassed that of squamous cell carcinoma around 1990. Squamous cell carcinoma, however, remains the more common esophageal cancer worldwide.

Affected patients generally present in their middle sixties. Approximately 95% of patients are Caucasian, and males outnumber females 5 : 1. Unlike patients with squamous cell carcinoma, those with adenocarcinoma do not generally have a history of heavy tobacco or alcohol use. A strong correlation with obesity and adenocarcinoma has also been consistently found.

Barrett's esophagus is the most important risk factor for adenocarcinoma. This condition occurs when normal stratified squamous mucosa is replaced by columnar-lined epithelium that extends upward from the gastroesophageal junction (see Figs. 2 and 3). It is an acquired metaplastic process in response to recurrent mucosal injury, commonly the result of long-standing gastroesophageal reflux. Additional undetermined factors clearly play a role in this process, as suggested by findings that fewer than 50% of patients with adenocarcinoma have pathologic evidence of esophagitis, and that neither antireflux pharmacologic therapy nor surgery has been shown to result in reversion of metaplastic epithelium to normal squamous epithelium. Because of the association between Barrett's esophagus and adenocarcinoma, endoscopic screening is recommended in patients with established Barrett's esophagus.

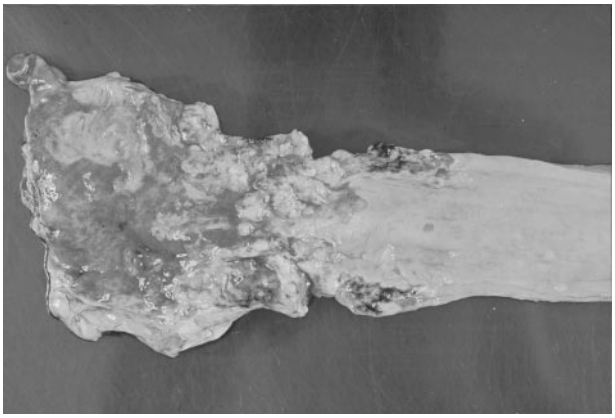


FIGURE 2 Gross view of an ulcerated adenocarcinoma of the distal esophagus, originating from Barrett's esophagus.

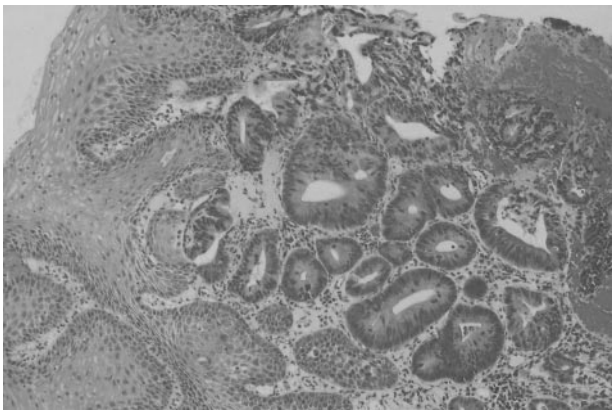


FIGURE 3 Low-power microscopic view of a well-differentiated adenocarcinoma with some distortion of glandular architecture.

Pathology

Adenocarcinomas typically arise from the distal third of the esophagus. Initially, they tend to appear as flat or raised patches of otherwise intact mucosa at or near the gastroesophageal junction. The general morphologic features of advanced disease are similar to those of squamous cell carcinoma; they can be polypoid, ulcerative, or infiltrating. Microscopically, most adenocarcinoma tumors are mucin-producing with intestinal-type features. As with squamous cell carcinoma, metastases to adjacent or regional lymph nodes occur early in the course of the disease process. Direct invasion of surrounding structures can also occur.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Squamous cell carcinoma and adenocarcinoma have similar clinical presentations. Obstruction of the lumen of the esophagus by tumor generally causes progressive food dysphagia. Patients may initially experience an intermittent sensation of friction or slow food passage, which may evolve into complete solid food intolerance. Accompanying weight loss, due to tumor anorexia and poor nutritional intake as a result of dysphagia, suggests advanced disease. The presence of odynophagia or retrosternal discomfort may indicate mediastinal involvement. Hoarseness may occur in the setting of recurrent laryngeal nerve involvement. Chronic gastrointestinal blood loss with iron deficiency anemia as a consequence of luminal epithelium disruption is common. Acute gastrointestinal blood loss is rare

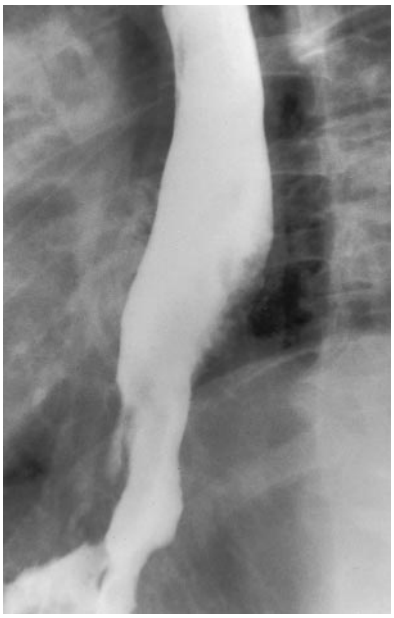


FIGURE 4 Barium esophogram, demonstrating a circumferential adenocarcinoma of the distal esophagus.

but has been reported in the context of tumor invasion into the aorta or pulmonary arteries. Aspiration pneumonias due to tracheobronchial fistula formation can occur as a late manifestation of tumor progression.

A diagnostic workup for esophageal cancer should be initiated in patients with dysphagia in the setting of risk factors for esophageal cancer. Barium studies are useful in detecting findings consistent with advanced lesions, such as high-grade strictures or fistulas, but often fail to identify early esophageal cancers (see Fig. 4). Endoscopy with biopsies is the most effective means of establishing the diagnosis. It must be performed even for patients with normal barium studies due to the limited sensitivity of barium studies. Multiple endoscopic biopsies increase the accuracy of diagnosis; when employed in conjunction with cytology brushings, the diagnostic yield exceeds 90%.

STAGING AND SURVIVAL

The ultimate prognosis is strongly dependent on the stage of disease at diagnosis. An accurate assessment of extent of disease is therefore necessary in the process of formulating treatment strategies. The initial step in staging typically begins with a CT scan of the chest and abdomen to evaluate for evidence of local and distant metastases. Patients without evidence of metastatic disease generally should have endoscopic ultrasound (EUS) in an effort to further delineate the extent of locoregional disease. For patients found to have proximal tumors, bronchoscopy is generally carried out to assess for evidence of spread to the trachea. Bone scanning is indicated for patients with clinical or laboratory findings suggestive of bone metastases.

EUS employs a high-frequency ultrasound transducer to assess the extent of invasion into the esophageal wall, which is depicted as a five-layered structure. EUS is also highly accurate in assessing for involvement of regional lymph nodes. Examination of the celiac nodes is particularly important, because evidence of spread to this group implies distant metastatic disease and therefore unresectability. The use of fine-needle aspiration (FNA) biopsy increases the staging accuracy of EUS by providing cytological evidence. With surgical resection specimens as a gold standard, the sensitivity for EUS FNA exceeds 85% in terms of detecting regional metastatic disease.

The TNM staging system of the American Joint Committee on Cancer (AJCC) for esophageal cancer is widely used (Table I). The important components of the system are extent of tumor invasion (T) and the presence of metastatic (M) disease in regional lymph nodes (N) or solid organs. Metastases in regional lymph nodes are either absent (N0) or present (N1).

TABLE I TMN Staging System for Cancer of the Esophagus

Primary tumor (T)			
TX: Primary tumor cannot be assessed			
T0: No evidence of primary tumor			
Tis: Carcinoma *in situ*			
T1: Tumor invades lamina propria or submucosa			
T2: Tumor invades muscularis propria			
T3: Tumor invades adventitia			
T4: Tumor invades adjacent structures			
Lymph node (N)			
NX: Regional lymph nodes cannot be assessed			
N0: No regional lymph node metastasis			
N1: Regional lymph node metastasis			
Distant metastasis (M)			
MX: Presence of distant metastasis cannot be assessed			
M0: No distant metastasis			
M1: Distant metastasis			
Stage grouping			
0	Tis	N0	M0
I	T1	N0	M0
IIA	T2	N0	M0
	T3	N0	M0
IIB	T1	N1	M0
	T2	N1	M0
III	T3	N1	M0
	T4	N1	M0
IV	Any T	Any N	M1

As well, organ metastases are either absent (M0) or present (M1). Resultant TNM classifications in turn correspond to disease stage. Esophageal cancer usually presents in an advanced stage, and, as a consequence, the overall prognosis is poor. Tumor histology does not appear to have an important impact on prognosis. Survival, moreover, is closely related to extent of tumor invasion and metastatic spread and thereby decreases sharply with increasing stage. Patients with T1 disease are reported to have a 5-year survival rate approaching 46% versus only 7% in patients with T4 disease. When stratified by TNM subtype, the 5-year survival rates for stages I, II, III, and IV disease are 60.4, 31.3, 19.9, and 4.1%, respectively.

TREATMENT STRATEGIES

Surgical resection with the intent to cure represents the only treatment strategy that has repeatedly been shown to result in prolonged survival. Radical surgery is the approach of choice for localized stage I and stage II disease. Surgery does not appear to have a role in advanced disease, particularly with distant metastases. Resection of the esophagus is achieved by various surgical techniques that differ in the type of incision, the extent of resection, the conduit for reconstruction, and the method of reanastomosis. Unfortunately, the majority of patients at presentation are not found to have resectable disease after staging. With potential resection candidates, a careful preoperative assessment of the patient's overall medical and performance status is essential. Even in high-volume centers, postoperative mortality and morbidity rates approach 10 and 75%, respectively.

Since the implementation of surgical resection, many postoperative patients have been found to develop systemic metastases without local recurrence. It has been proposed that this pattern of recurrent disease may be the consequence of present but undetectable metastatic disease at the time of diagnosis. Therefore, to improve overall survival in operative candidates, various chemotherapy and chemoradiation regimens have been investigated as a means of obliterating presumptive micrometastases. In general, clinical trial outcomes have been disappointing, because little additional benefit has been shown to be gained by the use of such treatment modalities. Single-modality chemotherapy or radiation therapy before or after surgery has not been shown to confer a significant survival benefit. There are, however, several small early-phase trials that suggest that neoadjuvant chemoradiation may improve short-term survival over surgery alone.

Because the majority of patients after staging are not candidates for a curative approach, palliation remains an important component to the management of esophageal cancer. In patients with obstructive symptoms, esophageal bypass surgery represents a means of reestablishing gastrointestinal continuity without the associated morbidity and mortality seen with surgical resection. In the less severe situation of esophageal stricture formation, repeated endoscopic dilations or expandable metal stent deployment can be undertaken to maintain patency. In selected patients, additional endoscopic modalities may be effective in treating dysphagia. These include photodynamic therapy, endoscopic laser therapy, and bipolar coagulation. Palliative radiation and chemotherapy also have important roles. Randomized trials have indicated that chemoradiation with cisplatin and fluorouracil-based regimens is superior to radiation alone. This approach has evolved into the standard of care for patients with unresectable disease.

SUMMARY

Over the past 30 years, there have been significant changes in the epidemiology of esophageal cancer. Adenocarcinoma is now the most common type of esophageal cancer in most developed countries. Nevertheless, extent of disease at diagnosis remains the most important prognostic factor. Advances with regard to imaging have greatly enhanced the accuracy of disease staging. In particular, endoscopic ultrasound has evolved into an important means of evaluating for locoregional spread of disease. In the minority of patients without advanced disease, radical surgery is a potential curative therapy. The utility of chemotherapy and/or radiotherapy in patients with either resectable or unresectable disease remains an area of debate. Trials examining new cytotoxic agents in different combined modality regimens are in process.

See Also the Following Articles

Barrett's Esophagus • Cancer, Overview • Dysphagia • Endoscopic Ultrasonography • Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD • Gastroesophageal Reflux Disease (GERD) • Smoking, Implications of

Further Reading

Allum, W., Griffin, S., Watson, A., and Colin-Jones, D. (2002). Guidelines for the management of esophageal Cancer. *Gut* **50**(Suppl. V), V1–V23.

Blot, W., and McLaughlin, J. (1999). The changing epidemiology of esophageal cancer. *Semin. Oncol.* **26**(Suppl. 15), 2–8.

DeMeester, T. (1997). Esophageal carcinoma: Current controversies. *Semin. Surg. Oncol.* **13**, 217–233.
El-Serag, H. (2002). The epidemic of esophageal adenocarcinoma. *Gastroenterol. Clin.* **31**, 421–440.
Entwistle, J., and Goldberg, M. (2002). Multimodality therapy for resectable cancer of the thoracic esophagus. *Ann. Thorac. Surg.* **73**, 1009–1015.
Lerut, T., Coosemans, W., De Leyn, P., Van Raemdonck, D., Deneffe, G., and Decker, G. (1999). Treatment of esophageal carcinoma. *Chest* **116**(Suppl. 6), 463S–465S.
Mallery, S., and Van Dam, J. (2000). EUS in the evaluation of esophageal carcinoma. *Gastrointest. Endosc.* **52**(Suppl. 6), S6–S11.
Meyenberger, C., and Fantin, A. (2000). Esophageal carcinoma: Current staging strategies. Recent results *Cancer Res.* **155**, 55–62.
Miller, D. (2000). Clinical trials for esophageal carcinoma. *Chest Surg. Clin. North Am.* **10**, 583–590.
Reed, C. (1999). Surgical management of esophageal carcinoma. *Oncologist* **4**, 95–105.

Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD

STUART JON SPECHLER
Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

Barrett's esophagus The condition in which an abnormal columnar epithelium replaces the stratified squamous epithelium that normally lines the distal esophagus.
dysplasia A constellation of histological abnormalities that suggest that one or more clones of cells have acquired genetic damage rendering them neoplastic and predisposed to malignancy.
gastroesophageal reflux disease A disease that occurs when the passage of stomach contents into the esophagus results in symptoms or signs such as chest pain, esophagitis, chronic cough, or recurrent pneumonia.

Barrett's esophagus develops when refluxed gastric juice injures the esophageal mucosa and the injury heals through a metaplastic process in which columnar cells replace reflux-damaged squamous cells. Specialized intestinal metaplasia, the epithelium that characterizes Barrett's esophagus, is composed of a variety of columnar cell types that can have gastric, small intestinal, and colonic features. This abnormal columnar epithelium may be more resistant to gastroesophageal reflux disease (GERD) than the native squamous mucosa, but the metaplastic cells are predisposed to malignancy. GERD and Barrett's esophagus are the major recognized risk factors for esophageal adenocarcinoma, a lethal malignancy that has increased profoundly in frequency over the past few decades in Western countries.

INTRODUCTION

The diagnosis of Barrett's esophagus is made by endoscopic evaluation. Columnar epithelium in the esophagus has a characteristic dull, reddish appearance that can be distinguished readily from the glossy, pale squamous epithelium. The diagnosis of Barrett's esophagus is established when biopsy specimens of the esophageal columnar epithelium show specialized intestinal metaplasia. Barrett's esophagus can be further categorized according to the extent of the metaplastic lining. Patients who have >3 cm of specialized intestinal metaplasia have long-segment Barrett's esophagus, a condition that usually is associated with severe gastroesophageal reflux disease (GERD). Patients with <3 cm of metaplasia have short-segment Barrett's esophagus, a disorder often associated with only mild GERD. Although it is not clear whether these two types of Barrett's esophagus have the same risk for carcinogenesis, patients with short- and long-segment Barrett's esophagus are managed similarly.

Published estimates on the annual risk of cancer in patients with Barrett's esophagus have ranged from 0.2 to 2.9%, with an average risk of approximately 1%. The studies on which those estimates were based involved primarily patients referred to tertiary care

 Copyright 2004, Elsevier (USA). All rights reserved.

DeMeester, T. (1997). Esophageal carcinoma: Current controversies. *Semin. Surg. Oncol.* **13**, 217–233.
El-Serag, H. (2002). The epidemic of esophageal adenocarcinoma. *Gastroenterol. Clin.* **31**, 421–440.
Entwistle, J., and Goldberg, M. (2002). Multimodality therapy for resectable cancer of the thoracic esophagus. *Ann. Thorac. Surg.* **73**, 1009–1015.
Lerut, T., Coosemans, W., De Leyn, P., Van Raemdonck, D., Deneffe, G., and Decker, G. (1999). Treatment of esophageal carcinoma. *Chest* **116**(Suppl. 6), 463S–465S.
Mallery, S., and Van Dam, J. (2000). EUS in the evaluation of esophageal carcinoma. *Gastrointest. Endosc.* **52**(Suppl. 6), S6–S11.
Meyenberger, C., and Fantin, A. (2000). Esophageal carcinoma: Current staging strategies. Recent results *Cancer Res.* **155**, 55–62.
Miller, D. (2000). Clinical trials for esophageal carcinoma. *Chest Surg. Clin. North Am.* **10**, 583–590.
Reed, C. (1999). Surgical management of esophageal carcinoma. *Oncologist* **4**, 95–105.

Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD

STUART JON SPECHLER
Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

Barrett's esophagus The condition in which an abnormal columnar epithelium replaces the stratified squamous epithelium that normally lines the distal esophagus.
dysplasia A constellation of histological abnormalities that suggest that one or more clones of cells have acquired genetic damage rendering them neoplastic and predisposed to malignancy.
gastroesophageal reflux disease A disease that occurs when the passage of stomach contents into the esophagus results in symptoms or signs such as chest pain, esophagitis, chronic cough, or recurrent pneumonia.

Barrett's esophagus develops when refluxed gastric juice injures the esophageal mucosa and the injury heals through a metaplastic process in which columnar cells replace reflux-damaged squamous cells. Specialized intestinal metaplasia, the epithelium that characterizes Barrett's esophagus, is composed of a variety of columnar cell types that can have gastric, small intestinal, and colonic features. This abnormal columnar epithelium may be more resistant to gastroesophageal reflux disease (GERD) than the native squamous mucosa, but the metaplastic cells are predisposed to malignancy. GERD and Barrett's esophagus are the major recognized risk factors for esophageal adenocarcinoma, a lethal malignancy that has increased profoundly in frequency over the past few decades in Western countries.

INTRODUCTION

The diagnosis of Barrett's esophagus is made by endoscopic evaluation. Columnar epithelium in the esophagus has a characteristic dull, reddish appearance that can be distinguished readily from the glossy, pale squamous epithelium. The diagnosis of Barrett's esophagus is established when biopsy specimens of the esophageal columnar epithelium show specialized intestinal metaplasia. Barrett's esophagus can be further categorized according to the extent of the metaplastic lining. Patients who have >3 cm of specialized intestinal metaplasia have long-segment Barrett's esophagus, a condition that usually is associated with severe gastroesophageal reflux disease (GERD). Patients with <3 cm of metaplasia have short-segment Barrett's esophagus, a disorder often associated with only mild GERD. Although it is not clear whether these two types of Barrett's esophagus have the same risk for carcinogenesis, patients with short- and long-segment Barrett's esophagus are managed similarly.

Published estimates on the annual risk of cancer in patients with Barrett's esophagus have ranged from 0.2 to 2.9%, with an average risk of approximately 1%. The studies on which those estimates were based involved primarily patients referred to tertiary care

 Copyright 2004, Elsevier (USA). All rights reserved.

centers; however, the patients whose cancer risk may have exceeded that for patients treated in primary care practices. Furthermore, a recent report has provided compelling evidence that the cancer risk in Barrett's esophagus has been overestimated for years because of publication bias, the selective reporting of studies that have positive or extreme results. Modern studies suggest that patients with Barrett's esophagus develop esophageal cancer at a rate of approximately 0.5% per year.

SCREENING AND SURVEILLANCE FOR PATIENTS WITH BARRETT'S ESOPHAGUS

To decrease mortality from esophageal adenocarcinoma, authorities have proposed that patients with chronic GERD symptoms should have endoscopic screening for Barrett's esophagus. Long-segment Barrett's esophagus can be found in 3 to 5% of such patients, whereas 10 to 15% have short-segment Barrett's esophagus. In patients with GERD, risk factors for adenocarcinoma in Barrett's esophagus include male gender, white ethnicity, obesity, advanced age, and long duration of GERD symptoms. However, any screening program that targets only such individuals can have only limited impact on cancer mortality rates because 40% of patients with esophageal adenocarcinoma have no history of GERD symptoms whatsoever. Currently, there is little evidence that screening programs have prevented deaths from esophageal adenocarcinoma. Among patients found to have these tumors, fewer than 5% are known to have had Barrett's esophagus before they presented with symptoms of esophageal cancer.

A number of medical societies, including the American College of Gastroenterology, have recommended regular endoscopic cancer surveillance for patients with Barrett's esophagus in order to decrease mortality from esophageal cancer and thereby prolong survival. The need for such surveillance has been questioned, however, because the practice has not been proved to decrease cancer mortality and available studies suggest that Barrett's esophagus does not even affect longevity. Studies on survival in Barrett's esophagus have consisted predominantly of older patients who often succumbed to comorbid illnesses rather than to esophageal adenocarcinoma. Proponents of surveillance argue that the results of those studies may not be applicable to younger, healthier patients with Barrett's esophagus.

Retrospective studies have documented that endoscopic surveillance can detect curable neoplasms in Barrett's esophagus and that cancers discovered during surveillance are less advanced than those found in patients who present with cancer symptoms such as dysphagia and weight loss. Furthermore, computer models have suggested that endoscopic surveillance for patients with Barrett's esophagus can prolong life, provided that certain assumptions are met. For example, in one Markov model of 50-year-old patients who had an assumed annual cancer incidence rate of 0.4%, endoscopic surveillance every 5 years was found to be the preferred strategy, costing $98,000 per quality-adjusted life year gained.

DYSPLASIA IN BARRETT'S ESOPHAGUS

As in other organs, cancers in Barrett's esophagus are thought to evolve through a sequence of genetic (DNA) alterations that give the cells certain growth advantages (Fig. 1). These same DNA abnormalities also can cause morphological changes in the tissue that the pathologist recognizes as dysplasia. Dysplasia is a constellation of histological abnormalities that suggest that one or more clones of cells have acquired genetic damage rendering them neoplastic and predisposed to malignancy. Dysplasia can be categorized as low- or high-grade depending on the extent of cytological and architectural changes. Endoscopic surveillance in Barrett's esophagus is performed primarily to seek dysplasia, with the rationale that resection or ablation of the esophagus lined by dysplastic epithelium may prevent the progression to invasive malignancy.

Unfortunately, dysplasia is an imperfect marker for malignancy. For example, the grading of dysplastic changes is a subjective skill. Among experienced pathologists, interobserver agreement for the diagnosis of low-grade dysplasia in Barrett's esophagus is less than 50%, whereas interobserver agreement for high-grade dysplasia is approximately 85%. Approximately one-third

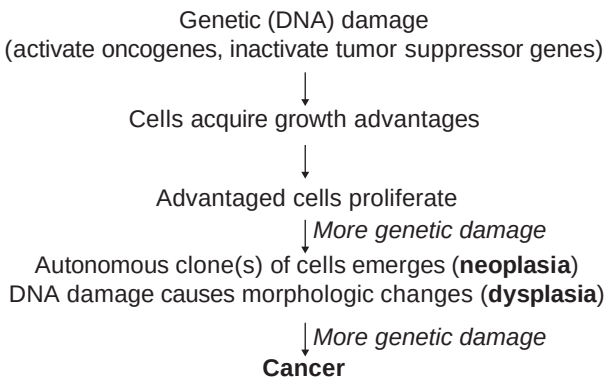


FIGURE 1 Carcinogenesis in Barrett's esophagus.

of patients who have esophagectomy for high-grade dysplasia have cancers found in the resected esophagus, cancers that were missed because of biopsy sampling error. Furthermore, the natural history of dysplasia in Barrett's esophagus is not well defined. Published estimates of the 5-year cumulative esophageal cancer incidence in patients with high-grade dysplasia range from 9 to 59%.

Researchers have been exploring alternative markers for cancer risk and preliminary data suggest that abnormalities in p53 expression and cellular DNA content (as assessed by flow cytometry) may be more sensitive and specific predictors of malignant potential than the histological finding of dysplasia. In addition, studies are evaluating the role of endoscopic techniques such as chromoendoscopy, endosonography, and fluorescence spectroscopy that enable the endoscopist to identify abnormal areas for biopsy sampling, rather than relying on random sampling to detect dysplasia. Currently, however, none of these tests or techniques has been shown to provide sufficient clinical information to justify its routine application in practice. Endoscopy with random biopsy sampling for dysplasia, despite all the shortcomings, remains the clinical standard for managing patients with Barrett's esophagus.

Patients who have verified high-grade dysplasia in Barrett's esophagus must choose among three management options: (1) esophagectomy, (2) endoscopic ablative therapy, or (3) intensive surveillance. Esophagectomy is the only therapy that clearly can prevent the progression from dysplasia to invasive cancer. Unfortunately, this procedure is associated with a 3 to 12% operative mortality rate and a 30 to 50% rate of serious operative complications. Endoscopic ablative therapies use thermal or photochemical energy to destroy the abnormal epithelium or localized lesions can be resected using a technique called endoscopic mucosal resection that is similar to colonoscopic polypectomy. Unfortunately, side effects of these therapies are frequent. For example, approximately one-third of the patients treated with photodynamic therapy using porfimer sodium develop esophageal strictures. Ablative therapies are expensive and they may not eradicate all of the tissue with neoplastic predisposition. As yet, no study has shown that these treatments decrease the long-term risk for cancer development and, currently, endoscopic ablative therapies should be considered experimental. Intensive endoscopic surveillance for high-grade dysplasia in Barrett's esophagus entails endoscopic examinations every 3 to 6 months, withholding invasive therapies until biopsy specimens reveal adenocarcinoma. Few published data directly support the safety and efficacy of this management strategy, however, and published reports suggest that 3 to 25% of cancers discovered in this fashion may be incurable.

MANAGEMENT RECOMMENDATIONS

The management of Barrett's esophagus that has been endorsed by the American College of Gastroenterology is as follows:

- Patients with Barrett's esophagus should have regular surveillance endoscopy to obtain esophageal biopsy specimens. GERD should be treated prior to surveillance to minimize confusion caused by inflammation in the interpretation of dysplasia.
- For patients who have had two consecutive endoscopies that show no dysplasia, surveillance endoscopy is recommended at an interval of every 3 years.
- If dysplasia is noted, the finding should be verified by consultation with another expert pathologist.
- For patients with verified low-grade dysplasia after extensive biopsy sampling, yearly surveillance endoscopy is recommended.
- For patients found to have high-grade dysplasia, another endoscopy should be performed with extensive biopsy sampling (especially from areas with mucosal irregularity) to look for invasive cancer and the histology slides should be interpreted by an expert pathologist. If there is focal high-grade dysplasia (defined as high-grade dysplastic changes involving fewer than five crypts), the condition may be followed with endoscopic surveillance performed at 3-month intervals. If there is verified multifocal high-grade dysplasia, intervention (e.g., esophagectomy) may be considered.

Although not specifically recommended in the practice guidelines, clinicians can consider the use of experimental ablative therapies such as photodynamic therapy for their patients with high-grade dysplasia in Barrett's esophagus, *provided the therapy is administered as part of an established, approved research protocol*. The use of ablative therapies outside of research protocols cannot be condoned at this time. Finally, the clinician should appreciate that no management strategy or patients with Barrett's esophagus has been verified by studies demonstrating that the strategy prolongs life.

See Also the Following Articles

Barrett's Esophagus • Cancer, Overview • Esophageal Cancer • Gastroesophageal Reflux Disease (GERD)

Further Reading

Corley, D. A., Levin, T. R., Habel, L. A., Weiss, N. S., and Buffler, P. A. (2002). Surveillance and survival in Barrett's adenocarcinomas: A population-based study. *Gastroenterology* **122**, 633–640.

Lagergren, J., Bergstrom, R., Lindgren, A., and Nyren, O. (1999). Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N. Engl. J. Med.* **340**, 825–831.

Morales, C. P., Souza, R. F., and Spechler, S. J. (2002). Hallmarks of cancer progression in Barrett's oesophagus. *Lancet* **360**, 1587–1589.

Provenzale, D., Schmitt, C., and Wong, J. B. (1999). Barrett's esophagus: A new look at surveillance based on emerging estimates of cancer risk. *Am. J. Gastroenterol* **94**, 2043–2053.

Sampliner, R. E., and The Practice Parameters Committee of the American College of Gastroenterology (2002). Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. *Am. J. Gastroenterol.* **97**, 1888–1895.

Shaheen, N. J., Crosby, M. A., Bozymski, E. M., and Sandler, R. S. (2000). Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* **119**, 333–338.

Soni, A., Sampliner, R. E., and Sonnenberg, A. (2000). Screening for high-grade dysplasia in gastroesophageal reflux disease: Is it cost-effective? *Am. J. Gastroenterol.* **95**, 2086–2093.

Spechler, S. J. (2002). Barrett's esophagus. *N. Engl. J. Med.* **346**, 836–842.

Esophageal Strictures

WILLIAM G. PATERSON
Queen's University, Kingston, Canada

bougie A mercury-filled rubber tube that is passed by mouth and down the esophagus to dilate esophageal strictures.

proton pump inhibitors A class of drugs that block the H^+,K^+-ATPase ("proton pump") in the gastric parietal cell, thereby suppressing gastric acid secretion.

An esophageal stricture is a circumscribed narrowing of the esophageal lumen that may impede the passage of food to the stomach, thereby producing dysphagia.

ETIOLOGY AND PATHOGENESIS

Table I summarizes the major types of esophageal strictures categorized according to their presumed etiology or pathogenesis. By far the most common cause of esophageal stricture is gastroesophageal reflux disease (peptic stricture), accounting for at least 80% of cases. Peptic strictures result from recurrent reflux of acidic gastric juices up into the distal esophagus, with resulting inflammation, ulceration, and collagen formation. Inflammation triggered by virtually any agent is capable of causing stricturing of the esophagus. Lumenal narrowing from inflammation is typically due to thickening of all layers of the esophageal wall, with collagen deposition within the submucosa. Another common cause of lumenal narrowing is mucosal rings or webs, the most common type being the so-called "Shatski's ring." This is a prominent circumferential fold of mucosa at the junction between the squamous epithelium of the esophagus and the columnar epithelium of the stomach. An important cause of esophageal stricturing is malignant disease, the most common types being squamous cell cancer and adenocarcinoma. It is for this reason that patients presenting with new onset dysphagia require timely investigation.

CLINICAL PRESENTATION AND MANAGEMENT

Patients with esophageal strictures typically present with solid food dysphagia. Before these symptoms occur, the esophageal lumen usually has narrowed to <15 mm. The pattern of dysphagia may help the clinician predict the type of pathology present. Patients with unpredictable intermittent dysphagia to both solids and liquids are likely to have an esophageal motor disorder rather than a stricture as the cause. Rapidly progressive

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Corley, D. A., Levin, T. R., Habel, L. A., Weiss, N. S., and Buffler, P. A. (2002). Surveillance and survival in Barrett's adenocarcinomas: A population-based study. *Gastroenterology* **122**, 633–640.

Lagergren, J., Bergstrom, R., Lindgren, A., and Nyren, O. (1999). Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N. Engl. J. Med.* **340**, 825–831.

Morales, C. P., Souza, R. F., and Spechler, S. J. (2002). Hallmarks of cancer progression in Barrett's oesophagus. *Lancet* **360**, 1587–1589.

Provenzale, D., Schmitt, C., and Wong, J. B. (1999). Barrett's esophagus: A new look at surveillance based on emerging estimates of cancer risk. *Am. J. Gastroenterol* **94**, 2043–2053.

Sampliner, R. E., and The Practice Parameters Committee of the American College of Gastroenterology (2002). Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. *Am. J. Gastroenterol.* **97**, 1888–1895.

Shaheen, N. J., Crosby, M. A., Bozymski, E. M., and Sandler, R. S. (2000). Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* **119**, 333–338.

Soni, A., Sampliner, R. E., and Sonnenberg, A. (2000). Screening for high-grade dysplasia in gastroesophageal reflux disease: Is it cost-effective? *Am. J. Gastroenterol.* **95**, 2086–2093.

Spechler, S. J. (2002). Barrett's esophagus. *N. Engl. J. Med.* **346**, 836–842.

Esophageal Strictures

WILLIAM G. PATERSON

Queen's University, Kingston, Canada

bougie A mercury-filled rubber tube that is passed by mouth and down the esophagus to dilate esophageal strictures.

proton pump inhibitors A class of drugs that block the H^+,K^+-ATPase ("proton pump") in the gastric parietal cell, thereby suppressing gastric acid secretion.

An esophageal stricture is a circumscribed narrowing of the esophageal lumen that may impede the passage of food to the stomach, thereby producing dysphagia.

ETIOLOGY AND PATHOGENESIS

Table I summarizes the major types of esophageal strictures categorized according to their presumed etiology or pathogenesis. By far the most common cause of esophageal stricture is gastroesophageal reflux disease (peptic stricture), accounting for at least 80% of cases. Peptic strictures result from recurrent reflux of acidic gastric juices up into the distal esophagus, with resulting inflammation, ulceration, and collagen formation. Inflammation triggered by virtually any agent is capable of causing stricturing of the esophagus. Lumenal narrowing from inflammation is typically due to thickening of all layers of the esophageal wall, with collagen deposition within the submucosa. Another common cause of lumenal narrowing is mucosal rings or webs, the most common type being the so-called "Shatski's ring." This is a prominent circumferential fold of mucosa at the junction between the squamous epithelium of the esophagus and the columnar epithelium of the stomach. An important cause of esophageal stricturing is malignant disease, the most common types being squamous cell cancer and adenocarcinoma. It is for this reason that patients presenting with new onset dysphagia require timely investigation.

CLINICAL PRESENTATION AND MANAGEMENT

Patients with esophageal strictures typically present with solid food dysphagia. Before these symptoms occur, the esophageal lumen usually has narrowed to <15 mm. The pattern of dysphagia may help the clinician predict the type of pathology present. Patients with unpredictable intermittent dysphagia to both solids and liquids are likely to have an esophageal motor disorder rather than a stricture as the cause. Rapidly progressive

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Esophageal Strictures

Congenital
Esophageal atresia
Tracheoesophageal fistula
Inflammatory
Peptic
Chemical injury—pill-induced, caustic ingestion (e.g., lye)
Allergic—eosinophilic esophagitis
Autoimmune—Crohn's disease, epidermolysis bullosa, graft versus host disease
Infectious
Monilial
Tuberculosis
Neoplastic
Benign—e.g., leiomyoma
Malignant—squamous cell carcinoma, adenocarcinoma
Iatrogenic
Postoperative
Radiation
Sclerotherapy
Idiopathic
Webs and rings—e.g., Shatski's ring, cervical esophageal web

solid food dysphagia with weight loss suggests malignant disease, whereas slowly progressive solid food dysphagia without weight loss is more in keeping with a peptic stricture. Sporadic dysphagia to meat suggests the presence of a Shatski's ring. Other symptoms may also be present depending on the underlying etiology of the stricture. For instance, patients with peptic strictures often relate a long history of heartburn and acid regurgitation before the onset of their dysphagia.

Esophageal strictures are diagnosed by performing barium contrast studies and/or endoscopy of the esophagus. In most instances, strictures diagnosed by X ray must also be assessed endoscopically, at which time mucosal biopsy may be performed. This is primarily to ensure that the stricture is not caused by a malignant neoplasm.

Many strictures can also be treated at the time of endoscopy by performing dilation with either a bougie or a balloon. This serves to stretch the lumen open, but it is important to then treat the underlying cause (if possible) in order to decrease the chance of the stricture recurring. For instance, to prevent recurrence of peptic strictures, patients should be placed on proton pump inhibitors to suppress reflux of gastric acid or undergo anti-reflux surgery.

See Also the Following Articles

Dysphagia • Gastroesophageal Reflux Disease (GERD) • Proton Pump Inhibitors • Swallowing

Further Reading

O Connor, J. B., and Richter, J. E. (1999). Esophageal strictures. *In* "The Esophagus" (D. O. Castell and J. E. Richter, eds.), 3rd Ed., pp. 473–483. Lippincott Williams and Wilkins, Philadelphia.

Spechler, S. J. (1999). AGA technical review on treatment of patients with dysphagia caused by benign disorders of the distal esophagus. *Gastroenterology* **117**, 233–254.

Spechler, S. J. (1999). American Gastroenterological Association medical position statement on treatment of patients with dysphagia caused by benign disorders of the distal esophagus. *Gastroenterology* **117**, 229–232.

Esophageal Surgery

YAEL REFAELY* AND RAPHAEL BUENO*,†

Brigham and Women's Hospital, Boston and †Harvard Medical School

achalasia Failure to relax.
Collis gastroplasty "Lengthens" the esophagus by transforming a portion of the upper stomach into an esophageal extension.
diverticula Small, balloon-like pouches that protrude from a hollow organ or structure.
esophagectomy Removal of the esophagus.
fundoplication Wrapping the stomach fundus around the esophagus.
gastroplasty Stomach-reducing surgery.

Operative techniques for treating esophageal disease have advanced considerably in recent years, both as a result of the improved understanding of esophageal anatomy and physiology and the successful introduction of minimally invasive approaches to esophageal surgery. Laparoscopic esophageal procedures for myotomy and antireflux procedures, as well as video-assisted thoracoscopic procedures for benign disorders and staging of esophageal cancer, have gained popularity and appear to be as effective as their open equivalents, while causing less postoperative morbidity. New anesthetic techniques, including the use of epidural catheters, and effective improvements in critical care have reduced perioperative mortality and made open surgical procedures safer. Despite the preponderance of minimally invasive techniques, open surgical procedures continue to make up an important component of esophageal surgery for the treatment of benign and malignant esophageal diseases.

SURGERY FOR BENIGN DISORDERS OF THE ESOPHAGUS

Antireflux Procedures

Gastroesophageal reflux disease (GERD) results from abnormal esophageal exposure to gastric acid. It is an exceedingly common disorder in the United States and usually responds well to abstinence from causative agents and pharmacological therapy. Patients with GERD refractory to maximal medical therapy may elect to undergo antireflux surgery. Whereas medical treatment concentrates on elevating the pH of the gastric acid and decreasing the incidence and duration of reflux episodes, surgical efforts are directed toward restoring normal function to the gastroesophageal junction (GEJ), which acts as a sphincter barrier between the acidic content of the stomach and the esophagus. This goal is achieved by reducing the GEJ to its normal intraabdominal position and tightening its sphincter function. A variety of techniques have been described to achieve these goals. The most common include (1) the Belsey–Mark IV antireflux repair, a partial fundoplication performed through a left thoracotomy, and (2) the Nissen fundoplication, which can be performed via a transthoracic, transabdominal, or laparoscopic approach. The Hill repair and other partial fundoplications are also surgical options that can be applied transabdominally.

Patients with severe, long-standing reflux disease often develop chronic inflammation and esophageal fibrosis that can lead to esophageal shortening. In these circumstances, the intraabdominal esophageal segment is insufficient and tension on the repair may lead to failure of the antireflux procedure. For patients with a foreshortened esophagus, a Collis gastroplasty is frequently performed to lengthen the intraabdominal esophagus and prevent postoperative failures.

Selection of the Optimal Antireflux Surgical Procedure

Patients with severe GERD who fail medical therapy require at the very least preoperative evaluation with endoscopy, manometry, and a pH probe. A barium swallow and gastric-emptying studies as well as careful psychological evaluation are occasionally warranted. For patients with normal esophageal motility and gastric emptying, most surgeons would opt to perform a laparoscopic transabdominal Nissen fundoplication. An open or laparoscopic drainage procedure is added for those patients with impaired gastric emptying. The transthoracic approach should be considered whenever any of the standard abdominal approaches carries an increased risk of failure or complication. For instance, the transthoracic approach would be appropriate in patients who have (1) associated esophageal motility disorder, (2) massive hiatal hernia, with the stomach pulled into the chest, (3) esophageal diverticula that require resection, or (4) profoundly foreshortened

 Copyright 2004, Elsevier (USA). All rights reserved.

esophagus, as well as (5) for patients who have undergone multiple previous abdominal operations or who have previously experienced failed antireflux procedures. The procedure of choice for patients with esophageal motility disorders, such as scleroderma or achalasia, is a partial as opposed to complete fundoplication in order to prevent postoperative obstruction. A partial fundoplication is any procedure in which the wrap is less than 360°, but more than 180°, such as the posterior Toupet fundoplication or the anterior Dor fundoplication.

Belsey–Mark IV Partial Fundoplication

The Belsey–Mark IV fundoplication is a 240° wrap that is usually performed through a left posterolateral thoracotomy incision, although it has also reportedly been performed thoracoscopically. The esophagus and both vagus nerves are exposed and mobilized followed by exposure of the right and left crura. The peritoneum is entered by dividing the phrenoesophageal membrane and the gastrohepatic ligament. The hiatal dissection is then completed, the hernia sac is excised, and a few short gastric arteries are divided. The esophagogastric junction is elevated into the chest for suture placement. When esophageal shortening is found, a Collis gastroplasty is performed utilizing a stapling device. The crura are approximated to close the hiatus snugly. The next step is the placement of three sutures in each of two to three rows or tiers to create a 240° fundoplication and concurrently reduce the wrap into the abdomen. In experienced hands, this operation is effective and durable. The major disadvantage is the requirement for thoracotomy.

Nissen Fundoplication

Nissen fundoplication (Fig. 1) is the most popular operation for GERD and considered to be the most effective. The procedure entails using a portion of the fundus of the stomach to create a 360° wrap around the lower 3–4 cm of the esophagus and then suturing it in place, so that the gastroesophageal sphincter passes through a short tunnel of the stomach. The wrap should be loose enough to accommodate a No. 48 to 56 French dilator within the esophagus after the procedure. To prevent axial slipping, the top of the wrap should be sutured to the side of the esophagus either separately or as part of the stitch securing the wrap. There are five major steps to this procedure: (1) crural dissection with the identification and preservation of both vagi, including the hepatic branch of the anterior vagus, (2) circumferential dissection of the esophagus, (3) crural closure, (4) fundic mobilization by division of the short gastric vessels, and (5) creation of a short (2–3 cm), loose fundoplication by enveloping the posterior and the anterior walls of the fundus. This operation can be performed transthoracically and transabdominally. The introduction of the laparoscopic approach has made this procedure far more common in the United States because of the minimal perioperative discomfort. Relief of typical reflux symptoms (i.e., heartburn, regurgitation, and dysphagia) is achieved in more than 90% of the patients following laparoscopic Nissen fundoplication. A lower incidence of symptom reduction is noted among patients with atypical reflux symptoms (i.e., cough, asthma, or laryngitis). These symptoms are relieved in only two-thirds of cases. However, recent reports indicate benefit in patients undergoing lung transplantation. Temporary dysphagia is common following this operation and may be seen in up to 50% of the patients; however, the dysphagia usually resolves after 3 months in most patients.

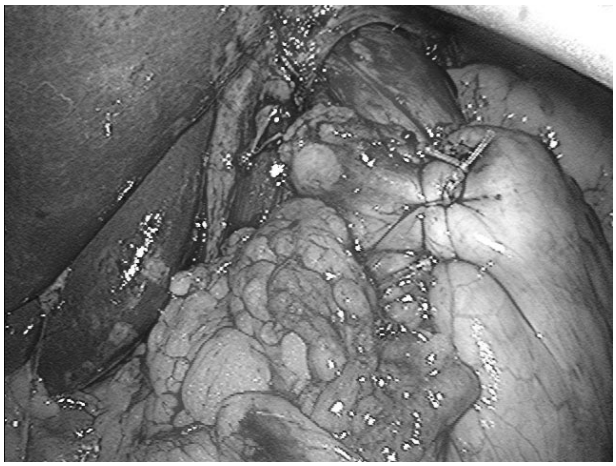


FIGURE 1 A picture of the Nissen wrap at the conclusion of a laparoscopic procedure. The sutures are evident in the middle of the wrap. The esophagus is underneath and cephalad to the wrap.

Collis Gastroplasty

In performing the Collis gastroplasty for a short esophagus, a stapler is used to form a 4- to 5-cm neoesophagus out of the proximal stomach, thereby effectively lengthening the esophagus and transposing the esophagogastric junction distally, below the diaphragm, where a high intraabdominal pressure, as opposed to a negative intrathoracic pressure, will reduce reflux. A large-caliber Maloney bougie (48 to 56 French) is placed in the esophagus to prevent narrowing of the lumen. The bougie is held against the lesser curvature, and the fundus is retracted away at a right angle to the esophagus. A stapler is applied immediately alongside

the bougie on the greater curvature side, simultaneously cutting and stapling the cardia. This procedure can also be performed thoracoscopically or laparoscopically.

Technical Complications of Antireflux Procedures

Early and late technical complications of the antireflux procedures are summarized in Table I. Late complications may cause failure of the repair and necessitate reevaluation and probable reoperation. Recurrent heartburn and regurgitation demand evaluation with contrast studies and esophagoscopy. Anatomic conditions determined to be responsible for failure of the fundoplication usually necessitate reoperation. Dysphagia unrelated to recurrent reflux or ulceration generally responds to dilatation and does not necessitate reoperation.

Surgery for Motility Disorders of the Esophagus

The act of swallowing is quite complex and requires the perfect coordination between the central nervous system and the various neural and muscular elements of the esophagus. Any discrepancy, such as hyperactivity or hypoactivity or discoordination, at any level of the esophagus can cause clinical symptoms. A number of motor disorders of the esophagus can cause a variety of adverse effects on swallowing; the most common disorders are described in the following sections.

TABLE I Technical Complications of Antireflux Procedures

Technical complications	Clinical manifestations
Early	**Early**
Injury to the spleen during short gastric ligation	Bleeding, splenectomy
Torsion of the fundus	Perforation and sepsis
Gastroplasty leak at the staple line	Perforation and sepsis
Late	**Late**
Injury to the vagi	Gastric dysfunction, gas-bloat syndrome
Poor crural approximation	Migration of the wrap into chest, paraesophageal hernia
Inadequate mobilization of the fundus	Tension on the wrap, later disruption, recurrent reflux
Inadequate fixation of wrap to esophagus (slipped Nissen)	Hourglass stomach; heartburn, regurgitation, dysphagia
Long wrap	Gas-bloat syndrome
Wrap that is too tight	Persistent dysphagia

Pharyngoesophageal (Zenker's) Diverticulum

The most common diverticulum of the esophagus, known as Zenker's diverticulum, usually occurs in elderly patients. It is a false diverticulum, with the esophageal mucosa (and not full-thickness esophagus) herniating between the oblique fibers of the inferior constrictor muscle of the pharynx and the transverse fibers of the cricopharyngeus muscle. Patients usually present with dysphagia and regurgitation. The radiographic appearance of the pharyngoesophageal diverticulum on a barium swallow is diagnostic. This disorder is thought to be caused by discoordination between relaxation and contraction of the pharynx and the upper esophageal sphincter. The preferred method of treatment is a single-stage resection with a concomitant myotomy. In this procedure, the esophagus is explored through a left neck incision by retracting the thyroid gland medially and the sternocleidomastoid muscle and carotid sheath laterally. Excess retraction is avoided to protect the recurrent laryngeal nerve. The mucosal sac is dissected up to its neck, such that the surrounding ring of the muscular defect is clearly defined. A curved clamp or a stapler is placed across the neck of the sac at a right angle to the long axis of the esophagus, and the sac is amputated. Interrupted fine silk sutures are placed in the mucosa as it is being incised. Alternatively, the diverticulum can be excised with a stapler. The operation carries a minimal risk and low recurrence rate. To prevent possible recurrences, a cricopharyngeal myotomy is routinely performed.

Epiphrenic Diverticulum

Epiphrenic diverticula are typically located within the distal third of the esophagus. This type of propulsive diverticulum arises because of abnormally elevated intraluminal esophageal pressures distally creating a functional obstruction. Patients usually present with dysphagia, regurgitation of previous meals, and frequently recurrent pneumonias. Relief of symptoms and prevention of aspiration pneumonia are the chief indications for resection of the diverticulum. The surgical approach is made through a left sixth or seventh interspace posterolateral thoracotomy. The diverticulum is mobilized to its base and amputated with a TA-30 stapler. A long, extramucosal esophagomyotomy is performed from the level of the diverticulum and sometimes from the aortic arch to the esophagogastric junction on the opposite wall of the esophagus. If an associated hiatal hernia or incompetent lower esophageal sphincter is found, an antireflux procedure should be added. Some surgeons advocate continuing

the myotomy below the GEJ and routinely add an antireflux procedure. A partial fundoplication procedure, such as the 240° Belsey—Mark IV, would be preferred, to obviate functional obstruction. The results of this surgery are usually excellent, even in the older patient population.

Cardiomyotomy for Achalasia (Heller Operation)

This operation was first described by Heller in 1914 for the treatment of esophageal achalasia. The operation originally consisted of two myotomies on opposite sides of the esophagus, performed through a laparotomy. Although some surgeons performed the operation through a laparotomy, others preferred a thoracotomy approach. The introduction of minimally invasive laparoscopic cardiomyotomy with its lower morbidity has resulted in the adoption of this minimally invasive technique by most surgeons. This operation is used primarily for the treatment of achalasia and is indicated in every patient diagnosed with achalasia who can tolerate general anesthesia. In the laparoscopic version, a myotomy is performed to a total length of approximately 3—5 cm. It is extended onto the stomach for a length of 1.5—2 cm. Its proximal extent is limited only by the operative exposure. The longitudinal and circular muscle fibers are divided while the esophagus is splayed over a bougie. Special attention is taken to protect the underlying mucosa from injury. The cut edge of the myotomy is gently dissected away, to be separated by at least one-third of the circumference of the esophagus, to prevent recurrence. Following completion of the myotomy, a partial antireflux procedure is usually added. This procedure can also be performed via a left open thoracotomy or a minimally invasive thoracoscopic approach. The myotomy is carried out from the level of the inferior pulmonary vein to just a few millimeters into the stomach and an antireflux procedure is not generally required.

The two most common fundoplications performed in conjunction with the Heller myotomy are, respectively, the posterior (Toupet) and anterior (Dor) 180° fundoplications. There are insufficient reports to measure the success rate of these fundoplication procedures, but it is clear that surgical myotomy is currently the best option for patients with achalasia. Relief of dysphagia and regurgitation may be expected in 85—90% of patients undergoing myotomy with or without fundoplication at 5 years. Chest pain is improved in 75—80%. Abnormal gastroesophageal reflux occurs in 15—25% of the patients.

Surgery for Benign Tumors of the Esophagus

Benign tumors of the esophagus are less common than malignant ones and usually are discovered incidentally. When symptomatic, large, or when the diagnosis is in question, these tumors should be surgically removed. Leiomyomas of the esophagus are the most common and usually grow within the muscular layer of the esophagus. They should not be endoscopically biopsied unless they have an intraluminal component because a mucosal injury during the biopsy will complicate surgical removal. These tumors can usually be enucleated without entering the esophageal lumen. Both open and minimally invasive thoracic approaches have been described. On rare occasions, esophageal resection is required for extirpation of esophageal leiomyomas.

Enteric (or esophageal) cysts are also common benign esophageal lesions. These are congenital and usually do not communicate with the esophageal lumen. Esophageal cysts, also known as esophageal duplications, may grow in size, become filled with mucus, and may hemorrhage or cause symptoms of pain, reflux, or obstruction. Unless their diagnosis is certain and there is no evidence of enlargement or symptoms, these should be removed, usually via video thoracoscopy.

SURGERY FOR CARCINOMA OF THE ESOPHAGUS

Esophageal Cancer

Esophageal cancer is a highly virulent malignancy; although rare, the incidence has recently been rising in the United States and Western Europe. In its earliest stages, esophageal cancer is readily curable by surgical treatment alone, with 5-year survival rates approaching 80%. By the time patients with esophageal cancer develop symptoms, however, this malignancy is usually moderately advanced in stage and more likely to be associated with metastatic lymph node involvement. The 5-year survival rates after surgical resection may be as high as 34% for patients with early nodal disease, but are more typically in the range of 10—15% for locally advanced lesions. It is unusual for patients with metastatic disease to survive 5 years after diagnosis.

There are two common histological types of esophageal cancer, squamous cell carcinoma and adenocarcinoma. Until recently, squamous cell carcinoma of the esophagus was the most common histological type seen. This disease is usually found in patients with the usual risk factors associated with other aerodigestive tract carcinomas. In particular, smoking and alcohol abuse are associated with a 5-fold increased risk of developing

squamous cell esophageal carcinoma. The combination of heavy smoking and drinking increases this risk 25- to 100-fold. African-Americans are five times more likely to develop squamous cell carcinoma than are members of any other racial or ethnic group, and the cancer is four to six times more likely to occur in males than females. Remarkably, a major shift has recently been observed in the histologic type of esophageal cancer. Since the early 1990s, the incidence of adenocarcinoma of the esophagus, located in the distal esophagus and gastroesophageal junction, has been gaining rapidly over the squamous cell variety. This change of prevalence is due to an increase in the incidence of adenocarcinoma without a change in the incidence of squamous cell carcinoma.

Surgical resection is the first-line therapy for the majority of patients with cancer of the esophagus and gastric cardia. Surgery remains the gold standard by which all other therapeutic modalities are measured. The main goals of surgery are (1) to extirpate the cancer by removing the primary tumor and lymph nodes with curative intention, (2) to relieve dysphagia, and (3) to maintain the continuity of the alimentary tract. The various types of surgical approaches may be combined with preoperative neoadjuvant chemotherapy and radiation therapy, although the data for their efficacy are limited at best. Chemotherapy or radiation therapy alone was not found to be effective in a neoadjuvant setting. Chemotherapy also has been used in many treatment protocols as an adjuvant therapy following surgery and has shown promise in cancer of the cardia and gastroesophageal junction.

Preoperative evaluation includes endoscopic documentation of cancer, chest and abdominal computer tomography (CT), a positron emission tomography (PET) scan to exclude metastatic disease, esophageal ultrasound for evaluation of the T and N stage, and a cardiorespiratory evaluation to ensure the patient's ability to withstand surgery. Laparoscopic and thoracoscopic surgical staging procedures have recently been advocated prior to esophagectomy or neoadjuvant therapy, given the limitation of other staging techniques.

The choice of operation for carcinoma of the esophagus depends on a number of factors, including preference of the surgeon, location of the tumor, body habitus, site of prior surgery, condition of the patient, choice of esophageal substitute, and history of prior radiation therapy. The two most important factors influencing the choice of procedure are the location of the tumor and the surgeon's preference. Cervical esophageal carcinoma is preferentially treated by chemotherapy and radiation at most centers. (Although some surgeons advocate pharyngolaryngectomy combined with extrathoracic esophagectomy and gastric interposition, these operations are beyond the scope of this article.) As for the more common cancers of the middle and lower thoracic esophagus and gastroesophageal junction, several types of operations are feasible, as listed in Table II. Esophageal resection should always be performed under careful monitoring that includes central venous pressure, arterial lines, electrocardiogram, end tidal CO_2, and O_2 saturation. A double-lumen endotracheal tube should be inserted in all cases, save for transhiatal esophagectomy, in order to collapse the lung on the operated side. Epidural analgesia is helpful for postoperative pain relief. A jejunostomy tube is often inserted for postoperative alimentation.

Options for Reconstruction After Esophagectomy

After a partial or complete esophagectomy, a conduit must be established to preserve the continuity of the digestive tract. The stomach, jejunum, and colon have all been successfully used as esophageal substitutes, but the stomach appears to be the conduit of choice because of ease of mobilization and ample vascular supply. The higher incidence of mortality reported with the use of the colon conduit is most probably related to the necessity of performing three anastomoses (i.e., coloesophagostomy, cologastrostomy, and

TABLE II Surgical Techniques to Resect and Replace the Esophagus in Malignant Disease

Technique	Surgeon (year)	Location of anastomosis
Left thoracoabdominal	Adams and Phemister (1938)	Left hemithorax
Ivor Lewis	Lewis (1946)	Right hemithorax
McKeown	McKeown (1976)	Left neck
Radical en bloc esophagogastrectomy	Skinner and co-workers (1982)	No limitations
Transhiatal esophagectomy	Orringer (1984)	Left neck
Three-hole esophagectomy	Swanson (2001)	Left neck

colocolostomy). The colon conduit is used if the patient has undergone partial or total gastrectomy previously, or if the tumor involves the stomach, precluding a 5-cm margin. Jejunal conduits can also be used, but they are least favorable due to their limited vascular supply and are therefore limited to short distal segmental replacement. A free jejunal segment may be used, usually in the neck, with reanastomosis of its blood supply and drainage to the carotid and jugular vessels. In rare circumstances, a musculocutaneous tube created from the pectoralis major muscle may be used as a replacement for the cervical esophagus.

Transhiatal Esophagectomy

The transhiatal esophagectomy technique was popularized by Orringer and Sloan in 1978, to avoid a thoracotomy and reduce the possibility of an intrathoracic anastomotic leak. The operation begins with the patient in a supine position, with the neck extended to the right side, exposing the left sternocleidomastoid muscle. The procedure is performed through left cervical and upper abdominal incisions. The resection begins with the abdominal dissection, followed by a blunt dissection of the thoracic esophagus from the abdominal and neck incisions until it is freed. The specimen is removed and a gastric tube is constructed and pulled up to the neck to create a cervical gastroesophageal anastomosis. The blind dissection increases the possibility of vascular or airway injury. It also limits the extent of the lymph node dissection. In expert hands, this procedure can be performed with morbidity and mortality as low as morbidity and mortality of other esophageal procedures and with similar long-term outcomes.

The steps in the abdominal dissection in esophagectomy are similar for most of the approaches. An upper midline incision is made from xiphoid to umbilicus. The peritoneal cavity is explored and inspected for metastases. If liver metastasis or other peritoneal spread is found, the resection is aborted and a jejunal feeding tube is inserted for palliative care of the patient. Following a negative inspection, total mobilization of the stomach is performed. The stomach is divided from its blood supply, preserving the right gastroepiploic artery. A gastric conduit based on this artery is created (see later) to replace the esophagus. The hiatus is then enlarged manually and the left lobe of the liver is mobilized from the diaphragm and gently retracted rightward for exposure. The distal esophagus is dissected under direct vision, and the middle esophagus is first palpable through the enlarged hiatus. Resectability is confirmed at this point. The most difficult area of dissection is the area immediately surrounding the tumor. Occasionally, a portion of the diaphragm is resected en bloc with the tumor. Local extension of the tumor involving the pleura, inferior pulmonary ligament, posterior pericardium, and so forth should be appreciated and treated accordingly. The cervical dissection is also similar for all neck approaches to the esophagus. The left cervical incision is made along the anteromedial border of the sternocleidomastoid muscle. The left is the preferred side of approach because the recurrent laryngeal nerve is less likely to be injured. A right-sided dissection can be substituted when needed for clinical reasons. The muscle and the carotid sheath are reflected laterally and no traction is placed on the trachea, to avoid injury to the left recurrent laryngeal nerve. The rest of the dissection is performed bluntly in a posterior dissection to get to the prevertebral fascia. The dissection then proceeds around the esophagus, which is freed from the trachea in the tracheoesophageal groove and encircled. The blunt dissection of the superior portion of the esophagus continues up to the level of the carina. Completion of the dissection is finally achieved through both incisions along the posterior mediastinum. The stomach conduit is than created by dividing the gastroesophageal junction along the lesser curvature of the stomach with multiple stapler firings. A Kocher maneuver is performed to allow the pylorus to reach the level of the hiatus and reduce tension. A pyloroplasty or pyloromyotomy is often performed to prevent future gastric-emptying difficulties, which are seen in up to 10% of patients undergoing esophagectomy. The specimen is removed from the field and the gastric conduit is then passed through the bed of the resected esophagus to the neck. The anastomosis is then performed in the neck. A feeding jejunostomy tube is placed routinely at the end of the operation.

The advantages to the transhiatal esophagectomy (THE) include the creation of a proximal surgical margin that is well away from the tumor site, an extrathoracic esophagogastric anastomosis that is easily accessible in the event of complications (e.g., anastomotic leak), and a reduced overall perioperative morbidity attributed to avoidance of the thoracotomy. The disadvantages to utilizing the THE are twofold: limited exposure increases the risk of injury to major structures in case of adherent or invasive tumor, and compromises are made in the mediastinal lymph node dissection (used mainly for staging and determination of response to treatment if given preoperatively).

Ivor Lewis Esophagectomy

The Ivor Lewis esophagectomy technique was proposed by Lewis in 1946 and immediately became the

most popular technique of esophageal resection. It combines two stages: abdominal laparotomy followed by right thoracotomy and intrathoracic anastomosis. The extent of resection is governed by the tumor location and includes at least a 5-cm margin proximal and distal to the tumor as well as to all the surrounding periesophageal tissue. This technique permits a partial or subtotal esophageal resection to be performed and provides exposure for complete regional lymph node dissection. The operation is performed in a specific order. First, the abdominal dissection is as for the THE. The thoracic dissection then follows. The patient is repositioned in a left lateral decubitus position for a standard right posterolateral thoracotomy. The esophagus is mobilized out of its bed and the tumor is carefully dissected off the posterior wall of the mainstem bronchus and pericardium. The esophagus is totally mobilized, beginning at the esophageal hiatus and extending just above the azygos vein. The already mobilized stomach is pulled into the thoracic cavity and the tumor is resected with at least a 5-cm margin. The lesser curvature of the stomach is divided and the gastric conduit is created. An end-to-side esophagogastric anastomosis is then performed, usually at the level of the azygos vein. The chest is closed in usual fashion, with a chest tube left in the pleural space and a nasogastric tube in the stomach conduit. This approach is ideal for lower and midesophageal lesions that are still lower than the carina, and provides excellent exposure for mediastinal lymph node dissection. It is limited only by the requirement of an intrathoracic anastomosis.

Left Thoracoabdominal Approach

The left thoracoabdominal approach is indicated for tumors in the lower third of the esophagus and the gastric cardia. The principal disadvantage is that the aortic arch limits the extent of the proximal dissection, rendering the intrathoracic anastomosis more difficult to perform. However, a third interspace left thoracotomy can always be added to gain additional proximal margin. This operation is frequently associated with significant esophagitis from bile reflux. Therefore, many surgeons prefer to resect the entire stomach and to create a Roux-en-Y jejunal interposition, which is anastomosed to the residual thoracic esophagus. The patient is positioned in the right lateral decubitus and a thoracotomy incision is made through the sixth intercostal interspace. The diaphragm is then incised peripherally, about 2 cm from the costal margins in order to reduce the risk of postoperative diaphragmatic paralysis. The pulmonary ligament is then divided and the esophagus is mobilized. Next, the stomach is mobilized as described above. Once adequately mobilized, the distal esophagus and stomach are resected and the anastomosis is carried out in the left hemithorax. The morbidity and mortality with this approach are similar to those of the Ivor Lewis approach. Gastric tip necrosis, which can be seen after a longer type esophagectomy, is less common with an intrathoracic anastomosis because the gastric tube is not unduly stretched.

The McKeown Operation and Three-Hole Esophagectomy (Abdominothoracicocervical Esophagectomy)

The conception of esophagectomy performed through three discrete incisions was first described by McKeown in 1972 and was aptly named the "three-stages esophagectomy," because it includes (1) an abdominal mobilization of the stomach, (2) a right thoracotomy to mobilize the esophagus, and (3) a cervical incision through which the esophagogastric anastomosis is performed. The first two stages are identical to those described for the Ivor Lewis approach except for mobilizing the entire intrathoracic esophagus. The third stage is carried out through a neck incision and includes passage of the stomach through the thoracic inlet and performance of cervical gastroesophageal anastomosis. This procedure results in a complete resection of the esophagus and its lymph node stations, with a "safe cervical anastomosis." A subtle variant of this approach has been recently popularized by the group at Brigham and Women's Hospital and has been named the "three-hole esophagectomy." It involves a right thoracotomy as an initial phase in the operation, with complete dissection of the intrathoracic esophagus with its lymph node stations, as described earlier. The chest is then closed and the remainder of the operation is performed as a transhiatal esophagectomy: Abdominal and cervical exploration and dissection are performed, the esophagus is divided proximally in the neck, and the stomach is divided to create a gastric tube that is pulled up to the neck for a cervical anastomosis. The advantage of this variant is in early determination of whether the tumors in the lower and midesophagus invade local structures, facilitating early assessment of tumor resectability. Video-thoracoscopic dissection can be substituted for the thoracotomy in order to minimize morbidity. A completely minimally invasive version of this procedure has also been described. The main benefits of both methods over THE include complete lymph node dissection and a safer visualized dissection of the intrathoracic esophagus.

Radical en Bloc Esophagectomy

Radical esophagogastrectomy was first described by Logan in 1963, but gained attention just after Skinner reported his experience in 1983. Skinner's results with 80 patients demonstrated that the operation can be performed safely, with an operative mortality of 11%, and that the survival rates are as good or better than previously achieved, with actuarial survival rates of 24% at 3 years and 18% at 5 years. The added benefit of radical esophageal resection is still controversial in Western countries and its use is limited to a few centers in the United States and Europe. In Japan, a more radical version of this approach is popular, including a three-field lymph node dissection (i.e., abdominal, thoracic, and cervical), as opposed to Skinner's technique, which involves only the abdominal and thoracic fields. Radical en bloc esophagectomy includes excision of the tumor-bearing esophagus, with 10-cm proximal and distal margins and a wide envelope of surrounding tissues, including the pericardium anteriorly, both pleural surfaces laterally, and the lymphatic tissues, inclusive of the thoracic duct wedged dorsally between the esophagus and the spine. Additionally, an upper mediastinal and abdominal lymphadenectomy is performed (two fields). Resection of tumors in the cardia or lower esophagus is performed through a thoracoabdominal incision, whereas for tumors in the thoracic esophagus, the operation is performed through a right thoracotomy and abdominal incision. Because the adequate proximal resection margin defined by this approach is 10 cm from the tumor and a clean margin from Barrett's esophagus, the anastomosis is performed in the neck whenever those two criteria meet high up in the thorax. The advantage of this procedure is the theoretical possibility of maximal locoregional control. Disadvantages include the facts that a survival benefit has not been proved and that the operation carries an increased risk for potential complications.

Complications of Esophageal Resection for Cancer

The operative mortality for esophagogastrectomy has significantly declined over the years, and in recent series it ranges between 2 and 10%. Centers with high volume report a 2% perioperative mortality. Other surgical complications are listed in Table III. Among them, the most common early complications are pulmonary (pneumonia and atelectasis), followed by cardiac, with arrhythmia occurring in 10% of the patients (most commonly atrial fibrillation). The occurrence of anastomotic leak ranges between 5 and 10% in different series. Although its incidence is not affected by the anastomotic site, it is easier to handle leakage from a neck anastomosis than from a thoracic anastomosis. Leak is more likely to occur in patients who are malnourished or diabetic and in those who have received preoperative radiation and have tension at the anastomosis site or a compromised blood supply to the conduit. Leak usually occurs within 10 days of the surgical procedure and most commonly these patients present with sepsis. For cervical anastomotic leakage, a conservative approach is advised, with open drainage of the cervical wound. Most of these patients ultimately develop esophageal stricture, requiring dilatation. In sharp contrast, intrathoracic anastomotic leaks are far more serious and carry a mortality of between 20 and 40%. Commonly, the patient will need another operation in order to inspect the anastomosis and drain the pleural space and mediastinum. In cases in which there is no necrosis and a relatively stable patient, attempts at repair with drainage are reasonable. Temporary exclusion is also possible. If a necrotic segment of stomach is discovered and the patient is in septic shock, debridement and takedown are performed. The viable stomach should be returned to the abdomen, and a cervical esophagostomy, as long as possible, is constructed for diversion. A decompressing gastrostomy is usually performed, and possible future reconstruction with substernal colon interposition will be required following recovery and absent recurrent cancer. Other complications include chylothorax from a thoracic duct injury, recurrent laryngeal nerve paresis, or paralysis and anastomotic strictures.

TABLE III Complications of Esophageal Resection

Early (perioperative)	Late
Atelectasis	Reflux
Acute respiratory distress syndrome	Regurgitation
Pneumonia	Dumping
Arrhythmia (usually atrial)	Anastomotic stricture
Myocardical infarction	Hiatal hernia
Anastomotic leak	
Conduit necrosis	
Pulmonary embolism	
Tracheobronchial injury	
Splenic injury	
Laryngeal nerve injury	
Chylothorax	

See Also the Following Articles

Achalasia • Esophageal Cancer • Gastroesophageal Reflux Disease (GERD) • Minimally Invasive Surgery • Zenker's Diverticulum

Further Reading

Pearson, F. G., Cooper, J. D., Deslauriers, J., Ginsberg, R. J., Hiebert, C. A., Patterson, G. A., and Urschel, H. C. (eds.) (2002). "Esophageal Surgery," 2nd Ed. Churchill-Livingstone, New York.

Sugarbaker, D. J., DeCamp, M. M., and Liptay, M. J. (1997). Surgical procedures to resect and replace the esophagus. *In* "Maingot's—Abdominal Operations" (M. J. Zinner, S. I. Schwartz, and H. Ellis, eds.), 10th Ed., Vol. I, pp. 885–912. Appleton & Lange, Stamford, CT.

Esophageal Trauma

SUBROTO PAUL* AND RAPHAEL BUENO*,†

**Brigham and Women's Hospital, Boston and †Harvard University*

esophagogastroduodenoscopy Examination of the esophagus, stomach, and duodenum with video-assisted technologies.

fundoplication Procedure in which the stomach fundus is wrapped around the esophagus.

transesophageal echocardiography Examination of the heart with an ultrasound probe placed through the esophagus.

Esophageal trauma is often a medical emergency requiring prompt attention and management. There are several types of esophageal injury. It can be iatrogenic in nature, resulting from endoscopic or other surgical procedures, it can occur spontaneously, or it can arise either from penetrating injury resulting in esophageal perforation or from caustic injury usually in the setting of attempted suicide. Prompt definitive treatment is required since delay can lead to decreased patient survival. This article focuses on the causes as well as the signs and symptoms of esophageal trauma. Management of the acute and chronic sequelae of these injuries is also discussed.

PERFORATING ESOPHAGEAL INJURY

Etiology

Esophageal perforation results from a variety of causes, including iatrogenic injury, spontaneous rupture, and both blunt and penetrating injuries. Iatrogenic injury typically results from injury sustained during endoscopic procedures, such as esophagogastroduodenoscopy (EGD) and transesophageal echocardiography. Iatrogenic injury is most commonly seen as a complication of endoscopic procedures. Spontaneous rupture typically occurs after prolonged vomiting. Perforation resulting from blunt and penetrating esophageal injuries is described in the trauma literature.

In patients with a normal esophagus, the cricopharyngeal portion of the proximal esophagus is the most commonly afflicted region. Mid and distal esophageal perforations typically result from endoscopic procedures including biopsy, stenting, and dilation used in the diagnosis or treatment of lesions of the esophagus, such as esophageal cancer. Less commonly, iatrogenic esophageal perforation may result from other types of instrumentation, such as nasogastric tube insertion, inadvertent esophageal intubation with an endotracheal tube, or injury during nonesophageal surgical procedures, such as tracheostomy and thyroidectomy.

Spontaneous rupture of the esophagus, known as Boerhaave's syndrome, is typically caused by prolonged vomiting usually in patients with a history of alcoholism. However, spontaneous perforation is also seen in patients with chronic esophageal injury resulting from severe reflux or severe candidal or herpetic infection. In these cases, esophageal perforation occurs posteriorly in the lower esophagus.

Penetrating injury to the neck can also result in esophageal perforation. Because of the posterior location of the esophagus in both the neck and chest, penetrating esophageal injury is rarely isolated and often accompanied by other more immediate life-threatening injuries to the trachea or carotid artery. Blunt trauma to the neck, such as might be sustained in high-speed motor vehicle accidents or blows to the neck, rarely

 Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Pearson, F. G., Cooper, J. D., Deslauriers, J., Ginsberg, R. J., Hiebert, C. A., Patterson, G. A., and Urschel, H. C. (eds.) (2002). "Esophageal Surgery," 2nd Ed. Churchill-Livingstone, New York.

Sugarbaker, D. J., DeCamp, M. M., and Liptay, M. J. (1997). Surgical procedures to resect and replace the esophagus. *In* "Maingot's—Abdominal Operations" (M. J. Zinner, S. I. Schwartz, and H. Ellis, eds.), 10th Ed., Vol. I, pp. 885–912. Appleton & Lange, Stamford, CT.

Esophageal Trauma

SUBROTO PAUL* AND RAPHAEL BUENO*,†

**Brigham and Women's Hospital, Boston and †Harvard University*

esophagogastroduodenoscopy Examination of the esophagus, stomach, and duodenum with video-assisted technologies.

fundoplication Procedure in which the stomach fundus is wrapped around the esophagus.

transesophageal echocardiography Examination of the heart with an ultrasound probe placed through the esophagus.

Esophageal trauma is often a medical emergency requiring prompt attention and management. There are several types of esophageal injury. It can be iatrogenic in nature, resulting from endoscopic or other surgical procedures, it can occur spontaneously, or it can arise either from penetrating injury resulting in esophageal perforation or from caustic injury usually in the setting of attempted suicide. Prompt definitive treatment is required since delay can lead to decreased patient survival. This article focuses on the causes as well as the signs and symptoms of esophageal trauma. Management of the acute and chronic sequelae of these injuries is also discussed.

PERFORATING ESOPHAGEAL INJURY

Etiology

Esophageal perforation results from a variety of causes, including iatrogenic injury, spontaneous rupture, and both blunt and penetrating injuries. Iatrogenic injury typically results from injury sustained during endoscopic procedures, such as esophagogastroduodenoscopy (EGD) and transesophageal echocardiography. Iatrogenic injury is most commonly seen as a complication of endoscopic procedures. Spontaneous rupture typically occurs after prolonged vomiting. Perforation resulting from blunt and penetrating esophageal injuries is described in the trauma literature.

In patients with a normal esophagus, the cricopharyngeal portion of the proximal esophagus is the most commonly afflicted region. Mid and distal esophageal perforations typically result from endoscopic procedures including biopsy, stenting, and dilation used in the diagnosis or treatment of lesions of the esophagus, such as esophageal cancer. Less commonly, iatrogenic esophageal perforation may result from other types of instrumentation, such as nasogastric tube insertion, inadvertent esophageal intubation with an endotracheal tube, or injury during nonesophageal surgical procedures, such as tracheostomy and thyroidectomy.

Spontaneous rupture of the esophagus, known as Boerhaave's syndrome, is typically caused by prolonged vomiting usually in patients with a history of alcoholism. However, spontaneous perforation is also seen in patients with chronic esophageal injury resulting from severe reflux or severe candidal or herpetic infection. In these cases, esophageal perforation occurs posteriorly in the lower esophagus.

Penetrating injury to the neck can also result in esophageal perforation. Because of the posterior location of the esophagus in both the neck and chest, penetrating esophageal injury is rarely isolated and often accompanied by other more immediate life-threatening injuries to the trachea or carotid artery. Blunt trauma to the neck, such as might be sustained in high-speed motor vehicle accidents or blows to the neck, rarely

 Copyright 2004, Elsevier (USA). All rights reserved.

creates perforating injury and most often leads to an intramucosal hematoma and subsequent dysphagia. However, penetrating injury in this setting can be associated with high morbidity and mortality, since it is often missed and when present may be accompanied by airway injury. Hence, a high index of suspicion is always required in the evaluation of trauma patients, for if unrecognized, esophageal injury has an extremely poor prognosis.

Diagnosis

Signs and symptoms of esophageal perforation are fairly similar and depend more on the site of perforation than the mechanism of the injury. Signs of perforation include tachycardia, fever, subcutaneous air, and, with injuries of the thoracic esophagus, chest dullness. Symptoms include pain, vomiting, dysphagia, dyspnea, and malaise. Cervical signs and symptoms include neck pain and occasionally subcutaneous air, whereas signs and symptoms of abdominal and thoracic esophageal perforation include, respectively, subxiphoid or epigastric pain and retrosternal pain occasionally radiating to the back. Tachycardia and dyspnea are usually uniformly present in patients with esophageal perforation. Sepsis, hypotension, and shock can also occur, usually later in the course of the disease.

The best diagnostic studies in symptomatic patients vary with the potential cause and the degree of clinical suspicion. Often in cases of iatrogenic endoscopic injury, no further diagnostic studies are required as the injury is sufficiently large for direct endoscopic visualization. When the diagnosis is less certain, chest x-ray suggests the diagnosis in over 90% of cases. pneumomediastinum, subcutaneous emphysema, and left-sided pleural effusion are all suggestive of esophageal injury. An upper gastrointestinal series (UGI) with water-soluble contrast will also confirm the diagnosis in 90% of patients. Barium is to be avoided in patients suspected of having esophageal perforation since leakage of barium out of the esophagus can result in chemical mediastinitis. Chest computerized tomography (CT) enhanced with enteral water-soluble contrast is currently the most sensitive test as it often demonstrates the site of perforation as well as the associated pneumomediastinum and pleural effusions. Chest CT is reported to be over 95% sensitive for the diagnosis of esophageal perforation in most studies. EGD can also be used to increase the diagnostic sensitivity; however, if used alone, esophagoscopy may miss the site of injury. Esophagoscopy is further limited by the difficulty of seeing the site of injury due to the usual absence of inflammation early after perforation and by the angulation of the esophagus. EGD may also exacerbate the injury and increase contamination of the area. Hence, it is not recommended as the primary modality of diagnosis. Chest CT and water-soluble UGI are the two recommended diagnostic modalities.

Management

Treatment of esophageal perforation is dependent on etiology, location, and the length of time between diagnosis and treatment. Patient prognosis worsens with increasing delay in diagnosis as a result of the release of inflammatory mediators and the onset of septic physiology. Iatrogenic endoscopic perforation carries the best prognosis as diagnosis is immediate, whereas strain-related perforation from vomiting (i.e., Boerhaave's syndrome) carries the worst prognosis as diagnosis is often delayed.

Anatomic location of the perforation influences survival as well. Cervical perforations are better contained and result in less spillage with a resulting mortality of less than 10%, whereas thoracic and abdominal perforations lead to widespread contamination of the mediastinum and abdomen with a mortality of over 50% in most series.

Initial management of esophageal perforation involves the prevention of oral intake, fluid resuscitation with intravenous solutions, histamine 2 (H_2) blockers, or proton pump inhibitors, and the administration of broad-spectrum antibiotics to cover oral and gastrointestinal flora including fungal organisms. Nasogastric tube placement is avoided until the perforation is repaired to prevent further escalation of injury if operative intervention is imminent; otherwise it is placed for gastric decompression.

Nonoperative management with antibiotics and parenteral hyperalimentation carries a 20–40% mortality rate in most series. However, in carefully selected patients with a well-contained, internally drained esophageal perforation with no signs of mediastinal or abdominal contamination and no septic physiology, nonoperative management may be attempted with a low acceptable mortality, especially in poor operative candidates. Typically, these esophageal leaks are located in the cervical esophagus, whereas most thoracic and abdominal perforations require operative intervention.

Surgical intervention remains the primary treatment of choice in the management of esophageal perforation. Operative intervention consists of (1) primary repair; (2) diversion with concomitant exclusion; (3) T-tube drainage; or (4) esophagectomy with reconstruction, which may be immediate or staged. Regardless of the

exact surgical method, the main principles of operative intervention include containing the leak and draining the injury site.

Primary repair for esophageal perforation is typically reserved for cervical injury, which can be repaired and subsequently drained. Primary repair for thoracic and abdominal perforations can be considered when there is minimal mediastinal and intra-abdominal contamination, respectively, and with minimal devitalized tissue in a stable patient. Primary repair is most successful within 24–48 h of injury. Repairs may be reinforced with healthy and viable tissues such as muscle—platysma and intercostal—thick pleura, pericardium, or adjacent stomach. In the repair of thoracic esophageal injuries, it is usually prudent to debride and drain the entire mediastinum initially. The esophagus is then mobilized around the area of perforation, which may be posterior with respect to the site of thoracotomy. Careful dissection will usually reveal the site. It is important to open the muscle layers over the perforation to identify the full extent of the injury, which is often subtle. A two-layer repair is usually recommended and intercostal muscle or other viable tissues are used to buttress the repair. The chest is then irrigated and extensively drained. In the case of very low esophageal perforations, a fundoplication or a Thal patch using the stomach can be accomplished as well.

In the presence of widespread contamination and substantial devitalized esophageal tissue, options include exclusion and diversion, T-tube drainage, or esophagectomy. The exclusion and diversion technique involves the creation of a cervical esophageal fistula and a gastrostomy and jejunostomy with delayed reconstruction 3 months to 1 year later. In this procedure, the mediastinum is debrided, the perforation is repaired, and the flow of saliva and bile is diverted. A modification in the thoracic procedure involves the repair and drainage of the perforation with stapling of the esophagus above and below the injury. In this procedure, a nasogastric tube is placed above the staple line and a gastrostomy is used for gastric decompression. Over an interval of a few months, the peristalsis of the esophagus usually allows for recanalization, obviating the need for additional surgery.

Another approach utilizes a biliary T-tube positioned inside the perforation with the hole debrided and sewn around it. This drainage with a Silastic tube creates a controlled esophagocutaneous fistula and allows for the removal of the T-tube 6 months to 1 year later after resolution of the inflammatory state. At that point, the patient can be reevaluated and the remaining esophagus studied to consider whether any further reconstructive options are needed.

Esophagectomy, with or without immediate reconstruction, should be considered for perforation in the setting of carcinoma, severe stricture or stenosis, or mega-esophagus. Reconstruction techniques include creating a "neo-esophagus" from the stomach, jejunum, colon, or platysma. If delayed reconstruction is chosen in the setting of a septic patient in whom a short operative procedure is desired, a cervical esophageal fistula and gastrostomy and jejunostomy are created, as in the diversion and exclusion technique. Esophageal stenting of small perforations has also been described, but large studies to determine its efficacy are lacking.

Postoperatively, these patients require enteral nutrition through feeding gastromies or jejunostomies. In the setting of the ileus, parenteral nutrition is required until the perforation has been shown to be healed, if primarily repaired, or the reconstructed neo-esophagus has been shown to be patent. Typically, water-soluble UGI studies are done to evaluate the operative result in these cases.

CAUSTIC ESOPHAGEAL INJURY

Etiology

Caustic esophageal injury usually results from suicide attempts and its severity depends on the type of chemical ingested, the quantity ingested, and the duration of exposure to the caustic agent. Acid (e.g., battery acid, bleach) exposure to esophageal tissue leads to coagulative necrosis, whereas alkaline agents (e.g., lye) cause liquefactive necrosis and a more severe injury pattern. Regardless of the causative agent, treatment must be initiated immediately to prevent the early and late-term sequelae of this extremely morbid and often fatal injury.

Diagnosis

Diagnosis of caustic injury should be suspected in all patients brought to the emergency room after attempted suicide. Signs and symptoms include oropharygneal pain from burn injury, emesis, drooling, and dysphagia. If perforation has occurred, signs and symptoms of perforation will appear as outlined above. Chest X ray and either a UGI series or a chest CT are required, especially if perforation is suspected. These studies should be followed by esophagoscopy 12–24 h postinjury if esophageal perforation is not suspected in order to examine the extent of esophageal involvement and depth of injury. Bronchoscopy is often required to evaluate the airways when involved by aspiration during initial ingestion or vomiting.

Management

Successful management of caustic injury depends on identification of the caustic agent and early detection of esophageal perforation, as these patients have the highest mortality of all perforating esophageal injuries. Esophageal perforation caused by caustic agents is managed similarly to perforation for other etiologies as outlined above. Most cases of esophageal perforation from caustic injury require either exclusion and diversion or esophagectomy, since extensive esophageal injury sustained in these cases is not amenable to primary repair or T-tube repair technique.

In cases where esophageal perforation has not resulted from caustic exposure, management consists of preventing oral intake, fluid resuscitation with intravenous solutions, H_2 blockers, or proton pump inhibitors, and the administration of broad-spectrum antibiotics. Steroids to reduce edema and down-regulate the inflammatory response in both the airway and the esophagus may also be used. However, no studies have shown a mortality benefit. Nasogastric tube placement is avoided until perforation is ruled out.

Patients are kept NPO (L. *non per os;* nothing by mouth) for 24–72 h until EGD and UGI series with water-soluble contrast show no evidence of esophageal leak, and at that time oral intake can be resumed. If more extensive injury is found at the time of endoscopy, the patient can be maintained NPO on intravenous fluids and antibiotics and receive nutrition enterally via nasogastric feeding tubes or parenterally until repeat endoscopy in 72 h to 1 week shows sufficient esophageal healing. If at any time the patient shows signs of systemic sepsis, an endoscopy, UGI, and chest CT should be performed to search for an esophageal perforation.

Late complications of caustic injury include tracheoesophageal fistula and strictures. Esophageal strictures can often be treated symptomatically with dilation and now stenting, but severe strictures often require esophagectomy and reconstruction as outlined above for perforation. Tracheoesophageal fistula requires operative intervention. Other late complications include severe reflux, often unrelieved by maximal medical therapy requiring surgical intervention with fundoplication, or rarely, esophagectomy and reconstruction. Patients surviving caustic esophageal injury also have an increased incidence of esophageal carcinoma and, hence, must be monitored for this complication with surveillance EGD and biopsy.

Esophageal trauma may be caused by perforating injury, usually the result of iatrogenic endoscopic injury; it may arise spontaneously in certain disorders associated with prolonged vomiting; or it may be the result of caustic injury usually associated with attempted suicide. Several good diagnostic tools and surgical techniques are available for managing esophageal trauma. Regardless of the etiology, esophageal trauma carries a high risk of mortality and morbidity if not diagnosed and treated rapidly.

See Also the Following Articles

Boerhaave's Syndrome • Esophageal Surgery • H2-Receptor Antagonists • Proton Pump Inhibitors

Further Reading

Goudy, S., Miller, F., and Bumpous, J. (2002). Neck crepitance: Evaluation and management of suspected upper aerodigestive tract injury. *Laryngoscope* **112**, 791–795.

Graeber, G., and Murray, G. (1992). Injuries of the esophagus. *Semin. Thorac. Cardiovasc. Surg.* **4**, 247–254.

Iannettoni, M., Vlessis, A., Whyte, R., and Orringer, M. (1997). Functional outcome after surgical treatment of esophageal perforation. *Ann. Thorac. Surg.* **64**, 1606–1609.

Sokolov, V., and Bagirov, M. (2001). Reconstructive surgery for combined tracheo-esophageal injuries and their sequelae. *Eur. J. Cardiothorac. Surg.* **20**, 1025–1029.

Thompson, E., Porter, J., and Fernandez, L. (2002). Penetrating neck trauma: An overview of management. *J. Oral Maxillofac. Surg.* **60**, 918–923.

Tomaselli, F., Maier, A., Pinter, H., and Smolle-Juttner, F. (2002). Management of iatrogenous esophagus perforation. *Thorac. Cardiovasc. Surg.* **50**, 168–173.

Wright, C., Mathisen, D., Wain, J., Moncure, A., Hilgenberg, A., and Grillo, H. (1995). Reinforced primary repair of thoracic esophageal perforation. *Ann. Thorac. Surg.* **60**, 245–248.

Zwischenberger, J., Savage, C., and Bidani, A. (2002). Surgical aspects of esophageal disease. *Am. J. Respir. Crit. Care Med.* **165**, 1037–1040.

Esophageal Ulcers

KANTHI YALAMANCHILI AND MARK FELDMAN
Presbyterian Hospital of Dallas

dysphagia Difficulty in swallowing.
odynophagia Pain experienced when swallowing food or liquid.
ulcer A hole in the mucosal lining of the esophagus that often results in pain or heartburn.

An esophageal ulcer is a consequence of disruption of esophageal mucosal integrity leading to a local mucosal defect or excavation, often associated with active inflammation. Esophageal ulceration results from a breakdown of esophageal mucosal defense and repair. Symptoms of esophageal ulcers include odynophagia, dysphagia, and heartburn or chest pain that may be exacerbated by eating. Etiologies of ulcers are multiple and are sometimes idiopathic.

INTRODUCTION

Esophageal ulcers may result from gastroesophageal reflux disease, immunosuppression/infections, medication or chemical ingestion, tumors, or certain medical treatments and procedures (Table I).

GASTROESOPHAGEAL REFLUX DISEASE

Gastroesophageal reflux disease (GERD) occurs most often due to impaired lower esophageal sphincter function. As a result, stomach contents reflux into the esophagus. If left untreated, GERD may progress from esophageal erosions to formation of discrete esophageal ulcers. In some cases, such acid reflux may lead to Barrett's metaplasia.

IMMUNOSUPPRESSION-ASSOCIATED ESOPHAGEAL ULCERS

Esophageal ulcers may develop in early stages of HIV infection, when transient fever, chills, malaise, skin rash, and lymphopenia often occur; the ulcers tend to be multiple small aphthoid lesions. Later in the course of HIV infection (AIDS), deep ulcers extending up to several centimeters in size may be seen. These larger ulcers can result in fistula formation, perforation, hemorrhage, or stricture formation. Opportunistic infections in AIDS patients from *Candida* species or certain viruses (cytomegalovirus, herpes simplex virus, and rarely Epstein–Barr virus) may also result in esophageal ulcer formation. The pathophysiology of idiopathic esophageal ulcers in HIV patients is poorly understood, although it has been speculated that the squamous epithelial cells lining the esophagus undergo apoptosis as an "innocent bystander" effect of nearby HIV-infected T cell activity.

TABLE I Classification of Esophageal Ulcers[a]

Related to GERD (i.e., peptic ulcers), often with Barrett's esophagus
Related to HIV infection and other conditions with immunosuppression
Candida species
HSV
CMV
EBV
Idiopathic
Infections (immunocompetent hosts)
HSV
EBV (rare)
Mycobacterium tuberculosis
Candida—sometimes soon after antibiotic therapy (rare)
Pill or medication—induced, or ingestion of caustics
Antibiotics (especially tetracyclines)
Potassium chloride tablets
Iron tablets
NSAIDs
Bisphosphonates (e.g., alendronate)
Quinidine
Antivirals (e.g., zalcitabine)
Strong acids or alkalis (caustics)
Tumors with ulcers
Squamous cell cancer
Adenocarcinoma (often with Barrett's esophagus)
Other (metastatic, melanoma)
Following sclerotherapy or band ligation for esophageal varices

[a] Abbreviations: GERD, gastroesophageal reflux disease; CMV, cytomegalovirus; HSV, herpes simplex virus; EBV, Epstein–Barr virus; NSAIDs, nonsteroidal antiinflammatory drugs.

Copyright 2004, Elsevier (USA). All rights reserved.

Organ transplant recipients and patients on potent antiinflammatory drugs (e.g., glucocorticoids and tumor necrosis factor inhibitors) can also develop opportunistic esophageal infections with ulcer formation.

IMMUNOCOMPETENCY-ASSOCIATED ESOPHAGEAL ULCERS

In immunocompetent persons, esophageal ulcerations can rarely complicate infectious mononucleosis. Herpes simplex virus (HSV) and *Mycobacterium tuberculosis* can also cause esophageal ulcers. These infections require confirmatory histopathologic diagnosis during esophagoscopy. HSV esophagitis and ulcers may occur spontaneously via reactivation and spread of HSV to the esophageal mucosa via the vagus nerve or may result from direct extension of the oropharyngeal infection into the esophageal mucosa.

MEDICATION-INDUCED ESOPHAGEAL ULCERS

Multiple medications have been linked to esophageal ulcers (Table I). Antibiotics taken as pills or tablets, especially doxycycline, potassium chloride, iron-containing pills, nonsteroidal antiinflammatory drugs (NSAIDs), quinidine, bisphosphonate medications, and antiviral drugs, comprise about 90% of all cases. Because the pharmacologic properties of these medications are so diverse, it is difficult to pinpoint a single unifying mechanism for pill-induced ulceration. Prolonged mucosal contact, the chemical nature of the drug, and drug solubility all contribute to the development of esophageal ulcers. The contact time is dependent on the size and shape of the pill and variations in esophageal anatomy. Coated pills, larger pills, pills taken at bedtime, and sustained-release formulations seem to enhance the chance of pill ulcerations. A preexisting swallowing disorder need not be present. However, there is a predisposition to pill ulceration in patients with preexisting esophageal strictures, tumors, and motility disorders such as achalasia and scleroderma. Symptoms of pill-induced ulcers of the esophagus include sudden onset of painful swallowing and retrosternal chest pain, which usually resolve 5 days to 2 weeks after discontinuation of the medication.

Diagnosis of NSAID-induced ulceration is one of exclusion based on elimination of reflux disease, cancer, or infections as possible causes in patients concurrently taking NSAIDs. Most common culprits are aspirin, ibuprofen, and naproxen. Bleeding is a common feature and there seems to be a prevalence of distal esophageal ulcers in these patients.

The bisphosphonates, which are often used for osteoporosis and Paget's disease may also cause esophagitis. Patients may report esophageal discomfort with the very first dose and a majority experience odynophagia within a week. Bleeding and perforation may result. This patient population tends to be older women who may not take precautions against esophageal injury, such as swallowing the pill with 6–8 ounces of water and then sitting upright for 30 minutes. Iron and potassium compounds in the sustained release preparations can also cause ulcerations. Patients usually do not experience pain with ingestion of iron and potassium and the onset of dysphagia is very slowly progressive.

Zalcitabine (dideoxycytidine, or ddC), a nucleoside analogue used as an antiviral agent, has been reported to cause oral and esophageal ulcers. The majority of these ulcerations resolve within 4 to 7 days even with continuation of the medication.

Caustic ingestions remain a worldwide problem. Two-thirds of household ingestions are accidental and occur in children younger than 6 years of age. Caustic agents may be alkaline or acidic, solid or granular. The concentration, quantity, and physical state of the agent along with the duration of exposure determine the degree of gastrointestinal injury, initially presenting as ulcers but sometimes progressing to severe bleeding, fibrosis with strictures, perforation, or (years later) carcinoma. In caustic ingestions, a poor correlation exists between upper endoscopy findings of gastrointestinal injury and clinical signs and symptoms, making upper endoscopy even more crucial to assess injury. However, the timing of an upper endoscopy after the caustic ingestion is debatable. Waiting 48 to 72 hours has been considered reasonable, although with flexible endoscopes, endoscopy may be performed even earlier if the patient is stable. Healing of injury is often associated with stricture formation.

TUMORS

Cancer of the esophagus with ulcer formation is relatively uncommon but extremely lethal if it occurs. Squamous cell esophageal carcinomas and adenocarcinomas cannot be distinguished endoscopically or radiographically. They mostly tend to cause symptoms in later stages when $>60\%$ of the esophageal lumen is infiltrated with the cancer. These malignancies cause

ragged, ulcerating changes in the esophageal mucosa. Therefore, esophagoscopy is crucial in visualizing the tumor and obtaining histopathologic confirmation.

ENDOSCOPIC BAND LIGATION, SCLEROTHERAPY, AND CHRONIC NASOGASTRIC TUBE PLACEMENT

Band ligation for treatment of esophageal varices in portal hypertension leads to superficial ulcers that occur where the bands are placed, but the frequency of deep esophageal injury is low. Many patients who undergo variceal sclerotherapy develop ulcers within the first few days of treatment because the injection of the sclerosant into and around the varices causes necrosis of esophageal tissue. Nasogastric tubes can cause local esophageal irritation and the ulcers that result may erode into adjacent structures such as major vessels. Studies have shown that even potent gastric acid suppressive therapies do not prevent esophageal damage associated with nasogastric tubes, incriminating local irritation as the key factor responsible.

TREATMENT AND PREVENTION

Proton pump inhibitors are the treatment of choice for esophageal ulcers due to GERD. High doses for 8–12 weeks may be needed to heal the ulcers. HIV associated idiopathic ulcers do not respond to antiviral or antifungal therapies but do respond to immunosuppressive regimens. Prednisone and thalidomide are the two most common agents used.

For pill induced ulcerations, patients with preexisting swallowing difficulties should take precautions such as sitting in the upright position, taking medications with food and water, and using liquid formulations if necessary. Lifestyle changes are key, such as taking pills with food or 6 ounces of water and avoiding recumbency immediately after pill ingestion. Speeding up healing is the only strategy for pill-induced ulcerations. Acid suppressive therapies may possibly allow for faster healing, but this has not been proved. In ulcerations related to ingestion of pill or tablet forms of antibiotics and other medications, epithelial inflammation resolves in 1 to 6 weeks after the medications are discontinued. Avoiding caffeine, alcohol, tobacco, and spicy foods may also possibly contribute to the healing of ulcers.

Infection-related esophageal ulcers respond to treatment of the specific infection. Goals of therapy after a caustic ingestion are to prevent perforation and avoid progressive fibrosis and stricture. Emergency surgery is the only option to treat perforation, but a variety of modalities, although unproved, have been speculated to prevent fibrosis and treat strictures; these include caustic neutralizing agents, corticosteroids, antibiotics, collagen synthesis inhibitors, heparin, early esophageal dilatation, and esophageal stents.

See Also the Following Articles

Barrett's Esophagus • Chest Pain, Non-Cardiac • Dysphagia • Fibrogenesis • Gastroesophageal Reflux Disease (GERD) • NSAID-Induced Injury • Proton Pump Inhibitors

Further Reading

Feldman, M., Friedman, L. S., and Sleisenger, M. H. (eds.) (2002). "Gastrointestinal and Liver Disease," 7th Ed., W. B. Saunders, Philadelphia.

Jaspersen, D. (2000). Drug-induced oesophageal disorders: Pathogenesis, incidence, prevention and management. *Drug Saf.* **22(3)**, 237–249.

Kikendall, J. W., Friedman, A. C., Oyewole, M. A., et al. (1983). Pill-induced esophageal injury: Case reports and review of medical literature. *Dig. Dis. Sci.* **28**, 174–181.

Kitchen, V. S., *et al.* (1990). Epstein–Barr virus associated oesophageal ulcers in AIDS. *Gut* **31**, 1223.

Tilbe, K. S., *et al.* (1986). A case of viral esophagitis. *J. Clin. Gastroenterol.* **8**, 494.

Ward, H. A. (2002).Cutaneous manifestations of antiretroviral therapy. *J. Am. Acad. Dermatol.* **46(2)**, 284–293.

Esophagus, Anatomy

LORI A. ORLANDO* AND ROY C. ORLANDO*,†

** Tulane University School of Medicine and †New Orleans Veterans Administration Hospital*

Auerbach's plexus Network of neurons located between the circular and longitudinal muscle layers of the gastrointestinal tract; primarily involved with the transmission of motor information between the central nervous system and the peripheral end organ. Idiopathic achalasia and diffuse esophageal spasm are two esophageal disorders believed to reflect damage to this plexus.

Barrett's esophagus Existence of specialized columnar epithelium instead of the stratified squamous epithelium that normally lines the distal esophagus; a premalignant lesion that is believed to occur by metaplastic change during the course of reflux damage to squamous epithelium.

Meissner's plexus Network of neurons located within the submucosa of the gastrointestinal tract; primarily involved with the transmission of sensory information between the central nervous system and the peripheral end organ.

sphincter Ringlike band of muscle fibers that constricts a passage or closes a natural orifice; may be skeletal, as in the upper esophageal and external anal sphincters, or smooth, as in the lower esophageal and internal anal sphincters.

The esophagus is designed as a delivery system to transport food from the mouth to the stomach. A hollow tube, it conducts food through two mechanisms, gravity and peristaltic contractions of the surrounding musculature. Retrograde flow of food and gastric acid is prevented by two sphincters, the upper and lower esophageal sphincters, which open in response to a swallowed bolus. This effectively creates a hollow conduit with a one-way flow of contents into the stomach. Located within the posterior mediastinum, immediately behind the trachea but anterior to the aorta, the body of the esophagus extends from the upper esophageal sphincter in the pharynx at the cricoid cartilage to the lower esophageal sphincter in the right crus of the diaphragm at the T10 vertebral level.

MUCOSA

The innermost mucosa of the esophagus surrounded by three outer layers, the submucosa, the muscularis propria (the inner circular and outer longitudinal muscular layers) (see Fig. 1), and outermost adventitia. Each layer is structurally unique, with varying contributions by the vasculature, musculature, and neural innervation. Grossly, the esophageal mucosa is a smooth, pink expanse that ends at the gastroesophageal junction demarcated by the Z-line, an irregular white line. Histologically, the mucosa is composed of the epithelial layer (nonkeratinized stratified squamous epithelium) and the lamina propria and muscularis mucosae layers (Fig. 2). Metaplastic change from stratified squamous epithelium to a specialized columnar epithelium occurs in Barrett's esophagus, a

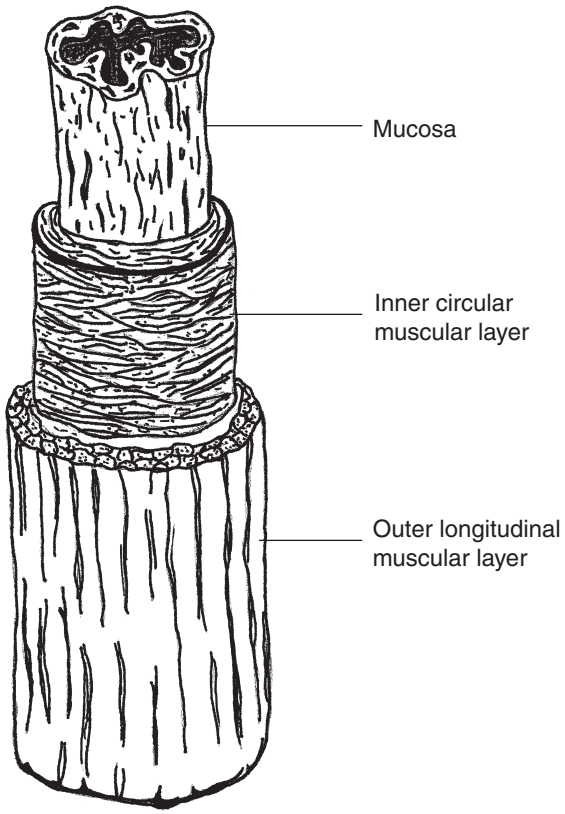


FIGURE 1 Anatomical relationship of the layers of the esophagus. The muscularis propria is made up of the inner circular and outer longitudinal muscular layers. Note the hollow lumen lined by the mucosa, which is collapsed between food boluses.

 Copyright 2004, Elsevier (USA). All rights reserved.

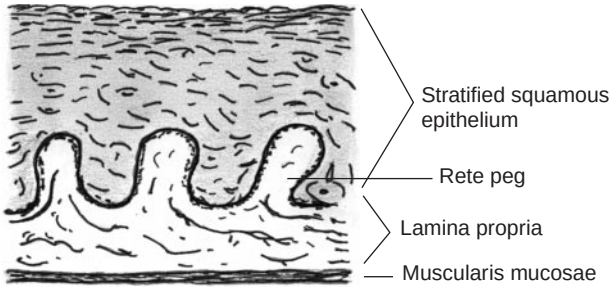


FIGURE 2 The three layers of the mucosa, depicting the superficial stratified squamous epithelium and the sequentially deeper lamina propria and muscularis mucosae. Note the rete pegs of the lamina propria projecting into the epithelial layer above.

premalignant condition. The lamina propria, a connective tissue and vascular layer, resides beneath the epithelium. Rete pegs, also known as dermal papillae, are rounded protrusions of the lamina propria into the epithelial layer above. Although rete pegs are normal, extension into the upper half of the epithelium is a marker for gastroesophageal reflux disease. The final mucosal layer, the muscularis mucosae, is composed of smooth muscle cells. It demarcates the mucosa from the submucosa. Neural structures are present in the mucosa, with sensory tufts extending into the epithelium; arterial and venous structures are present in the lamina propria (see later for discussion of venous drainage of the esophagus).

SUBMUCOSA

The submucosa, a dense network of connective tissue, blood vessels, lymphatics, neurons, and esophageal glands, primarily functions as a secretory layer. The glands are acinar, composed of cuboidal cells that, through a collecting system, secrete into the esophageal lumen mucus, bicarbonate, and epidermal growth factor. The mucus is primarily for lubrication because, unlike in the stomach, it does not form a definable viscoelastic protective layer over the squamous epithelium. The secreted bicarbonate protects the lumen by neutralizing refluxed acid, thereby raising pH to normal. Once esophageal damage has occurred, epidermal growth factor can bind to cell receptors and stimulate esophageal repair. Meissner's plexus, a neural network located within the submucosa, transmits sensory (afferent) stimuli to the central nervous system through both parasympathetic and sympathetic pathways.

MUSCULARIS PROPRIA

The muscularis propria provides the peristaltic contractions necessary to propel a food bolus from the oral cavity to the stomach. The proximal third of the muscularis propria is composed exclusively of skeletal (striated) muscle and the distal third is composed exclusively of smooth muscle; the middle third is a combination of both. Two layers, an inner circular muscular layer and outer longitudinal muscular layer, traverse the length of the muscularis propria (Fig. 1). The amplitude of peristaltic contractions, as measured by manometry, correlates with the efficacy of bolus clearance. Furthermore, inefficient peristaltic contractions can lead to dysphagia, as occurs in the swallowing disorder, achalasia. Innervation of the muscularis propria is complex and differs for the striated and smooth muscle cells. Vagal control of peristalsis occurs at two different nuclei, the dorsal motor nucleus for smooth muscle cells and the nucleus ambiguous for skeletal muscle cells. The Auerbach's (myenteric) plexus, located between the inner circular and outer longitudinal layers, provides preganglionic vagal parasympathetic innervation to smooth muscle cells, whereas the postganglionic vagal nerves end directly on the motor endplate of the skeletal muscle cells. Within the myenteric plexus, excitatory and inhibitory neurons coexist. The excitatory neurons have muscarinic M_2 receptors, which contract both layers of smooth muscle in response to acetylcholine stimulation. Inhibitory neurons induce inner circular smooth muscle cells to relax, probably through the release of nitric oxide and vasoactive intestinal peptide.

ADVENTITIA

The final layer of the esophagus, the adventitia, is a loose connective tissue network. Although all other digestive tract structures contain a serosal layer, the esophagus does not.

UPPER ESOPHAGEAL SPHINCTER

The upper esophageal sphincter (UES) marks the upper boundary and thus the beginning of the esophagus. It is created by the convergence of the inferior pharyngeal constrictor muscle and the cricopharyngeus muscle (Fig. 3). This band of striated muscle, innervated by the pharyngeal branch of the vagus nerve, is contracted at rest, creating a high-pressure zone about 1–2 cm in length. The primary function of the UES is to prevent aspiration of air into the esophagus during normal respiration. Relaxation occurs during the pharyngeal phase

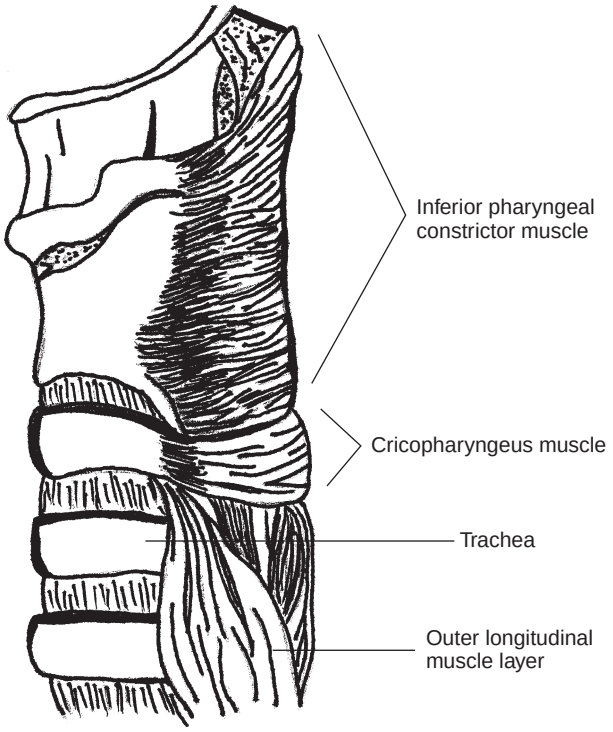


FIGURE 3 Anatomy of the upper esophageal sphincter.

of the swallow response, when elevation of the larynx and hyoid simultaneously close the airway and open the esophageal sphincter.

LOWER ESOPHAGEAL SPHINCTER

The lower esophageal sphincter (LES) forms the lower boundary of the esophagus, creating a barrier between it and the gastric fundus. Located within the right diaphragmatic crus, it is composed of asymmetrically thickened smooth muscle about 3–4 cm in length (Fig. 4). In addition, the fortuitous positioning of the LES within the diaphragm allows diaphragmatic contractions to make a significant contribution to the high-pressure zone at the esophagogastric junction. Esophageal sliding is limited by the phrenoesophageal ligament, which attaches the lower esophagus to the diaphragmatic fascia. Despite this preventive measure, the esophagus occasionally migrates orad, permitting a portion of the fundus to enter the thorax, the so-called sliding hiatal hernia, and allowing reductions in LES pressure. Normally, the LES is tonically contracted, resulting in manometric pressures that are about 20 mmHg greater than in the stomach. This creates a high-pressure zone that combats the reflux of gastric acid. Relaxation occurs in response to a swallowed food bolus, lasting about 5–7 seconds. Relaxation also occurs in response to gastric distension, as in the belch reflex, but lasts longer, at 10–20 seconds. Myenteric neurons convey parasympathetic vagal impulses to the LES, transmitting excitatory and inhibitory signals analogous to their actions within the muscularis propria. Neural innervation contributes greatly to LES pressure; however, many substances can influence pressure, either increasing or decreasing it. These substances range from food to medications to hormones.

CIRCULATION

The esophageal arterial blood supply arises from multiple vessels, which anastamose within the submucosa, forming extensive networks. The superior and inferior thyroid arteries supply the upper esophagus. The descending aorta and bronchial and right intercostal arteries supply the midesophagus, and the left gastric, left inferior phrenic, and splenic arteries supply the distal esophagus. The venous drainage is executed by a network of vessels within the esophageal wall (Fig. 5). Small intraepithelial vessels drain from the stratified epithelium of the mucosa into the superficial venous plexus within the two deeper layers, the lamina propria and muscularis mucosae. From the superficial plexus, blood is drained into the deep intrinsic veins located within the submucosa. Perforating veins traverse laterally through the muscularis propria to connect the deep intrinsic veins to the adventitial extrinsic

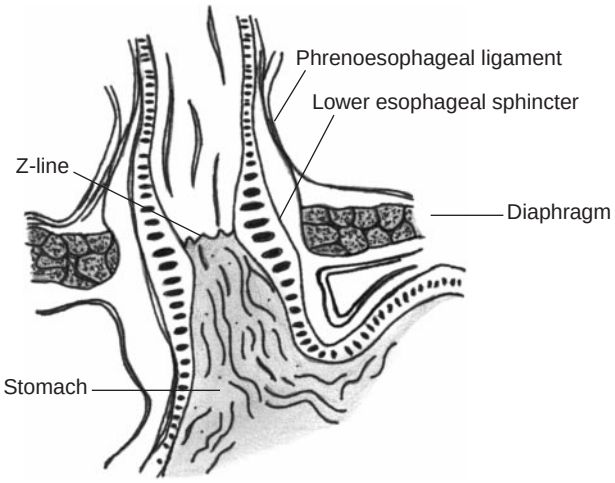


FIGURE 4 Anatomy of the lower esophageal sphincter. Note the phrenoesophageal ligament; it attaches the esophagus to the diaphragmatic fascia.

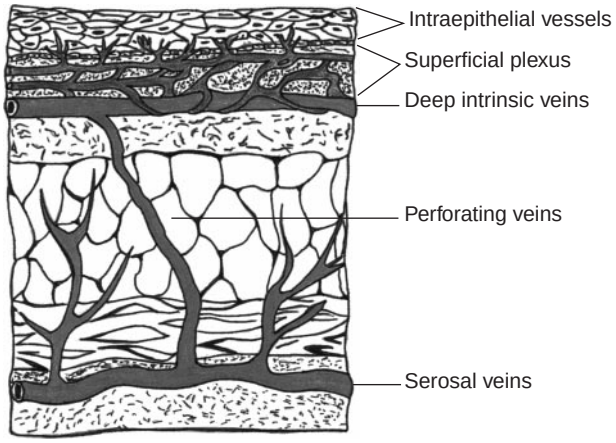


FIGURE 5 Venous drainage of the esophagus.

venous drainage system, which is composed of the serosal and periesophageal veins. Thereafter, drainage is variable, depending on the location. The upper esophagus drains to the superior vena cava, the mid-esophagus drains to the azygous veins, and the distal esophagus drains to the portal vein. Portal hypertension leads to dilation of the deep intrinsic veins, eventually displacing more superficial structures and protruding into the esophageal lumen. The lymphatic system is also segmental, with extensive submucosal anastomoses that account for the distant spread of most esophageal cancers.

See Also the Following Articles

Gastrointestinal Tract Anatomy, Overview • Stomach, Anatomy

Further Reading

Biancani, P, Harnett, K., and Behar, J. (1999). Esophagus. *In* "Textbook of Gastroenterology" (T. Yamada, D. Alpers, L. Laine, C. Owyang, and D. Powell, eds.), 3rd Ed., Vol. 1, pp. 165–181. Lippincott Williams & Wilkins, Philadelphia.

Ergun, G. A., and Kahrilas, P. J. (1997). Esophageal anatomy and physiology. *In* "Gastroenterology and Hepatology. The Comprehensive Visual Reference" (M. Feldman and R. C. Orlando, eds.), Vol. 5, pp. 2–19. Current Medicine, Philadelphia.

Long, J. C., and Orlando, R. C. (1998). Esophageal anatomy and development. *In* "Sleisenger & Fortran's Gastrointestinal and Liver Disease" (M. Feldman and B. F. Scharschmidt, eds.), 6th Ed., Vol. 1, pp. 457–459. W. B. Saunders, Philadelphia.

Esophagus, Development

LORI A. ORLANDO* AND ROY C. ORLANDO*,†

*Tulane University School of Medicine and †New Orleans Veterans Administration Hospital

atresia The congenital absence or closure of a normal body orifice or tubular organ, e.g., esophageal or biliary atresia.

diverticulum A circumscribed pouch or sac of variable size that may occur normally, as in embryologic development of the respiratory bud off the foregut, or pathologically, resulting from the herniation of the lining mucous membrane through a defect in the muscular wall of a tubular organ.

dysphagia A symptom in which the patient experiences trouble or difficulty swallowing. When the symptom occurs with solids only, it usually reflects a lumen-narrowing obstruction and when it occurs with both liquids and solids, it usually reflects an oropharyngeal or esophageal motor disorder.

Schatzki ring A thin circumferential ingrowth of mucosa at the squamocolumnar junction between esophagus and stomach. When the lumen is compromised to <15 mm, the ring may result in intermittent dysphagia for solids or an episode of acute esophageal food impaction.

sphincter A ringlike band of muscle fibers that constricts a passage or closes a natural orifice. The muscle may be skeletal as in the upper esophageal and external anal sphincters or smooth as in the lower esophageal and internal anal sphincters.

stenosis A pathologic narrowing or stricture of a duct or passage way that may be either acquired, e.g., by radiation injury to the esophagus, or congenital, e.g., esophageal stenosis.

Organ development is a complex and intricate process. A general understanding of esophageal embryology provides key information regarding esophageal disorders, as several commonly encountered abnormalities can be traced back to fetal life. This article describes the general process by which the esophagus emerges from precursor tissues and grows throughout childhood. Several important developmental anomalies are also reviewed.

EMBRYOLOGY

In the embryo, invagination of the endoderm-lined yolk sac into a hollow cylinder creates the primitive gut, which is separated into the pharyngeal, gut, foregut, midgut, and hindgut. Each segment ultimately matures into a separate portion of the digestive tract. The upper digestive tract and lower respiratory tract develop from the foregut. At gestational week 4, the respiratory bud appears as a diverticulum on the ventral surface of the foregut. The diverticulum elongates into a parallel hollow tube, which is gradually isolated from the dorsal portion by a thin membrane, the esophagotracheal septum. The ventral tube, the primordial respiratory tract, goes on to develop into the bronchi, bronchioles, alveoli, and visceral pleura. Meanwhile, the esophagus, the dorsal foregut tube, undergoes lumenal obliteration by proliferating ciliated columnar epithelium. Later, at week 10, the lumen recanalizes through the formation and coalescence of vacuoles. At week 16, stratified squamous epithelium replaces the columnar epithelial lining and the transformation is complete.

CHILDHOOD

Within a few days of birth, the neonate is able to reflexively swallow and breathe without choking, in part due to a fully functional and competent upper esophageal sphincter. The lower esophageal sphincter (LES) is another story. A majority of infants have gastroesophageal reflux, varying from asymptomatic to severe reflux esophagitis, as a consequence of LES incompetence. Factors associated with lower sphincter pressures include the angle of the gastroesophageal junction and the positioning of the sphincter within the abdomen instead of its correct location, being straddled by the diaphragm. In addition, the smaller capacity of the esophagus promotes vomiting in response to reflux more frequently in children than in adults. As the infant grows, the LES angle and location correct their position, leading to a fully competent sphincter and resolution of pathologic gastroesophageal reflux.

DEVELOPMENTAL ANOMALIES

Congenital anomalies of the esophagus are common, occurring in 1 in 3000 live births. They are associated with either genetic defects or intrauterine stresses that impede fetal maturation. Premature infants are prone to

 Copyright 2004, Elsevier (USA). All rights reserved.

developmental anomalies, with up to 50% having involvement of the vertebral, anal, cardiac, tracheal, esophageal, renal, and limb systems. Esophageal anomalies are particularly frequent.

Esophageal Atresia and Tracheoesophageal Fistulas

Esophageal atresia and tracheoesophageal fistulas are the most common developmental anomalies of the esophagus. The former results from the failure of the foregut to recanalize and the latter result from the failure of the lung bud to separate completely. Esophageal atresia occurs alone in 7% of cases, with tracheoesophageal fistulas accompanying the remainder of cases. In both forms, the esophagus is divided into two separate unconnected segments: the upper esophagus, which ends in a blind pouch, and the lower esophagus. In isolated atresia, the lower esophagus connects to the stomach, and in tracheoesophageal fistulas, it communicates with the trachea.

Congenital Stenosis

Esophageal stenosis is a rare anomaly, occurring in 1 in 25,000 live births. The stenotic segment, located in the distal two-thirds of the esophagus, varies from 2 to 20 cm in length. Although the etiology is unknown, one suggestion is that there is incomplete separation of the lung bud from the primitive foregut. Other hypotheses include fibromuscular hypertrophy or loss of muscle-relaxing myenteric neurons.

Esophageal Rings

The distal esophagus may contain two "rings," the A and B rings, which bound the proximal and distal esophageal vestibule and constrict the esophageal lumen. The A ring, a 4 mm band of hypertrophied muscle, is rare and marks the proximal border of the vestibule at the upper end of the LES. The B ring, also called the Schatzki ring, is common, found in up to 14% of the general population. A thin 2 mm membrane marking the distal border of the esophageal vestibule at the gastric cardia, it is associated with hiatal hernias. When present, it also demarcates the squamocolumnar junction. Classically, they remain asymptomatic until the esophageal lumen is narrowed to less than 15 mm, at which point intermittent dysphagia for solids and acute solid-food impactions can occur.

Esophageal Webs

Esophageal webs are congenital anomalies characterized by one or more thin horizontal membranes of stratified squamous epithelium within the upper and mid esophagus that protrude from the anterior wall and extend laterally. Ninety-five percent are symptomatic, usually with dysphagia. An association between cervical esophageal webs, dysphagia, and iron deficiency anemia occurs in Plummer-Vinson (or Paterson-Kelly) syndrome. Correction of iron deficiency may result in both anatomical correction of the web and symptom resolution.

See Also the Following Articles

Development, Overview • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Neonatal Tracheoesphageal Anomalies • Webs

Further Reading

Herbst, J. J. (1996). The esophagus. *In* "Nelson Textbook of Pediatrics" (W. E. Nelson, R. E. Behrman, R. M. Kliegman, and A. M. Arvin, eds.), 15th Ed., Vol. 1, pp. 1051–1058. W. B. Saunders Co., Philadelphia, PA.

Long, J. C., and Orlando, R. C. (1998). Esophageal anatomy and development. *In* "Sleisenger and Fortran's Gastrointestinal and Liver Disease" (M. Feldman and B. F. Scharschmidt, eds.), 6th Ed., Vol. 1, pp. 459–465. W. B. Saunders Co., Philadelphia, PA.

Exocrine Pancreas

NINA W. FLAY AND FRED S. GORELICK
Yale University and Veterans Affairs Medical Center, Connecticut

acinar cells Pyramidal-shaped epithelial cells in the pancreas that synthesize and secrete digestive enzymes.
acinus Functional unit of the exocrine pancreas composed of acinar, centroacinar, and duct cells.
centroacinar cells Terminal duct cells bordering the acinar lumen that secrete bicarbonate into the lumen.
exocytosis Process by which secretory granules fuse with the plasma membrane and release their contents.
goblet cells Mucus-synthesizing and secreting cells found in the epithelium of pancreatic ducts.
islet–acinar portal system Specialized vascular system connecting the islets of Langerhans and the exocrine pancreas that delivers islet hormones to the acinar cells.
pancreatic stellate cells Cells of stellate morphology located between acini that contain vitamin A and respond to cytokines with the production of collagen, leading to fibrosis.
zymogen granules Mature secretory granules in acinar cells containing stored digestive enzymes or zymogens.

The pancreas is a mixed endocrine/exocrine gland enveloped by a thin layer of connective tissue. Connective tissue septa invade the gland, dividing it into lobules, serving as a support for tissue and providing a pathway for vasculature, nerve fibers, and lymphatic ducts. The pancreas lacks a complete fibrous capsule. As a result, cancerous tumor cells can spread to all of the structures located posteriorly to the pancreas by traveling along nerve sheaths. They facilitate their passage by generating proteases that degrade connective tissue. The exocrine portion of the pancreas takes up about 84% of the total organ; 2% of the organ is composed of the endocrine cells (islets of Langerhans). Blood vessels comprise about 4% of the pancreas and the remaining 10% of the pancreas is composed of the extracellular matrix (collagen fibers, reticular fibers, fibroblasts). Enhanced deposition of connective tissue within the pancreas is a feature of chronic pancreatitis of various etiologies. The collagen is probably synthesized and secreted from stimulated fibroblasts and a special class of vitamin A-storing cells known as pancreatic stellate cells. Also, under normal conditions, there is a considerable amount of fat scattered throughout the parenchyma of the human pancreas. Notably, this is the first tissue to show injury during the inflammatory disease of the pancreas known as pancreatitis.

FUNCTIONAL MORPHOLOGY OF THE EXOCRINE PANCREAS

The acinus is the basic functional unit of the exocrine pancreas. It is composed of acinar cells, centroacinar cells, and duct cells (Fig. 1). A connective tissue matrix, including a basal lamina, surrounds the acinus, but does not insert itself between individual acinar cells. Nerve terminals in the exocrine pancreas differ among species, but in general they end near blood vessels, acinar cells, and duct cells. In mammals, they terminate near the acinar cell plasma membrane. In pancreatic duct cells, nerve terminals are usually on or near the basement membrane. The presence of nerve terminals so close to the basement membrane of these cells is critical because regulation of pancreatic acinar cell secretion is partially under the control of cholecystokinin (CCK) released from proximal intestinal cells in response to food intake. CCK has an indirect effect on pancreatic secretion by stimulating the release of acetylcholine (ACh) and other neurotransmitters from nerves. ACh, in turn, activates muscarinic receptors on the acinar cell to stimulate pancreatic secretion.

The glandular structure of the exocrine pancreas differs slightly from that of other digestive glands: the pancreas is composed of true glands with blind endings, similar to salivary gland structure, but also has other

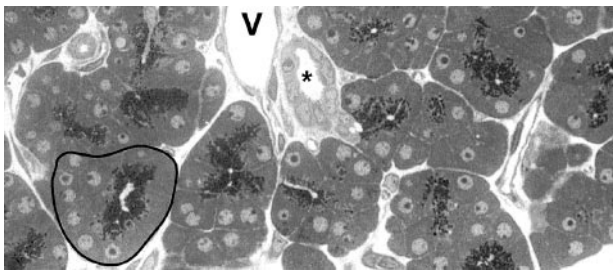


FIGURE 1 Exocrine pancreas structure. A portion of a rat pancreatic lobule containing individual acini is visible. An intralobular duct (asterisk) alongside a blood vessel (V) is visible at the top of the field. A single acinus (dark outline) is made up of several acinar cells. The polarized cells reveal large nuclei in the basolateral domains and zymogen granules near the apical surface. The clear area in the center of the acinus is the lumen.

 Copyright 2004, Elsevier (USA). All rights reserved.

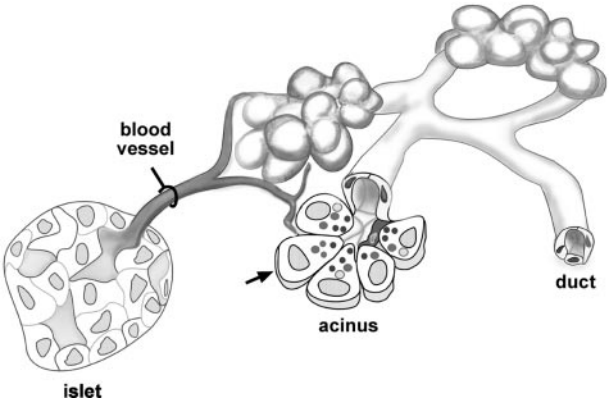


FIGURE 2 Acinar–islet portal system and duct structure. Note the blood vessel directly connecting the islet of Langerhans to the group of acini. Two duct structures (blind-ended and looped) are illustrated. The small gray shaded cell adjacent to an acinar cell at the beginning of the duct system represents a centroacinar cell. The arrow points to a single acinar cell in the acinus.

"looped" ducts seen going into, and emerging from, groups of acini (Fig. 2).

ACINAR PORTAL SYSTEM

The islet–acinar portal system is an arterial supply network that connects the islet cells with the exocrine cells (Fig. 2). The system surrounds and invades the islet cells (endocrine system), receives hormones and other peptides from islet cells, and then directly delivers its contents to the neighboring acinar cells (exocrine system). These endocrine hormones may modulate the function of the exocrine pancreas. For example, in the region adjacent to the pancreatic islet cells, the acinar cells contain more cytoplasm, a larger nucleus, and an increased amount of zymogen granules compared to those that are more distal to the islet. Islet hormones may also modulate both basal and stimulated pancreatic exocrine secretion.

THE ACINUS

The three-dimensional architecture of the acinus has been revealed using confocal analysis and scanning electron microscopy, which show the relationship of the connective tissue elements, vasculature, and excretory ducts to the acinar and centroacinar cells. Collagen and reticular fibers form a scaffold that surrounds the acini. The reticular fibers form a network with fenestrations, allowing for a continuous space between ducts, acini, vasculature, and lymph vessels. Techniques that use tissue disruption and scanning electron microscopy reveal that most pancreatic acini contain 30 to 40 large cells, but some smaller groups of acini of 10 to 20 cells are also present. However, electrical coupling experiments show that the functional coupled subunit is much larger and consists of about 500 cells. A possible advantage of an increased size of the functional acinus might be an enhanced secretory response when only a limited number of cells are stimulated.

Acinar Cells

Pancreatic acinar cells are pyramidal-shaped, polarized epithelial cells. The base of the cell rests on a basement membrane and the apex opens onto the lumen of the acinus. Junctional complexes are found between the cells, dividing each cell into an apical and a basolateral zone. The junctional complex is composed of four structures. The most apically located, the zonula occludens (tight junction), has two functions: it serves as a barrier to water and solutes and it restricts the molecular movement between apical and basolateral domains. The next three structures in order toward the basal portion of the cell are the zonula adherens, the macula adherens (desmosomes), and the macula communicantes (gap junctions). Gap junctions can selectively conduct small molecules (molecular weight <1000) between cells and may help to coordinate the response of an acinus to physiologic stimuli and to injury.

The pancreatic acinar cell is compartmentalized into distinct regions (Fig. 3). In the apical area, the cell contains zymogen granules and apical microvilli. Just deep to the apex, in the perinuclear region, there are condensing vacuoles and a Golgi apparatus. The basolateral zone contains the nucleus and a rich supply of rough endoplasmic reticulum (rER) (Fig. 4). This imparts the

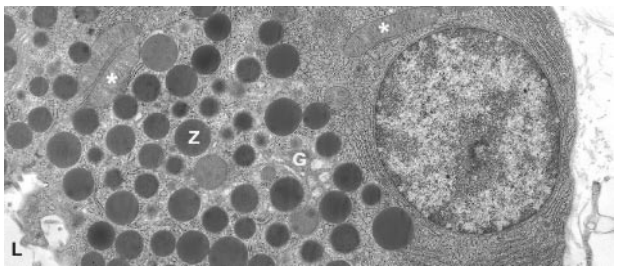


FIGURE 3 Pancreatic acinar cell. The apical surface of the cell lines the lumen (L). A small number of microvilli are evident at the lumen. Note the apical zone containing zymogen granules (Z) and a mitochondrion (asterisk). The Golgi apparatus (G) is in the perinuclear location. The nucleus and rough endoplasmic reticulum are in the basolateral zone. Some collagen fibrils are present in the connective tissue deep into the basal portion of the acinar cell (rat, $\times 5600$ original magnification).

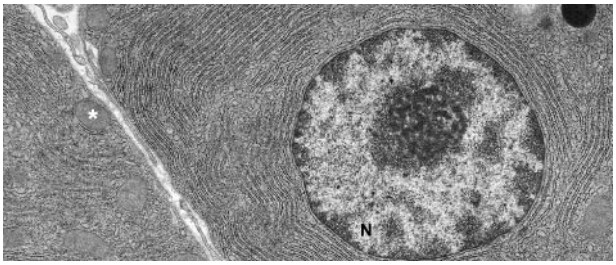


FIGURE 4 The basolateral region of the pancreatic acinar cell. Note the large nucleus (N) containing a small amount of heterochromatin, a great amount of euchromatin, and a prominent endoplasmatic reticulum and an elaborate nucleolus. Several mitochondria, one of which is labeled with an asterisk, are present (rat, ×5600 original magnification).

characteristic basophilia visible under light microscopy. Lysosomes, which are part of the late endosomal pathway, are scattered throughout the cytoplasm. Mitochondria are spatially organized as three distinct groups. One group is near the basolateral membrane, where active transport across the basement membrane is occurring. Another group of mitochondria is in the perinuclear region (Fig. 3). The third group resides in the apical region, separating zymogen granules from the basolateral area. The apically distributed mitochondria may have a role in generating calcium oscillations that are responsible for stimulated zymogen secretion.

The morphology of the acinar cell faithfully reveals its function in the synthesis, storage, and exocytosis of both active and inactive digestive enzymes. Pancreatic acinar cell enzymes are synthesized in the rER, traverse the late ER–Golgi compartment through vesicular transport, and enter the Golgi, where they are modified and sorted. In the Golgi, lysosomal enzymes, the ultimate destination of which are lysosomes, receive the mannose-6-phosphate (M6P) residue. Binding of the enzyme–M6P molecule to the M6P receptor (M6PR) results in targeting of the molecule to the late endosomal compartment. In late endosomes, a decline in pH initiates the release of lysosomal enzymes from the M6PR. The receptor is recycled to the Golgi or to the plasma membrane, and the lysosomal enzyme is transported to the lysosome. Additionally, M6PRs found on condensing vacuoles may retrieve lysosomal enzymes that have been sorted to the secretory pathway and return them to the Golgi complex.

In contrast to lysosomal proteins, secretory proteins are destined for exocytosis. On their exit from the Golgi, they enter membrane-bound, acidic condensing vacuoles, located near the nucleus (Fig. 5). Condensing vacuoles maintain the lowest pH compartment within the secretory pathway. As these vacuoles mature, their protein contents are condensed and the membrane is pinched off. The diameter of the resulting pancreatic zymogen granule is about two-thirds that of the condensing vacuole. Zymogen granules are stored in the apical one-third of the acinar, awaiting signals for exocytosis.

When the acinar cell is stimulated, contents of the zymogen granules are released into the acinar lumen. This process, called exocytosis, involves several steps:

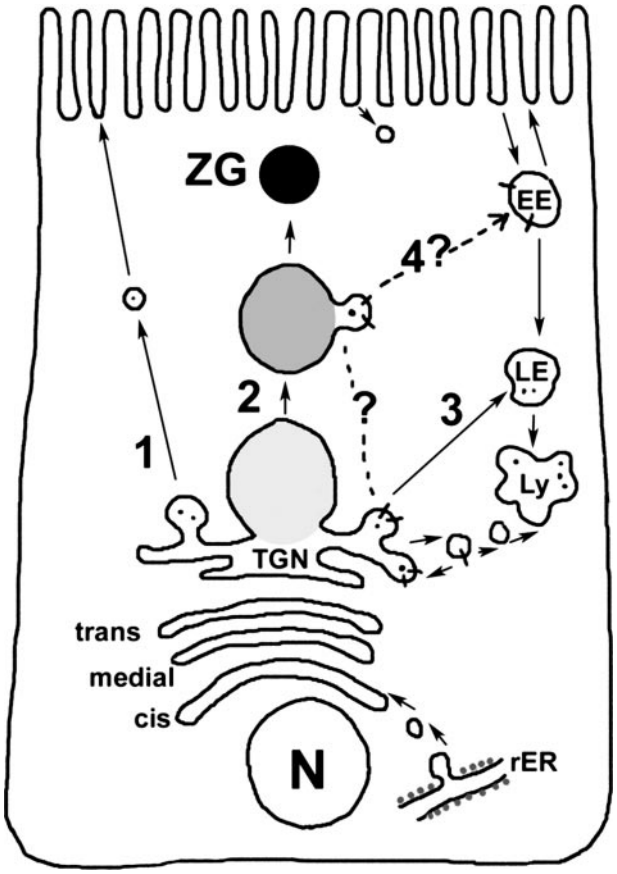


FIGURE 5 Vesicular trafficking pathways in the pancreatic acinar cell. Four main pathways are shown. (1) Constitutive secretion of proteins. A small fraction of newly synthesized proteins are secreted in an unregulated pathway in vesicles that bud off the trans-Golgi network (TGN). (2) Secretory pathway for zymogen granule release. The dark cores in vesicles budding from the TGN represent condensing vacuoles. As they mature, the membrane is pinched off and the content of the granule condenses to form zymogen granules (ZG). (3) Endocytic pathway for lysosomal enzyme delivery to lysosomes (short lines, mannose-6-phosphate receptors; EE, early endosome; LE, late endosome; Ly, lysosome). (4) A possible pathway for retrieval of lysosomal enzymes from the secretory pathway to the endocytic pathway. The open circles represent recycling membrane vesicles. Golgi domains are represented as trans, medial, and cis; rER, rough endoplasmic reticulum.

movement of granules to apical membrane, fusion of granules with the apical membrane, and the release of contents into the lumen. Actin is a major component of the apical cytoskeleton and plays a pivotal role in exocytosis. One study suggests that prior to fusion, zymogen granules become coated with actin, which might constrain their distribution to the apical part of the acinar cell and facilitate the movement of the zymogen granules across the subapical actin network toward the targeted site of fusion. A second actin network, located just below the apical membrane, may act as a barrier, blocking fusion of the zymogen granules with the apical plasma membrane. When the cell is stimulated, the actin barrier is reorganized and secretory granules can contact the apical plasma membrane. Fusion of the granules with the plasma membrane probably requires the interactions of the soluble *N*-ethylmaleimide-sensitive attachment protein receptor (SNARE) complex, a special group of proteins found on secretory granules and apical plasma membranes, along with other cytosolic factors.

Recent research, using atomic force microscopy, has identified a change in plasma membrane structure during exocytosis. In acinar cells, increases in the size of membrane depressions correlate with an increase in amylase secretion, and depression dynamics and amylase release are partly regulated by actin. This may indicate that there are predefined sites for fusion on the apical membrane. Following exocytosis, activation of the apical endocytic pathway leads to the subsequent retrieval of the secretory granule membrane via clathrin-coated retrieval vesicles. Therefore, an intact apical cytoskeleton may be a requirement for regulating normal exocytosis, maintaining organelle polarity, and membrane retrieval.

Calcium is the major secondary messenger for regulating protein secretion from pancreatic acinar cells. The major calcium storage compartments are in the ER and mitochondria. In acinar cells, polarized (apical to basal) waves of calcium are induced by the serial activation of two intracellular calcium channels that are principally responsible for calcium signaling: the apically located inositol 1,4,5-trisphosphate (IP_3) and the ryanodine receptor (RyR) located in the basolateral domain. Calcium waves begin in the apical domain of the cell, in the region of the type III IP_3 receptor (IP_3R), the most heavily expressed IP_3R isoform. Subcellular release of calcium coordinates the action of the IP_3R and RyR calcium channels. When acinar cells are stimulated with neurotransmitters, growth factors, or hormones, stored calcium is released from the ER by the action of IP_3. IP_3 binds to the IP_3R receptor, an ER calcium channel, to release calcium into the cytosol.

Centroacinar Cells

The location, morphology, and biochemistry of centroacinar cells reflect their role in electrolyte transport. Seen in cross-section, the centroacinar cells are centrally located within the acinus, and mark the beginning of ducts (Figs. 2 and 6). Centroacinar cells are small, pale-staining cells with microvilli on their apical surfaces. The pale-staining character reflects the sparse cytoplasm, small amount of rER, small size of the Golgi apparatus, and the lack of zymogen granules within the cell. Large, abundant, elongated mitochondria, suggesting an active role in ion secretion, are seen in the basal portion of these cells. The basal lamina staining characteristics of centroacinar cells differs from those of acinar cells. In centroacinar cells, the basal lamina contains more abundant, fixed anionic binding sites, suggestive of a role in ion transport. The luminal surface of the centroacinar cell is stained with silver, a marker of anions (in this case, bicarbonates). Unlike acinar cells, centroacinar cells contain carbonic anhydrase, which catalyzes the generation of HCO_3^- from CO_2. The cells also contain sialic-rich acid proteins, which work as cation filters, on their basolateral surfaces.

Pancreatic Duct Cells and Goblet Cells

Pancreatic duct cells exhibit morphologic heterogeneity along the length of the entire pancreatic duct; functional studies indicate that the types of solute-transport proteins within duct cells correlate to differences in duct cell location and morphology. For instance, duct cells most proximal to the acinus are squamous or low cuboidal in shape, having few cytoplasmic vesicles but a great many mitochondria (Fig. 6). These morphological features correspond to the function played by proximal duct cells in fluid and

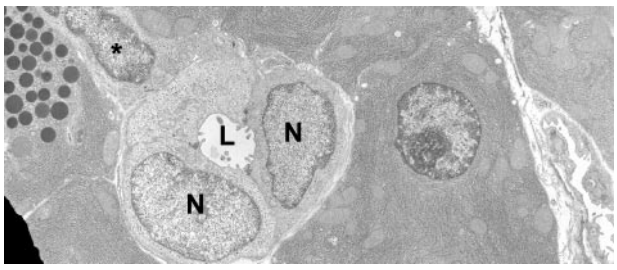


FIGURE 6 Pancreatic duct cells. In the center of the field, three low cuboidal duct cells surround a central lumen (L), forming a small intralobular duct. In two of the cells, nuclei (N) are visible. The duct is located close to the basolateral zones of five acinar cells. In the left upper field, a centroacinar cell with an elongated nucleus (asterisk) is visible. Note the pale-staining cytoplasm of the duct and centroacinar cells (rat, $\times 3300$ original magnification).

electrolyte transport. More distal duct cells are cuboidal to columnar in shape, and they contain more cytoplasmic vesicles and granules, which corresponds to the fact that more distal duct cells are capable not only of fluid and electrolyte transport, but also of protein secretion. Pancreatic duct cells express high levels of carbonic anhydrase and presumably play a role in HCO_3^- secretion. The duct and centroacinar cells secrete bicarbonate, which serves to neutralize gastric acid in the duodenum.

HCO_3^- secretion requires the activity of the cystic fibrosis transmembrane regulator (CFTR). The CFTR protein is concentrated on the apical membrane of proximal pancreatic duct cells. CFTR is a cAMP-regulated Cl^- conductance located in the luminal membrane and may be the major bicarbonate conductive pathway. In cystic fibrosis, a CFTR gene mutation results in the loss of CFTR expression at the plasmalemma and disrupts apical transport in pancreatic duct cells. There is a decrease in the secretion of bicarbonate and water, leading to duct obstruction, inflammation, and fibrosis.

Goblet cells are also found among the pancreatic duct cells. At the head of the main pancreatic duct (near the ampulla), goblet cells comprise about 25–30% of the duct epithelium. More proximally, the goblet cell population decreases to 5–10% of the duct cell population. Goblet cells secrete mucin (high-molecular-weight glycoproteins) production. When hydrated, mucins form mucus, which lubricates, hydrates, and mechanically protects surface epithelial cells. Mucins may also combat pancreatic infections by binding to pathogens and interacting with immune-competent cells.

Pancreatic Stellate Cells

Stellate cells have recently been isolated and characterized in the pancreas. They are similar to the stellate cells first identified in the liver. Hepatic stellate cells are known to play a major role in production of hepatic fibrosis and are the source of collagen and other extracellular matrix proteins in liver disease. Pancreatic stellate cells (PSCs) are located in the interlobular area, or the interacinar region, of the pancreas. The cells are triangular in shape and contain a prominent ER with cisternae containing flocculent materials (which reflect the presence of collagen and other extracellular matrix proteins), microfilaments, a large nucleus, few mitochondria, and cytoplasmic lipid droplets. In culture, PSCs have characteristics that are identical to those of hepatic stellate cells, with abundant vitamin A-containing lipid droplets in the cytoplasm. PSCs respond to cytokines (platelet-derived growth factor, transforming growth factor-β) by proliferating and synthesizing large amounts of extracellular matrix proteins (collagen, fibronectin, laminin). In addition, activated PSCs undergo a transformation into a myofibroblast-like phenotype, exhibiting positive staining for the cytoskeletal marker protein α-smooth muscle actin (α-SMA). PSCs are also activated by ethanol exposure and the ethanol effect is mediated by metabolism to acetaldehyde and generation of oxidant stress within the cells. The presence of PSCs in fibrotic areas is suggestive of their participation in pancreatic fibrosis development, a pathological feature of chronic pancreatitis.

In summary, the functional subunit of the exocrine pancreas is the acinus. It is composed of several major cell types (acinar, centroacinar, duct, goblet, and PSC), connective tissue, vasculature, and lymphatics. Although there are subspecializations among the cell types, in general, the acinus synthesizes and secretes active and inactive digestive enzymes that ultimately enter the descending duodenum, where they are activated and digest foodstuffs. The duct and centroacinar cells secrete bicarbonate, which serves to neutralize gastric acid in the duodenum.

See Also the Following Articles

Endocrine Pancreas • Exocytosis • Pancreas, Anatomy • Pancreas, Development • Pancreatic Bicarbonate Secretion • Pancreatic Enzyme Secretion (Physiology)

Further Reading

Akao, S., Bockman, D. E., Lechene de la Porte, P., and Sarles, H. (1986). Three dimensional pattern of ductuloacinar associations in normal and pathological human pancreas. *Gastroenterology* **90**, 661–668.

Apte, M. V., Haber, P. S., Darby, S. J., Rodgers, S. C., McCaugan, G. W., Korsten, M. A., Pirolak, R. C., and Wilson, J. S. (1999). Pancreatic stellate cells are activated by proinflammatory cytokines: Implications for pancreatic fibrogenesis. *Gut* **44**(4), 534–541.

Bachem, M. G., Schneider, E., Gross, H., Weidenbach, H., Schmid, R. M., Menke, A., Siech, M., Beger, H., Grunert, A., and Adler, G. (1998). Identification, culture and characterization of pancreatic stellate cells in rats and humans. *Gastroenterology* **115**(2), 421–432.

Hootman, S. R., and de Ondarza, J. (1995). Regulation of goblet cell degranulation in isolated pancreatic ducts. *Am. J. Phys.* **68**(1, Pt. 1), G24–G32.

Leite, M. F., Burgstahler, A. D., and Nathanson, M. H. (2002). Ca^{2+} waves require sequential activation of inositol triphosphate receptors and ryanodine receptors in pancreatic acini. *Gastroenterology* **122**(2), 415–427.

Marino C. R., Matovcik, L. M., Gorelick, F. S., and Cohn, J. A. (1991). Localization of the cystic fibrosis transmembrane conductance regulator in pancreas. *J. Clin. Invest.* **88**, 712–716.

Motto, P. M., Macchiarelli, G., Nottola, S. A., and Correr, S. (1997). Histology of the exocrine pancreas. *Microsc. Res. Tech.* **37**, 384–398.

Murakami, T., Hitomi, S., Ohtsuka, A., Taguchi, T., and Fujita, T. (1997). Pancreatic insulo-acinar portal systems in humans, rats, and some other mammals: Scanning electron microscopy of vascular casts. *Microsc. Res. Tech.* **37**, 478–488.

Park, M. K., Ashby, M. C., Erdemli, G., Petersen, O. H., and Tepikin, A. V. (2001). Perinuclear, perigranular and sub-plasmalemmal mitochondria have distinct functions in the regulation of cellular calcium transport. *EMBO J.* **20**(8), 1863–1874.

Raeder, M. G. (1992). The origin of and subcellular mechanisms causing pancreatic bicarbonate secretion. *Gastroenterology* **103**(5), 1674–1684.

Schneider, S. W., Sritharan, K. C., Geibel, J. P., Oberleithner, H., and Jena, B. P. (1997). Surface dynamics in living acinar cells imaged by atomic force microscopy: Identification of plasma membrane structures involved in exocytosis. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 316–321.

Valentijn, K. M., Gumkowski, F. D., and Jamieson, J. D. (1999). The subapical actin cytoskeleton regulates secretion and membrane retrieval in pancreatic acinar cells. *J. Cell Sci.* **112**(Pt. 1), 81–96.

Williams, J. A. (2001). Intracellular signaling mechanisms activated by cholecystokinin-regulating synthesis and secretion of digestive enzymes in pancreatic acinar cells. *Annu. Rev. Physiol.* **63**, 77–97.

Exocrine Pancreatic Insufficiency

EUGENE P. DIMAGNO
Mayo Clinic

chronic pancreatitis Continual inflammatory disease of the pancreas.

enteric-coated Oral medication that is coated or encapsulated to avoid acid damage in the stomach but dissolve in the intestine.

malabsorption The failure to digest and absorb dietary nutrients.

proton pump inhibitors Drugs that block gastric acid secretion by inhibiting the ATPase enzyme that transports protons.

steatorrhea Increased fat in stool due to malabsorption.

Exocrine pancreatic insufficiency is the reduced secretion of exocrine pancreatic digestive enzymes from the pancreas into the duodenum. It arises from the destruction of pancreatic acinar cells that synthesize and secrete enzymes (pancreatitis, cystic fibrosis), obstruction of the ducts leading from the pancreas to the duodenum (pancreatic cancer), or removal of part of or the entire pancreas (pancreatic surgery). Mild to moderate exocrine insufficiency may be asymptomatic, but severe exocrine insufficiency causes malabsorption and malnutrition, which impacts on morbidity and mortality. For example, treatment of exocrine insufficiency in cystic fibrosis slows the decline of pulmonary function and prevents growth retardation.

PATHOPHYSIOLOGY AND NATURAL HISTORY

Malabsorption of fat, protein, or carbohydrate does not occur until maximal secretion of enzymes is reduced by $>90\%$. Fat malabsorption is the greatest problem, because lipase secretion declines more rapidly than that of other enzymes, causing steatorrhea to occur before other forms of malabsorption. It is also more difficult to treat than carbohydrate or protein malabsorption. The time course to onset of steatorrhea varies among diseases. Exocrine insufficiency is usually present at the time of diagnosis in cystic fibrosis and pancreatic cancer and occurs in 75–80% of patients. However, in alcoholic chronic pancreatitis, early-onset and late-onset idiopathic chronic pancreatitis, it takes a mean of 13, 26, and 17 years, respectively, to develop severe exocrine insufficiency.

Carbohydrate malabsorption and protein malabsorption are easier to treat than steatorrhea because (1) they occur later and are less severe than steatorrhea; (2) salivary amylase, gastric proteases, and intestinal brush border enzymes that digest carbohydrate and peptides partially compensate for the lack of pancreatic amylase and proteases; and (3) amylase and proteases are

 Copyright 2004, Elsevier (USA). All rights reserved.

Murakami, T., Hitomi, S., Ohtsuka, A., Taguchi, T., and Fujita, T. (1997). Pancreatic insulo-acinar portal systems in humans, rats, and some other mammals: Scanning electron microscopy of vascular casts. *Microsc. Res. Tech.* **37**, 478–488.

Park, M. K., Ashby, M. C., Erdemli, G., Petersen, O. H., and Tepikin, A. V. (2001). Perinuclear, perigranular and sub-plasmalemmal mitochondria have distinct functions in the regulation of cellular calcium transport. *EMBO J.* **20**(8), 1863–1874.

Raeder, M. G. (1992). The origin of and subcellular mechanisms causing pancreatic bicarbonate secretion. *Gastroenterology* **103**(5), 1674–1684.

Schneider, S. W., Sritharan, K. C., Geibel, J. P., Oberleithner, H., and Jena, B. P. (1997). Surface dynamics in living acinar cells imaged by atomic force microscopy: Identification of plasma membrane structures involved in exocytosis. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 316–321.

Valentijn, K. M., Gumkowski, F. D., and Jamieson, J. D. (1999). The subapical actin cytoskeleton regulates secretion and membrane retrieval in pancreatic acinar cells. *J. Cell Sci.* **112**(Pt. 1), 81–96.

Williams, J. A. (2001). Intracellular signaling mechanisms activated by cholecystokinin-regulating synthesis and secretion of digestive enzymes in pancreatic acinar cells. *Annu. Rev. Physiol.* **63**, 77–97.

Exocrine Pancreatic Insufficiency

EUGENE P. DIMAGNO
Mayo Clinic

chronic pancreatitis Continual inflammatory disease of the pancreas.

enteric-coated Oral medication that is coated or encapsulated to avoid acid damage in the stomach but dissolve in the intestine.

malabsorption The failure to digest and absorb dietary nutrients.

proton pump inhibitors Drugs that block gastric acid secretion by inhibiting the ATPase enzyme that transports protons.

steatorrhea Increased fat in stool due to malabsorption.

Exocrine pancreatic insufficiency is the reduced secretion of exocrine pancreatic digestive enzymes from the pancreas into the duodenum. It arises from the destruction of pancreatic acinar cells that synthesize and secrete enzymes (pancreatitis, cystic fibrosis), obstruction of the ducts leading from the pancreas to the duodenum (pancreatic cancer), or removal of part of or the entire pancreas (pancreatic surgery). Mild to moderate exocrine insufficiency may be asymptomatic, but severe exocrine insufficiency causes malabsorption and malnutrition, which impacts on morbidity and mortality. For example, treatment of exocrine insufficiency in cystic fibrosis slows the decline of pulmonary function and prevents growth retardation.

PATHOPHYSIOLOGY AND NATURAL HISTORY

Malabsorption of fat, protein, or carbohydrate does not occur until maximal secretion of enzymes is reduced by $>90\%$. Fat malabsorption is the greatest problem, because lipase secretion declines more rapidly than that of other enzymes, causing steatorrhea to occur before other forms of malabsorption. It is also more difficult to treat than carbohydrate or protein malabsorption. The time course to onset of steatorrhea varies among diseases. Exocrine insufficiency is usually present at the time of diagnosis in cystic fibrosis and pancreatic cancer and occurs in 75–80% of patients. However, in alcoholic chronic pancreatitis, early-onset and late-onset idiopathic chronic pancreatitis, it takes a mean of 13, 26, and 17 years, respectively, to develop severe exocrine insufficiency.

Carbohydrate malabsorption and protein malabsorption are easier to treat than steatorrhea because (1) they occur later and are less severe than steatorrhea; (2) salivary amylase, gastric proteases, and intestinal brush border enzymes that digest carbohydrate and peptides partially compensate for the lack of pancreatic amylase and proteases; and (3) amylase and proteases are

 Copyright 2004, Elsevier (USA). All rights reserved.

more resistant to acid denaturation than lipase. In contrast, fat digestion depends mainly on the pancreatic lipase.

TREATMENT OF EXOCRINE PANCREATIC INSUFFICIENCY

The current clinical dictum is that oral pancreatic enzyme replacement should be started when malabsorption occurs and the principle is simple—replace inadequate secretion of pancreatic enzymes by giving enzymes by mouth. Unfortunately, although carbohydrate and protein malabsorption is mostly reversed by pancreatic enzymes, fat malabsorption rarely is, because inadequate amounts of pancreatic enzymes may be ingested and because improper timing and schedule of taking enzymes and lipase may be inactivated by intragastric and intraduodenal acidity (irreversible inactivation at pH ≤ 4). Up to 90% of the ingested enzymes are destroyed by acid in the stomach and by acid and proteolytic digestion by chymotrypsin and trypsin in the small intestine.

Dosage

To abolish malabsorption, ~10% of normal enzyme output needs to be delivered to the duodenum. Because of misinterpretation of the experimental data, most pharmaceutical companies recommend a dose of 30,000 USP (United States Pharmacopoeia) lipase units. However, a unit of lipolytic activity in the experimental studies of the authors is equivalent to 3 USP units. Therefore, at least 90,000 USP units of lipase is needed. Although this dose rarely abolishes steatorrhea, it usually decreases steatorrhea by 50%. Inadequate dosing may occur because the amounts of enzymes in preparations vary among lots.

Pancreatic enzymes should be given with meals (e.g., 25% of the dose after first few bites, 50% of the dose during the meal, and 25% of the dose immediately at the end of the meal). Although taking enzymes every hour for 18 h (same total dose as prandial administration) is as effective as taking enzymes with meals, the prandial schedule is more practical. The hourly schedule may be more effective when postprandial gastric pH is >4.0 for more than 1 h, which occurs in some chronic pancreatitis patients.

Choice of Pancreatic Enzymes: Non-Enteric-Coated or Enteric-Coated?

Enteric-coated preparations should traverse the stomach and duodenum without lipase inactivation. However, when adequate dosages of lipase are used, enteric-coated and non-enteric-coated enzymes reduce steatorrhea to a similar extent. Better delivery of lipase into the small intestine occurs with enteric-coated microspheres (ECMS), but reduced steatorrhea due to smaller pellets is unconfirmed. In addition, ECMS are expensive and may be associated with fibrosing colopathy because of release of the enzymes and the polymeric coating in the distal small intestine. ECMS may be appropriate if patients do not respond to conventional enzymes with or without acid suppression therapy, are unable to tolerate conventional preparations, are unable to swallow tablets, have undergone gastric surgery or bypass surgery that might impair mixing, or have documented persistent gastric and duodenal pH of <4.0 despite acid suppression therapy. ECMS should not be used if there is hypochlorhydria (e.g., total gastrectomy or long-term acid suppression therapy for other gastrointestinal diseases).

Adjunctive Acid Suppression Therapy

Acid inactivation of lipase can be prevented by neutralization or suppression of gastric acid secretion. However, magnesium-containing antacids and calcium carbonate are not of benefit. Aluminum hydroxide is slightly beneficial. Low doses (2.5 g) of sodium bicarbonate are ineffective and large doses (16.6 g/day given with meals) are minimally effective. In contrast, adjunctive treatments with histamine-2 (H2) receptor antagonists or proton pump inhibitors (PPIs) are very effective in reducing steatorrhea because they inhibit gastric acid secretion and also enhance lipase and bile acid concentrations by reducing duodenal volume flow. Effective preparations in at least 40% of patients include the H2 receptor antagonists cimetidine, ranitidine, and famotidine and the PPIs omeprazole and lansoprazole. Although adjunctive acid suppression therapy is beneficial, routine use may not be justified because of cost, drug interactions, and long-term safety. It should be used only when abolition of steatorrhea is not achieved with conventional enzymes.

DIET MODIFICATION

Currently it is recommended that low-fat diets (50−75 g/day) should be used if steatorrhea is symptomatic. This recommendation may undergo change as recent studies have demonstrated in pancreatic-insufficient dogs that fat absorption was greater when pancreatic enzymes were given with high-fat diets. However, the effect of high-fat diets with pancreatic enzymes on fat absorption has not been tested in patients with exocrine insufficiency.

SUMMARY

Treatment of exocrine pancreatic insufficiency is necessary if there are symptoms of malabsorption secondary to severe exocrine pancreatic insufficiency, steatorrhea, diarrhea, or weight loss. Initially, patients are treated with 90,000 USP units of non-enteric-coated enzymes with meals. If patients continue to have steatorrhea and are symptomatic, dietary fat should be reduced to 50–75 g/day. If steatorrhea does not decrease, adjuvant therapy with H2 receptor antagonists or a PPI should be added, which will abolish steatorrhea in approximately half of the patients. Adjuvant acid suppression therapy is preferred over enteric-coated enzymes because it increases intraduodenal concentrations of lipase and decreases bile acid malabsorption.

Reducing steatorrhea by 50% (but not abolishing it) improves bowel symptoms, nutritional status, the sense of well-being, and weight gain. However, the optimal goal of pancreatic enzyme replacement is the complete abolition of steatorrhea because this improves growth in children and because adults with exocrine insufficiency have a significantly shortened life span, partly due to an increase in atherosclerotic cardiovascular disease. The cause of cardiovascular disease is unknown, but malabsorption, malnutrition, and metabolic derangement might be involved.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Cystic Fibrosis • Exocrine Pancreas • Malabsorption • Pancreatic Cancer • Pancreatic Digestive Enzymes • Pancreatic Enzyme Secretion (Physiology)

Further Reading

DiMagno, E. P., Go, V. L., and Summerskill, W. H. (1973). Relations between pancreatic enzyme outputs and malabsorption in severe pancreatic insufficiency. *N. Engl. J. Med.* **288**(16), 813–815.

DiMagno, E. P., Layer, P., and Clain, J. E. (1993). Chronic pancreatitis. *In* "The Pancreas: Biology, Pathobiology, and Diseases" (V. L. W. Go, E. P. DiMagno, J. D. Gardner, E. Lebenthal, H. A. Reber, and G. A. Scheele, eds.), pp. 665–706. Raven Press, New York.

Dutta, S. K., Rubin, J., and Harvey, J. (1983). Comparative evaluation of the therapeutic efficacy of a pH-sensitive enteric coated pancreatic enzyme preparation with conventional pancreatic enzyme therapy in the treatment of exocrine pancreatic insufficiency. *Gastroenterology* **84**(3), 476–482.

Gaskin, K., Gurwitz, D., Durie, P., Corey, M., Levison, H., and Forstner, G. (1982). Improved respiratory prognosis in cystic fibrosis patients with normal fat absorption. *J. Pediatr.* **100**, 857–862.

Graham, D. Y. (1982). Pancreatic enzyme replacement: The effect of antacids or cimetidine. *Digest. Dis. Sci.* **27**(6), 485–490.

Hendeles, L., Dorf, A., Steckenko, A., and Weinberger, M. (1990). Treatment failure after substitution of generic pancrealipase capsules: Correlation with in vitro lipase activity. *J. Am. Med. Assoc.* **263**, 2459–2461.

Layer, P., Go, V. L., and DiMagno, E. P. (1986). Fate of pancreatic enzymes during small intestinal aboral transit in humans. *Am. J. Physiol.* **251**(4, Pt. 1), G475–G480.

Regan, P. T., Malagelada, J. R., DiMagno, E. P., Glanzman, S. L., and Go, V. L. (1977). Comparative effects of antacids, cimetidine and enteric coating on the therapeutic response to oral enzymes in severe pancreatic insufficiency. *N. Engl. J. Med.* **297**(16), 854–858.

Suzuki, A., Mizumoto, A., Sarr, M. G., and DiMagno, E. P. (1997). Bacterial lipase and high-fat diets in canine exocrine pancreatic insufficiency: A new therapy of steatorrhea? *Gastroenterology* **112**(6), 2048–2055.

Exocytosis

XUEQUN CHEN AND JOHN A. WILLIAMS
University of Michigan, Ann Arbor

exocytosis The process by which membrane-enclosed intracellular vesicles fuse with the plasma membrane and then open and release their contents to the extracellular space.
membrane fusion The process by which membrane lipid bilayers merge.
NSF A cytosolic protein required for vesicle fusion that is sensitive to *N*-ethylmaleimide.
Rab A family of small monomeric GTP-binding proteins with an overall structure similar to that of Ras and which regulate vesicle trafficking and fusion.
secretory granule Membrane-bound vesicle containing condensed secretory material.
SNAREs Membrane proteins that serve as receptors for soluble *N*-ethylmaleimide-sensitive factor attachment proteins and that participate in vesicular docking and fusion.

Exocytosis is the process by which membrane-enclosed intracellular vesicles fuse with the plasma membrane and then open and release their contents to the extracellular space. Many cellular processes involve exocytosis: for example, in the gastrointestinal tract, the release of neurotransmitters in synaptic vesicles from presynaptic enteric neurons, the secretion of digestive enzymes in zymogen granules from gastric chief cells and pancreatic and parotid acinar cells, the release of mediators from mast and enterochromaffin-like cells, and the release of peptide hormones from endocrine cells such as the pancreatic islets. The fusion of vesicles with the plasma membrane also mediates the translocation of proteins such as Na^+,K^+-ATPase and glucose transporters (Glut 4) to the plasma membrane. Exocytosis, especially when coupled with endocytosis, also contributes to the turnover of plasma membrane.

Some exocytotic processes occur continually in the cell and are termed constitutive, whereas other exocytotic processes are regulated. The two pathways diverge in the *trans*-Golgi network. Many soluble proteins are continually secreted from the cell by the constitutive secretory pathway, such as plasma proteins secreted by the liver and basement membrane components secreted by epithelial cells. All cells require this constitutive secretory pathway, which involves the movement of small vesicles to the plasma membrane. This pathway also supplies the plasma membrane with newly synthesized lipids and proteins. Specialized secretory cells, however, have a second pathway—the regulated exocytosis pathway, in which selected proteins are stored in secretory vesicles (frequently called secretory granules or dense core vesicles because they have dense cores when viewed in the electron microscope) for later release triggered by extracellular signals.

MORPHOLOGY OF EXOCYTOSIS

Secretory vesicles bud off from the Golgi network, undergo maturation, and translocate toward their destination plasma membrane. The sizes of mature secretory vesicles vary in different cell types, with their diameters ranging from ~ 100 nm in synaptic vesicles to ~ 1 μm in zymogen granules. The cytoskeleton, especially microtubules, contributes to the transport of newly synthesized secretory vesicles from the Golgi to the cell surface and also may direct them toward a specific plasma membrane domain.

As shown in Fig. 1, several stages have been recognized in exocytosis. After translocation from the Golgi network to the periphery of cell surface (step 1), the secretory vesicles are tethered in the vicinity of the cell surface (step 2). This is followed by docking, where the granules are in close apposition to the membrane (step 3), which is followed by the fusion process itself

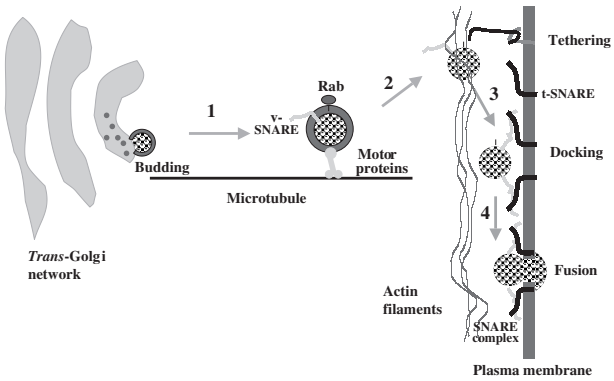


FIGURE 1 Tentative model of the steps in exocytosis of secretory granules including trafficking, docking, and fusion.

 Copyright 2004, Elsevier (USA). All rights reserved.

(step 4). Docking by itself is not a sufficient condition for fusion and some biochemical events, described as priming reactions, are likely to occur, rendering the granules competent for fusion. On fusion, the secretory vesicle membranes fuse with the plasma membrane, fusion pores open, and the contents of the vesicles are released through the fusion pores. Secretory vesicles fused with the plasma membrane maintain their Ω-shaped profile for a short period of time after releasing their soluble contents and this Ω-shaped profile becomes the morphological hallmark of membrane fusion.

In many types of secreting cells, such as chromaffin cells and pancreatic acinar cells, actin filaments form a cortical actin network under the membrane. This actin web serves as a barrier and regulates the access of granules to the docking sites. Recent evidence indicates that the cortical actin network may play multiple roles in regulating exocytosis.

MOLECULAR MECHANISM OF EXOCYTOSIS

Exocytosis is believed to share basic mechanisms with other vesicular trafficking events in the secretory pathway that are common in organisms ranging from yeast to human. To ensure the specificity of membrane fusion, transport vesicles must be highly selective in recognizing the correct target membrane with which to fuse. In the case of exocytosis, the secretory vesicles must recognize and fuse specifically with the target plasma membranes. Current models of membrane targeting and fusion have been dominated by consideration of two types of proteins, small G-proteins of the Rab family and SNARE [soluble *N*-ethylmaleimide (NEM)-sensitive factor attachment protein receptor] proteins. Early studies had shown that vesicular fusion events required a cytosolic protein sensitive to NEM, termed NSF for NEM-sensitive factor. Subsequent studies identified membrane proteins that interacted with NSF and were termed SNARE proteins. In the SNARE hypothesis, transport vesicles in the secretory pathway possess a set of vesicle v-SNAREs that interact with the complementary target membrane t-SNAREs and a stable v-SNARE/t-SNARE complex is assembled through coiled-coil interactions of α-helices between proteins inserted in opposing membranes. This pairing of a v-SNARE and a corresponding t-SNARE is thought to have a central role in vesicle docking and the fusion of the vesicle with the target membrane.

SNAREs have been best characterized in nerve cells, where they mediate the docking and fusion of synaptic vesicles at the nerve terminal plasma membrane. The principal v-SNARE in synaptic vesicles is VAMP (vesicle-associated membrane protein), also called synaptobrevin. The principal t-SNAREs in the plasma membrane of axon terminals are syntaxin and SNAP25 (soluble NSF attachment protein 25). Four long α-helices (two from SNAP25 and one each from VAMP and syntaxin) form a very stable SNARE complex between this pair of v- and t-SNAREs. The SNARE complexes at neuron terminals are the targets of powerful neurotoxins that are secreted by the bacteria that cause tetanus and botulism. These toxins are highly specific proteases that cleave SNARE proteins in the nerve terminals and thereby block synaptic transmissions.

In addition to SNAREs, several other cytosolic proteins are required for vesicle fusion. NSF is an ATPase that cycles between the membrane and the cytosol, and other proteins called α-, β-, and γ-SNAPs are required for NSF to bind to the vesicle membrane. SNAP in cooperation with NSF catalyzes the disassembly of *cis*-SNARE complexes (residing on the same membrane) using the energy of ATP. The requirement of NSF-mediated reactivation of SNAREs allows the cells to control when and where the membrane fusion events take place. In the case of synaptic vesicle exocytosis, which is regulated by calcium, another protein, synaptotagmin, is believed to serve as a calcium sensor. There are a large number of synaptotagmin analogues, some of which are calcium insensitive. Whether a synaptotagmin is required for all exocytotic events remains unclear. Another protein variously known as nSec1 or Munc 18 may serve as an inhibitory protein, preventing fusion until appropriate regulatory signals occur.

Once a transport vesicle has recognized its target membrane and docked, it unloads its cargo by membrane fusion. In regulated exocytosis, fusion is delayed until it is triggered by a specific extracellular signal. Docking and fusion are two distinct and separable processes. Docking requires only that the two membranes come close enough for proteins to interact. Fusion requires a much closer approach, bringing lipid bilayers to within 1.5 nm of each other so that they can join. For this close approach, water must be displaced from the hydrophilic surface of the membrane—a process that is energetically highly unfavorable. In one model, the formation of the SNARE complex may work like a winch, using energy that is freed when the interacting helices wrap around each other to pull the two membranes together, while simultaneously squeezing out water molecules from the interface.

Rab proteins, a family of monomeric GTP-binding proteins, participate in the control of the specificity of vesicular transport. All Rab proteins contain

approximately 200 amino acids and have an overall structure similar to that of Ras. Like the SNAREs, each Rab protein has a characteristic distribution on cell membranes and every organelle has at least one Rab protein on its cytosolic surface. For example, Rab5 is associated with the early endosome, Rab3A is associated with synaptic vesicles, and Rab3D is associated with secretory granules in exocrine glands. Rab proteins act as molecular switches that cycle between the inactive GDP-bound and active GTP-bound forms. The conversion to the GTP-bound form is stimulated by a Rab guanine nucleotide exchange factor and GTP hydrolysis is catalyzed by a Rab GTPase-activating protein. In their active form, Rab proteins interact with their effector proteins, which facilitate and regulate the rate of vesicle docking. Although the Rab proteins and their effectors use a range of different molecular mechanisms to influence vesicular transport at different sites in the secretory pathway, they have a common function. They help concentrate and tether vesicles near their target site and trigger the release of SNARE control proteins. In this way, Rab proteins speed the process by which appropriate SNARE proteins in two membrane find each other.

See Also the Following Articles

Chief Cells • Enterochromaffin-like (ECL) Cells • Mast Cells

Further Reading

Martin, T. F. (2002). Prime movers of synaptic vesicle exocytosis. *Neuron* **34**, 9–12.

Pelham, H. R. (2001). SNAREs and the specificity of membrane fusion. *Trends Cell Biol.* **11**, 99–101.

Pfeffer, S. R. (2001). Rab GTPases: Specifying and deciphering organelle identity and function. *Trends Cell Biol.* **11**, 487–491.

Rizo, J., and Sudhof, T. C. (2002). Snares and Munc18 in synaptic vesicle fusion. *Nat. Rev. Neurosci.* **3**, 641–653.

Segev, N. (2001). Ypt/Rab GTPases: Regulators of protein trafficking. *Science STKE*, re11.

F

Familial Adenomatous Polyposis (FAP)

MARCIA CRUZ-CORREA AND FRANCIS M. GIARDIELLO
The Johns Hopkins University School of Medicine

autosomal dominant Mendelian gene that always manifests phenotypically; all affected individuals have at least one affected parent, and the phenotype affects males and females equally.

germ-line mutations Genetic alterations that occur in the cells that are of direct descent, from the zygote to gamete; are transmitted to progeny.

Osteomas Benign, slow-growing tumor composed of well-differentiated, densely sclerotic, compact bone, usually arising in the membrane bones, particularly the skull and fascial bones.

Penetrance Frequency of expression of a genotype; a trait has reduced penetrance if it is expressed less than 100%.

Familial adenomatous polyposis (FAP) is an autosomal dominantly inherited syndrome that accounts for approximately 1% of the total colorectal cancer cases in the United States. FAP arises from germ-line mutations of the adenomatous polyposis coli (*APC*) gene located on chromosome 5q21. In 22–46% of cases, this disorder arises as spontaneous mutations, without an associated family history. It is estimated that FAP affects 1 in 10,000 individuals and is nearly 100% penetrant. This syndrome occurs worldwide and affects both sexes equally. The disease is characterized by the formation at early age of adenomatous polyps, primarily in the large intestine, and is associated with a virtual 100% risk of colorectal cancer. The mean age of adenoma development is 15 years. The average age of colorectal carcinoma in FAP ranges from 34.5 to 43 years. Colorectal cancer, primarily in the left colon, is inevitable in FAP patients if colectomy is not performed.

CLINICAL CHARACTERISTICS

FAP is an autosomal dominant disease clinically characterized by the occurrence of hundreds to thousands of adenomas throughout the colorectum at an early age (Fig. 1). The size of polyps in FAP is usually <1 cm; they may be pedunculated or sessile, with tubular, tubulovillous, or villous histology. Inevitably, colorectal cancer occurs usually by the fifth decade of life if colectomy is not performed. In addition, patients can acquire a variety of benign and malignant extracolonic manifestations (Table I).

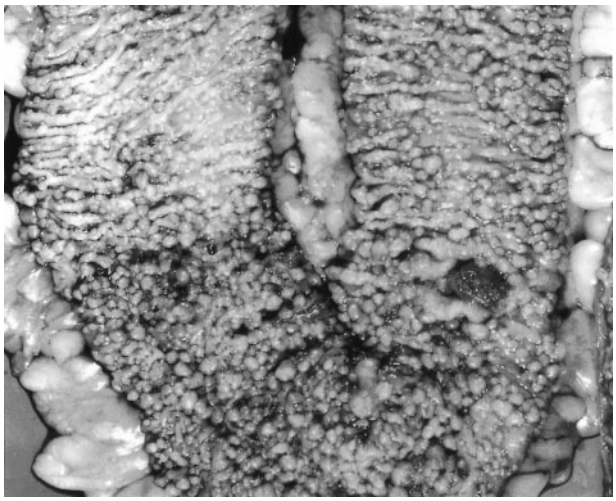


FIGURE 1 Photograph of a colectomy specimen in a patient with FAP, demonstrating hundreds of adenomatous polyps.

Benign Extracolonic Manifestations

Extracolonic Polyps

Adenomas of the small intestinal, primarily in the duodenum clustered around the papilla of vater, are found in 24–90% of FAP patients. Adenomas elsewhere in the small intestine occur occasionally. Two types of polyps are noted in the stomach: adenomas and fundic gland retention polyps. Adenomas, which can diffusely affect the stomach, are noted in up to 50% of FAP patients. Fundic gland retention polyps occur in 50% of FAP patients, usually localized in the fundus and body of the stomach while sparing the antrum. Fundic gland retention polyps are generally considered nonneoplastic, though dysplasia has been noted on histologic review in some individuals.

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Extracolonic Features in FAP

Cancers (lifetime risk)	Other lesions
Duodenal (5–11%)	CHRPE[b]
Pancreatic (2%)	Nasopharyngeal angiofibroma
Thyroid (2%)	Osteomas
Brain (medulloblastoma)[a] (<1%)	Radio-opaque jaw lesions
Hepatoblastoma (0.7% of children <5 years old)	Supernumerary teeth
	Lipomas, fibromas, epidermoid cysts
	Desmoid tumors
	Gastric adenomas/fundic gland polyps
	Duodenal, jejunal, and ileal adenomas

[a] Crail's syndrome is characterized by medulloblastoma associated with adenomatous polyposis.
[b] CHRPE, Congenital hypertrophy of the retinal pigment epithelium.

Desmoid Tumors

Benign desmoid tumors arise from the mesenchymal primordial germ cell layer and are usually slow growing. As fibromatous lesions occurring in extremities, the abdominal wall, and the mesentery of about 10% of FAP patients, desmoid tumors are the cause of death in about 1% of patients. About 80% of desmoid tumors are localized in the small bowel mesentery, whereas 20–30% develop within the abdominal wall or on the extremities. Desmoid tumors are flat or lobulated masses that often have intralesional hemorrhages and may undergo cystic degeneration. Desmoids, particularly of the mesentery, can grow locally, causing obstructive complications. Desmoid formation is associated with surgical trauma, familial aggregation, and specific *APC* gene mutations (see later).

Cutaneous Lesions

Epidermoid cysts, subcutaneous lesions typically located on the extremities, scalp, and face, develop in the teenage years. Fibromas, lipomas, and sebaceous cysts can also occur. Sebaceous cysts in association with FAP was previously called "Oldfield syndrome."

Osteomas are benign lesions that develop in the skull, long bones, and characteristically in the mandible at the angle of the jaw. These lesions, readily appreciated by physical examination, are usually asymptomatic but occasionally grow, causing complaints of local pain or discomfort. Occult radio-opaque jaw lesions (ORJLs) are osteosclerotic, asymptomatic lesions seen on panoramic jaw radiographs. These markers predict the development of adenomatous polyposis. Dental abnormalities are noted in 17% of FAP patients and include supranumerary and impacted teeth. Nasopharyngeal angiofibroma is a highly vascular, locally invasive tumor occurring almost exclusively in the nares or nasopharynx of male adolescents.

Malignant Manifestations

Adenocarcinoma of the duodenum and periampullary region is second only to the colorectum as a site of malignancy in patients with FAP. A high percentage of gastric cancer has also been reported in some polyposis registries worldwide. However, an analysis of American patients with FAP reveals a strikingly high relative risk of duodenal and ampullary adenocarcinoma but no increased risk of gastric or nonduodenal small intestinal cancer compared with the general population.

Hepatoblastoma is a rare, malignant, embryonal liver cancer occurring in the first 5 years of life in about 0.7% of the offspring of parents with FAP. This rapidly progressive tumor is potentially curable by removal, but carries a grave prognosis when malignancy has spread beyond surgical resection.

FAP has been associated with extracolonic malignancies of the thyroid gland, adrenal gland, biliary tree, pancreas, and brain. The relative risks of thyroid and pancreatic cancer are almost eight and five times that of the general population, respectively. Papillary thyroid carcinomas are associated with activation of the *RET* proto-oncogene and have a good prognosis.

Pigmented ocular fundus lesions (POFLs), also called congenital hypertrophy of the retinal pigment epithelium (CHRPE), are discrete, round to oval, darkly pigmented areas from 0.1 to 1.0 optic-disc diametersin size; they can be detected on the retina by indirect ophthalmoscopy. CHRPE consists of multiple hyperplastic layers of retinal pigment epithelium with hypertrophied cells filled with large spherical melanosomes, often in clusters. Although CHRPE causes no symptoms, it can

be a useful marker in identifying asymptomatic carriers in families with FAP. Also, the presence or absence of CHRPE lesions in an FAP patient correlates with the specific location of *APC* gene mutation.

FAP is also known by other terminology; the diagnosis of familial polyposis includes patients without extracolonic manifestations, and in Gardner syndrome, patients have various extracolonic manifestations (i.e., osteomas, epidermoid cysts, skin fibromas, desmoids, and jaw cysts). Other variants of FAP include Crail's syndrome, defined as typical FAP together with central nervous system malignancies (medulloblastoma), and attenuated familial adenomatous polyposis coli (AFAP). The clinical characteristics of AFAP include oligopolyposis (fewer than 100 colorectal adenomas at presentation) and a delayed onset of colorectal cancer, occurring on average 12 years later than in classic FAP.

GENETIC DEFECT

FAP is caused by germ-line mutations of the *APC* gene on chromosome 5q21. *APC* is a large gene, encompassing 15 exons, with an open reading frame of 8538 base pairs. The *APC* gene is a tumor suppressor gene encoding a 2843-amino-acid protein with a putative role in cell adhesion, signal transduction, and transcriptional activation. *APC* gene mutations can be found in 80–90% of families with FAP. More than 300 different disease-causing mutations of the *APC* gene have been identified. Most mutations are frameshifts that result from the insertion or deletion of one to eight base pairs. A higher frequency of mutations has been found in the 5′ region of exon 15 between codons 1000 and 1600; this is designated the mutation cluster region, comprising 20% of the entire gene. About 30% of *APC* mutations are at codon 1061, 1309, or 1465.

Several phenotypic characteristics in FAP patients correlate to the site of the *APC* gene mutation. Mutations between codons 169 and 1393 result in classic FAP, whereas more 5′ and 3′ mutations cause attenuated FAP. Alterations between codons 1250 and 1464 result in profuse colorectal polyposis (thousands rather than hundreds of colorectal adenomas), and retinal lesions occur only with mutations between codons 463 and 1444. *APC* mutations between codons 1445 and 1578 have been associated with severe desmoids, osteomas, epidermoid cysts, and upper gastrointestinal polyps. Nonetheless, considerable phenotypic variability may occur even among individuals and families with identical genotypic mutations.

The recently identified mutation in the *APC* gene, APC I1307K, was discovered as a cause for an unidentified proportion of familial colorectal cancer in individuals of Ashkenazi Jewish descent. Individuals with this mutation do not display the extracolonic manifestations of FAP but have an increased risk of development of colorectal cancer, with an estimated odds ratio of 1.4 : 1.9.

DIAGNOSIS

The diagnosis of FAP can be made clinically by the identification of hundreds to thousands of colorectal adenomatous polyps on colonoscopic examination. Individuals with attenuated FAP exhibit oligopolyposis (fewer than 100 colorectal adenomas). Histology is the key to differentiating FAP from the other known polyposis syndromes, such as lymphoid hyperplasia or hyperplastic polyposis, which may mimic FAP endoscopically. Also, the diagnosis of FAP can be confirmed by genetic testing. Additionally, in at-risk individuals, the presence of more than three pigmented ocular fundus lesions on ophthalmologic examination confirms the diagnosis of FAP.

MANAGEMENT

Screening

First-degree relatives of patients with FAP should undergo screening for FAP between ages 10 and 12 years old. The screening test of choice is genetic testing for the *APC* gene mutation. Indications for *APC* gene testing are found in Table II. The *APC* gene mutation responsible for the disorder in the pedigree can be identified in 80–90% of families with FAP. Genetic counseling is an essential part of genetic testing. Genetic counseling should include patient education, screening, management recommendations, possible consequences of genetic testing, and written, informed consent for genetic testing obtained from the patient and/or parents. Consequently, it may be prudent to refer relevant families to a regional high-risk colon cancer program for initial evaluation, where trained personnel are available to perform genetic testing and pedigree research.

TABLE II Indications for *APC* Gene Testing

≥100 colorectal adenomas
First-degree relatives of patients with FAP[a]
≥20 cumulative colorectal adenomas[b]
First-degree relatives of patients with attenuated FAP[a]

[a] Customarily, individuals older than 10 years are offered gene testing.

[b] Suspected attenuated FAP.

TABLE III Screening/Surveillance Guidelines in Patients with FAP

Patient group	Surveillance
At-risk individuals	Genotyping
	APC gene mutation (+), flexible sigmoidoscopy annually starting at age 12 years
	APC gene mutation (−), flexible sigmoidoscopy age 25 years
	If genotyping not available
	Flexible sigmoidoscopy annually starting age 12 years, then every 2 years starting at age 25, then every 3 years starting age 35, then as per the guidelines for average-risk individuals starting at age 50 years
Affected individuals	Upper gastrointestinal tract surveillance every 3−4 years, and annually if upper-tract polyps
	If retained rectum or J-pouch, flexible sigmoidoscopy every 6 months or 1−2 years, respectively
	Annual physical exam and routine blood tests

Once the disease-causing mutation is identified in an individual affected with FAP, other family members can be tested and endoscopic surveillance can be directed only at those who test positive for the mutation. If the pedigree mutation is not found or if informative genetic testing cannot be done, all first-degree family members should undergo endoscopic screening. Current screening recommendations include yearly sigmoidoscopy starting at 12 years old, reducing screening frequency with each subsequent decade up to age 50 years, after which screening should conform to the American Cancer Society guidelines for average-risk patients. Due to the increased risk of hepatoblastoma in patients with FAP, screening with α-fetoprotein levels and ultrasound imaging of the liver may be prudent in children of affected parents from infancy to 7 years of age (Table III).

Treatment

Surgical Treatment

All patients with FAP require surgical therapy. Colectomy is the recommended treatment to eliminate the risk of colorectal cancer. Prophylactic surgery should be performed shortly after the diagnosis of FAP is established clinically by endoscopy. There are several surgical options available, including subtotal colectomy with ileorectal anastomosis, total proctocolectomy with Brooke ileostomy, and total proctocolectomy with mucosal proctectomy and ileoanal pull-through (with pouch formation). Due to the risk of cancer in the retained rectal segment, surgical approaches that eliminate the rectum are advocated.

Patients with subtotal colectomy require routine endoscopic surveillance of the remaining rectum about every 6 months for recurrent adenomas and/or carcinomas. The cumulative incidence of rectal cancer in patients with FAP is approximately 25% at 25 years after colectomy with ileorectal anastomosis. Factors associated with increased risk of subsequent rectal cancer include high number of rectal polyps, long rectal segment (>10 to 15 cm), inadequate endoscopic surveillance, and colon cancer at the time of colectomy. About 16% of patients with an ileorectal anastomosis will eventually need proctocolectomy. Although the long-term risk of neoplastic transformation in the ileoanal pouch of patients with restorative proctocolectomies is low, adenomas in the pouch have been reported. In such cases, some experts have recommended that endoscopic biopsy surveillance should be considered.

Most authorities recommend upper endoscopic surveillance (with biopsy and brushing) of the stomach, duodenum, and periampullary region with front- and/or side-viewing endoscopes, every 6 months to 4 years, depending on polyp burden. Annual physical examination of the thyroid is warranted, along with consideration for ultrasonography.

Medical Treatment

Use of nonselective or selective cyclooxygenase-2 (COX-2) inhibitors (sulindac and celecoxib, respectively) to prevent or induce regression of polyps in the retained rectum of patients with FAP has been shown to be effective in short-term trials, and in long-term sulindac studies. Currently, celecoxib is approved by the Food and Drug Administration (FDA) for this indication. Administration of sulindac for primary chemoprevention of FAP has failed to prevent the development of adenomatous polyposis in gene mutation carriers. Moreover, use of sulindac has not shown to be effective in regression of duodenal adenomas.

The mechanisms by which nonsteroidal antiinflammatory agents (NSAIDs) mediate polyp regression are not completely understood. However, prostaglandins are believed to promote tumorigenesis by increasing cellular proliferation and inhibiting apoptosis, and NSAIDs (such as sulindac) reduce production of prostaglandins from arachidonic acid via cyclooxygenenase inhibition.

Acknowledgments

Supported by grants from the Clayton Fund, John R. Rangos, Sr. Charitable Foundation, Merck Company, National Institutes of Health grants CA 53801, 1K-07 CA92445-01, and 5P50 CA62924.

See Also the Following Articles

Colectomy • Colorectal Adenocarcinoma • Colorectal Adenomas • Familial Risk of Gastrointestinal Cancers • Genetic Counseling and Testing

Further Reading

Cruz-Correa, M., Hylind, L. M., Romans, K. E., Booker, S. V., and Giardiello, F. M. (2002). Long-term treatment with sulindac in familial adenomatous polyposis: A prospective cohort study. *Gastroenterology* **122**(3), 641–645.

Giardiello, F. M., Brensinger, J. D., Luce, M. C., Petersen, G. M., Cayouette, M. C., Krush, A. J., Bacon, J. A., Booker, S. V., Bufill, J. A., and Hamilton, S. R. (1997). Phenotypic expression in adenomatous polyposis families with mutation in the 5′ region of the adenomatous polyposis coli gene. *Ann. Intern. Med.* **126**, 514–519.

Giardiello, F. M., Brensinger, J. D., Petersen, G. M., Luce, M. C., Hylind, L. M., Abacon, J. A., Booker, S. V., Parker, R. D., and Hamilton, S. R. (1997). The use and interpretation of commercial *APC* gene testing for familial adenomatous polyposis. *N. Engl. J. Med.* **336**, 823–827.

Giardiello, F. M., Brensinger, J. D., and Petersen, G. (2001). American Gastroenterological Association technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* **121**, 198–213.

Giardiello, F. M., Brensinger, J. D., and Petersen, G. (2001). American Gastroenterological Association medical position statement: Hereditary colorectal cancer and genetic testing. *Gastroenterology* **121**, 195–197.

Giardiello, F. M., Hamilton, S. R., Krush, A. J., Piantadosi, S., Hylind, L. M., Celano, P., Booker, S. V., Robinson, C. R., and Offerhaus, G. J. A. (1993). Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. *N. Engl. J. Med* **328**, 1313–1316.

Kinzler, K. W., Nilbert, M. C., Su, L. K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., Mckechnie, D., Finniear, R., Markham, A., Groffen, J., Bogushki, M. S., Altschul, S. F., Horri, A., Ando, H., Miyoshi, Y., Mike, Y., Nishisho, I., and Nakamura, Y. (1991). Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661–665.

Laken, S. J., Papadopoulos, N., Petersen, G., Gruber, S. G., Hamilton, S. R., Giardiello, F. M., Brensinger, J. D., Vogelstein, B., and Kinzler, K. W. (1999). Analysis of masked mutations in familial adenomatous polyposis. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2322–2326.

Nagase, H., Miyoshi, Y., Horii, A., Aoki, T., Ogawa, M., Utsunomiya, J., Baba, S., Sasazuki, T., and Nakamura, Y. (1992). Correlation between the location of germ-line mutations in the *APC* gene and the number of colorectal polyps in familial adenomatous polyposis patients. *Cancer Res.* **52**, 4055–4057.

Nugent, K. P., and Northover, J. (1994). Total colectomy and ileorectal anastomosis. *In* "Familial Adenomatous Polyposis and Other Polyposis Syndromes" (R. K. S. Phillips, A. D. Spigelman, and J. P. Thomson, eds.), pp. 79–91. Edward Arnold, London.

Familial Risk of Gastrointestinal Cancers

JASON C. WILLS AND RANDALL W. BURT
University of Utah Health Science Center, Salt Lake City

adenomatous polyp A polyp derived from the surface epithelium and exhibiting cellular and morphologic dysplasia.
attenuated A less intense form of a condition or disease.
extracolonic Pertaining to areas outside the colon itself.
familial A disorder or trait with excess familial occurrence.
genotype The genetic constitution of an individual.
hamartomatous polyp A benign polyp that arises from the overgrowth of some constituent of the lamina propria, submucosa, or muscular tissue. The surface mucosa or epithelium is normal and corresponds to the location of the polyp within the gastrointestinal tract. Some conditions with hamartomatous polyps have a malignant predisposition.
hereditary Transferred via genes from parent to child.
hyperplastic polyp A nonneoplastic epithelial polyp that is almost always found in the colon. These polyps are characterized by a saw-toothed appearance of surface epithelium thought to be secondary to inadequate sloughing of the cells. Endoscopically, they are usually one to several millimeters in diameter and appear to have little if any malignant potential. Only rarely do they grow larger. These tiny polyps account for approximately half of the polyps found in the colon and their primary importance is distinguishing them from adenomatous polyps. The syndrome of hyperplastic polyposis, on the other hand, does have some risk of malignancy.
inflammatory polyp A pseudo-polyp, consisting of normal tissue, often with increased inflammatory elements. These polyps are mostly associated with the chronic inflammatory conditions of the bowel including ulcerative colitis and Crohn's disease. They do not have malignancy risk, but the inflammatory conditions themselves do.
mismatch repair gene A gene responsible for repairing small DNA errors or mismatches that occur during DNA replication.
mutation A permanent transmissible change in the genetic material, usually in a single gene.
phenotype The physical characteristics of an individual as determined by the expression of that individual's genes.
polyp An abnormal growth of tissue. Polyps are defined histologically within two broad categories, epithelial or hamartomatous. Polyps may be pedunculated (having a stalk) or sessile (flat) and the surface if often described as either smooth or lobulated.
polyposis A condition in which numerous polyps develop in various areas of the gastrointestinal tract.

Gastrointestinal (GI) syndromes include those inherited diseases of the GI tract that predispose to cancer. This article will discuss the risk of gastrointestinal cancer in known inherited syndromes as well as the increased familial risk of colorectal cancer that is not yet genetically defined. Many inherited colon cancer syndromes belong to the category of "polyposis syndromes" in which multiple polyps of the GI tract form and precede colorectal cancer. The polyposis syndromes are categorized by polyp histology and are discussed in detail. More recently, elucidation of the underlying genetic etiology of these conditions has allowed further refinement of syndrome classification. Several polyposis syndromes that are not premalignant and others that are not inherited are also discussed, as they must be distinguished from the inherited, precancerous syndromes. In addition, several miscellaneous inherited conditions that lead to GI cancers but do not involve polyps will be discussed.

THE POLYPOSIS SYNDROMES

Conditions in Which Adenomatous Polyps Develop

Adenomatous polyps are derived from surface epithelium and are divided into three subtypes: tubular, tubulo-villous, and villous. Adenomas are precancerous growths in which the surface epithelium of the gastrointestinal (GI) tract shows features of dysplasia. The lamina propria, submucosa, and muscularis features of the tissue remain normal. Dysplasia is characterized by branching of the microscopic glands and loss of goblet cells and cellular features, including loss of basilar polarity of the nucleus, an increased nuclear : cytoplasmic ratio, increased basophilia of the cytoplasm, and loss of cytoplasmic glycogen. Villous changes are seen as the presence of elongated villi at the surface of the polyp. Each of the three subtypes of adenomatous polyps may also be subclassified by the degree of dysplasia: mild, moderate, and severe. The risk of an

 Copyright 2004, Elsevier (USA). All rights reserved.

adenoma containing malignant tissue or becoming malignant, especially in the colon, relates to several variables including polyp size, the amount of villous tissue in the polyp, and the degree of dysplasia.

Endoscopically, polyps become visible at 1 to 2 mm in diameter. As they enlarge to centimeter size, they usually become more pedunculated, although they may remain sessile. Approximately 5% of adenomas eventually become malignant in accordance with the polyp → cancer sequence. The change to malignancy takes, on average, 10 years, although the variation in duration is substantial. Thirty to 50% of adults eventually develop small adenomas of the colon. Approximately 5% of these colonic adenomas enlarge and become malignant. Approximately half of colonic polyps less than 5 mm in diameter are adenomas. Ninety-five percent of colonic polyps larger than 1 cm are adenomas, as hyperplastic polyps rarely grow larger and hamartomatous polyps are uncommon.

The inherited adenoma syndromes of colon cancer account for only a small fraction of persons with colon cancer overall (approximately 3 to 5%). Nonetheless, the inherited conditions remain a very important part of clinical medicine because of the high risk of colon cancer in affected persons and the inherited nature of the conditions in families. Additionally, the inherited syndromes have revealed important clues to understanding the molecular pathogenesis of all colon cancers.

Familial Adenomatous Polyposis

Disease phenotype Familial adenomatous polyposis (FAP) is characterized by the development of hundreds to thousands of colonic adenomatous polyps. It occurs in between 1 in 8000 and 1 in 10,000 births. The average patient age of adenoma development is 16 years, whereas the average patient age of colon cancer is 39 years. Colon cancer will inevitably develop in all FAP patients if the colon is not removed. If the affected colon is left intact, 87% of FAP patients will develop colon cancer by age 45 and 93% will develop colon cancer by the age of 50. Symptoms such as bloody bowel movements, change in bowel habits, diarrhea, abdominal pain, and weight loss will occur, on average, at age 36. Unfortunately, by the time symptoms occur, two-thirds of persons will already have colon cancer. This syndrome accounts for less than 0.5% of colon cancers in the United States.

Extracolonic gastrointestinal findings are common in FAP. Gastric polyps occur in up to 90% of patients but rarely cause symptoms. The polyps are fundic gland polyps histologically with very little malignant risk (0.5%).

Duodenal polyps occur in up to 90% of FAP cases. These are adenomatous polyps and do have malignant potential with 5−10% of persons with FAP developing duodenal cancer. The duodenal papilla is particularly prone to the development of adenomatous polyps and cancer. Small bowel adenomas also occur distal to the duodenum but the risk of malignancy appears to be very low.

Extraintestinal lesions are common and appear to correlate with the location of the mutation in the adenomatous polyposis coli (APC) gene. Osteomas occur in 20% of FAP patients, most frequently at the angle of the mandible and on the skull. They also may occur on any bone surface of the body, including the long bones. Epidermoid cysts occur in 20% of patients typically along the legs, face, scalp, and arms but potentially on any cutaneous site. The osteomas and epidermoid cysts are always benign but may cause cosmetic problems as they vary in size from millimeters to centimeters. Asymptomatic sclerotic bone lesions of the mandible or maxilla occur in over 90% of FAP patients. Supernumerary teeth and odontomas are unusual.

Congenital hypertrophy of the retinal pigment epithelium (CHRPE) occurs in some families but is always asymptomatic. Desmoids, benign fibroblastic tissue growths, occur in 10% of patients. Symptoms occur in half of these patients through local invasion or compression of other organs, particularly in the abdomen. Lesions in the mesentery are often referred to as mesenteric fibromatosis. Stimulated by surgery, trauma, and estrogens, these lesions can lead to significant morbidity and even some mortality. Extraintestinal cancers are sometimes observed in FAP. Pancreatic and thyroid cancers occur later in life in 2% of patients. Hepatoblastomas develop in 1.6% of children usually within the first 5 years of life. In the FAP variant Turcot syndrome, central nervous system (CNS) tumors occur. These are usually cerebellar medulloblastomas and may occur at any age but in $<1\%$ of patients.

Disease variants Gardner syndrome is a variant of FAP that involves the presence of prominent extracolonic manifestations together with typical colonic findings of FAP. These extracolonic manifestations include osteomas, odontomas, epidermoid cysts, fibromas, and desmoid tumors. This variant is mainly of historical significance, as it is now known to arise from mutations of the same gene, with extraintestinal manifestations determined to some extent by the location of the mutation on the APC gene.

Turcot syndrome is another variant of FAP, with the same colonic findings but coupled with CNS malignancies. Two-thirds of Turcot syndrome families are a subset of FAP and typically have cerebellar

medulloblastomas. The other one-third of Turcot families develop glioblastomas. The colonic lesions of this smaller subset arise not from mutations in the FAP gene, but from mutations in mismatch repair genes.

In attenuated adenomatous polyposis coli (AAPC), the average number of colonic adenomas is only 30. Although these polyps can form throughout the colon, they tend to develop in the proximal colon. The appearance of both colonic adenomas and cancer is delayed approximately 10 years compared with FAP. The risk of cancer is lower than in FAP, but still approaches 80% over a lifetime. The upper GI findings do not appear to be attenuated.

Genetics FAP is inherited as an autosomal dominant disease with near 100% penetrance in terms of colonic polyps and cancer. At least 95% of affected families arise from mutations in the APC gene on chromosome 5q21. The APC gene is a tumor suppressor gene whose normal function is to control cell proliferation and apoptosis. Most mutations reported give rise to a premature stop codon through either a single base-pair substitution or one or two base-pair deletions or additions. Some large chromosomal deletions have been reported. Approximately 30% of newly diagnosed cases, not belonging to families with known disease, appear to represent new mutations.

Normally, the APC gene functions as a negative regulator in the WNT signaling pathway. It binds to soluble β-catenin together with several other molecules and phosphorylates it, leading to cytosolic degradation of the β-catenin. The stimulation of the WNT signaling pathway gives rise to uncoupling of the β-catenin to the complex and therefore no phosphorylation occurs. β-Catenin then traverses the nuclear membrane and binds to Tcf-Lef transcription regulators, giving rise to the transcription of various proteins (c-Myc, cyclin D1, and others). The result is increased cell proliferation and decreased cell apoptosis. Therefore, aberrant function with the mutant APC gene results in a constant failure of β-catenin/phosphorylation and unregulated nuclear transcriptional stimulation, leading to unregulated cell proliferation and unregulated suppression of apoptosis.

Genetic testing Genetic testing is now considered part of the standard management of families with FAP. Testing is used in two settings: (1) to confirm the diagnosis of FAP in suspected cases and (2) to determine the gene carriers in families with FAP.

A person known to have the disease is tested first. Many methods are used for genetic testing in FAP. *In vitro* protein truncation assay detects the presence of truncated mutations *in vitro*. It detects a mutation in 80–90% of affected families known to have FAP. Once the mutation has been found in an affected person, testing is nearly 100% effective in detecting mutations in family members. A second method, gene sequencing, is often preceded by single-strand conformational polymorphism (SSCP) or denaturing gradient gel electrophoresis (DGGE) to narrow the area of the gene where sequencing is to be performed. Sequencing is up to 95% effective at finding a disease-causing mutation if one is present and, once a mutation is found, sequencing is nearly 100% effective at detecting the mutation in family members. If these two methods are unsuccessful, genetic linkage testing can be performed. Two or more affected persons from two generations must be living for DNA to be obtained and linkage to be performed. Linkage is effective in $>95\%$ of families with $>98\%$ accuracy.

Genotype–phenotype correlations have not yet been found to be of precise use in clinical settings. At this time, the following correlations have been made:

1. CHRPE is present in families with mutations distal to exon 9 of the APC gene;
2. Dense polyposis is present with mutations in the midportion of exon 15;
3. AAPC is found with mutations in the extreme proximal or distal end of the APC gene; and
4. Osteomas and desmoids are more commonly found with mutations in the distal portion of exon 15.

Clinical management Genetic testing should be considered between the ages of 10 and 12 years when it will first be clinically useful. There may be a role for earlier testing to determine who should be screened for hepatoblastoma.

Gastrointestinal tract screening Sigmoidoscopy should be performed every 1–2 years in gene carriers or in all at-risk persons if genetic testing is not performed or not informative. This should begin at age 10 to 12 years. Colonoscopy should be performed in families with AAPC every 2 years beginning by age 21 years, or earlier, depending on the age of polyp emergence in other family members.

Upper GI endoscopy should begin when colon polyps emerge or by age 25 years and repeated every 1–3 years depending on polyp size, number, and histology. Side viewing should be performed as part of the examination to carefully identify and examine the duodenal papilla. Small bowel X ray should be carried out if numerous or large adenomas are present in the duodenum as well as before planned colectomy. The number and size of polyps found determine follow-up. Overall cancer-screening recommendations for FAP are found in Table I.

Colon surgery Surgery should be considered once polyps emerge, although surgery can often be de-

TABLE I Cancer Risk and Screening Recommendations in Familial Adenomatous Polyposis

Cancer	Cancer risk	Screening recommendation
Colon	Nearly 100%	Sigmoidoscopy annually, beginning at age 10–12 years
Duodenal or peri-ampullary	5–10%	Upper GI endoscopy (including side-viewing exam) every 1–3 years, starting at age 20–25 years
Pancreatic	~2%	Possibly periodic abdominal ultrasound after age 20 years
Thyroid	~2%	Annual thyroid exam, starting at age 10–12 years
Gastric	~0.5%	Same as for duodenal cancer
CNS, usually cerebellar meduloblastoma (Turcot syndrome)	<1%	Annual physical exam, possibly periodic head CT scan in affected families
Hepatoblastoma	1.6% of children <5 years of age	Possible liver palpation, hepatic ultrasound, α-fetoprotein annually during first decade of life

CT, computed tomography.

layed until a patient finishes high school. Surgery should not be delayed if the number of polyps is near 100, if any polyps approach 1 cm in diameter, or if advanced histology is detected. If surgery is delayed, colonoscopy should be performed annually.

Surgical options include subtotal colectomy with ileorectal anastomosis or total colectomy with mucosal proctocolectomy, ileal pouch construction, and ileo-anal anastomosis. The former is single stage and much less complicated but the latter is a more definitive surgery that eliminates the risk of rectal cancer. The subtotal colectomy still requires surveillance of the remaining rectum every 6–12 months for polyp ablation. The total colectomy also requires surveillance of the ileal pouch every 2 years as polyps sometimes recur there. Because the total colectomy is more a complicated surgery, it should be performed only by specialized surgeons.

Chemoprevention Chemoprevention as primary therapy in the gut is still experimental. Nonsteroidal anti-inflammatory drugs are known to suppress rectal adenomas and possibly colonic adenomas. Sulindac, in particular, has a modest effect but the side effects in the upper GI are of concern. Other, more selective agents, such as the cyclooxygenase-2 inhibitors or sulindac metabolites, are currently being evaluated as they also decrease the size and number of polyps throughout the colon. At this time, only celecoxib is approved for preventive treatment of the remaining rectum in persons with a subtotal colectomy.

Hereditary Nonpolyposis Colorectal Cancer

Disease phenotype Hereditary nonpolyposis colorectal cancer (HNPCC), also termed Lynch syndrome, accounts for 2–3% of all colon cancer cases. It is characterized by multiple family members with colorectal and other cancers. This is not a polyposis syndrome in the strict sense, as multiple adenomatous polyps may form but most usually only a few will appear. It is inherited as an autosomal dominant disease of primarily colon and endometrial cancer. The colonic polyps that form are larger, occur earlier, and more often contain advanced histology when compared to age-matched controls. Although usually only one or several adenomatous polyps form, the average lifetime risk of developing colon cancer is 80% for those affected by HNPCC. The average age of cancer diagnosis is 44 years.

Diagnosis The clinical criteria for the diagnosis of HNPCC are called the Amsterdam criteria. They are as follows:

Classic or Amsterdam I Criteria There should be at least three relatives with colorectal cancer with the following criteria also met:

1. One should be a first-degree relative of the other two.
2. At least two successive generations should be affected.
3. At least one colorectal cancer should be diagnosed before the age of 50 years.
4. Familial adenomatous polyposis should be excluded.
5. Tumors should be verified by pathological examination.

Amsterdam II Criteria (a recent modification of the original criteria) There should be at least three relatives with an HNPCC-associated cancer (colorectal, endometrial, small bowel, ureteral, and renal pelvis).

The same additional criteria as stated for the Amsterdam I Criteria also need to be met.

Clinical Implications of Meeting the Criteria When the classic Amsterdam criteria are met, 50–70% of families will be found to have a mutation of one of the mismatch repair genes. Approximately 8% of families with multiple cases of colon cancer, but who do not meet the classic Amsterdam Criteria, will be found to have a mutation in one of the mismatch repair genes and, therefore, will have HNPCC. An elevated risk

for cancer at many other sites exists and screening is recommended (see Table II).

Disease variants Muire-Torre syndrome is typical HNPCC together with cutaneous tumors, such as sebaceous adenomas or epitheliomas, basal cell epitheliomas, keratoacanthomas, or sebaceous carcinomas (usually of the eyelid). This arises from mutations in the MSH2 mismatch repair gene.

Turcot syndrome is HNPCC together with CNS tumors, especially glioblastomas. This represents one-third of Turcot syndrome families.

Patients with MSH6 syndrome carry a 60% risk of developing endometrial cancer and a 40% risk of colon cancer. Some families reported have tumor onset later than typical HNPCC. The syndrome is associated with mutations in the MSH6 mismatch repair gene.

Genetics Hereditary nonpolyposis colorectal cancer is inherited in an autosomal dominant manner with at least 80% penetrance by age 70. HNPCC arises from mutations in any one of the mismatch repair (MMR) genes, of which six mutations are known. Two gene mutations, MLH1 and MSH2, are present in over 95% of those families in whom mutations have been found. Genetic errors that accumulate when the MMR genes are mutated and dysfunctional are quite specific and include genes such as TGF-β and BAX.

Mismatch repair genes These genes are responsible for the repair of errors that occur during DNA replication. When one of these genes is damaged, a certain type of DNA mutation, called a replication error, accumulates throughout the genome of the involved tumors. Replication errors are common during cell division but are usually repaired by the MMR system. When an MMR gene itself is damaged or inactivated, replication errors persist and accumulate through repeated cell divisions. Such mutations are most easily identified in segments of DNA called microsatellites, which are sequences of repeating DNA bases found throughout the human genome. When multiple microsatellite errors are present, the tumor tissue is said to exhibit microsatellite instability (MSI). Almost all (>90%) colon cancers in HNPCC exhibit microsatellite instability. MSI is found in only 15% of sporadic colon cancer and occurs by a different mechanism. Because MSI is easily detected in tumor tissue, it is used as a marker for HNPCC. It has been suggested that MSI testing on tumors should be performed when any of the "Bethesda Criteria" are met (Table III). The Bethesda criteria are a more complex and inclusive set of rules but it is hoped that a greater number of HNPCC families will be detected by the use of these criteria.

Genetic testing Testing for MSI instability is carried out on the tumor tissue. To find mutations in the mismatch repair genes, sequencing is most commonly used. Gene sequencing is often preceded by SSCP or DGGE to narrow the area of the gene where sequencing is to be performed. This is successful in 50 to 70% of families who meet the Amsterdam criteria. Once a disease-causing mutation is found in an index case, testing in other family members approaches 100% accuracy.

Genetic testing should be performed to confirm a diagnosis of HNPCC. It should also be offered to members of families in which there is a known mutation or that meet the Amsterdam criteria. In families not meeting the Amsterdam criteria, however, the issue is problematic. The phenotype is not distinct in a single individual as it is in other polyposis syndromes. Yet, a percentage of families with a family history of colon cancer but not meeting the Amsterdam criteria will,

TABLE II Cancer Risk and Screening Recommendations for Hereditary Nonpolyposis Colorectal Cancer

Cancer	Cancer risk	Screening recommendations
Colon	>80%	Colonoscopy every 1–2 years starting at age 20–25 years or 10 years younger than earliest case in the family, whichever is earlier
Endometrial	43–60%	Pelvic exam, transvaginal ultrasound and/or endometrial aspirate every 1–2 years starting at age 25–35 years
Ovarian	9–12%	Same as for endometrial cancer
Gastric	13–19%	Upper GI endoscopy every 1–2 years starting age 30–35 years
Urinary tract	4–10%	Ultrasound and urinalysis every 1–2 years starting age 30–35 years
Renal cell adenocarcinoma	3.3%	Same as for urinary tract cancer
Biliary tract and gallbladder	2.0–18%	Uncertain, possibly liver chemistries annually after age 30 years
Central nervous system, usually glioblastoma (Turcot syndrome)	3.7%	Uncertain, possibly annual physical exam and periodic head CT scan in affected families
Small bowel	1–4%	Uncertain, at least small bowel X ray if symptoms occur

Note. CT, computed tomography.

TABLE III The Modified Bethesda Criteria for Hereditary Nonpolyposis Colorectal Cancer

Individuals with cancer in families that meet the Amsterdam criteria
Individuals with two synchronous or metachronous HNPCC-related cancers[a]
Individuals with colorectal cancer *and*
—A first-degree relative with colorectal cancer and/or HNPCC-related extracolonic cancer[a] and/or colorectal adenoma *and*
—A cancer diagnosed before age 50 years *or*
—An adenoma diagnosed before age 40 years
Individuals with colorectal and endometrial cancer diagnosed before age 45 years
Individuals with right-sided undifferentiated colon cancer diagnosed before age 45 years
Individuals with signet-ring cell type colorectal cancer diagnosed before age 45 years
Individuals with adenomas diagnosed before age 40 years

[a] Colon, rectum, stomach, small bowel, endometrium, ovary, ureter, and other cancers.

indeed, have HNPCC. Three approaches have been suggested to determine which families that do not meet the Amsterdam criteria should nonetheless have genetic testing for HNPCC.

1. Apply MSI testing to the colon cancer tissue in the following situations and when positive, perform testing on the DNA from peripheral blood to find MMR mutations:
   a. Colon cancer diagnosis < 50 years;
   b. Colon cancer plus one first-degree relative with colon or colorectal cancer;
   c. Colon cancer plus a previous colon or endometrial cancer.

   With this method, 24% of colon cancer cases will undergo MSI testing of the tumor tissue and 4% of colon cancer cases will have MMR mutations found.
2. Use a specific logistic model applied to an extended family that includes kindred structure and known cancer cases.
   a. If the model predicts $> 20\%$ chance of HNPCC, go directly to mutation finding.
   b. If the model predicts $< 20\%$ chance of HNPCC, do MSI first and go to mutation finding only if MSI is positive on tumor tissue.
3. Go directly to MMR mutation finding if one of the first three Bethesda criteria is positive, but use age < 50 years rather than 45 years.

   In one study, this approach had 94% sensitivity and 49% specificity.

Clinical management Colon screening should be performed in all at-risk persons within a family known to have HNPCC either by genetic testing or by the Amsterdam criteria, even if a disease-causing mutation cannot be found. Screening should be carried out by colonoscopy every 2 years starting at age 25, or 10 years earlier than the youngest colon cancer diagnosis in the family, whichever comes first.

Screening recommendations and cancer risk in other organs are listed in Table II.

Surgical management A subtotal colectomy is the procedure of choice to prevent colon cancer and should be performed if colon cancer is diagnosed or an advanced adenomatous polyp that cannot be adequately treated endoscopically is found. Prophylactic colectomy in persons known to have a disease-causing mutation in one of the MMR genes is currently being debated. Hysterectomy is indicated if any dysplasia is found with endometrial aspirate.

Conditions in Which Hamartomatous Polyps Develop

The most common hamartomas involved in inherited conditions include Peutz-Jeghers polyps (arborizing pattern of muscular tissue) and juvenile polyps (overgrowth of the lamina propria often with cysts). Other hamartomatous polyps that occasionally occur include lipomas, leiomyomas, neurofibromas, and ganglioneuromas. All of these polyps may be either sessile or pedunculated.

Although the hamartomatous polyps themselves are considered nonneoplastic and therefore benign lesions, most of the hamartomatous conditions exhibit substantial malignancy risk. Furthermore, adenomatous change has been noted to sometimes occur in both Peutz-Jeghers and juvenile polyps.

Hamartomatous polyps are much less common than adenomatous polyps but more commonly involve extracolonic sites of the GI tract. Hamartomas account for substantially less than 5% of colonic polyps overall. Two to 3% of children will develop sporadic juvenile polyps of the colon, whereas sporadic Peutz-Jeghers polyps are very unusual. The hamartoma syndromes are likewise much less common than adenoma syndromes.

Peutz-Jeghers Syndrome

Disease phenotype Peutz-Jeghers is characterized by melanin pigment spots, often in the perioral area, and distinctive polyps throughout the gastrointestinal tract but most commonly in the small bowel. Polyps are sessile or pedunculated, ranging in size from a few millimeters to centimeters. The small bowel is most commonly affected (78%), followed by the colon (42%), stomach (38%), and rectum 28%.

During the first three decades of life, benign complications of bleeding, obstruction, and intussusception occur but malignant transformation of the polyps is the main concern thereafter. GI and non-GI cancers are common, with a combined frequency of 93% by age 65 years. The characteristic pigmentation is seen in >95% of individuals, most commonly the perioral and buccal areas. They also occur around the eyes, palm, soles of the feet, and perineum. Pigmentation appears in infancy and begins fading at puberty except on the buccal mucosa, which provides a clinical feature for diagnosis throughout life.

Genetics Peutz-Jeghers is an autosomal dominantly inherited syndrome occurring in 1 in 200,000 births. It arises from mutations of the STK11 gene (also called LKB1). Only half of families with the disease have this mutation, suggesting that there is another responsible gene. Genetic testing for this specific mutation is commercially available.

Clinical management Management involves screening for cancer prevention and is reviewed in Table IV.

Juvenile Polyposis

Disease phenotype Juvenile polyps occur in 2% of children. The diagnosis of juvenile polyposis (JP) is made with the presence of 10 or more juvenile polyps, which occur most commonly in the colon but which can occur anywhere in the GI tract. The full-blown syndrome may be characterized by hundreds of polyps. These hamartomatous polyps involve overgrowth of the lamina propria, often with cysts. Endoscopically they appear smooth, sessile, or pedunculated and often are covered with exudate. They range in size from millimeters to centimeters. Polyps in JP differ from sporadic juvenile polyps in that new polyps almost always form after removal and polyps always occur in adults. Benign problems, particularly colonic bleeding or anemia, usually occur in the first decade of life. The risk of colon cancer is substantial, with the average age of diagnosis being 34 years. Cancer in the stomach, small bowel, and pancreas has been reported but the

TABLE IV Cancer Risk and Screening Recommendations for Peutz-Jeghers Syndrome

Cancer type	Cancer risk[a]	Screening recommendation
GI Cancer		
Colon	39%	Colonoscopy, beginning with symptoms or in late teens if no symptoms occur; interval determined by number of polyps but at least every 3 years once begun
Pancreatic	36%	Endoscopic or abdominal ultrasound every 1–2 years starting at age 30 years
Stomach	29%	Upper GI endoscopy every 2 years starting at age 10 years
Small bowel	13%	Annual hemoglobin, small bowel X ray every 2 years, both starting at age 10 years
Esophagus	0.5%	None given
Non-GI cancer		
Breast	54%	Annual breast exam and mammography every 2–3 years, both starting at age 25 years
Ovarian	21%	Annual pelvic exam with pap smear, and annual pelvic or vaginal ultrasound and/or uterine washings, all starting at age 20 years
Uterine	9%	
Adenoma malignum (cervix)	Rare	
Sex cord tumor with annular, tubules, in almost all women	20% become malignant	
Sertoli cell tumor (males), rare	10–20% become malignant	Annual testicular exam, starting at age 10 years, testicular ultrasound if feminizing features occur
Lung	15%	None given

[a] Cumulative cancer risks, ages 15–64 years.

TABLE V Screening Recommendations for Juvenile Polyposis

Cancer	Cancer risk	Screening recommendations
Colon	May be as high as 50%	Colonoscopy, beginning with symptoms or in early teens if no symptoms occur; interval determined by number of polyps but at least every 3 years once begun
Gastric and duodenum	Rare	Upper GI every 3 years, starting in early teens (mainly to avoid complications of benign polyps)

association is not certain. Congenital defects are seen with the nonfamilial form of the disease and include cardiac and cranial abnormalities, cleft palate, polydactyly, and bowel malrotations. Cancer risk and screening recommendations are outlined in Table V. Half of the affected families have a mutation of the SMAD4 gene (also called DPC4) on chromosome 18 or the BMPRA1 gene on chromosome 10. Other genes may be involved. Testing for the SMAD4 gene mutation is commercially available.

Clinical management Clinical management consists mainly of prevention of benign and malignant complications with empiric guidelines as described under the screening recommendations given in Table V.

Cowden Syndrome

Disease phenotype Multiple hamartomatous polyps occur in the colon and throughout the GI tract in Cowden syndrome (CS). A number of different types of hamartomas occur. Juvenile polyps are by far the most common but lipomas, inflammatory polyps, ganglioneuromas, and lymphoid hyperplasia are also seen.

Skin lesions commonly occur in CS. The hallmark is the presence of multiple facial trichilemmomas. They most commonly occur around the mouth, nose, and eyes. Café-au-lait spots, vitiligo, cysts, and squamous cell or basal cell carcinomas have less commonly been described.

Oral mucosal lesions similar to trichilemmomas develop a few years after skin growths in approximately 85% of individuals. They appear as pinpoint, red, flat-topped papules on the outer lips and small, flat, papillomatous or verrucous papules on the oral mucosa, gingiva, and tongue.

Thyroid abnormalities occur in two-thirds of patients. Goiter arising from nodular hyperplasia or follicular adenomas is seen histologically. The risk of thyroid carcinoma is 10%.

Breast lesions, including fibrocystic disease or fibroadenomas, occur in three-fourths of affected females. The reported incidence of breast cancer is 50% with frequent bilateral occurrence. Median age at diagnosis is just 41 years.

Additional benign soft tissue and visceral tumors, such as hemangiomas, lipomas, lymphangiomas, neurofibromas, uterine leiomyomas, and meningiomas, have been observed. Developmental and congenital abnormalities also occur. These include hypoplastic mandible, prominent forehead, and a high arched palate.

Diagnosis The International Cowden Consortium for the diagnosis of CS has suggested specific clinical diagnostic criteria, as shown in Table VI.

The diagnosis can be made in an individual based on mucocutaneous lesions alone if:

- there are $\geq$6 facial papules, of which $\geq$3 are trichilemmomas;
- **or** if there are $\geq$6 palmoplantar keratoses;
- **or** if oral mucosal papillomatosis and acral keratosis are found;
- **or** if there are cutaneous facial papules with oral mucosal papillomatosis.

If mucocutaneous lesions are absent, the diagnosis can be made under the following conditions:

1. two major criteria (one must be macrocephaly or LDD) are met; or
2. one major criterion and three minor criteria are met; or
3. four minor criteria are met.

Operational diagnosis in a family in which one person is diagnostic for Cowden syndrome is made with the presence of pathognomonic criteria: any one major criterion with or without minor criteria or any two minor criteria must be met.

A related syndrome, Bannayan-Riley-Ruvalcaba (BRR) syndrome, is believed to be allelic to Cowden syndrome, arising from mutations of the PTEN gene. It is characterized by macrocephaly, lipomas, pigmented macules of the glans penis, and other features of Cowden syndrome.

Genetics Cowden syndrome is inherited in an autosomal dominant manner and occurs in 1 in 200,000 individuals. The mutation of the PTEN gene, a tumor suppressor gene, is found in 80% of those who meet the diagnostic criteria for Cowden syndrome and

TABLE VI The *International Cowden Consortium's* Clinical Diagnostic Criteria for the Diagnosis of CS

Pathognomonic criteria—mucocutaneous lesions
a. Trichilemmomas, facial
b. Acral keratoses
c. Papillomatous papules
d. Mucosal lesion
Major criteria
a. Breast carcinoma
b. Thyroid carcinoma (nonmedullary), especially follicular thyroid carcinoma
c. Macrocephaly (megalencephaly $>$ 95th percentile)
d. Lhermitte-Duclos disease (cerebellar dysplastic gangliocytoma)
e. Endometrial carcinoma
Minor criteria
a. Other thyroid lesions (e.g., adenoma or multinodular goiter)
b. Mental retardation (IQ $\leq$ 75)
c. GI hamartomas
d. Fibrocystic breast disease
e. Lipomas
f. Fibromas
g. GU tumor (renal cell carcinoma, uterine fibroids) or malformation

approximately 50% of those who meet the criteria for BRR syndrome.

Management There is little if any risk of colon cancer and therefore no screening is needed. Thyroid cancer risk is 3−10% and it is recommended that patients have annual thyroid exams starting in the teenage years. Breast cancer risk is 25−50%. Annual breast exams starting at 25 years and annual mammography starting at 30 years are recommended. The uterine and ovarian cancer risks are possibly elevated but screening recommendations are uncertain.

Hereditary Mixed Polyposis Syndrome

There have been three families described with links to a locus on chromosome 6. Affected persons exhibit primarily juvenile polyps, but also adenomatous polyps, hyperplastic polyps, and sometimes polyps containing mixed histology.

Gorlin's Syndrome

Gorlin's syndrome is also called nevoid basal cell carcinoma syndrome. This is an autosomal dominant disorder with an estimated occurrence of 1 in 56,000. It primarily involves multiple basal cell carcinomas, odontogenic jaw cysts, broad facies, congenital skeletal anomalies, ectopic calcification of the falx cerebri, and characteristic pits in the skin of the palms and soles. Gastric hamartomatous polyps have been reported. It occurs secondary to mutation of the tumor suppressor gene PTCH on chromosome 9q22.3.

Cronkhite-Canada Syndrome

This is an acquired condition characterized by the rapid onset of generalized polyposis, cutaneous hyperpigmentation, hair loss, nail atrophy, diarrhea, weight loss, and hypogeusia (the dominant symptom in most patients).

It is extremely rare with a worldwide distribution. The average age of onset is 59 years with a range of 31 to 86 years. Males are affected 60% of the time but no familiar occurrences have been observed. It is rapidly progressive and can be fatal within a few months, although a protracted course is also observed. Diarrhea and protein-losing enteropathy may be severe, leading to malnutrition and its complications.

The polyps vary in size from millimeters to centimeters and are always sessile. The esophagus appears to be spared. The mucosa between visible polyps also shows changes typical of juvenile polyps. Adenomas and cancer have been reported in the hamartomatous polyps. The incidence of colon cancer is 12%.

The etiology is unclear but may have nutritional, infectious, and immunological associations as variable success has been seen with hyperalimentation, antibiotics, and corticosteroids. Surgery may be necessary for complications of polyps, cancer, and protein-losing enteropathy.

Neurofibromatosis Type I

Neurofibromatosis type I (NFI) is also called Von Recklinghausen disease. Clinical features include greater than five cutaneous café-au-lait spots and frequent neurofibromas of the skin. Approximately 25% will exhibit multiple intestinal polypoid neurofibromas and less commonly ganglioneuromas. The small bowel is most commonly affected, followed by the stomach and colon. It is autosomal dominant, arising from mutations of the NFI gene on chromosome 17q.

Multiple Endocrine Neoplasia Type I

Multiple endocrine neoplasia type I is also called multiple neuroma syndrome. In addition to the presence of medullary thyroid carcinoma, pheochromocytoma, and parathyroid disease, there are enlarged and nodular lips (from ganglioneuromas), marfanoid habitus, and ganglioneuromatosis of the entire GI tract from lip to anus, although most commonly in the colon and rectum. Dysmotility, as a result of the ganglioneuromatosis, causes diarrhea and/or constipation. The syndrome arises from mutations of the RET proto-oncogene on chromosome 10q11.2 but half of the

cases are considered to be new mutations. Diagnosis is made by biopsy.

Devon Polyposis

Devon polyposis is also known as Devon family syndrome. Three generations of females in a family were found to have multiple inflammatory fibroid polyps of the ileum. The polyps varied in size from 0.5 to 8.0 cm. Similar polyps were found in the gastric antrum in one patient. Each affected person experienced intussusception or small bowel obstruction. The polyps were a benign proliferation of histiocytes.

Miscellaneous Conditions with Multiple Polypoid Lesions of the Gastrointestinal Tract

Hyperplastic Polyposis Syndrome

Hyperplastic polyposis (HPP) is usually defined as one of three phenotypes:

1. ≥ 5 hyperplastic polyps proximal to the sigmoid colon with at least 2 that are > 10 mm in size; or
2. any number of hyperplastic polyps proximal to the sigmoid in an individual with a first-degree relative with HPP; or
3. > 30 hyperplastic polyps of any size distributed throughout the colon.

All these phenotypes appear to have an increased risk of colon cancer. Cancer appears in small and large polyps with or without dysplasia. Patients are encouraged to undergo colectomy and asymptomatic family members should probably undergo colonoscopy. Families with HPP have been reported but account for a small minority of cases. The underlying genetic etiology is unknown.

Inflammatory Polyps in Inflammatory Bowel Disease

Multiple inflammatory polyps are frequently found in ulcerative colitis (UC) and Crohn's disease. The polyps represent remaining normal tissue, with some inflammatory elements that persist during the healing phases of the diseases. The polyps themselves have no malignant potential, although both UC and Crohn's disease have a colon malignancy risk, which parallels the extent of colonic involvement and the duration of the disease.

Nodular Lymphoid Hyperplasia

Lymphoid nodules are a normal part of the GI tract, but are usually not visible during colonoscopy. Prominent lymphoid nodules may sometimes be observed in the following situations:

1. occurring in association with lymphoma;
2. occurring in association with common variable immunodeficiency;
3. occurring in younger children and due to an uncertain cause, particularly when occurring in the terminal ileum;
4. occurring in adults and due to an uncertain cause, particularly when occurring in the terminal ileum.

Lymphomatous Polyposis: Two Types

Multiple lymphomatous polyposis This is a non-Hodgkin's B-cell lymphoma, which is the GI counterpart of mantle cell lymphoma. It arises from lymphocytes that have homing receptors for the lymphoid tissue of the GI tract, thus giving rise to diffuse polyposis. Mucosa-associated lymphoid tissue lymphomas, follicular lymphomas, and primary T-cell lymphomas have now also been described and are morphologically similar to multiple lymphomatous polyposis (MLP) in that they exhibit a multiple polypoid appearance in the gastrointestinal tract and are thus considered types of MLP.

Immunoproliferative small intestinal disease Immunoproliferative small intestinal disease (IPSID), previously called both Mediterranean-type lymphoma and α-heavy chain disease, can also exhibit multiple nodular lesions of the small bowel. These lymphomatous lesions result from an intense proliferation of the plasma cells of the lamina propria of the GI tract. A paraprotein is usually present and it is usually classified as a plasma cell tumor or an immunoblastic sarcoma. The disease most often occurs in young people with predominance in the Mediterranean region and an association with malabsorption.

Leiomyomatosis and Lipomatosis

Leiomyomatosis and lipomatosis have been reported but usually only a single leiomyoma or lipoma occurs.

Pneumatosis Cystoides Intestinalis

Pneumatosis cystoides intestinalis is characterized by multiple air-filled cysts of the wall of the GI tract, usually from endoscopy, trauma, surgery, infection, or systemic disease, but sometimes the condition is idiopathic.

FAMILIAL RISK AND COLON CANCER

Risk of Colon Cancer in Relatives

The majority of colon cancer cases appear to be sporadic with no evidence of an inherited disorder.

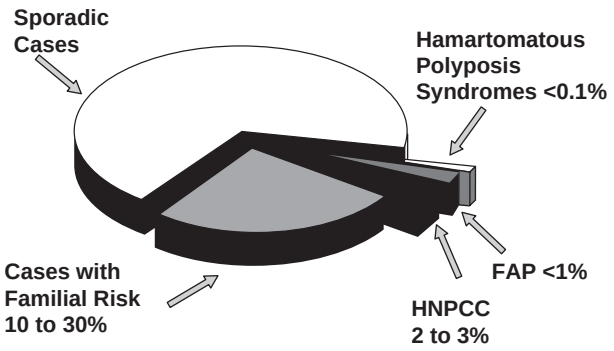


FIGURE 1 The fractions of colon cancer cases that arise in various family risk settings.

Roughly 20–30% of cases can be attributed to familial risk, with known inherited syndromes accounting for 1–3% (Fig. 1). The average American has an approximately 6% lifetime risk of developing colorectal cancer but first-degree relatives of persons with colon cancer have a two- to threefold increased risk of large bowel malignancy compared with controls or the general population. A high-risk family history is defined as a first-degree relative with colorectal cancer or adenoma diagnosed earlier than age 60 years or two first-degree relatives diagnosed with colorectal cancer at any age.

Many studies examining the familial risk of colon cancer also relate certain clinical findings to the severity of risk. The findings most consistently predictive of severity were the number of immediate relatives with colon cancer and the age of cancer diagnosis. If two or more first-degree relatives had colon cancer, the risk of large bowel malignancy for other family members is consistently higher than if only one first-degree relative was affected. The risk is higher if the first-degree relative is diagnosed at a younger age (Table VII). Colon cancer even in a second-degree relative (grandparent, aunt, uncle) or third-degree relative (great-grandparent, cousin) increases an individual's risk for colon cancer, but only approximately 50% above the average risk.

Some familial risk stratification findings related to adenomatous polyps in relatives have also been made. In the National Polyp Study, the risk of colorectal cancer in siblings and parents of persons with any adenomatous polyp was 1.78 (95% CI, 1.18–2.67). The risk of colon cancer, however, was 2.59 (95% CI, 1.64–4.58) in siblings of an adenoma patient diagnosed at an age <60 years compared with siblings of an adenoma patient diagnosed at an age ≥ 60 years. Therefore, first-degree relatives of patients with adenomatous polyps have an increased risk of colon cancer. Similarly, first-degree relatives of persons with colon cancer have an increased risk of adenomatous polyps. Given the current knowledge about the adenoma–carcinoma sequence, this would seem logical.

Genetics

The etiology of the commonly observed familial risk is generally unknown. It likely occurs from both inherited susceptibility and share environmental factors, with genetic determinants making one more susceptible to deleterious environmental agents. Spouses of persons with colon cancer do not exhibit the increased risk found in first-degree relatives, indicating the small contribution of shared environment. Kindred studies have further found that common familial risk probably arises from mildly to moderately penetrant inherited susceptibility factors. A number of genes and chromosomal loci that seem to be involved in this manner have been described. The I1307K APC mutation in the Ashkenazi Jewish population gives rise to a milder form of colon cancer predisposition than observed in FAP. The degree of predisposition is low to modest at most, with no difference in phenotype from sporadic colorectal cancer.

Disease-causing mutations of the MSH6 gene (a mismatch repair gene) are found in approximately 7% of patients with a positive family history of colon cancer. This mutation may be responsible for a substantial

TABLE VII Summary of Familial Risk of Colon Cancer

Familiar setting	Approximate lifetime risk of colon cancer
General population	6%
One first-degree relative with colon cancer	2- to 3-fold increase
Two first-degree relatives with colon cancer	3- to 4-fold increase
First-degree relative with colon cancer diagnosed at ≤ 50 years	3- to 4-fold increase
One second- or third-degree relative with colon cancer	$\sim$1.5-fold increase
Two second-degree relatives with colon cancer	$\sim$2- to 3-fold increase
One first-degree relative with an adenomatous polyp	$\sim$2-fold increase

Note. First-degree relatives include parents, siblings, and children. Second-degree relatives include grandparents, aunts, and uncles. Third-degree relatives include great-grandparents and cousins.

number of familial colon cancers that occur at somewhat older ages but do not fit into the syndrome of HNPCC.

A type I transforming growth factor receptor allele, TβR-1(6A), has been found in a higher fraction of colon cancer patients, both homozygotes and heterozygotes.

A family whose members exhibit frequent colonic adenomas, villous adenomas, serrated adenomas, and colon cancer was found to link to a locus on chromosome 15q. Finally, certain polymorphisms of genes involved in the metabolism of both protective and deleterious environmental agents have been associated with predisposition to colon cancer. Such genes include methylenetetrahydrofolate reductase and *N*-acetyltransferases 1 and 2 (NAT1 and NAT2).

Screening

The presence of strong familial risk dictates a more aggressive screening compared with average-risk individuals beginning at a younger age. The American Cancer Society recommends full colonoscopy for persons with first-degree relatives with colorectal cancer or adenoma diagnosed prior to age 60 years and for those with two or more first-degree relatives with colorectal cancer. This screening should start at age 40 years or 10 years earlier than the age of diagnosis of the youngest affected relative. Full colonoscopy is recommended every 5 years if no neoplasms are found. Patients with a less severe family history should have average risk screening, but starting at age 40 years, rather than 50 years. Current screening recommendations for persons with increased familial risks are outlined in Table VIII.

OTHER CANCERS OF THE GASTROINTESTINAL TRACT

Hereditary Diffuse Gastric Cancer

A small fraction of gastric cancers occur as part of inherited syndromes in which it is not the predominant feature. These include HNPCC, Peutz-Jeghers, FAP, juvenile polyposis, and possibly Cowden syndrome, as well as subsets of families with Li-Fraumeni Syndrome (LFS). However, hereditary diffuse gastric cancer (HDGC) is a rare syndrome, accounting for a substantial fraction of inherited gastric cancers. It is an autosomal dominant disease with approximately 70% penetrance arising from mutations of the DCH1 gene encoding the intracellular adhesion receptor molecule, E-cadherin. The average age of diagnosis is 38 years, with several cases reported to occur under the age of 18, but not younger than 15 years. A "*linitis plastica*" picture develops with mucosal abnormalities usually occurring late. The histology is always of the diffuse type.

Screening is recommended with endoscopy with biopsies every 6–12 months in gene carriers. Prophylactic surgery should be discussed with all gene carriers but the high morbidity and mortality of total gastrectomy should be weighed against the very high risk of developing this highly incurable cancer. Additional cancers of the breast, colon, and prostate have been reported in HDGC families but the association of these cancers is uncertain. Even so, screening for these other cancers is recommended in gene carriers.

Familial Pancreatic Cancer

A number of families with multiple cases of pancreatic cancer have now been identified. Risk factors for pancreatic cancer are typically environmental but genetic etiology relates to several syndromes in which pancreatic cancer is observed and also families in which a specific syndrome has not been identified. These include the following:

1. familial atypical multiple mole–melanoma (FAMM), with mutations of CDKN2A;
2. Breast–Ovarian Cancer Family syndrome (BRCA1/BRCA2);
3. hereditary pancreatitis;
4. Von Hippel-Lindau syndrome; and
5. isolated pancreatic cancer families.

TABLE VIII Colon Cancer Screening Recommendations for Persons with Familial Risk

Familial risk category	Screening recommendation
Two or more first-degree relatives with colon cancer **or** One first-degree relative with colon cancer or adenomatous polyps diagnosed at age <60 years	Colonoscopy every 5 years beginning at age 40 years or 10 years younger than earliest diagnosis in the family, whichever is earlier; double-contrast barium enema may be substituted but colonoscopy is preferred
Second- or third-degree relatives with colon cancer **or**	Same as for average-risk individual
First-degree relative with colon cancer or adenomatous polyp diagnosed ***after*** age 60 years	Same as for average-risk individual but begin at age 40 years

Hereditary Hemochromatosis

Disease Phenotype

Hereditary hemochromatosis (HH) is characterized by increased iron absorption with resultant tissue iron deposition. Liver cirrhosis along with a risk of hepatic cancer, hepatoma, occurs in advanced cases. Most persons are asymptomatic, with symptoms developing with advanced iron overload. Symptoms correlate to the organ involved and include abdominal pain (hepatomegaly), arthralgias (arthritis-chondrocalcinosis), loss of libido and impotence (pituitary, cirrhosis), cardiac arrhythmias and congestive heart failure (heart), amenorrhea (cirrhosis), and diabetes (pancreas). Nonspecific symptoms include weakness, fatigue, lethargy, apathy, and weight loss.

Findings on physical exam range from no findings or mild hepatomegaly in the asymptomatic patient to a broad range in the symptomatic patient including arthritis and joint swelling, increased skin pigmentation, edema, and other signs of congestive heart failure, testicular atrophy, hypogonadism, or findings of hypothyroidism. Typical stigmata of end-stage liver disease are found, such as hepatosplenomegaly, ascites, encephalopathy, and cutaneous stigmata-like angiomata and loss of hair on the extremities.

Genetics

This is a common autosomal recessively inherited disorder affecting between 1 in 200 and 1 in 400 persons of northern European descent. The disease arises from mutations of the HFE gene; two mutations have been identified, C282Y and H63D. Homozygosity for the C282Y mutation is found in 64–100% of hemochromatosis patients. Compound heterozygotes C282Y/H63D account for a very small percentage of patients. Other mutations yet to be defined are responsible for 10 to 15% of hemochromatosis patients. Genetic testing for these mutations should be performed in any individual with iron overload and family members once appropriate mutations are confirmed.

Pathophysiology

The HFE protein is found in the intestinal crypt cell of the duodenum and is associated with β2-microglobulin and the transferrin receptor. HFE protein facilitates transferrin receptor-dependent iron uptake into crypt cells. Mutant HFEs lose this ability, leading to iron deficiency within the duodenal crypt cells. This results in the increased expression of an iron transport protein called divalent metal ion transporter 1 (DMT-1, also called DCT-a or Nramp2). This is responsible for unregulated dietary iron absorption in the villous cells of the small intestine. Up-regulation of DMT-1 expression has been confirmed in the HFE knockout mouse and in humans with hereditary hemochromatosis.

Diagnosis

Any persons suspected of having hemochromatosis should first have iron studies performed that include serum total iron and iron-binding capacity, transferrin saturation, and ferritin. If transferrin saturation or ferritin is elevated, a liver biopsy should be performed to establish the diagnosis of hemochromatosis by histology using the hepatic iron concentration (HIC) and then calculating the hepatic iron index, which takes into account the progressive increase in HIC with age. Generally, a hepatic iron index >1.9 is diagnostic of hereditary hemochromatosis but heterozygotes can have a lower value (1.5–1.8). Liver biopsy is also helpful to establish the presence of cirrhosis although it is unlikely in those with a serum ferritin level <1000 ng/ml, normal liver enzymes, and no hepatomegaly and in those who are <40 years old.

It is very important to distinguish HH from other conditions of iron overload because the treatment varies. Acquired iron overload can result from ineffective erythropoiesis as in β-thalassemia, sideroblastic anemia, aplastic anemia, pyruvate kinase deficiency, and pyridoxine-responsive anemia. Liver diseases associated with alcohol, hepatitis B and C, nonalcoholic steatohepatitis, porphyria cutanea tarda, or postportocaval shunting all can result in iron overload. Transfusions of red blood cells or parenteral iron either from iron injections or with hemodialysis also may result in iron overload. In addition, miscellaneous conditions can be responsible: iron overload in individuals in sub-Saharan Africa, neonatal iron overload, aceruloplasminemia, and congenital atransferrinemia.

Treatment

When a patient with HH has been found to have iron overload, weekly phlebotomies of 500 ml of whole blood should be carried out (equivalent to 200–250 mg iron). This is continued until the hemoglobin levels do not recover before the next treatment. Monitoring serum transferrin saturation (TS) and ferritin levels every 3 months also is helpful with continued phlebotomy until TS is $<50\%$ and ferritin is <50 ng/ml. Patients diagnosed and treated before cirrhosis have a normal expected survival but not so in those with cirrhosis. In those persons with established cirrhosis, hepatoma screening with liver ultrasound and serum α-fetoprotein should be performed every 6–12 months.

Liver transplantation has been performed but because the abnormal gene is in the enterocyte and

not the hepatocyte, iron reaccumulates in the allograft. Undiagnosed cancer, infection, and coexisting cardiac disease contribute to postoperative morbidity and mortality.

See Also the Following Articles

Cancer, Overview • Colorectal Cancer Screening • Familial Adenomatous Polyposis (FAP) • Genetic Counseling and Testing • Hamartomatous Polyposis Syndromes • Hereditary Hemochromatosis • Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

Further Reading

Bacon, B. R. (2001). Hemochromatosis: Diagnosis and management. *Gastroenterology* **120**, 718–725.

Burt, R., and Jass, J. R. (2000). Hyperplastic polyposis. *In* "World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Digestive System" (S. R. Hamilton, L. A. Aaltonen, eds.), pp. 135–136. IARC Press, Lyon, France.

Burt, R. W. (2000). Colon cancer screening. *Gastroenterology* **119**, 837–853.

Caldas, C., Carneiro, F., Lynch, H. T., *et al.* (1999). Familial gastric cancer: Overview and guidelines for management. *J. Med. Genet.* **36**, 873–880.

Chung, D. C. (2000). The genetic basis of colorectal cancer: Insights into critical pathways of tumorigenesis. *Gastroenterology* **119**, 854–865.

Desai, D. C., Murday, V., Phillips, R. K. S., Neale, K. F., Milla, P., and Hodgson, S. V. (1998). A survey of phenotypic features in juvenile polyposis. *J. Med. Genet.* **35**, 476–481.

Eng, C. (2000). Will the real Cowden syndrome please stand up: Revised diagnostic criteria. *J. Med. Genet.* **37**, 828–830.

Fine, K. D., and Stone, M. J. (1999). Alpha-heavy chain disease, Mediterranean lymphoma, and immunoproliferative small intestinal disease: A review of clinicopathological features, pathogenesis, and differential diagnosis. *Am. J. Gastroenterol.* **94**, 1139–1152.

Friedl, W., *et al.* (2001). Can APC mutation analysis contribute to therapeutic decisions in familial adenomatous polyposis? Experience from 680 FAP families. *Gut* **48**, 515–521.

Giardiello, F. M., Brensinger, J. D., and Petersen, G. M. (2001). AGA Technical Review on hereditary colorectal cancer and genetic testing. *Gastroenterology* **121**, 198–213.

Giardiello, F. M., Bresinger, J. D., Tersmette, A. C., *et al.* (2000). Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* **119**, 656–660.

Guillem, J. G., Smith, A. J., Culle, J., and Ruo, L. (1999). Gastrointestinal polyposis syndromes. *Curr. Prob. Surg.* **36**, 801–818.

Hampel, H., and Peltomaki, P. (2000). Hereditary colorectal cancer: Risk assessment and management. *Clin. Genet.* **58**, 89–97.

Kadmon, M., Tandara, A., and Herfarth, C. (2001). Duodenal adenomatosis in familial adenomatous polyposis coli: A review of the literature and results from the Heidelberg Polyposis Register. *Int. J. Colorectal Dis.* **16**, 63–75.

Kim, H. R., and Kim, Y. J. (1998). Neurofibromatosis of the colon and rectum combined with other manifestations of von Recklinghausen's disease: Report of a case. *Dis. Colon Rectum* **41**, 1187–1192.

Kolodner, R. D., Tytell, J. D., Schmeits, J. L., *et al.* (1999). Germ-line MSH6 mutations in colorectal cancer families. *Cancer Res.* **59**, 5068–5074.

Laken, S. J., *et al.* (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* **17**, 79–83.

Loukola, A., de la Chapelle, A., and Aaltonen, L. A. (1999). Strategies for screening for hereditary non-polyposis colorectal cancer. *J. Med. Genet.* **36**, 819–822.

Lynch, H. T., and Lynch, J. (2000). Lynch syndrome: Genetics, natural history, genetic counseling, and prevention. *J. Clin. Oncol.* **18**, 19S–31S.

Marsh, D. J., Coulon, V., *et al.* (1998). Mutation spectrum and genotype–phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. *Hum. Mol. Genet.* **7**, 507–515.

McGarrity, T. J., Kulin, H. E., and Saino, F. J. (2000). Peutz-Jeghers syndrome. *Am. J. Gastroenterol.* **95**, 596–604.

Murata, I., Yoshikawa, I., *et al.* (2000). Cronkhite-Canada syndrome: Report of two cases. *J. Gastroenterol.* **35**, 706–711.

Nowak, D. A., *et al.* (2001). Lhermitte-Duclos disease (dysplastic gangliocytoma of the cerebellum). *Clin. Neurol. Neurosurg.* **103**, 105–110.

Syngal, S., Fox, E. A., Eng, C., Kolodner, R. D., and Garber, J. E. (2000). Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. *H. Med. Gene* **37**, 641–645.

Vasen, H. F. A. (2000). Clinical diagnosis and management of hereditary colorectal cancer syndromes. *J. Clin. Oncol.* **18**, 81S–92S.

Wirtzfeld, D. A., Petrelli, N. J., *et al.* (2001). Hamartomatous polyposis syndromes: Molecular genetics, neoplastic risk, and surveillance recommendations. *Ann. Surg. Oncol.* **8**, 319–327.

Fast-Track Surgery

DOUGLAS W. WILMORE
Brigham and Women's Hospital and Harvard Medical School

epidural anesthesia Technique in which local anesthetics are placed in the space around the spinal cord and its protective membrane, blocking nerve transmission.
hypothermia Body temperature below normal.
ileus Situation in which the intestinal tract fails to propel food along its surface.
meta-analysis Statistical approach to reach a conclusion using combined data from multiple trials.
regional blocking procedure Method of injecting local anesthesia under the skin, but around the nerves that are supplied to a specific area.
spinal anesthesia Technique in which local anesthetics are placed within the fluid-filled sack around the spinal cord, blocking nerve impulses.

Fast-track surgery is the combined use of a variety of relatively new approaches in operative techniques and perioperative care to enhance the outcome of the elective surgical patient. These methods greatly reduce the stress of an operative procedure, controlling pain and reducing complications. As a result, the operation may be performed as outpatient surgery, or the patient may be required to stay in the hospital for only a day or two.

INTRODUCTION

The new operative and perioperative methods applied in fast-track surgery include epidural or regional anesthesia, minimally invasive operative techniques, optimal pain control, and aggressive postoperative rehabilitation, including early oral nutrition and enforced ambulation. These approaches, when combined, reduce the deleterious effects of the stress of an operation; treating the stress of surgery limits organ dysfunction and the potential for complications and allows shortened convalescence and an earlier recovery. The following measures are generally utilized in most rapid-recovery programs.

OPTIMIZING PREOPERATIVE PHYSIOLOGY FUNCTION AND PATIENT EDUCATION

Before any elective operation, a patient is assessed and comorbid conditions are treated. For example, the patient with diabetes mellitus has blood sugar control maximized and the patient with mild congestive heart failure receives medications (diuretics and cardiotonic drugs) to assure optimal cardiac function. In addition to these and other conditions, it has been shown that if pharmacological means can be used to enforce abstinence in alcohol misusers (even in those without signs of liver impairment), mortality rate will decrease and shortened recovery can be achieved. Prolonged smoking cessation (for 1–2 months) is necessary in the preoperative period to reduce postoperative pulmonary complications.

Classic studies have demonstrated that informed patients require less pain medication in the postoperative period and have shorter hospital stays. Nurses or physician assistants can help the patient understand various aspects of the hospital procedures and may aid by teaching techniques to reduce pain and anxiety following the operation.

ANESTHESIA

Even though a patient receives general anesthesia and is sleeping during surgery, the incision and surgical procedure are perceived by the central nervous system. This results in a variety of reflex responses transmitted via spinal nerves to other organs (liver and intestinal tract, for example), activates the autonomic nervous system (which activates heart rate and increases vascular resistance, for example), and stimulates the hypothalamic–pituitary–adrenal axis (to release glucocorticoids, a major "alarm" hormone). These responses, when combined, are termed the "stress response," and if uncontrolled or exaggerated, they contribute to debility following an operation.

How can the stress response be modified? Local anesthetics and other agents can be placed within the subarachnoid or epidural space (techniques that provide spinal or epidural anesthesia, respectively), thus blocking nerve transmission from the operative site and diminishing the stress response. The effects of blocking nervous impulses from the surgical site are greatest when procedures are performed in the

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Effect of Regional Anesthetic/Analgesic Techniques Compared with General Anesthesia and Systemic Analgesics on Postoperative Morbidity

Complications	Reduction in morbidity
Pulmonary infectious complications	~30%
Respiratory depression	~40%
Pulmonary embolism	~50%
Myocardial infarction	~30%
Other thromboembolic complications	~40%
Ileus	2 days
Blood loss and transfusion requirements	~20–30%
Cerebral complications	No effect
Renal failure	~30%
Other infectious complications (wound, etc.)	No effect

lower body (the lower abdomen, pelvis, and lower extremities), compared with upper abdominal and thoracic operations. This approach improves postoperative pulmonary function, decreases cardiac demands, reduces ileus, and improves pain relief. A recent meta-analysis comparing neural axial blockade (e.g., spinal or epidural anesthesia) with general anesthesia, involving 141 trials including 9559 patients, showed improved morbidity and mortality with the blocking techniques (Table I).

Not all procedures require spinal or epidural anesthesia, and regional blocking procedures have proved quite satisfactory for patients undergoing mastectomy or inguinal hernioraphy.

IN THE OPERATING ROOM

The use of minimally invasive surgical techniques, such as laparoscopic cholecystectomy, has greatly reduced the stress of an operation. When carefully studied, minimally invasive surgical techniques have been found to reduce the inflammatory responses following an operation, although they fail to attenuate early metabolic responses to surgery. Pulmonary function is improved and the postoperative ileus is reduced with the minimally invasive approach. These techniques are now being extended to a variety of procedures in the abdomen and are also used for cardiovascular, thoracic, cerebral, and major orthopedic procedures.

Operating rooms are cold (21°–24°C), and semiclad patients are further disadvantaged by the use of drugs or techniques that inhibit normal responses to cold exposure, such as shivering. This has traditionally resulted in mild hypothermia, which increases the patient's stress response following the procedure. Newer techniques of keeping patients warm in the operating room have resulted in a threefold decrease in the rates of wound infections, a reduction in operative blood loss, a decrease in untoward cardiac events (including ventricular tachycardia), and a reduction in protein loss and patient discomfort.

MODIFYING POSTOPERATIVE CARE

Little evidence supports the routine use of nasogastric tubes in the postoperative period, and, in fact, this approach may be detrimental by increasing the incidence of pneumonia. In addition, there is little scientific basis to the use of drains after cholecystectomy, joint replacements, colon resection, thyroidectomy, and radical hysterectomy.

Oral intake is commonly limited to the postoperative period, but with epidural anesthesia and minimally invasive surgical techniques, ileus is minimized and early feeding is possible. If postoperative nausea and vomiting occur, drugs are available for effective treatment.

Bed rest is frequently recommended following an operation, but with adequate pain control and the use of short-acting anesthetic agents, early ambulation is possible. Exercise also enhances early oral feedings, reduces venous stasis and the potential complications of thrombosis, and stimulates skeletal muscle protein synthesis.

Adequate pain control is essential following an operation and this may require special training of the hospital staff or organization of an acute-care pain service with expertise in the multifaceted aspects of pain control. After minor to moderate operations, patients should receive nonopioid analgesics such as nonsteroidal antiinflammatory agents. This avoids the potential side effects of narcotics, which prolong recovery. More complicated operations are associated with greater pain intensity; using epidural anesthesia for the next 2–3 days reduces the stress of the operation and enhances recovery.

After the second postoperative day, recovery depends on resolution of pain and fatigue. The latter may be related to sleep disturbances that occur in the hospital setting because of noise, drugs, and possibly inflammatory factors. Loss of strength is related to muscle weakness secondary to inactivity and reduced food intake. By attenuating surgical stress and emphasizing early mobility and food intake, this problem may be greatly reduced and convalescence shortened.

TABLE II Recent Data on Fast-Track Surgery from Single-Center Studies

Operation	Hospital stay
Laparoscopic cholecystectomy	Ambulatory procedure
Laparoscopic or vaginal hysterectomy	Ambulatory procedure, 1 day
Laparoscopic gastroesophageal reflux surgery	Ambulatory procedure, 1 day
Elective surgery for aortic aneurysm	3–4 days
Carotid endarterectomy	1–2 days
Mastectomy	Ambulatory procedure, 1 day
Lung lobectomy	1–2 days
Prostatectomy	1–2 days
Partial colectomy	2 days

THE FUTURE

Many medical units are now initiating some or all of the fast-track approaches in the surgical patient. Early results, usually from single-center trials, are encouraging (Table II). When patient satisfaction is studied, it has been high and exceeds that measured with the usual postoperative approach. Rather than emphasizing monitoring and high-technology interventions in the postoperative period, the fast-track approach emphasizes rehabilitative care, with emphasis on pain relief, mobilization, and nutrition.

The trend in the future will be to greatly reduce the length of the hospital stay following operations, because patients recover from their operations sooner. This will be related to less operative stress because of the use of fast-track and other techniques. In addition, morbidity and mortality will decrease because the stress response to a specific surgical procedure is diminished. Further work is needed to understand perioperative pathophysiology, to reduce hospital convalescence, and to improve operative outcome. Physician education is also needed to institute fast-track care.

See Also the Following Articles

Laparoscopy • Minimally Invasive Surgery

Further Reading

Kehlet, H. (1999). Acute pain control and accelerated postoperative surgical recovery. *Surg. Clin. North Am.* **79**, 431–433.

Rogers, A., Walker, N., Schug, S., *et al.* (2000). Reduction of postoperative mortality and morbidity with epidural or spinal anesthesia: Results from overview of randomized trials. *Brit. Med. J.* **321**, 1–12.

Sessler, D. I. (1997). Mild perioperative hypothermia. *N. Engl. J. Med.* **336**, 1730–1737.

Wilmore, D. W, and Kehlet, H. (2001). Management of patients in fast track surgery. *Brit. Med. J.* **322**, 473–476.

Fat Digestion and Absorption

CHARLES M. MANSBACH
The University of Tennessee Health Science Center and Veteran's Affairs Medical Center, Memphis

apolipoprotein B-48 Product of translation of 48% of the total apoB-100 transcript.
chylomicron Lipoprotein uniquely made in the intestine; contains a central core of triglyceride, cholesterol ester, and phospholipid surrounded by a phospholipid coat and apolipoproteins.
enterocytes Mature absorptive cells that line the villi of the small intestine.
exocytosis Budding of membranes in which a selected particle is pushed out into the pericellular space by rupture of the bud.
lipase Enzyme that hydrolyzes glycerol esters of fatty acids.
micelles Association of amphiphilic compounds in which hydrophilic surfaces face outward into the aqueous medium and hydrophobic surfaces face the interior.
phospholipase Enzyme that hydrolyzes phospholipids at specific sites on the molecule.
triglyceride Fat in which fatty acids are esterified to each of the three alcoholic groups.
very low-density lipoprotein Triglyceride-carrying lipoprotein that is like a chylomicron except that it is made both in the intestine and the liver and is smaller.

Absorption of lipid by the intestine is a series of events complicated by the hydrophobic nature of dietary fat. Ingested fat must first be processed to water-interactive lipid products such as fatty acids, which can then be dispersed in the aqueous environment of the intestinal lumen. The lipid products are absorbed into the enterocytes of the small intestine and are resynthesized as triglyceride; if this process does not occur, the fatty acids act as soaps, which can dissolve cellular membranes. A process in which the triglyceride is packaged in chylomicrons enables the fat ultimately to be directed to its final destination, mostly muscle and adipose tissue.

FAT ABSORPTION

Humans can absorb and process large amounts of daily dietary fat in an extremely efficient manner. In the gastrointestinal tract, fat, which is water insoluble, can be made to interact with water, enabling absorption. Once absorbed, hydrolyzed lipids are resynthesized as triacylglycerol and participate in the formation of the chylomicron. The nascent chylomicron is moved from its site of formation in the endoplasmic reticulum to the Golgi for further processing and then is transported from the Golgi to the basolateral membrane of the enterocyte, where the mature chylomicron moves into the lamina propria by reverse exocytosis and into the lymph. The metabolism of chylomicrons is determined by lipoproteins that are present on the chylomicron surface.

FAT DIGESTION

The normal American daily diet contains 100–150 g of fat. This fat is absorbed with 95% efficiency, and even immense amounts of fat, up to 500 g/day, can be absorbed with the same efficiency. It is evident that large variations in fat intake can be handled easily. Although dietary fat has a bad connotation, both because of obesity and because of the relationship between blood cholesterol and the development of atherosclerosis, it should be remembered that fat makes most foods tastier and that some fatty acids are essential for normal body functioning. The so-called essential fatty acids are not synthesized by humans and therefore must be supplied in the diet and appropriately absorbed. Cholesterol, too, is a necessary constituent of biomembranes. In considering the complexities of fat digestion and absorption, it should be kept in mind that the body operates in an aqueous environment but that ingested lipids are most often in a water-insoluble state. If humans did not possess mechanisms for breaking down water-insoluble compounds, there would be no absorption of lipids, which are necessary for their caloric content and for their specific, essential fatty acids.

Gastric Lipase

Fat digestion begins in the stomach. The chief cells of the gastric mucosa produce gastric lipase (GL), the preduodenal lipase of humans. This lipase primarily hydrolyzes fatty acids esterified at the *sn*-3 position on the glycerol backbone of triacylglycerol (TAG). This is especially true if the fatty acid (FA) at the *sn*-3 position is of medium carbon chain length, an important factor in the digestion of milk, the TAG of which

 Copyright 2004, Elsevier (USA). All rights reserved.

is mainly composed of a medium-chain FA at the *sn*-3 position. In neonates, the pancreas is not well developed and pancreatic lipase is secreted in greatly reduced amounts compared to adult levels. As a result, GL is the primary enzyme that hydrolyzes fat in the intestine at this time of life. The major products of lipolysis by gastric lipase are one fatty acid and *sn*-1,2-diacylglycerol (DAG), although the enzyme can hydrolyze both of the primary alcoholic groups of TAG, resulting in *sn*-2-monoacylglycerol (MAG) and two fatty acids. The effect of GL is to provide FAs, which are surface active, and DAG, which dissolves into the interior of lipid droplets and awaits further hydrolysis. The surface-active FAs have the effect of reducing the diameter of the lipid droplets, greatly increasing their surface area with respect to their interaction with water (Fig. 1). This has been shown to increase the rate at which lipid is absorbed in the small intestine. In patients with pancreatic insufficiency, GL provides the majority of lipolytic activity in the intestine. In cases of atrophic gastritis, in which little if any GL should be produced, no defect in overall lipid absorption is found. This points to the lack of requirement of GL for lipid absorption.

Gastric lipase is about the same molecular weight as pancreatic triacylglycerol lipase (PTL) but differs from it in significant ways. The pH optimum for GL is, as would be expected, more acidic, 4–6 versus the pH optimum for PTL of 8.5. GL resists proteolysis by pepsin whereas PTL is susceptible. GL is 15–18% glycosylated whereas PTL is only minimally so. The purified enzyme is very active (1.2 mmol/mg protein/min) against an optimal substrate, tributyrin.

Pancreatic Lipase

Pancreatic triacylglycerol lipase is the single most important determinant of lipid absorption. In its absence, only 30% of an ingested lipid load is absorbed. The enzyme is secreted by the acinar glands of the pancreas into the pancreatic duct and then into the intestine in response to ingestion of a fatty meal. The hormonal signal that initiates the process is cholecystokinin (CCK). CCK originates in enteroendocrine cells in the duodenum. PTL, unlike all other pancreatic hydrolases, does not require activation prior to attaining maximal activity.

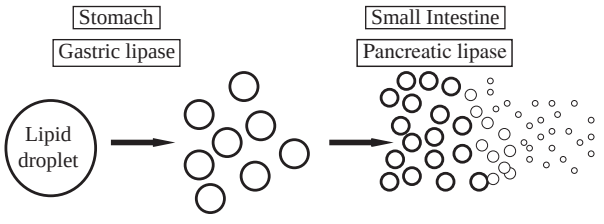


FIGURE 1 Gastric and pancreatic lipases, by producing surface-active fatty acids, progressively increase the surface area of the lipid droplets by decreasing their diameter.

The physiology of PTL has been extensively studied. The enzyme operates only at an oil/water interface; it does not hydrolyze TAG substrate that is soluble in water. An interfacial recognition site on PTL assures that it is correctly aligned with its substrate, TAG, at an oil/water interface. At this point, a hinge mechanism lifts a "lid," exposing a serine hydrolase within the hydrolytic site. Unfortunately, bile salts, which are present in the normal digestive tract after a meal, remove the lipase from the oil/water interface and render the PTL inactive. An associated protein, secreted by the pancreas in an inactive form as pro-colipase, is activated by trypsin, and the active protein, colipase, is able to hold PTL at the oil/water interface even in the presence of bile acids. The colipase activation peptide, called enterostatin, is absorbed intact and has been shown to be a satiety factor. The result of continued lipolysis is to further reduce the size of the lipid droplets, further increasing the surface area for additional lipolytic activity (Fig. 1).

PTL hydrolyzes TAG at primary ester groups, resulting in the release of two FAs and MAG. If the MAG isomerizes to the *sn*-1,3 position, then the MAG isomer becomes a target for further lipolysis, with the final products being three FAs and glycerol. The rate of isomerization is slow, however. These products are constantly being swept away from the oil/water interface, so there is no product inhibition of lipolytic activity. Against the ideal substrate tributyrin, the purified enzyme is even more active (4.5 mmol/mg protein/min) than GL.

Pancreatic Phospholipase A_2

Pancreatic phospholipase A_2 (PLA_2) is secreted by the pancreas from the pancreatic acinar glands in response to the hormonal signal, CCK. It is secreted as a proenzyme that must be activated by trypsin prior to achieving maximal activity. The proenzyme can hydrolyze its phospholipid substrates at a reduced rate but cannot recognize and be activated by an oil/water interface, as can the active enzyme. The specific hydrolytic site of the enzyme is the FA esterified at the *sn*-2 position, resulting in release of an *sn*-1-lysophospholipid and one FA, usually an unsaturated one because that is the commonest FA esterified at that location. The resultant products can be absorbed easily; unhydrolyzed phospholipid is not absorbed.

The effect of PLA_2 on lipid absorption is to remove phospholipids from the surface of fat emulsions, enabling PTL to more quickly penetrate the surface of the lipid droplet and begin hydrolysis. In the presence of phospholipids, PTL does not hydrolyze the neutral TAG until some of the phospholipid has been hydrolyzed, the so-called lag time. After the phospholipid is partially removed from the emulsion surface, PTL activity increases exponentially.

It is interesting to consider why PLA_2 does not hydrolyze the apical membrane phospholipids that it comes into contact with after its activation. Venom phospholipases A_2 are readily able to attack membrane phospholipids. Snake venom PLA_2 injected into the paw of a rabbit causes swelling and inflammation, whereas there is little effect if pancreatic PLA_2 is injected. Similarly, red cells are rapidly lysed when exposed to snake venom PLA_2 but not following exposure to pancreatic PLA_2. The reason for this is the relative abilities of the two PLA_2s to penetrate the membrane surface layer of phospholipids. The lateral surface pressure of the phospholipids in membranes is on the order of 30 dynes/cm^2, a pressure that venom PLA_2 can easily penetrate. This is far above the ability of pancreatic PLA_2, which has a maximal activity at 12 dynes/cm^2.

Cholesterol Esterase

Cholesterol esterase (CE) (bile salt-stimulated esterase or carboxyl ester lipase) primarily hydrolyzes cholesterol esters. It is synthesized in the pancreas and is released in response to CCK in a fully active form. Studies from CE knockout mice show that this enzyme has no effect on free cholesterol absorption but does dramatically lower cholesterol ester absorption, pointing to its key role in hydrolyzing cholesterol ester to free cholesterol and FA. Its nonspecific hydrolyzing ability means that it can hydrolyze TAG as well as MAG. It would appear that its effect on TAG is not clinically significant, whereas the hydrolysis of MAG would reduce the efficiency of the reesterification process to TAG that occurs in the enterocyte (see later).

INTRALUMINAL PROCESSES

After the emulsion lipid droplets are broken down into small-sized, stable emulsion particles, pancreatic lipase, in conjunction with colipase and PLA_2, attack their substrates at the oil/water interface. The released hydrolytic products are removed from the surface by their solubilization in bile salt micelles or in liquid crystals of MAG and FA (Fig. 2). Bile salts, by their self-associating qualities above a certain concentration, i.e., the critical micellar concentration, form a micelle, arranged such that the hydrophobic sides of the bile salts face each other and the hydrophilic sides face the aqueous solution. This forms a hydrophobic interior into which FA and MAG acyl chains and cholesterol penetrate, forming a bile salt mixed micelle. These micelles diffuse toward the apical surface of the enterocytes. FAs and MAG have a finite water solubility and these monomers diffuse as well toward the apical surface. Because diffusion rates are inversely proportional to the square of the radius of the diffusing moiety, the micelles diffuse much more slowly than the monomer species. However, the concentration of FAs and MAG in the micelles is so great that it more than makes up for the slower diffusion rate in terms of the ability of the micelles to deliver lipids efficiently to the apical enterocyte surface (Fig. 2). These micelles, the monomers, and the liquid crystals diffuse across the unstirred layer next to the apical surface of the intestine. The enterocytes secrete H^+, making the juxtaapical surface more acidic compared to the bulk luminal fluid, the so-called acidic microclimate. The effect of the reduction in juxtaapical pH to ≈ 6.0 is to partially protonize the FA, which has an effective pK_a of 4.5. The protonated FA then can passively traverse the apical membrane into the enterocyte. Once in the enterocyte, the more neutral pH causes the FA to again become ionized, effectively trapping the FA inside the cell. FA absorption may also be mediated in part by membrane fatty acid transporters. Several have been described: A plasma membrane-associated fatty acid binding protein (FABPpm) is present in the human colon carcinoma cell line, Caco-2 cells, which at confluence have many of the characteristics of intestinal absorptive cells. Antibodies to this transporter do not inhibit the uptake of FA. A fatty acid transporter protein (FATP) but not FAT/CD36 is expressed in Caco-2 cells. Treatment of these cells with trypsin to hydrolyze surface potential FA transporters reduces the uptake of FA into Caco-2 cells by $\approx 30\%$, suggesting that at least some FAs are absorbed by a FA transporter, presumably

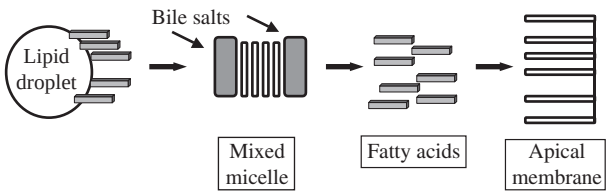


FIGURE 2 Lipase action produces fatty acids, which combine with bile salts to form mixed micelles. The micelles diffuse toward the cell surface; the fatty acids solubilize in the aqueous fluid and penetrate the apical surface membrane.

FATP. The role of the micelles and liquid crystals is to act as reservoirs for monomer FAs and MAG so that the monomer concentration of both lipids remains maximal in the unstirred layer.

RESYNTHESIS OF TRIACYLGLYCERIDES IN ENTEROCYTES

Interaction of Absorbed Split Lipids with Fatty Acid Binding Proteins

FAs and MAGs that desorb from the cytosolic face of the apical membrane are in part solubilized in the cytosol in monomer form. In addition, the intestine expresses two lipid-binding proteins, the intestine and liver fatty acid binding proteins (I-FABP and L-FABP). These proteins comprise a large amount of the total proteins in the cytosol, up to 2% each. Each I-FABP can bind only one FA, whereas each L-FABP can bind two FAs as well as MAG. The role of the FABPs in promoting TAG resynthesis is not clear, but it has been shown that FAs bound to FABP move more quickly in the cytosol than do FAs alone. It is postulated that L-FABP acts as a reservoir for intracellularly bound FA and MAG, delivering its cargo to the site of TAG resynthesis by diffusion across the apical membrane–endoplasmic reticulum (ER) space. By contrast, I-FABP delivers its FA cargo by actual contact with the ER membrane.

Synthesis of TAG in the ER

The products of intestinal luminal lipolysis are resynthesized to TAGs, phospholipids, and cholesterol esters by the ER of the enterocyte. The enterocyte, which has no effective control over the entry rate of FAs, performs this function very quickly or risks having toxic concentrations of FAs present in the cytosol. FA is a soap and as such greatly disturbs cellular membranes or solubilizes them completely, depending on the FA concentrations. The two ways that the enterocyte has to dispose of FAs is by binding them to FABP or by reforming TAGs, which are essentially physicochemically inert. The enterocyte can perform the resynthetic function very quickly. In 30 seconds, large amounts of FAs, of which 79% is already reesterified to TAGs, can be absorbed from the lumen of the intestine. Multiple enzymes perform this function. First the FA must be activated to form FA–coenzyme A (CoA) by the enzyme FA–CoA ligase. This enzyme is located on the cytosolic hemileaflet of the ER membrane. The activated FA can then be used by a series of two acyltransferases to form TAG. First DAG is formed from MAG by MAG acyltransferase (MGAT) and subsequently TAG is formed from the DAG by DAG acyltransferase (DGAT). If MAG is not available, then TAG can be resynthesized by the so-called *de novo*, or Kennedy, pathway. This begins with *sn*-3-glycerol phosphate, which is acylated by glycerol phosphate acyltransferase to lysophosphatidic acid (LPA). LPA, in turn, is acylated by LPA acyltransferase to phosphatidic acid (PA). PA phosphatase, the limiting enzyme on this pathway, then hydrolyzes the PA to DAG. This DAG, in contrast to the DAG synthesized from MAG, can be utilized for phospholipid synthesis. The DAG can then be converted to TAG by DGAT. In this reaction, another enzyme appears to be important, acyl-CoA acyltransferase (AAT). The role of AAT is not yet clear, but its removal from the TAG synthetic complex stops TAG synthesis, at least in liver.

Movement of TAG to the ER Lumen

It is not certain on what side of the ER membrane the lipid resynthesis occurs. Some resynthesis clearly occurs on the cytosolic face of the ER membrane. The TAG thus produced must cross the ER membrane to gain access to the ER lumen, where the forming chylomicron is situated. There is a finite solubility of TAG in membranes of 3%, thus, depending on the rapidity of TAG movement across the membrane; a variable proportion of cytosolically synthesized TAG could be made available for chylomicron formation. Alternatively, the TAG could be formed on the lumenal side of the ER membrane, which would do away with the requirement for TAG to traverse the ER membrane. If DGAT and possibly MGAT are located on the luminal side of the ER membrane, however, they must have access to their substrate FA-CoA. FA-CoA cannot cross the ER membrane and CoA concentrations in the ER lumen are very low. Therefore, a mechanism must exist for the FA-CoA to cross the ER lumen. Current data indicate that a majority of the TAG that is synthesized by the enterocyte for chylomicron formation is synthesized by DGAT on the luminal surface of the ER membrane.

CHYLOMICRON FORMATION AND MOVEMENT OF MATURE PARTICLES

Formation of the Chylomicron

Chylomicrons are formed in the ER lumen by a two-step process. In the first step, apolipoprotein B-48 (apoB-48), which is only 48% of the translated full-length apoB transcript, is pulled across the ER

membrane through its translocon by the microsomal triacylglycerol transport protein (MTP). MTP, acting as a chaperone, is made up of a large (97 kDa) and a small (55 kDa) component; the smaller component is protein disulfide isomerase (PDI). PDI is utilized for proper folding of proteins because its function is to link together intramolecular cysteines. ApoB-48, once inside the ER lumen, combines with phospholipid, mainly phosphatidylcholine (PC), and some TAG to form a dense, small, primordial chylomicron. The majority of the TAG that will eventually enter the chylomicron forms a lipid aggregate, mediated by MTP, more distally in the ER, mainly the smooth ER. How the dense chylomicron becomes lipidated by the TAG is unclear. However, the chylomicron is formed and results in a very large particle that is 100−500 nm in diameter. The forming chylomicron is composed of a TAG and CE core with a surrounding coat of phospholipid, mainly PC, cholesterol, apoB-48, apolipoprotein A-IV, and some C lipoproteins. This "mature" ER chylomicron is now ready to be transported to the Golgi (Fig. 3).

Movement of Chylomicrons from the ER to the Golgi

The movement of chylomicrons out of the ER has been postulated to be the rate-limiting step in the transit of TAG across the enterocyte. Consistent with this hypothesis, when the load of dietary TAG is increased, the amount of TAG remaining in the ER increases in direct proportion. In fact, if the dietary load is very great, at least in rats, some TAG never makes it across the ER membrane but remains on the cytosolic surface of the ER.

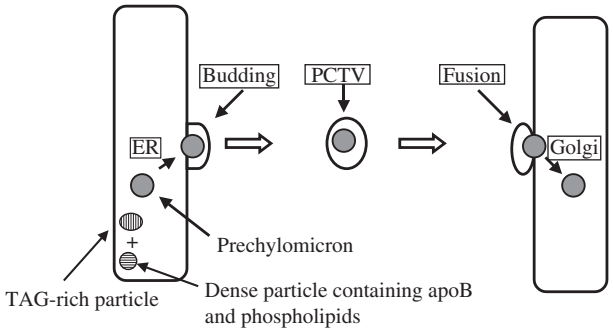


FIGURE 3 Prechylomicrons are cargo in the prechylomicron transport vesicle (PCTV), which buds from the endoplasmic reticulum (ER) membrane, traverses the intracellular space, and fuses with the Golgi, delivering its cargo to the Golgi lumen. TAG, Triacylglycerol; apoB, apolipoprotein B.

The movement of chylomicrons out of the ER is just beginning to be understood. It appears certain that the chylomicrons move from the ER to the Golgi in a prechylomicron transport vesicle (PCTV) (Fig. 3). The mechanism by which the prechylomicron is selected as cargo in the vesicle is not known. What is known is that the group of proteins known as the coatomer II proteins, or COPII proteins, which are crucial to the export of proteins from the ER to the Golgi, are not involved with the budding of the PCTV from the surface of the ER (Fig. 3). Indeed, inhibition of the protein vesicles causes an increase in PCTV formation by 6- to 10-fold. The actual proteins involved in budding from the ER are not known, but the COPII proteins appear to be present on the PCTV, not for the purpose of budding but rather to enable the PCTV to acquire the proteins requisite for targeting the PCTV to the cis-Golgi, tethering it to the Golgi, docking, and then fusion of the PCTV with the Golgi membrane. In this way, the very large chylomicron can be delivered across the Golgi membrane for the additional processing that takes place in the Golgi (Fig. 3). This process appears to be unique to the intestine in that PCTV cannot deliver its cargo to liver Golgi. In addition, the vesicle is sealed, which precludes entry of the cytosolic lipase (PTL) present in the cytosol of the intestine. This lipase, which is pancreatic PTL but expressed in the intestine, would hydrolyze the chylomicron TAG if it had access to the TAG present in chylomicrons. The proteins involved in targeting the PCTV to the Golgi and involved with its fusion with the cis-Golgi appear to be the typical fusion machinery used for protein vesicles, but the details are not yet worked out. In sum, the effect of this delivery mechanism is to move the very large prechylomicron out of the ER, across the cytosolic space, direct it to the Golgi, and to get it across the Golgi membrane to the Golgi lumen.

Apolipoprotein A-I (apoA-I), a constituent of the mature chylomicron particle, is not found in the PCTV but is found in both the ER and Golgi. It is well established that apoA-I secretion can be divorced from apoB-48 secretion as well as TAG secretion. Therefore, the lack of apoA-I in the PCTV offers a mechanism whereby apoA-I could get from the ER to the Golgi by an independent mechanism, presumably via the protein vesicles.

Movement of Chylomicrons from the Golgi to the Lymph

Once inside the Golgi lumen, the chylomicrons are destined to go into the lymph. The mechanism by which the prechylomicrons move across the Golgi stack to the

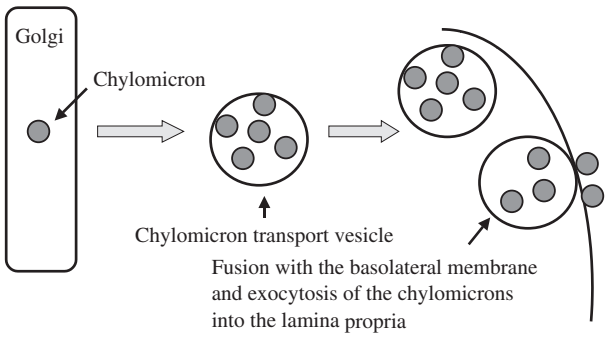


FIGURE 4 Chylomicron transport vesicles bud off the trans-Golgi network and go to the basolateral membrane, where they fuse and deliver their chylomicron cargo to the lamina propria.

trans-Golgi network is unknown. Also not clear is how the chylomicron is budded off the Golgi and enters a vesicle that transports several chylomicrons at a time from the Golgi to the basolateral membrane (Fig. 4). Once at the basolateral membrane (BLM), the vesicle fuses with the BLM and the chylomicrons are released into the intercellular space by reverse exocytosis. Once in the lamina propria, the rapidity of movement of the chylomicrons into the lymph is governed by the extent of tissue hydration. Under conditions in which hydration is poor, the chylomicron is slowed. By contrast, when the tissue is well hydrated, chylomicron movement is enhanced.

Effect of Phosphatidylcholine on Chylomicron Export

For many years it has been known that PC availability governs the rate at which chylomicrons are delivered to the lymph. For example, in rats infused intraduodenally with triolein (135 μmol/hour), the output rate varies from 50% in bile duct-intact rats to 85% of the input rate in rats in which PC was included in the triolein infusion. The output rate is even less in rats with a bile fistula; in these rats, if choline is infused into the duodenal lumen along with triolein, there is a modest increase in chylomicron output into the lymph. However, if lyso-PC is infused, then chylomicron output is restored to control (bile duct-intact) levels. These data suggest that *de novo* synthesis of PC cannot keep up with demand, but that when lyso-PC is provided, adequate amounts of PC for chylomicron export can be synthesized. At what step PC dependency occurs is not clear. It could be the utilization for PC for chylomicron membrane formation or it could be PC utilization for PCTV membranes. Bile PC appears to be favored over other forms of PC.

INTESTINAL PRODUCTION OF VLDL

The intestine also produces very low-density lipoproteins (VLDLs), which are operationally of greater density than chylomicrons. The vast majority of TAG that is exported from the intestine is as chylomicrons vs. VLDLs. If the dietary load of TAG is increased, the amount of TAG appearing in the chylomicrons and in the VLDLs progressively increases until a point is reached in which the amount in the VLDLs does not increase further, although additional output in chylomicrons continues. At least at low levels of dietary intake, the saturated FAs (palmitate) tend to be transported out of the intestine in VLDLs, whereas the unsaturated ones (oleate) appear in chylomicrons. Even under fasting conditions, TAG continues to be secreted by the intestine as both VLDLs and chylomicrons. Most of this TAG comes from acyl groups derived from the bile PC. In rats with a bile fistula, hardly any TAG is exported either in VLDLs or chylomicrons.

DIFFERENCES IN TRIGLYCERIDE TRANSPORT

The ileum is perfectly capable of absorbing lipid and resynthesizing TAG from it. The question is how it is transported from the gut. Interestingly, for reasons that are not clear, in the rat, lipid delivered to the distal intestine appears to be poorly transported by chylomicrons. Even if the distal intestine is perfused with lipid for a week to be certain that the lack of delivery of dietary lipid to the distal intestine as a causation of this effect is overcome, the lipid is still not well transported in chylomicrons. However, if the ileum is transposed to the proximal intestine and lipid is delivered to it, then the TAG is exported in chylomicrons. The reasons for these effects are not known nor is it known if these rat data are transferable to humans.

TWO POOLS OF INTESTINAL TRIGLYCERIDE

The TAG that is synthesized by the enterocytes can distribute into two different pools. The first pool, the prechylomicron pool, is composed mainly of dietary FAs and of TAGs synthesized predominantly from dietary MAGs. The result is a pool of TAG that closely reflects the acyl constituents of the diet. The data predict that chylomicrons, which utilize the TAGs from this pool, would consist of TAGs that mimic the diet. When tested, this prediction held true. It is also predicted that

this TAG would be sequestered in the lumen of either the ER and/or the Golgi. This prediction has also proved true.

The TAG in the second pool, the storage pool, has a completely different fate. Unlike the liver, in which TAG enters a storage pool prior to its secretion in very low-density lipoproteins (the major TAG transport vehicle of the liver), the intestine appears to transport absorbed dietary fat directly into chylomicrons, without passing through a storage pool. The storage pool acyl groups come predominantly from endogenous sources such as circulating FAs. If the circulating FAs are radiolabeled and traced to the intestine, it is found that at radiolabel steady state, most of the labeled FA is in TAG and that the mucosal TAG has a much higher radiolabel-specific activity compared to TAG that is being secreted into the lymph. In fact, the lymph TAG has a lower specific activity than does the mucosal TAG, the opposite of what occurs when dietary TAG is radiolabeled. These studies demonstrate that the TAG in the storage pool is not selected for export into the lymph. Studies show, however, that the TAG in this pool does leave the intestine. Because the TAG does not leave in the lymph, and it has been shown that endogenous TAG and especially FA appear in the portal vein, it is very likely that the TAG in this pool leaves via the portal vein, either as TAG or, if hydrolyzed by the intestinal lipase, as FA.

The two TAG pools likely split at the level of the ER. The rationale for this is that TAG within the ER lumen has a radiolabel-specific activity that is similar to that found in chylomicrons and approaches that of the dietary TAG. By contrast, the TAG on the cytosolic face of the ER has a very low radiolabel-specific activity when dietary TAG is radiolabeled. Because these two pools appear to be separated only by the ER membrane, it is clear that this is where the division between the two pools occurs.

DELIVERY OF FAT TO THE PORTAL VEIN SYSTEM

It has been established that 10–15% of absorbed FA enters the portal vein during normal lipid absorption, with the unsaturated FAs being favored. When studied in a more quantitative way, however, under conditions of large intake loads, the portal vein can transport up to 39% of the absorbed lipid as FA and some as TAG. Where do these FAs and TAGs come from? One potential is that these lipids are the ones that never cross the ER membrane but are sequestered on the cytosolic face of the ER. If ER in which this distribution of lipid occurs is incubated with cytosol containing lipase, the TAG on the outside of the ER is rapidly hydrolyzed, whereas the TAG in the ER lumen is not touched. The amount of FA and TAG delivered to the portal vein is inversely proportional to the amount of lipid appearing in the lymph. For example, in rats receiving PC in an intraduodenal triolein infusion, almost all the TAG infused comes out into the lymph. In this case, very little appears in the portal vein. Similarly, if the input load of triolein is greatly reduced, again, very little dietary lipid appears in the portal vein. It is not clear in what particle the dietary portal vein TAG is in. In snakes that lack a lymphatic system, portomicrons are generated. A similar situation may exist in humans, but this has not yet been demonstrated.

PHYSIOLOGICAL CONSEQUENCES OF LYMPHATIC/PORTAL VEIN DELIVERY

TAG that is synthesized in the intestine can be exported in either the portal vein or the lymph. Two factors that are known to influence this process are PC and dietary load, as discussed previously. The physiologically important question is the consequences of the TAG exiting the intestine by either path. If the TAGs exit in chylomicrons, their metabolism in the periphery, mostly by adipose tissue and muscle, results in a much smaller particle, the chylomicron remnant. Because of the greatly reduced surface area of the remnant versus the original chylomicron particle, the surface lipids bud off and become PC-rich disks containing predominantly apoA-I. When these disks become loaded with cholesterol, they become high-density lipoprotein (HDL) particles and initiate reverse cholesterol transport. The importance of this process to overall circulating HDL levels has recently become clearer. Mice who are mouse apoB knockouts have a lethal mutation. However, if they are made transgenic for the human apoB gene, they survive. The apoB gene is expressed only in the liver, however, and the intestine has no apoB. In this case, the mice do not mount a chylomicronemia after eating a lipid-loaded diet. These mice have only 36% of the levels of circulating HDL of normal mice. It is unclear if the same data would hold true in humans but it is clear that HDL does form after chylomicron metabolism. By contrast, if dietary lipids enter the portal vein, they go to the liver. Once in the liver, the lipids can either be stored or oxidized. If the lipids are stored, they eventually exit the hepatocyte in the liver's TAG-rich lipoprotein, VLDL. This lipoprotein circulates and, on its metabolism, it first becomes intermediate-density lipoprotein (IDL) and then low-density lipoprotein (LDL). LDL is the predominant cholesterol transporter and the recognized

lipoprotein particle associated with the development of atherosclerosis. Thus, forcing TAG to be exported in the lymph is likely to be beneficial to humans with respect to atherosclerosis.

See Also the Following Articles

Barrier Function in Lipid Absorption • Carbohydrate Digestion and Absorption • Cholesterol Absorption • Digestion, Overview • Pancreatic Enzyme Secretion (Physiology) • Pancreatic Triglyceride Lipase • Protein Digestion and Absorption of Amino Acids and Peptides • Small Intestine, Absorption and Secretion

Further Reading

Carey, M. C., and Hernell, O. (1992). Digestion and absorption of fat. *Semin. Gastrointest. Dis.* **3**, 189–208.

Carriere, F., Rogalska, E., Cudrey, C., *et al.* (1997). *In vivo* and *in vitro* studies on the stereoselective hydrolysis of tri- and diglycerides by gastric and pancreatic lipases. *Bioorg. Med. Chem.* **429**, 429–435.

Ho, S. Y., and Storch, J. (2001). Common mechanisms of monoacylglycerol and fatty acid uptake by human intestinal Caco-2 cells. *Am. J. Physiol.* **281**, C1106–C1117.

Hsu, K.-T., and Storch, J. (1996). Fatty acid transfer from liver and intestinal fatty acid-binding proteins to membranes occurs by different mechanisms. *J. Biol. Chem.* **271**, 13317–13323.

Kumar, N. S., and Mansbach II, C. M. (1997). Determinants of triacylglycerol transport from the endoplasmic reticulum to the Golgi in intestine. *Am. J. Physiol.* **273**, G18–G30.

Kumar, N. S., and Mansbach II, C. M. (1999). Prechylomicron transport vesicle: Isolation and characterization. *Am. J. Physiol.* **276**, G378–G86.

Lehner, R., and Kuksis, A. (1996). Biosynthesis of triacylglycerols. *Prog. Lipid Res.* **35**, 169–201.

Lowe, M. E. (1997). Structure and function of pancreatic lipase and co-lipase (review). *Annu. Rev. Nutr.* **17**, 141–158.

Mahan, J. T., Heda, G.D., Rao, R. H., and Mansbach II, C. M. (2001). The intestine expresses pancreatic triacylglycerol lipase: Regulation by dietary lipid. *Am. J. Physiol.* **280**, G1187–G1196.

Mansbach II, C. M., and Dowell, R. (2000). The effect of increasing lipid loads on the ability of the endoplasmic reticulum to transport lipid to the Golgi. *J. Lipid Res.* **41**, 605–612.

Mansbach II, C. M., and Nevin, P. (1998). Intracellular movement of triacylglycerols in the intestine. *J. Lipid Res.* **39**, 963–968.

Siddiqi, S. A., and Mansbach II, C. M. (2001). Pre-chylomicron transport vesicle (PCTV) budding from the ER is COPII machinery independent. *Gastroenterology* **120**(Suppl. 1), A1-765–AA1-88.

Tsuchiya, T., Kalogeris, T. J., and Tso, P. (1996). Ileal transportation into the upper jejunum affects lipid and bile salt absorption in rats. *Am. J. Physiol.* **271**, G681–G691.

Washington, L., Cook, G., and Mansbach, C. M. (2003). Inhibition of carnitine palmitoyltransferase in the rat small intestine reduces export of triacylglycerol into the lymph. *J. Lipid Res.* in press.

Yang, Y., and Lowe, M. E. (2000). The open lid mediates pancreatic lipase function. *J. Lipid Res.* **41**, 48–57.

Young, S. G., Cham, C. M., Pitas, R. E., *et al.* (1995). A genetic model for absent chylomicron formation: Mice producing apolipoprotein B in the liver, but not the intestine. *J. Clin. Invest.* **96**, 2932–2946.

Fecal Incontinence

WILLIAM E. WHITEHEAD
University of North Carolina at Chapel Hill

gluteal flap transposition When the external anal sphincter is damaged or denervated, an innervated and vascularized segment of the gluteus maximus muscle may be wrapped around the anal canal to substitute for the external anal sphincter.

gracilis muscle Located on the inner aspect of the thigh; may be cut at its distal end and transposed around the anal canal to substitute for a damaged or denervated external anal sphincter. The transposed muscle is often electrically stimulated to maintain a state of contraction.

levator ani muscle Part of the pelvic floor situated anterior to the anal canal; it can be voluntarily contracted to pinch off the rectum from the anal canal.

puborectalis muscle Sling muscle that forms part of the pelvic floor; it loops around the posterior aspect of the rectum and anchors anteriorly to the symphysis pubis. It maintains an angle between the rectum and anal canal and can be voluntarily contracted to further pinch off the rectum from the anal canal.

pudendal nerve motor latency Test for the integrity of the pudendal nerve; involves electrically stimulating the nerve with a finger-mounted electrode and measuring the time elapsed before the external anal sphincter contracts, as detected by electromyographic activity.

sampling reflex Internal and external anal sphincters spontaneously relax for brief periods, allowing the contents of the rectum to be exposed to the sensory receptors in the upper anal canal.

Fecal incontinence is defined as recurring, unintentional loss of fecal material in an individual with a developmental age of at least 4 years. Continence depends on rectal reservoir capacity, sensation of rectal filling, and sufficient strength in the striated pelvic floor muscles. Rectal or pelvic inflammation or injury and diseases that impair nerve conduction may cause fecal incontinence.

INTRODUCTION

In defining fecal incontinence, some clinicians believe that involuntary passage of flatus should be included, because this significantly impairs quality of life. However, involuntary passage of flatus may occur normally up to six times per day, making it difficult to distinguish health from disease. Therefore, patients whose only complaint is uncontrollable flatus should be distinguished from patients who involuntarily lose liquid or formed stool. The usual amount of incontinence should be noted: small-volume incontinence (less than two teaspoons) or only staining of underwear is five to seven times more common than large-volume incontinence and is likely to have a different etiology (e.g., hemorrhoids or irritable bowel syndrome versus sphincter muscle injury or pudendal nerve injury).

Several scales have been developed to measure the severity of fecal incontinence. Most of these scales ask patients to rate the frequency of gas, liquid, and solid stool incontinence; a summary score is calculated based on the frequency of gas incontinence plus the frequency of liquid incontinence plus the frequency of solid incontinence. Some severity scales also include questions on use of pads and impact on quality of life.

EPIDEMIOLOGY

The prevalence of fecal incontinence in community-dwelling people is 2.2–6.9%, but in nursing homes, 45–47% of residents are fecally incontinent. The total prevalence of fecal incontinence, when community-dwelling and institutionalized individuals are combined, is approximately 15% of the adult population. When surveys distinguish minor (small volume) from major (larger volume) fecal incontinence, minor incontinence is more common (6% minor incontinence vs. 1% major incontinence in one large survey). The etiologies for minor and major incontinence are believed to differ: approximately 50% of patients with hemorrhoids report minor incontinence and about 20% of patients with irritable bowel syndrome report occasional, minor incontinence. Major incontinence is more likely to involve obstetrical trauma, pudendal nerve injury, diabetes, dementia, or constipation with overflow.

The prevalence of fecal incontinence increases with age in both genders. This is partly explained by the association of incontinence with dementia and mobility impairment in older people. As a consequence, 97%

 Copyright 2004, Elsevier (USA). All rights reserved.

of patients with fecal incontinence in nursing homes also experience urinary incontinence whereas isolated fecal incontinence is more common in younger and community-living people.

Obstetrical trauma resulting in injuries to the external anal sphincter and/or its innervation during vaginal childbirth is a well-recognized etiology for fecal incontinence, and gynecologists report that fecal incontinence is eight times more prevalent in women than in men. However, surveys of representative community samples show approximately equal representation in men and women.

PATHOPHYSIOLOGY

Continence is preserved by a learned, voluntary behavior; in response to the perception of fecal material entering the rectum, voluntarily contraction of the external anal sphincter, puborectalis muscles, and perhaps other muscles of the pelvic floor is maintained for long enough to allow the rectum to relax to accommodate an increased volume. Thus, continence depends on three factors: the reservoir capacity of the rectum, sensation for rectal filling, and sufficient strength in the striated pelvic floor muscles to postpone stool passage. These functions and the muscles on which they depend are depicted in Fig. 1.

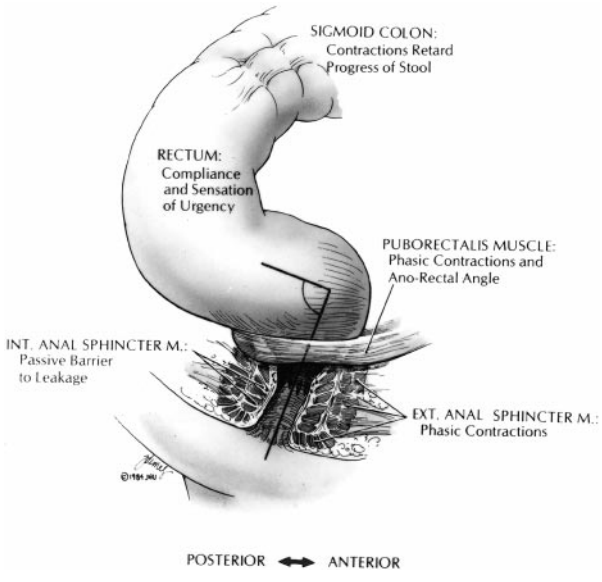


FIGURE 1 Continence depends on three factors: the reservoir capacity of the rectum, sensation for rectal filling, and sufficient strength in the striated pelvic floor muscles to postpone stool passage.

Rectal Compliance

As fecal material enters the rectum, the rectum is able to relax to accommodate increased volume, and there is little or no increase in intrarectal pressure. This accommodation continues up to a critical threshold at which the rectum reflexly contracts. Voluntary effort may also precipitate rectal contraction, but this is poorly studied. Noncompliance of the rectum due to inflammation (e.g., ulcerative proctitis) or scarring (e.g., radiation injury) is associated with an increased risk of incontinence.

Rectal Sensation

Relatively small distensions of the rectum are consciously perceived, allowing a choice to be made for voluntary contraction of the pelvic floor muscles and appropriate timing of defecation. Deficits in rectal sensation, which often occur in association with diabetes mellitus or spinal cord injury, may result in incontinence.

Anal Sensation and Sampling Reflex

The internal and external anal sphincters spontaneously relax for brief periods (sampling reflex), allowing the contents of the rectum to enter the upper anal canal. Afferent nerve endings there may contribute to continence by allowing an individual to discriminate whether the rectum contains gas, liquid, or formed stool. Anal sensation is less well studied than rectal sensation.

Pelvic Floor Muscles

The pelvic floor is composed of multiple muscles, including the puborectalis and levator ani muscles. The puborectalis is a sling muscle that anchors anteriorly to the symphysis pubis and loops around the rectum. Normal resting tone is responsible for preserving the approximately 90° angle between the rectum and anal canal. The puborectalis can be voluntarily contracted to further pinch off the rectum from the anal canal and prevent accidental passage of formed stool. The levator ani muscles can also be voluntarily contracted to pinch off the rectum from the anal canal. Injuries to the innervation of the puborectalis and levator ani are less common during childbirth than are injuries to the nerves supplying the external anal canal, and for this reason, surgeons often plicate these muscles to improve continence following obstetrical injuries to the external anal sphincter.

External Anal Sphincter

The external anal sphincter is a striated muscle surrounding the anal canal; it has no bony attachments and functions like a purse-string to close off the anal canal with voluntary effort. It is believed to be more effective than the puborectalis for preventing liquid and gas incontinence. Injuries to the external anal sphincter or its innervation are commonly associated with childbirth and increase the risk of incontinence.

Internal Anal Sphincter

The internal anal sphincter is a smooth muscle sphincter that normally stays closed to maintain anal canal pressure higher than rectal pressure. It is not under voluntary control and serves as a passive barrier to leakage of liquid and gas. The most common injury to the anal canal is an obstetrical tear, but decreases in internal anal sphincter tone may also occur for unknown reasons and may compromise continence.

DIAGNOSTIC ASSESSMENT

Medical History

A patient's medical history should include responses to questions about (1) typical bowel habits (diarrhea or constipation), (2) surgical and obstetrical history, (3) other medical conditions such as diabetes mellitus, inflammatory bowel disease, hemorrhoids, or irritable bowel syndrome, (4) volume and type of stool lost, and (5) circumstances in which incontinence occurs.

Necessary Tests

Anal canal ultrasound is the gold standard for assessing the structural integrity of the internal and external anal sphincters. In addition, anorectal manometry should be used to assess anal canal squeeze pressure, anal canal resting pressure, rectal sensation, and compliance of the rectum.

Optional Tests

Pudendal nerve motor latencies are a measure of the time elapsed between electrical stimulation of the pudendal nerve and contraction of the external anal sphincter. This is a relatively specific method of detecting pudendal nerve injuries, but it lacks sensitivity. Needle electromyographic (EMG) activity can detect subtle nerve injuries that may be missed by functional tests such as anorectal manometry. Surface EMG activity detected by perianal or intraanal electrodes may be useful for detecting nerve injuries and for biofeedback training.

TREATMENT

Medical Treatment

Antidiarrheal agents such as loperamide are used when incontinence is caused or exacerbated by diarrhea. Laxatives may be used in patients with overflow incontinence secondary to constipation with fecal impaction.

Biofeedback

This is a behavioral treatment in which patients are taught to improve anal canal squeeze pressures and/or rectal sensory thresholds with the help of electronic devices that amplify and display small contractions of pelvic floor muscles. This is often considered the second line of treatment (after medical therapy) because it involves minimal risk. Published reviews indicate that biofeedback benefits approximately 75% of patients, including approximately 50% who become completely continent and another 25% who have fewer accidents. However, randomized controlled trials are lacking.

Surgery

A variety of surgical techniques have been described for the repair of separated or damaged sphincters. The simplest is to identify the separated ends of the internal and/or external anal sphincter, juxtapose or overlap them, and suture them together. This may be combined with plication of the levator ani muscles or the puborectalis muscles. The success rate is approximately 68% at 1 year. When sphincteroplasty is not practical because the sphincter damage is too extensive or the sphincter is denervated, a flap of gracilis or gluteal muscle may be transposed to create a new sphincter. The transposed muscle may be electrically stimulated to maintain it in a contracted state until defecation is desired. In a multicenter randomized controlled trial of gracilis transfer surgery, 85% of patients reported initial improvement and 66% maintained this improvement for 2 years.

Less Commonly Employed Treatments

Artificial sphincters similar to those used for the treatment of urinary incontinence are reported to be successful in small series of patients. Electrical stimulation of the sacral nerve roots with implantable stimulators is also employed but is still regarded as experimental. Antegrade colonic enemas administered through a cecal conduit have been used successfully to

treat fecal incontinence in children with spina bifida, and may find application in older patients.

See Also the Following Articles

Anal Canal • Anal Sphincter • Constipation • Defecation • Flatulence • Sphincters

Further Reading

Diamant, N. E., Kamm, M. A., Wald, A., and Whitehead, W. E. (1999). AGA technical review on anorectal testing techniques. *Gastroenterology* **116**, 735–760.

Heymen, S., Jones, K. R., Ringel, Y., Scarlett, Y., and Whitehead, W. E. (2001). Biofeedback treatment of fecal incontinence: A critical review. *Dis. Colon Rectum* **44**, 728–736.

Jorge, J. M., and Wexner, S. D. (1993). Etiology and management of fecal incontinence. *Dis. Colon Rectum* **36**, 77–97.

McGrath, M. L., Mellon, M. W., and Murphy, L. (2000). Empirically supported treatments in pediatric psychology: Constipation and encopresis. *J. Pediatr. Psych.* **25**, 225–254.

Rotholtz, N. A., and Wexner S. D. (2001). Surgical treatment of constipation and fecal incontinence. *Gastroenterol. Clin. North Am.* **30**, 131–166.

Whitehead, W. E., Wald, A., and Norton, N. J. (2001). Treatment options for fecal incontinence. *Dis. Colon Rectum* **44**, 131–144.

Fibrogenesis

MEENA B. BANSAL AND SCOTT L. FRIEDMAN
Mount Sinai School of Medicine, New York

cytokine Nonantibody proteins secreted by multiple cell types that act as intercellular mediators.
extracellular matrix Array of macromolecules constituting the scaffolding for maintenance of liver architecture.
polymorphism Germ-line sequence alteration present in at least 1% of the population.

The fibrotic response, which underlies all the complications of end-stage liver disease, is deleterious both by its effects on cellular function (synthetic dysfunction, impaired metabolic activity, and encephalopathy) and by its mechanical contribution to increased portal resistance (ascites and variceal bleeding). Therefore, therapies that are able to retard and reverse the fibrotic response will have a dramatic impact on the treatment of patients with chronic liver disease. This article will review the current understanding of the cellular basis of hepatic fibrosis and how these insights are leading to advances in the diagnosis and treatment of chronic liver disease.

INTRODUCTION

Hepatic fibrosis is a reversible wound healing response characterized by the accumulation of extracellular matrix or "scar" that occurs in almost all patients with chronic liver injury. Ultimately, hepatic fibrosis leads to cirrhosis, characterized by nodule formation and organ contraction. The causes of cirrhosis are multiple and include congenital, metabolic, inflammatory, and toxic liver disease (Table I).

EXTRACELLULAR MATRIX COMPOSITION IN NORMAL LIVER AND HEPATIC SCAR

Extracellular matrix (ECM) refers to the array of macromolecules constituting the scaffolding of normal and fibrotic liver. The components of hepatic extracellular matrix include several families of structural and supporting molecules: collagens, noncollagen glycoproteins, matrix-bound growth factors, glycosaminoglycans, proteoglycans, and matricellular proteins.

In the normal liver, so-called "fibril-forming" collagens (types I, III, V, and XI) are largely confined to the capsule, around large vessels, and in the portal triad, with only scattered fibrils containing types I and III collagen in the subendothelial space. Smaller amounts of other collagens (including types VI, XIV, and XVIII), glycoproteins, matricellular proteins, and proteoglycans (consisting primarily of heparan sulfate) are also present.

Copyright 2004, Elsevier (USA). All rights reserved.

treat fecal incontinence in children with spina bifida, and may find application in older patients.

See Also the Following Articles

Anal Canal • Anal Sphincter • Constipation • Defecation • Flatulence • Sphincters

Further Reading

Diamant, N. E., Kamm, M. A., Wald, A., and Whitehead, W. E. (1999). AGA technical review on anorectal testing techniques. *Gastroenterology* **116**, 735–760.

Heymen, S., Jones, K. R., Ringel, Y., Scarlett, Y., and Whitehead, W. E. (2001). Biofeedback treatment of fecal incontinence: A critical review. *Dis. Colon Rectum* **44**, 728–736.

Jorge, J. M., and Wexner, S. D. (1993). Etiology and management of fecal incontinence. *Dis. Colon Rectum* **36**, 77–97.

McGrath, M. L., Mellon, M. W., and Murphy, L. (2000). Empirically supported treatments in pediatric psychology: Constipation and encopresis. *J. Pediatr. Psych.* **25**, 225–254.

Rotholtz, N. A., and Wexner S. D. (2001). Surgical treatment of constipation and fecal incontinence. *Gastroenterol. Clin. North Am.* **30**, 131–166.

Whitehead, W. E., Wald, A., and Norton, N. J. (2001). Treatment options for fecal incontinence. *Dis. Colon Rectum* **44**, 131–144.

Fibrogenesis

MEENA B. BANSAL AND SCOTT L. FRIEDMAN
Mount Sinai School of Medicine, New York

cytokine Nonantibody proteins secreted by multiple cell types that act as intercellular mediators.
extracellular matrix Array of macromolecules constituting the scaffolding for maintenance of liver architecture.
polymorphism Germ-line sequence alteration present in at least 1% of the population.

The fibrotic response, which underlies all the complications of end-stage liver disease, is deleterious both by its effects on cellular function (synthetic dysfunction, impaired metabolic activity, and encephalopathy) and by its mechanical contribution to increased portal resistance (ascites and variceal bleeding). Therefore, therapies that are able to retard and reverse the fibrotic response will have a dramatic impact on the treatment of patients with chronic liver disease. This article will review the current understanding of the cellular basis of hepatic fibrosis and how these insights are leading to advances in the diagnosis and treatment of chronic liver disease.

INTRODUCTION

Hepatic fibrosis is a reversible wound healing response characterized by the accumulation of extracellular matrix or "scar" that occurs in almost all patients with chronic liver injury. Ultimately, hepatic fibrosis leads to cirrhosis, characterized by nodule formation and organ contraction. The causes of cirrhosis are multiple and include congenital, metabolic, inflammatory, and toxic liver disease (Table I).

EXTRACELLULAR MATRIX COMPOSITION IN NORMAL LIVER AND HEPATIC SCAR

Extracellular matrix (ECM) refers to the array of macromolecules constituting the scaffolding of normal and fibrotic liver. The components of hepatic extracellular matrix include several families of structural and supporting molecules: collagens, noncollagen glycoproteins, matrix-bound growth factors, glycosaminoglycans, proteoglycans, and matricellular proteins.

In the normal liver, so-called "fibril-forming" collagens (types I, III, V, and XI) are largely confined to the capsule, around large vessels, and in the portal triad, with only scattered fibrils containing types I and III collagen in the subendothelial space. Smaller amounts of other collagens (including types VI, XIV, and XVIII), glycoproteins, matricellular proteins, and proteoglycans (consisting primarily of heparan sulfate) are also present.

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Causes of Fibrosis and Cirrhosis

Presinusoidal fibrosis
Schistosomiasis
Idiopathic portal fibrosis
Parenchymal fibrosis
Drugs and toxins
Alcohol
Methotrexate
Isoniazid
Vitamin A
Amiodarone
Perhexilene maleate
α-Methyldopa
Oxyphenisatin
Infections
Chronic hepatitis C virus, hepatitis B virus
Brucellosis
Echinococcus
Congenital or tertiary syphillis
Autoimmune
Chronic autoimmune hepatitis
Vascular abnormalities
Chronic passive congestion
Hereditary hemorrhagic telangiectasia
Metabolic/genetic diseases
Wilson's disease
Genetic hemochromatosis
α1-Antitrypsin deficiency
Carbohydrate metabolism disorders
Lipid metabolism disorders
Urea cycle defects
Porphyria
Amino acid metabolism disorders
Bile acid disorders
Biliary obstruction
Primary biliary cirrhosis
Secondary biliary cirrhosis
Cystic fibrosis
Biliary atresia/neonatal hepatitis
Congential biliary cysts
Idiopathic/miscellaneous
Nonalcoholic steatohepatitis
Indian childhood cirrhosis
Granulomatous liver disease
Polycystic liver disease
Postsinusoidal fibrosis
Veno-occlusive disease

Reprinted with permission from Friedman, S. (2002). Hepatic fibrosis: Consequences of liver disease. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 9th ed. Copyright Lippincott Williams & Wilkins 2002.

With progressive fibrosis, there is both a quantitative and qualitative shift in the matrix composition compared to normals and these changes are similar regardless of the type of liver injury. Total collagen content increases 3- to 10-fold, although the collagen itself is not "abnormal" in sequence or structure. Overall, there is a marked increase in "interstitial matrix" typical of the healing wound, which includes fibril-forming collagens (types I, III, and V) and some non-fibril-forming collagens (types IV and VI), several glycoproteins, as well as a large number of proteoglycans and glycosaminoglycans. In particular, there is a shift from proteoglycans containing heparan sulfate to those containing chondroitin and dermatan sulfates. This shift in ECM in turn contributes to the loss of hepatocyte microvilli and sinusoidal endothelial fenestrae, which results in deterioration of hepatic function (Fig. 1).

STELLATE CELL ACTIVATION: THE CENTRAL EVENT IN HEPATIC FIBROSIS

The hepatic stellate cell (previously called lipocyte, Ito, fat-storing, or perisinusoidal cell) is the primary source of the extracellular matrix in normal and fibrotic liver. Hepatic stellate cells are resident perisinusoidal cells in the subendothelial space between hepatocytes and sinusoidal endothelial cells. They are the primary site for storing retinoids and therefore can be recognized by their vitamin A autofluorescence. In addition, their perisinusoidal orientation and expression of the cytoskeletal proteins desmin and glial acidic fibrillary protein allow for their *in situ* identification.

Studies in both animals and humans with progressive injury have defined a gradient of changes within stellate cells that collectively are termed "activation" (Fig. 2). Stellate cell activation refers to the transition from a quiescent vitamin A-rich cell to a highly fibrogenic cell characterized morphologically by enlargement of rough endoplasmic reticulum, diminution of vitamin A droplets, ruffled nuclear membrane, appearance of contractile filaments, and proliferation. Proliferation of stellate cells generally occurs in regions of greatest injury, which is typically preceded by an influx of inflammatory cells and is associated with subsequent extracellular matrix accumulation.

Stellate cell activation, the central event in hepatic fibrosis, is thought to occur in two stages: initiation and perpetuation. Initiation refers to early events encompassing rapid changes in gene expression and phenotype that render the cells responsive to cytokines and other stimuli. It results from paracrine stimulation due to rapid, disruptive effects of liver injury on the homeostasis of neighboring cells and from early changes in ECM composition. Perpetuation incorporates those cellular events that amplify the activated phenotype through enhanced cytokine expression and responsiveness and involves at least seven discrete changes in cell

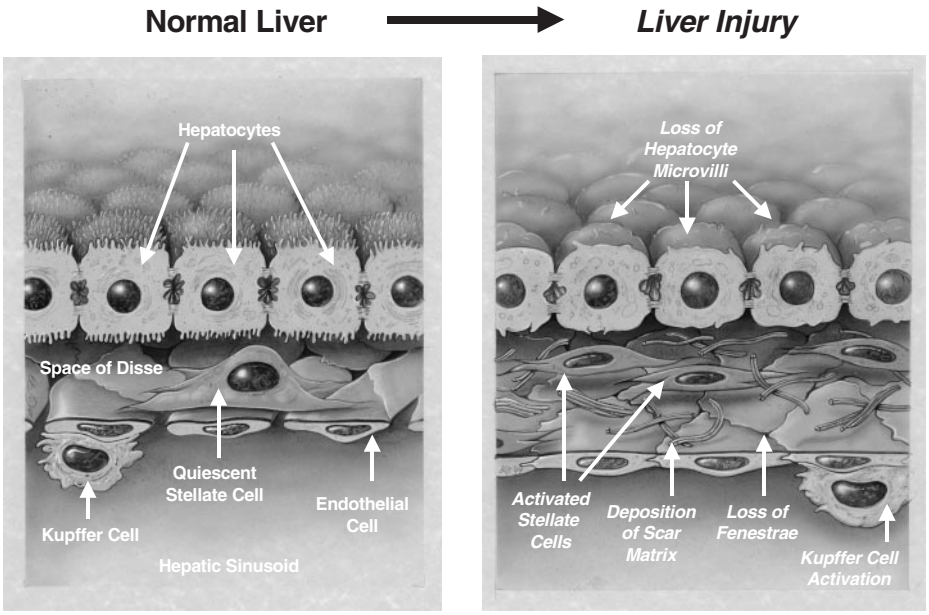


FIGURE 1 Matrix and cellular alterations in hepatic fibrosis. Changes in the subendothelial space of Disse and sinusoid as fibrosis develops in response to liver injury include alterations in both cellular responses and extracellular matrix composition. Stellate cell activation leads to the accumulation of scar (fibril-forming) matrix. This in turn contributes to the loss of hepatocyte microvilli and sinusoidal endothelial fenestrae, which result in the deterioration of hepatic function. Kupffer cell (macrophage) activation accompanies liver injury and contributes to paracrine activation of stellate cells. Reprinted from Friedman (2000), with permission of The American Society for Biochemistry and Molecular Biology.

behavior: (1) proliferation; (2) chemotaxis; (3) fibrogenesis; (4) contractility; (5) matrix degradation; (6) retinoid loss; and (7) white blood cell chemoattractant and cytokine release. Either directly or indirectly, the net effect of these changes is the accumulation of extracellular matrix.

CLINICAL ASPECTS OF HEPATIC FIBROSIS

Fibrosis Progression and Reversibility

The rate of progression of fibrosis in an individual patient with chronic liver disease cannot be predicted with certainty. However, some general rules apply:

1. Fibrosis usually requires at least several months to years of ongoing insult;
2. Severity of inflammation and injury usually correlate with rate of progression;
3. Concurrent hepatic insult by more than one agent is synergistic for the progression of fibrosis;
4. The exact moment at which fibrosis becomes irreversible is not known;
5. Host genotype is an intrinsic determinant of fibrosis.

Diagnosis and Assessment of Hepatic Fibrosis

Accurate assessment of the extent of fibrosis is essential to guide management and predict prognosis in patients with chronic liver injury. Histologic assessment of a liver biopsy specimen remains the "gold standard" for quantifying fibrosis, with increasing interest in the use of noninvasive markers to allow more frequent sampling and avoid the risks of percutaneous biopsy.

Histologic and Morphometric Methods

Several semiquantitative morphologic methods that evaluate extracellular matrix in biopsy specimens stained with either hematoxylin and eosin or connective tissue stains such as Masson's Trichrome, reticulin silver impregnation, or Van Gieson have been described. These methods can be prone to sampling error if the fibrosis is not uniformly distributed. Semiquantitative methods include the Knodell-Ishak score, the French Metavir system, and others. The systems correlate well with one another and employ a 4- to 6-point scale that grades fibrosis based on its distribution and amount. Standard methods may be complemented by more accurate morphometric approaches using image analysis,

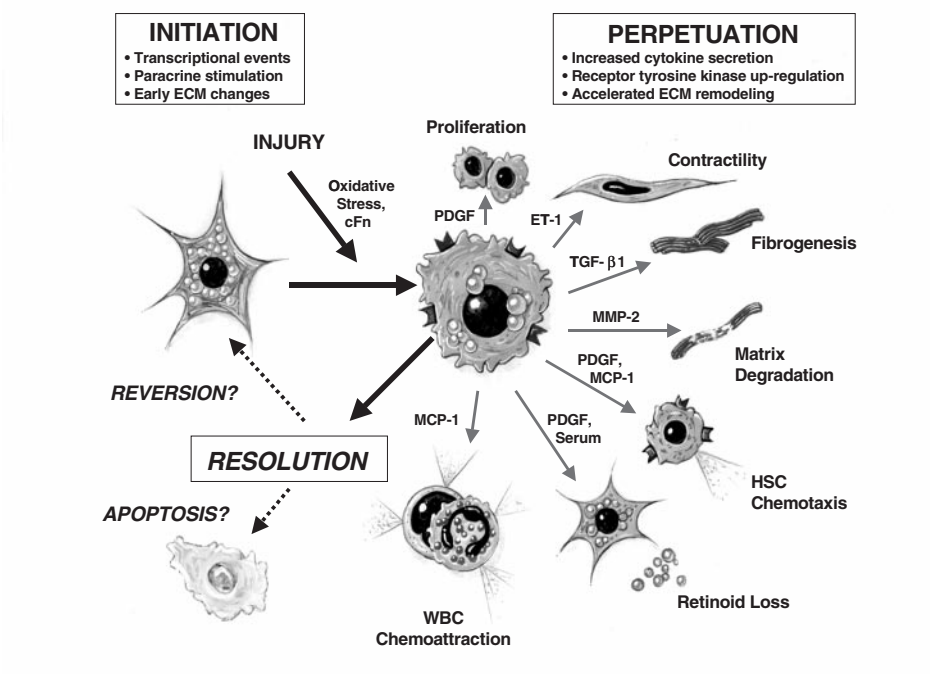


FIGURE 2 Phenotypic features of hepatic stellate cell activation during liver injury and resolution. Following liver injury, hepatic stellate cells undergo "activation," which connotes a transition from quiescent vitamin A-rich cells into proliferative, fibrogenic, and contractile myofibroblasts. The major phenotypic changes after activation include proliferation, contractility, fibrogenesis, matrix degradation, chemotaxis, retinoid loss, and white blood cell (WBC) chemoattraction. Key mediators underlying these effects are shown. The fate of activated stellate cells during resolution of liver injury is uncertain but may include reversion to a quiescent phenotype and/or selective clearance by apoptosis. PDGF, platelet-derived growth factor; MCP-1, monocyte chemoattractant protein-1. Reprinted from Friedman (2000), with permission of The American Society for Biochemistry and Molecular Biology.

in which tissue is stained with picrosirius red, which binds type I collagen.

Noninvasive Methods

There has been considerable effort to identify serum markers as noninvasive measures of hepatic fibrosis. Although their accuracy and predictive value are improving, they cannot yet supplant direct analysis of liver. Several classes of molecules can be measured in serum: (1) enzymes involved in extracellular matrix production or modification (lysyl oxidase or prolyl hydroxylase); (2) matrix molecules (hyaluronic acid, types IV and VI collagen, or laminin); (3) components of matrix molecules that are cleaved off before incorporation of the parent molecule into the fibrotic bundle (carboxy-terminal propeptide of type I collagen); (4) fibrogenic or proliferative cytokines (transforming growth factor-β1 or fibroblast growth factor); and (5) enzymes involved in matrix degradation (MMP-2, TIMP-1, and TIMP-2). Thus far, no single test has emerged as the perfect marker of fibrosis, although it remains possible that a battery of tests used together may prove useful.

Several caveats must be overcome for serum markers of fibrosis to become useful: (1) they typically reflect the rate of matrix turnover, not deposition, and thus tend to be more elevated when there is high inflammatory activity. Conversely, extensive matrix deposition can go undetected if there is minimal inflammation. (2) None of the molecules are liver-specific, so that concurrent sites of inflammation may contribute to serum levels. (3) Serum levels are affected by clearance rates that may be impaired due to either sinusoidal endothelial cell dysfunction or impaired biliary excretion.

In summary, there remains a compelling need for noninvasive markers that accurately reflect the matrix content of tissue and have better prognostic accuracy than standard clinical and laboratory indices such as the

Child-Pugh or MELD classification. Such markers will be invaluable as anti-fibrotic therapies undergo clinical trials in the coming years.

THERAPY OF HEPATIC FIBROSIS

The improved understanding of mechanisms underlying hepatic fibrosis makes effective anti-fibrotic therapy an emerging reality. Treatment will remain a challenging task, however, and thus far no drugs have been approved as anti-fibrotic agents in humans. Therapies will need to be well tolerated over decades, demonstrate liver-specific targeting, and have limited adverse effects on other tissues. Putative agents must have direct anti-fibrotic effects rather than indirect effects by abrogating the injury and must be effective in reversing already established liver disease, which more accurately mimics the clinical scenario in which they would be utilized. Given the long natural history of the fibrotic response and the fact that it is the scarring, not the injury, that usually leads to liver failure, anti-fibrotic therapies capable of even slowing the progression of the fibrotic response could lead to a dramatic improvement in morbidity and mortality from chronic liver disease. The paradigm of stellate cell activation provides an important framework to define sites of anti-fibrotic therapy. These include the following: (1) cure the primary disease to prevent injury; (2) reduce inflammation or the host response in order to avoid stimulating stellate cell activation; (3) directly down-regulate stellate cell activation; (4) neutralize proliferative, fibrogenic, contractile, and/or pro-inflammatory responses of stellate cells; (5) stimulate apoptosis of stellate cells; and (6) increase the degradation of scar matrix, by stimulating cells that produce matrix proteases, by down-regulating their inhibitors, or by directly administering matrix proteases.

FUTURE PROSPECTS

Continued progress can be anticipated in the molecular regulation of fibrosis and its treatment. Rapid advances in gene therapy, tissue-specific targeting, and high-throughput small-molecule screening of cytokine inhibitors are likely to benefit diagnosis and therapy of hepatic fibrosis. Methods have been developed for stellate cell-specific targeting in animal models, which could lead to successful targeting to minimize toxicity of anti-fibrotics, and for use as novel diagnostics. New insights into the regulation of growth and apoptosis could have direct implications for stellate cell behavior in liver injury. Sequencing of the human genome and use of microarrays may yield genetic polymorphisms that predict the rate of fibrosis prospectively and patterns of multigene expression that have clinical or therapeutic implications. Additionally, there is tremendous interest in herbal and natural anti-fibrotic remedies, particularly in the Far East, where many such compounds are undergoing clinical trials. Future therapies may emerge from these efforts as well.

Acknowledgment

Sections of this article were previously published in Friedman, S. (2002). Hepatic fibrosis: Consequences of liver disease. *In* "Schiff's Diseases of the Liver," 9th ed. Adapted with permission of the publisher Lippincott Williams & Wilkins.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Cirrhosis • Liver Biopsy

Further Reading

Bataller, R., and Brenner, D. A. (2001). Stellate cells as a target for treatment of liver fibrosis. *Semin. Liver Dis.* **21**(3), 437–452.

Friedman, S. L. (2000). Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J. Biol. Chem.* **275**(4), 2247–2250.

Friedman, S. L. (2001). Hepatic stellate cells. *Semin. Liver Dis.* **21**(3), 307–452.

Gressner, A. M., and Bachem, M. G. (1990). Cellular sources of noncollagenous matrix proteins: Role of fat-storing cells in fibrogenesis. *Semin. Liver Dis.* **10**(1), 30–46.

Powell, E. E., Edwards-Smith, C. J., Hay, J. L., *et al.* (2000). Host genetic factors influence disease progression in chronic hepatitis C. *Hepatology* **31**(4), 828–833.

Poynard, T., Bedossa, P., and Opolon, P. (1997). Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSIRVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* **349**(9055), 825–832.

Rojkind, M., Giambrone, M. A., and Biempica, L. (1979). Collagen types in normal and cirrhotic liver. *Gastroenterology* **76**(4), 710–719.

Schuppan, D., Ruehl, M., Somasundaram, R., and Hahn, E. G. (2001). Matrix as modulator of stellate cell and hepatic fibrogenesis. *Semin. Liver Dis.* **21**(3), 351–372.

Wu, J., and Danielsson, A. (1995). Detection of hepatic fibrogenesis: A review of available techniques. *Scand. J. Gastroenterol.* **30**(9), 817–825.

Fistula

THOMAS E. CLANCY AND DAVID B. LAUTZ
Brigham and Women's Hospital, Boston

fistula An abnormal connection between two epithelialized surfaces, such as two hollow viscera or one hollow viscera and the skin ("enterocutaneous fistula").

A fistula is most accurately described as an abnormal connection between two epithelialized organs or between an epithelialized organ and the skin. This article will discuss gastrointestinal fistulas only—those involving some portion of the alimentary tract. Although congenital causes of fistula exist, the majority are acquired. Furthermore, the majority of acquired fistulae are a complication of some surgical or endoscopic manipulation of the stomach or intestine. Drainage of intestinal contents through the fistula tract typically leads to severe metabolic and nutritional abnormalities, as well as local and systemic sepsis. Historical reports from the 1960s cite a mortality of approximately 40% from gastrointestinal fistulas. With improved management of the multiple nutritional and metabolic derangements of these patients and with improved control of sepsis, mortality is significantly decreased but still remains substantial. Today mortality ranges from 5 to 20%, depending on anatomic location of the individual fistula studied and etiology. Management of patients with gastrointestinal fistulas has evolved to emphasize metabolic resuscitation, nutritional support, and control of infection. This supportive, nonoperative approach can be prolonged. In most cases, surgical intervention is reserved for control of overwhelming contamination or for the failure of a long period of conservative management. With this algorithm, the vast majority of fistulas will close without an operation. Notable exceptions include fistulas involving a malignancy or complete disruption of intestinal continuity.

CLASSIFICATION OF FISTULAS

Fistulae are most commonly classified according to their anatomic localization. The general term "enterocutaneous fistula" usually refers to a fistula between the small intestine and the skin. Less common are gastric and duodenal fistulas, colocutaneous fistulas, pancreatic fistulas, and colonic fistulas. Anatomic localization is usually derived from a combination of plain contrast films (fistulogram) and computed tomography (CT). Anatomic information is further combined with etiologic information to characterize fistulas as simple acquired postoperative fistulas or as spontaneous fistulas, such as those linked to radiation therapy, inflammation, or malignancy.

Enterocutaneous fistulae are typically described as "low-output" if daily output from the cutaneous fistula is less than <200 cc per 24 h. Similarly, "high-output" fistulas are those with >500 cc output per 24 h. The link between output and prognosis is not accepted by all clinicians, although many feel that site-for-site, high-output fistulas are associated with increased mortality and decreased chance of successful spontaneous closure.

Less common than the external fistulas described above are "internal" fistulas between two viscera, such as enteroenteric (i.e., between two portions of the intestine) or enterovesical (between the intestine and the bladder). Internal fistulas are more difficult to recognize and have a far lower rate of spontaneous closure.

GENERAL MANAGEMENT OF FISTULAS

Stabilization

Fluids and Electrolytes

The most important aspect of the initial management of the patient with a gastrointestinal (GI) fistula is the recognition and treatment of existing metabolic abnormalities. Loss of enteric contents through the fistula leads to dehydration, significant electrolyte abnormalities, and often acidosis (although patients with gastric fistulas can be alkalotic). Intravenous access and resuscitation with isotonic crystalloid solution is the mainstay, with invasive hemodynamic monitoring as indicated. Oral intake generally increases flow through the GI tract and thus output from an enterocutaneous fistula, so it is typically withheld. The endpoint for fluid resuscitation should be the correction of electrolyte and acid–base derangements.

Nutrition

A significant number of patients with gastrointestinal fistulas present with malnutrition. In malnourished

 Copyright 2004, Elsevier (USA). All rights reserved.

patients, it is well recognized that complications and mortality are increased and chances of spontaneous fistula closure are decreased. Malnutrition is particularly common in high-output small bowel fistulas. As noted above, oral intake must be withheld to avoid stimulating further fluid, electrolytes, and protein. Enteral nutrition is occasionally appropriate for low-output fistulas from the distal intestinal tract in which adequate absorption may be obtained or with proximal fistulas with distal feeding via a jejunostomy. Thus, most GI fistula patients require some form of parenteral nutrition.

All patients with GI fistulas must undergo early assessment of nutritional status. Common methods include determination of nitrogen balance, laboratory data including albumin and prealbumin, and sometimes indirect calorimetry. An estimation of baseline required kilocalories can be obtained using the Harris-Benedict equation. Of note, energy requirements are increased in the setting of stress, and fistula patients need to receive up to 1.5 times their basal estimated energy expenditure in order to achieve positive nitrogen balance. It is appropriate in the nonseptic patient to obtain a form of central intravenous access that will allow the patient to receive weeks to months of parenteral nutrition.

Control of Sepsis

Loss of intestinal continuity is followed by localized infection that involved adjacent structures and subsequently drainage of an infected focus in the form of a fistula. Infection may remain localized, but may also evolve into generalized infection with peritonitis or sepsis. Sepsis control involves evacuation and washout of grossly contaminated debris. Inability to control infection remains a major cause of morbidity and mortality from GI fistulas. Without control of infection, successful resolution of fistulas is very rare. Identification and localization of a septic focus are crucial and CT scan is the best way to visualize intra-abdominal collections. For stable patients, attempted percutaneous drainage of an abscess under radiographic guidance is a reasonable fist step in management. The purpose of these percutaneous drains is to create a controlled fistula whereby output can be quantified, skin can be protected, and infection can be drained.

Early operative drainage is typically reserved for worsening infection. Aggressive intervention is not without risks, due to significant intra-abdominal inflammation and associated early adhesions. Particular indications for early operative management include the presence of multiple collections, an open anastamosis, incompletely drained fluid collections, or extensive cellulitis. Drains should be placed next to the fistula whenever possible.

Skin Care

External leakage of enteric fluids can lead to significant breakdown in skin integrity. Multiple strategies exist to reduce skin exposure to toxic intestinal contents, including the use of ostomy appliances, liberal use of powders and creams, consultation with an enterostomal therapist or wound expert, and the use of histamine receptor blockers. High-output fistulas often require the use of ostomy appliances for adequate control of effluent. Skin breakdown will complicate abdominal wall closure and interferes significantly with patient satisfaction.

Investigation/Localization of Fistula

The most useful initial radiographic study is the CT scan. CT allows preliminary information on anatomic localization to be acquired, but more importantly allows an assessment of local gastrointestinal spillage in the abdominal cavity and the need for operative or percutaneous drainage. Water-soluble oral contrast is most helpful and barium should be avoided due to the severe peritoneal inflammation it causes.

Additional gastrointestinal contrast studies are useful for further management once the patient has been stabilized. A "fistulogram" can be obtained by directly injecting water-soluble contrast into the fistula orifice. Small bowel follow-through or contrast enemas may be indicated as well. Precise anatomic localization of the fistula should be performed prior to any operative intervention if at all possible.

Definitive Management

Nonoperative Management

Factors associated with successful nonoperative management With adequate control of infection, as well as nutritional and metabolic support, most (approximately 90%) of gastrointestinal fistulas will close without operative management. A number of factors are associated with high spontaneous closure rate. Colonic and pancreatic fistulas commonly heal without operation. Fistulas that are acute, have a long fistula tract, have no distal obstruction, and have low output are more likely to heal spontaneously. Well-nourished patients also have greater success with nonoperative management.

Factors associated with nonhealing fistulas Fistulas arising in an area of previously irradiated tissue

or in the presence of ongoing infection will rarely close. Those fistula involving actively diseased bowel, such as those involving a malignancy or inflammatory bowel disease are also less likely to close. Likewise, a fistulogram showing interruption of bowel continuity suggests that spontaneous closure is unlikely. Local mechanical factors such as a foreign body or distal obstruction also promote nonhealing. Finally, chronicity of a fistula and epithelialization of the fistula tract are associated with decreased spontaneous closure.

Operative Management

With infection controlled, most fistulas should close in 4–6 weeks, so surgical closure is generally not recommended until at least 6–8 weeks after the onset of the fistula. Earlier operative intervention should be limited to control of infection and fecal diversion, if indicated. Early operation is associated with iatrogenic bowel injury and it is best to wait until acute inflammation has decreased in the operative field. Delayed operative management should involve meticulous sharp dissection and resection of the involved segment of bowel with primary anastamosis.

SITE-SPECIFIC ISSUES FOR FISTULAS

Small Bowel

The small intestine is the source of most of what are commonly referred to as "enterocutaneous fistulas" and most of the above principles apply here. Most fistulas are iatrogenic and postoperative; of these, approximately one-half are due to anastamotic leak and one-half are due to unrecognized bowel injury at the time of laparotomy. Approximately one-fifth of small intestinal fistulas are spontaneous, involving inflammatory bowel disease, radiation, or malignancy. Those due to radiation or malignancy are particularly unlikely to heal with conservative management.

Stomach/Duodenum

These represent a small proportion of GI fistulas and the vast majority are postsurgical. Duodenal fistulas have a high rate of spontaneous closure. Gastric fistulae, however, are less likely to close spontaneously, possibly due to the high percentage associated with ongoing disease such as malignancy. Histamine receptor blockers are particularly helpful in decreasing output and promoting closure although somatostatin analogues have not had similar success.

Colon/Rectum

Postoperative colonic fistulas have a high rate of spontaneous closure. Low-output fistula often heal while the patient continues to tolerate enteral feeding. With local infection, proximal diversion of the fecal stream may be indicated. In such cases, ileostomy is preferred to transverse colostomy to avoid distal bowel ischemia.

Colovesical fistulas rarely heal without surgery and require resection of involved colon in a one-stage procedure. Diagnosis is most reliable with cystoscopy, although sigmoidoscopy is also required to rule out the presence of cancer, which would indicate an *en bloc* partial cystectomy at the time of colectomy.

Radiation-associated fistulas may present years after radiation treatment of pelvic malignancy. These are often complex and may involve multiple organs. Treatment is primarily by proximal bowel diversion, without attempts at fistula resection.

Esophagus

Postoperative esophageal fistulas after esophageal anastamoses are more common than those after anastamoses in other portions of the GI tract, with an up to 10% leak rate from these anastamoses. Esophagogram is best for diagnosis. Leaks in the mediastinum require thoracotomy and diversion to control sepsis, whereas a cervical anastamosis permits cervical drainage and nonoperative management.

Particularly troublesome are fistula between the esophagus and trachea due to malignancy. These involve persistent contamination of the respiratory tract, leading to pneumonia and sepsis. Palliation is the goal, although resection of the fistula is complicated by a high rate of morbidity and mortality. A new technique is esophageal intubation with endoscopic stents to allow swallowing while preventing aspiration.

Pancreatic

Pancreatic fistulas result from disruption of the main pancreatic duct or one of its branches after trauma, pancreatitis, pancreatic surgery, or splenectomy. Diagnosis is suspected with a peripancreatic fluid collection on CT scan and is confirmed in the external fistula with high amylase content in the effluent. Endoscopic retrograde cholagiopancreatography is useful to define ductal anatomy and to identify ductal obstruction. Bowel rest is critical to management and most of these fistulas will close with nonoperative management. Somatostatin analogues may decrease fistula output but do not affect the rate of spontaneous closure.

See Also the Following Articles

Bile Duct Injuries and Fistulas • Endoscopy, Complications of • Malnutrition • Nutritional Assessment • Parenteral Nutrition

Further Reading

Berry, S. M., and Fischer, J. (1994). Enterocutaneous fistulas. *Curr. Prob. Surg.* **31**, 469.

Campos, A. C. L., Meguid, M. M., and Coelho, J. C. (1996). Factors influencing outcome in patients with gastrointestinal fistula. *Surg. Clin. North Am.* **76**, 1191–1198.

Denham, D. W., and Fabri, P. J. (2001). Enterocutaneous fistula. *In* "Current Surgical Therapy" (Cameron, ed.). Mosby, St. Louis, MO.

Edmunds, L. H., Williams, G. M., and Welch, C. E. (1960). External fistulas arising from the gastrointestinal tract. *Ann. Surg.* **152**, 445–471.

Pickhardt, P. J., Bhalla, S., and Balfe, D. M. (2002). Acquired gastrointestinal fistulas: Classification, etiologies, and imaging evaluation. *Radiology* **224**, 9–23.

Flatulence

CHRISTOPHER J. MCNEILL
Veterans Affairs Medical Center, San Diego

aerophagia The swallowing of excessive amounts of air.
flatulence The passage of colonic gas from the rectum.

The discomfort and embarrassment of intestinal "gas" have been a concern of humans for thousands of years. Hippocrates authored a treatise entitled "The Winds," which reviewed various illnesses that he postulated to be related to this malady. Over 200 years ago, Benjamin Franklin, with his droll sense of humor, published the pamphlet "Fart Proudly," in which he suggested dietary changes as a cure to "escaped wind." To this day, patients continue to experience gastrointestinal symptoms that they attribute to excess intestinal gas and, although the physiology of flatulence has been better defined, medical knowledge has yet to describe a cure.

SOURCE OF COLONIC GAS

Flatulence is not by itself a marker for intestinal disease and is in fact a perfectly normal occurrence in healthy adults. Studies of gas passage by normal adults on a regular diet have shown great variability in the volume of gas passed during a 24 h period (476–1496 ml with a median of 705 ml) and similar variability in the frequency of passage (10 to 20 times per day). For the purposes of this discussion, aerophagia is not considered a significant source of flatulence and will not be considered further. There are five odorless gases that account for 99% of gas passed per rectum: nitrogen, oxygen, carbon dioxide, hydrogen, and methane. The percentage of each gas is highly variable from subject to subject: N_2 (11–92%); O_2 (0–11%); CO_2 (3–54%); H_2 (0–86%); and CH_4 (0–54%). Other gases, present only in trace amounts, are odoriferous and account for the noxious smell of flatus, which in the case of hydrogen sulfide (H_2S) can be detected by the human nose in concentrations as low as one-half part per billion! As above, the gas composition of flatus is highly variable from person to person and is the net result of bacterial fermentation of partially digested foodstuffs entering the colon and to a lesser extent the passive diffusion of gases between blood and the colonic lumen.

Many vegetables, especially legumes and most notably beans, contain nonabsorbable oligosaccharides that enter the colon and provide a tasty substrate for gas-producing bacteria. The lactose in milk and milk products likewise is inadequately absorbed in many adults due to a deficiency of the enzyme lactase (β-galactosidase), which splits this sugar into glucose and galactose in the small intestine. Of commonly ingested flours, only rice and gluten-free wheat flour are completely absorbed and thus noncontributory to gas production. Sorbitol and xylitol, found in many sugar-free gums and candies, are poorly absorbed and add to flatulence. Figure 1 shows the fate of

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Bile Duct Injuries and Fistulas • Endoscopy, Complications of • Malnutrition • Nutritional Assessment • Parenteral Nutrition

Further Reading

Berry, S. M., and Fischer, J. (1994). Enterocutaneous fistulas. *Curr. Prob. Surg.* **31**, 469.

Campos, A. C. L., Meguid, M. M., and Coelho, J. C. (1996). Factors influencing outcome in patients with gastrointestinal fistula. *Surg. Clin. North Am.* **76**, 1191–1198.

Denham, D. W., and Fabri, P. J. (2001). Enterocutaneous fistula. *In* "Current Surgical Therapy" (Cameron, ed.). Mosby, St. Louis, MO.

Edmunds, L. H., Williams, G. M., and Welch, C. E. (1960). External fistulas arising from the gastrointestinal tract. *Ann. Surg.* **152**, 445–471.

Pickhardt, P. J., Bhalla, S., and Balfe, D. M. (2002). Acquired gastrointestinal fistulas: Classification, etiologies, and imaging evaluation. *Radiology* **224**, 9–23.

Flatulence

CHRISTOPHER J. MCNEILL
Veterans Affairs Medical Center, San Diego

aerophagia The swallowing of excessive amounts of air.
flatulence The passage of colonic gas from the rectum.

The discomfort and embarrassment of intestinal "gas" have been a concern of humans for thousands of years. Hippocrates authored a treatise entitled "The Winds," which reviewed various illnesses that he postulated to be related to this malady. Over 200 years ago, Benjamin Franklin, with his droll sense of humor, published the pamphlet "Fart Proudly," in which he suggested dietary changes as a cure to "escaped wind." To this day, patients continue to experience gastrointestinal symptoms that they attribute to excess intestinal gas and, although the physiology of flatulence has been better defined, medical knowledge has yet to describe a cure.

SOURCE OF COLONIC GAS

Flatulence is not by itself a marker for intestinal disease and is in fact a perfectly normal occurrence in healthy adults. Studies of gas passage by normal adults on a regular diet have shown great variability in the volume of gas passed during a 24 h period (476–1496 ml with a median of 705 ml) and similar variability in the frequency of passage (10 to 20 times per day). For the purposes of this discussion, aerophagia is not considered a significant source of flatulence and will not be considered further. There are five odorless gases that account for 99% of gas passed per rectum: nitrogen, oxygen, carbon dioxide, hydrogen, and methane. The percentage of each gas is highly variable from subject to subject: N_2 (11–92%); O_2 (0–11%); CO_2 (3–54%); H_2 (0–86%); and CH_4 (0–54%). Other gases, present only in trace amounts, are odoriferous and account for the noxious smell of flatus, which in the case of hydrogen sulfide (H_2S) can be detected by the human nose in concentrations as low as one-half part per billion! As above, the gas composition of flatus is highly variable from person to person and is the net result of bacterial fermentation of partially digested foodstuffs entering the colon and to a lesser extent the passive diffusion of gases between blood and the colonic lumen.

Many vegetables, especially legumes and most notably beans, contain nonabsorbable oligosaccharides that enter the colon and provide a tasty substrate for gas-producing bacteria. The lactose in milk and milk products likewise is inadequately absorbed in many adults due to a deficiency of the enzyme lactase (β-galactosidase), which splits this sugar into glucose and galactose in the small intestine. Of commonly ingested flours, only rice and gluten-free wheat flour are completely absorbed and thus noncontributory to gas production. Sorbitol and xylitol, found in many sugar-free gums and candies, are poorly absorbed and add to flatulence. Figure 1 shows the fate of

Copyright 2004, Elsevier (USA). All rights reserved.

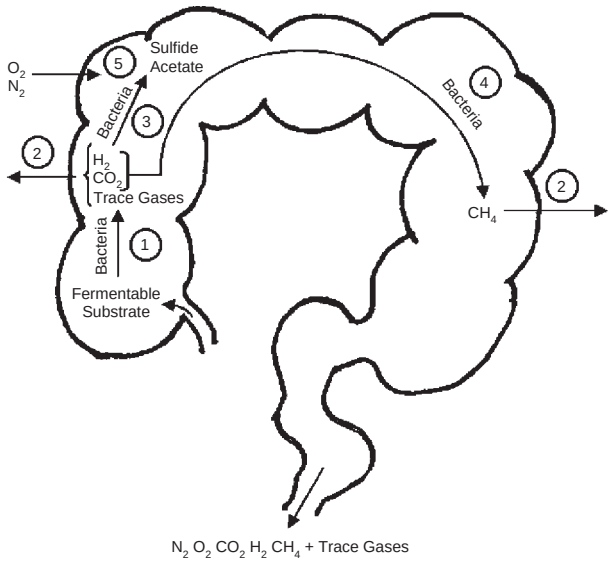


FIGURE 1 (1) In the colon, malabsorbed ingested material and mucus are fermented by bacteria that subsequently release trace gases (some of which are odoriferous), CO_2, and H_2. (2) A fraction of the bacterial gases are absorbed into the blood perfusing the colon. (3) In the right colon, H_2 is consumed by bacteria in the process of reducing sulfate to sulfide and converting CO_2 to acetate. (4) In addition, H_2 is consumed in the left colon by methanogens in the process of reducing CO_2 to CH_4. (5) N_2 and O_2 diffuse from the blood into the colonic lumen down a gradient created by the production of gas by bacteria. The net result of all of the aforementioned processes determines the composition and rate of excretion of gas per rectum. Adapted from Strocchi and Levitt, with permission.

these nonabsorbable fermentable substrates when they meet up with billions of hungry anaerobic bacteria in the colon. It is noteworthy that production of odorless methane gas seems limited to only one-third of adults and, when present in sufficient quantities, leads to stools that float. The combustibility of both methane and hydrogen can have explosive consequences for the immature individual who attempts to ignite his flatus.

THERAPEUTIC INTERVENTIONS

Treatable gastrointestinal pathology such as bacterial overgrowth, chronic exocrine pancreatic insufficiency, celiac disease, parasitic infestations, especially *Giardia lamblia*, or other causes of maldigestion and malabsorption should be identified and managed first. In general, pharmacological interventions with anticholinergics, activated charcoal, simethicone, probiotics, and antibiotics have proved ineffective. The mainstay of therapy has been the elimination of common offending flatugenic foods such as milk or milk products, beans, and sugar-free gums and candies. Commercially available β-D-galactosidase (LactAid drops) may be used to prophylactically treat milk or be ingested at mealtimes in lactase-deficient individuals to normalize lactose absorption. Soaking beans for up to 12 h and discarding the water before cooking or the use of α-D-galactosidase (Beano) at mealtime will enhance digestion and reduce the flatulence associated with the ingestion of beans. The selective avoidance of many other fruits and vegetables that are cited as problematic such as apricots, bananas, bell peppers, broccoli, Brussels sprouts, cabbage, carrots, celery, corn, onions, prunes, and raisins may also be helpful. Sulfur-rich foods such as eggs and cauliflower are notorious for promoting bacterial H_2S production and malodorous flatulence. Cooking or consuming products with rice or gluten-free wheat flour, which are well absorbed, will lessen gas production in the colon. In the end, if all else fails, walking briskly around will disperse the sounds and smells of flatulence and avoid the embarrassment of detection.

See Also the Following Articles

Belching • Carbohydrate and Lactose Malabsorption • Giardiasis • Malabsorption • Over-the-Counter Drugs

Further Reading

Berk, J. E. (1995). Gas. *In* "Bockus Gastroenterology" (W. S. Haubrick and F. Schaffner, eds.), pp. 113–128. W. B. Saunders, Philadelphia, PA.

Levitt, M. D., and Duane, W. C. (1972). Floating stools—Flatus versus fat. *N. Engl. J. Med.* **286**, 973.

Strocchi, A., and Levitt, M. D. (1992). Intestinal gas. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., p. 157.

Sutalf, L. O., and Levitt, M. D. (1979). Follow-up of a flatulent patient. *Digest. Dis. Sci.* **24**, 654.

Folate Deficiency

JUDITH K. SHABERT
Harvard Medical School

meningocele/meningomelocele Failure of closure of the spinal cord in the lower back that results in a protrusion of skin containing spinal tissue.
neural tube defect Congenital defect of the neural tissue affecting the spinal cord, spine, brain, and skull, resulting in miscarriage, fetal or neonatal death, or disability.
vitamer Active analogue and isomer of a vitamin.
vitamin Organic substances necessary in minute quantities to maintain normal metabolism, usually working as cofactors by regulating biochemical processes.

Folate, or pteroylglutamate, is a member of a family of naturally occurring compounds that serve as essential vitamins in the normal metabolism of animals and humans. Folic acid, or pteroylglutamic acid, is the synthetically derived form of folate that is included in nutritional supplements and enriched foods; its absorption is about 50% more efficient than the 100 or so vitamers (e.g., folate compounds) that occur naturally in foods, particularly in leafy vegetables such as spinach.

BIOCHEMICAL PHYSIOLOGY

Absorption

Naturally occurring folate in food is present in the polyglutamate forms and its absorption is dependent on enzymatic deconjugation on the brush border of the jejunum to produce folyl monoglutamate. Folate-binding proteins sequester folate and it is transported across the membrane through an active transport process. When intake of the vitamin exceeds active transport mechanisms, it passively diffuses across the cell membrane.

The 5-methyltetrahydromonoglutamyl form of folate is the most abundant folate in the blood. However, it must be demethylated to tetrahydromonoglutamate for cellular assimilation. It undergoes polyglutamation in a series of steps that require one molecule of ATP for the addition of each glutamate. The polyglutamyl form represents both the functional and the storage forms of folate, with over one-half of the vitamin stores found in the liver.

Biochemical Role in Methylation

The primary function of folate is to serve as a cofactor for enzymatic reactions involving the transfer of single carbons. Folic acid is essential as a one-carbon donor (a methyl donor) and is required for the *de novo* synthesis of DNA and RNA by transferring a one-carbon unit for purine and thymidylate synthesis. Folate is also required for the *de novo* synthesis of the amino acids choline, serine, and methionine.

Folate and vitamins B_6 and B_{12} are necessary for one-carbon metabolism that synthesizes methionine from homocysteine and produce *S*-adenosylmethionine (SAM), the universal methyl donor. A deficiency of B_{12} produces an excess of 5-methyltetrahydrofolate, creating a functional folate deficiency known as the "methylfolate trap."

GENETICS

Several hereditary enzymatic defects have been reported related to the regeneration of methionine from homocysteine, which can result in elevated blood and urine concentrations of homocysteine. The homozygous state for defects in cystathionine synthase and methionine synthase results in a characteristic syndrome of mental retardation, detached lens of the eyes, osteoporosis, and severe atherosclerotic and/or thromboembolic events, resulting in premature death.

A more common defect substitutes a thymidine for cytosine on the gene that produces 5,10-methylenetetrahydrofolate reductase (MTHFR), an enzyme essential for the activation of folate. This defect (C677T mutation) results in the substitution of the amino acid valine for alanine in the protein, which greatly reduces the activity of the enzyme, resulting in a folic acid-deficient state. The homozygotic form occurs in approximately 12% of the North American population. Approximately 25% of the population is heterozygous for this gene mutation. In general, this enzymatic defect can be overcome by the administration of increased quantities of folic acid (see Fig. 1).

 Copyright 2004, Elsevier (USA). All rights reserved.

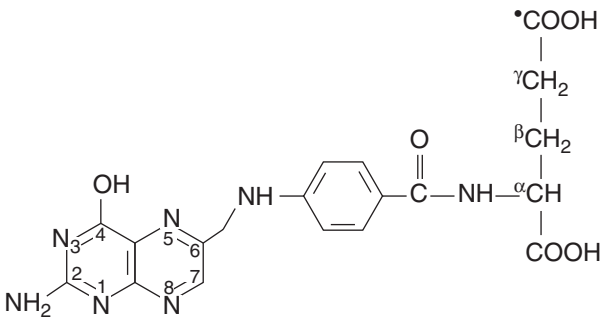


FIGURE 1 Folic acid.

REQUIREMENTS

The Dietary Reference Intake (DRI) established in 1998 for folate for children and adults recommends nearly double the recommendations established for folate in 1989. The recommendation for infants ranges from 65 to 80 μg/day. Children aged 1 to 13 years old require 150−300 μg/day, depending on their age. The folate recommendation for adult men and women is 400 μg/day; for pregnant women, the recommended intake is 600 μg/day, decreasing during lactation to 500 μg/day. This recommendation for an increased intake of folate was a result of extensive research that established the role of folate in the prevention of neural tube defects. Thus, it is recommended that all females who could become pregnant consume 400 μg/day of synthetic folic acid in addition to the amount of folate obtained in a well-balanced diet.

Folate requirements are increased during times of rapid cell turnover and increased metabolism,such as with tissue growth, infection, and recovery from catabolic states. Alcohol consumption increases folate requirements because alcohol interferes with folate utilization. As previously noted, folate exerts it effects in conjunction with other vitamins, specifically B_6 and B_{12}. Thus, the requirement for folate is predicated on satisfying requirements for these other nutritional factors.

SOURCES

Spinach and other green leafy vegetables are rich sources of folate along with other vegetables, fruit, Brewer's yeast, legumes, and especially liver. In the American diet, the most common food from which people obtain natural folate is orange juice. Some but not all breakfast cereals are fortified with folic acid to supply the total daily value of folate (400 μg/day) per serving. Folate is heat labile and up to 90% can be destroyed during cooking.

In 1996, the United States Food and Drug Administration mandated that, beginning in January of 1998, enriched cereal grain products, including wheat flour, pasta, breads, corn meal, and rice, be fortified with folic acid. Each 100 g of cereal grain is supplemented with 140 μg of folic acid. This mandate was instituted to increase folate consumption in women of childbearing age and thereby reduce the risk of having a pregnancy affected by neural tube birth defect.

DEFICIENCIES

It is estimated that at least 10% of the United States population has low folate stores, and this estimate may increase to 50% in disadvantaged populations. Deficiency can result from increased requirements, from decreased intake, from poor utilization, and from a functional folate deficiency associated with a deficiency of vitamin B_{12}.

Some anticonvulsant medication (dilantin, phenytoin, carbamazepine, and diphenylhydantoin), oral contraceptives, some antibiotics (trimethoprim, and triamterene), and excessive alcohol intake may also cause folate insufficiency. Due to the increased requirements of DNA and RNA for cell division, states of folate deficiency are most obvious in tissues with rapid cell turnover: hematopoietic cells, the gastrointestinal tract, the dermis, and germinal cells.

Hematopoietic System

Folate deficiency causes anemia, and stages of folate inadequacy are most evident in the hematopoietic system. Stage I (negative folate balance) is manifest with lowered plasma or serum folate <3ng/ml. In stage II folate deficiency, the red blood cell level is <160 ng/ml. Hypersegmented granulocytes (>3.5 lobes/cell) are seen with stage III folate deficiency. Stage IV deficiency is manifest by a full-blown anemia, with a hemoglobin level of <10 g/dl and macrocytic red blood cells.

Pregnancy

Approximately 4000 pregnancies are affected by neural tube defects (NTDs) annually in the United States. The major types of NTDs are anencephaly and spinal bifida. Half of all cases of NTDs produce anencephaly, which results in spontaneous abortion, fetal demise, or newborn death. Spina bifida affects cases of NTDs, with 10% of infants born with meningocele and 90% born with myelomeningocele.

Adequate folate intake in early pregnancy could prevent 40−70% of NTDs. The neural tube is formed between 17 and 30 days postconception (4−6 weeks

after the last menstrual period). Because approximately one-half of all pregnancies in the United States are unplanned, the U.S. Public Health Service recommended in 1992 that all women of childbearing age consume 400 μg/day of folic acid, even if pregnancy is not planned. The Food and Nutrition Board of the Institute of Medicine recognized that synthetic folic acid is more easily absorbed than folate from food, and in 1989 they recommended that 400 μg/day of synthetic folic acid be consumed by all women capable of becoming pregnant.

Despite the recommendation that women of childbearing age consume 400 μg/day of folic acid, a 1998 Gallop poll survey of women of childbearing age demonstrated that only 13% knew that folic acid reduces the risk for NTDs and only 7% knew that folic acid should be taken before conception; less than one-third of the women in this survey took folic acid on a regular basis.

The prevalence of NTDs in the United States has declined by 19% since mandatory fortification of cereal grain products was implemented in 1998.

Cardiovascular Disease

The association of hyperhomocysteinemia and atherosclerotic vascular disease was proposed in 1969. It is now accepted that hyperhomocysteinemia is a modifiable, independent risk factor for coronary artery disease (CAD) and peripheral vascular disease.

Adults with elevated blood homocysteine are 30% more likely to have premature vascular disease, compared to individuals with concentrations of <10 μmol/liter. In the Framingham Study, CAD and hyperhomocysteinemia showed an inverse correlation with serum folate concentration. Hyperhomocysteinemia has both genetic and nongenetic influences. Individuals who are homozygous for MTHFR and have low serum folate concentrations are at highest risk for elevated blood homocysteine. In experimental animals, elevated homocysteine levels cause an increase in coagulation of the blood and may produce a deleterious effect on the vascular endothelium.

The adverse effect of homocysteine on vessels appears to be a graded response. In patients with CAD, the risk of death 4–5 years after diagnosis is proportional to plasma homocysteine levels. Research has demonstrated a reduction of serum homocysteine in populations who receive folate and vitamins B_{12} and B_6. Despite these associations, no double-blind placebo controlled trial has demonstrated that the provision of folate, B_6, and B_{12} reduces the risk of atherosclerotic vascular disease, and the American Heart Association does not endorse the use of these vitamin supplements to treat elevated homocysteine.

Colon Cancer

Folate inadequacy has been implicated in the development of cancer, especially cancer of the colon. Folate supplementation above what is presently considered to be the basal requirement may have a protective effect. A 20-year epidemiological followup study of folate status and colon cancer risk was determined from the first National Health and Nutrition Evaluation Study (NHANES I). It was demonstrated that the risk for colon cancer in men decreases in a dose–response manner as folate intake increases, especially for men who do not drink alcohol and who eat a high-methionine diet. This association was not found in women. In the Nurses Health Study, 15 years of synthetic folic acid supplementation was associated with a decreased risk of colon cancer.

Inflammatory Bowel and Celiac Disease

Patients with both types of inflammatory bowel disease (IBD)—ulcerative colitis and Crohn's disease—are at risk for nutrient deficiencies, including folic acid. Reasons for this include increased folate requirements secondary to inflammation and catabolism, decreased absorption from malabsorption and inflammation of the gastrointestinal mucosa, and the effect of drugs on folate metabolism. Both methotrexate and sulfasalazine, used to treat IBD, are folate antagonists. There are numerous case reports of anemia and macrocytosis secondary to folate deficiency in patients on these medications. It appears that there may be an increased prevalence of a MTHFR variant in patients diagnosed with IBD compared to the general population. This may also explain the increased incidence of thrombosis in patients with IBD.

Individuals with celiac disease who have not been diagnosed or those who do not follow a gluten-free diet may develop folate deficiency secondary to malabsorption.

Seizure Disorders

The majority of antiseizure drugs are folate antagonists. Hyperhomocysteinemia with depressed folate concentrations can be seen in patients on antiseizure medications and it is associated with the MTHFR polymorphism. It is recommended that folic acid be prescribed along with antiseizure medications to prevent folate deficiency, although the provision of excessive folic acid may lower the seizure threshold. For some patients, individual titration of medication and folic acid may be necessary.

Cognitive Changes

Vitamin B_{12} and to a lesser extent folic acid have been associated with dementia and impairment of cognitive function in the elderly. Elevations in plasma homocysteine with normal serum folate have been demonstrated in patients with Alzheimer's disease. Cerebral spinal fluid folate concentrations are significantly lower in patients with late-onset Alzheimer's disease compared to age-matched controls. Improvement of cognitive function has been demonstrated in elderly patients with elevated homocysteine treated with folate and B_{12}. Folate is critical in brain metabolic pathways due to its role in the synthesis of choline and in the metabolism of neurotransmitters. The prevalence of borderline low or deficient serum or red blood cell folate levels in individuals diagnosed with depression is 15–38%. There is evidence that individuals refractory to antidepressant medication have inadequate folate status as well.

See Also the Following Articles

Celiac Disease • Cobalamin Deficiency • Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Dietary Reference Intakes (DRI): Concepts and Implementation • Vitamin B_{12}: Absorption, Metabolism, and Deficiency • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Centers for Disease Control web site. "Folic Acid Now." Retrieved June 19, 2003 at http://www.cdc.gov/ncbddd/folicacid.

Food Allergy

M. CECILIA BERIN
Mount Sinai School of Medicine, New York

anaphylaxis The term is used here specifically as systemic anaphylactic shock, affecting the cardiovascular system.

atopic A hereditary predisposition to immunoglobulin E-mediated allergic reactions against innocuous antigens (allergens).

RAST (radioallergosorbent test) Test that measures allergen-specific immunoglobulin E in the serum of patients.

skin-prick test A small needle or lancet is pressed through a commercial extract of a food into the epidermis. The presence of immunoglobulin E against the food tested results in a measurable wheal and flare reaction.

Food allergy is defined as an adverse immunologically mediated reaction to a food protein that not only can result in gastrointestinal symptoms but also can affect the skin, respiratory tract, and cardiovascular system. Food allergy can occur via different mechanisms, leading to very diverse clinical outcomes, from anaphylactic shock to enterocolitis. A limited number of foods are responsible for the majority of allergic reactions. The clinical spectrum and immune mechanisms of food allergy are discussed in this article.

INTRODUCTION

Adverse food reactions, of which food allergy is one category, can occur by toxic or nontoxic mechanisms. Toxic reactions are those that can occur in anyone, given a sufficient dose. Nontoxic reactions occur due to host factors, either by mounting an inappropriate immune response to a food protein (food allergy) or by nonimmune mechanisms, such as lactase deficiency resulting in intolerance to foods containing lactose. Thus, food allergy is only one mechanism by which adverse reactions to food can occur and refers only to those that are immunologically mediated.

Food allergies can be broadly grouped into two categories: immunoglobulin E (IgE)-mediated and

Copyright 2004, Elsevier (USA). All rights reserved.

Cognitive Changes

Vitamin B_{12} and to a lesser extent folic acid have been associated with dementia and impairment of cognitive function in the elderly. Elevations in plasma homocysteine with normal serum folate have been demonstrated in patients with Alzheimer's disease. Cerebral spinal fluid folate concentrations are significantly lower in patients with late-onset Alzheimer's disease compared to age-matched controls. Improvement of cognitive function has been demonstrated in elderly patients with elevated homocysteine treated with folate and B_{12}. Folate is critical in brain metabolic pathways due to its role in the synthesis of choline and in the metabolism of neurotransmitters. The prevalence of borderline low or deficient serum or red blood cell folate levels in individuals diagnosed with depression is 15–38%. There is evidence that individuals refractory to antidepressant medication have inadequate folate status as well.

See Also the Following Articles

Celiac Disease • Cobalamin Deficiency • Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Dietary Reference Intakes (DRI): Concepts and Implementation • Vitamin B_{12}: Absorption, Metabolism, and Deficiency • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Centers for Disease Control web site. "Folic Acid Now." Retrieved June 19, 2003 at http://www.cdc.gov/ncbddd/folicacid.

Food Allergy

M. CECILIA BERIN
Mount Sinai School of Medicine, New York

anaphylaxis The term is used here specifically as systemic anaphylactic shock, affecting the cardiovascular system.

atopic A hereditary predisposition to immunoglobulin E-mediated allergic reactions against innocuous antigens (allergens).

RAST (radioallergosorbent test) Test that measures allergen-specific immunoglobulin E in the serum of patients.

skin-prick test A small needle or lancet is pressed through a commercial extract of a food into the epidermis. The presence of immunoglobulin E against the food tested results in a measurable wheal and flare reaction.

Food allergy is defined as an adverse immunologically mediated reaction to a food protein that not only can result in gastrointestinal symptoms but also can affect the skin, respiratory tract, and cardiovascular system. Food allergy can occur via different mechanisms, leading to very diverse clinical outcomes, from anaphylactic shock to enterocolitis. A limited number of foods are responsible for the majority of allergic reactions. The clinical spectrum and immune mechanisms of food allergy are discussed in this article.

INTRODUCTION

Adverse food reactions, of which food allergy is one category, can occur by toxic or nontoxic mechanisms. Toxic reactions are those that can occur in anyone, given a sufficient dose. Nontoxic reactions occur due to host factors, either by mounting an inappropriate immune response to a food protein (food allergy) or by nonimmune mechanisms, such as lactase deficiency resulting in intolerance to foods containing lactose. Thus, food allergy is only one mechanism by which adverse reactions to food can occur and refers only to those that are immunologically mediated.

Food allergies can be broadly grouped into two categories: immunoglobulin E (IgE)-mediated and

Copyright 2004, Elsevier (USA). All rights reserved.

non-IgE-mediated hypersensitivity, also referred to as cell-mediated or delayed hypersensitivity. Symptoms can affect the gastrointestinal tract (nausea, vomiting, abdominal pain, malabsorption), the skin (urticaria, atopic dermatitis), the respiratory tract (bronchospasm, rhinitis), and the systemic circulation (anaphylactic shock). Although most rapid reactions that occur within minutes to 1–2 h after ingestion of the allergen are mediated by IgE, chronic disorders affecting the gastrointestinal tract include both IgE- and non-IgE-mediated mechanisms.

EPIDEMIOLOGY

Food allergy is most common in children and in atopic individuals. Although it is common for children to outgrow their hypersensitivity to food proteins, food allergies remain a significant problem in the adult population, with an estimated prevalence of 2%. In children under 3 years of age, 8% were found to have food allergy confirmed by oral allergen challenge. Food allergens affecting the pediatric and adult populations differ, with cow's milk, egg, wheat, peanut, and soy being dominant in early childhood and peanut, tree nuts, and shellfish constituting the main allergens of concern in the adult population. Peanut and tree nut allergies in particular are not frequently outgrown, providing an explanation for their dominance as allergens in the adult population. A strong genetic component has been observed in food allergy, as with other atopic disorders.

CLINICAL SPECTRUM

The first category of food allergic syndromes includes those that are IgE-mediated and immediate in onset. These include immediate gastrointestinal hypersensitivity reactions, with nausea, diarrhea, abdominal pain, or vomiting within minutes up to 1–2 h after ingestion of the allergen. This disorder is often associated with involvement of other organ systems, including the respiratory tract, skin, or cardiovascular system. IgE-mediated immediate reactions also include oral allergy syndrome, characterized as edema and pruritis limited to the oral cavity. These latter reactions usually occur in individuals allergic to pollens and who cross-react to proteins found in certain fresh fruits and vegetables.

The second group of food allergic disorders includes those that are IgE-associated, cell-mediated, and delayed in onset. These include the eosinophilic gastroenteropathies, characterized by eosinophilic infiltration of the gastrointestinal mucosa. Symptoms include postprandial nausea, diarrhea, vomiting, and protein-losing enteropathy. Atopic dermatitis in children is another disorder frequently associated with food allergy. Oral challenge exacerbates skin symptoms and allergen elimination can improve clinical symptoms. Food-specific T lymphocytes in eczematous lesions are thought to mediate this link between food allergens and skin symptoms. IgE can play a role by facilitating antigen uptake for presentation to T cells by keratinocytes.

Non-IgE-mediated disorders affecting the gastrointestinal tract include dietary protein-induced proctitis or proctocolitis, enteropathy, and enterocolitis. These disorders affect infants and are most often associated with reactivity to cow's milk or soy in the absence of detectable IgE by skin-prick test or RAST (radioallergosorbent test). Symptoms, depending on the site of involvement, include vomiting, diarrhea, rectal bleeding, growth failure, and hypoproteinemia. Histological findings reveal eosinophilia and eosinophilic abscesses in the colon or villous atrophy in the small intestine. A switch to hydrolysate formula in formula-fed infants or maternal dietary restriction in exclusively breast-fed infants is associated with clinical improvement. Celiac disease can also be regarded as a non-IgE-mediated food protein-induced enteropathy.

DIAGNOSIS

The gold standard for diagnosis of food allergy is the double-blind, placebo-controlled, food challenge, which eliminates the bias inherent in self-reporting of adverse reactions to food. The diagnostic approach begins with a detailed history that may suggest an IgE- or non-IgE-mediated reaction based on the suspected food, timing of reactions, and organ involvement. IgE against particular allergens can be measured by skin-prick testing or RAST, which measures allergen-specific IgE in serum. An elimination diet excluding potential allergens identified by laboratory tests should abolish symptoms and a follow-up oral challenge using individual allergens is used to identify those allergens that are responsible for clinical reactivity. This is important as the presence of allergen-specific IgE does not necessarily indicate clinical reactivity. There are currently no well-validated laboratory tests available for assessing non-IgE-mediated food allergies. Endoscopy can be used to support a diagnosis of food allergy in non-IgE-mediated allergic disorders, but does not identify the allergen. A diagnostic allergen elimination diet is required to identify potential allergens in non-IgE-mediated disorders and a follow-up oral challenge is used to confirm the diagnosis.

MANAGEMENT

Currently, allergen avoidance is the only effective strategy for management of food allergy. However,

accidental ingestion is common in patients with food allergy due to inadequate ingredient labeling, contamination during the manufacturing process of foods, or the lack of appropriate ingredient information in restaurants. Food allergies are the leading cause of generalized anaphylaxis seen in hospital emergency rooms and acute severe reactions require immediate management of respiratory and cardiovascular symptoms. There are no established preventative therapies, but emerging therapies under investigation include the use of anti-IgE antibodies, immunotherapy with mutated peptides or proteins, immunostimulatory oligonucleotide sequences, probiotics, and complementary medicines such as Chinese herbs.

ALLERGENS

A small number of allergens account for the majority of food-induced allergic reactions. Typically these allergens are 10–70 kDa in size and are heat-stable glycoproteins, which may make them resistant to digestion. Many food allergens have been identified, sequenced, and cloned. These have then been used to identify IgE-binding sites on the protein. One of the goals of this area of research is the development of mutated peptides or proteins that lack IgE-binding activity, but retain T-cell immunogenicity. These could therefore be used for immunotherapy without the high risk of anaphylaxis observed in food allergic patients.

MECHANISMS OF FOOD ALLERGY

A schematic illustrating the mechanisms involved in allergen-induced gastrointestinal dysfunction is shown in Fig. 1.

Immediate IgE-Mediated Gastrointestinal Hypersensitivity

Animal models of food allergy have typically relied on a model food allergen, such as ovalbumin (OVA) from chicken egg, which when administered systemically together with adjuvants such as alum or pertussis toxin induces the production of OVA-specific IgE antibodies in mice or rats. Antigen challenge leads to increased epithelial permeability and ion secretion (a driving force for water secretion and thus diarrhea) and changes in smooth muscle function, underlying motility changes. Antigen crosses the intestinal epithelium and cross-links IgE bound to the surface of mast cells in the lamina propria, inducing the release of mast cell mediators including histamine, serotonin, and prostaglandins. These mediators can then act directly on epithelial cells and smooth muscle cells to induce the immediate

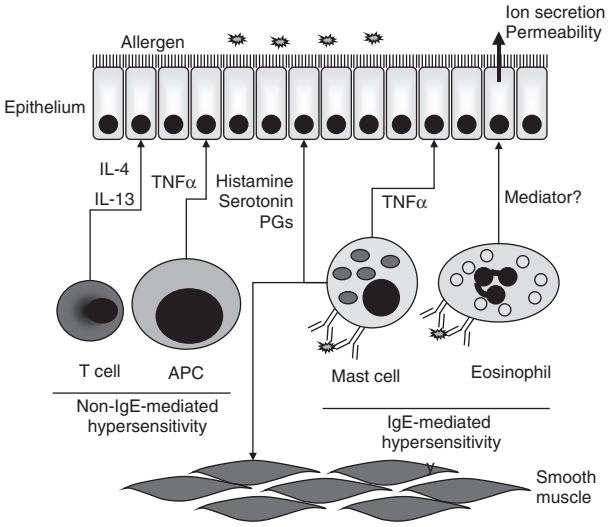


FIGURE 1 Schematic illustrating the mechanisms of allergen-induced gastrointestinal dysfunction in food allergy. Allergen present in the lumen must first cross the epithelial barrier. Interaction with IgE bound to the surface of mast cells or eosinophils can activate these cells to release mediators [histamine, serotonin, prostaglandins (PGs)] that can have an immediate action on intestinal epithelial or smooth muscle function. In addition, the release of cytokines such as TNFα from mast cells, or as yet unidentified mediators from eosinophils, can induce a late-phase reaction characterized by mucosal injury and cellular infiltration into the mucosa. Non-IgE-mediated reactions are believed to be mediated by presentation of antigen to T cells within the intestinal mucosa by antigen-presenting cells (APCs). Cytokines released by this interaction can induce chronic changes in mucosal architecture and epithelial function.

gastrointestinal symptoms observed after allergen challenge. Mast cell activation has been shown to occur in the intestine of patients with food allergy after administration of allergen directly in the duodenum.

It has been suggested that barrier function is reduced in patients with food allergy. In sensitized mice and rats, epithelial cells transport antigen from the lumen to the lamina propria by a mechanism involving IgE and its low-affinity receptor, CD23. This results in uptake of antigen in greater quantities and at a faster rate than in nonallergic animals, potentially explaining the rapid nature of intestinal hypersensitivity reactions. CD23 has also been demonstrated on human intestinal epithelial cells obtained from children with cow's milk enteropathy, suggesting that this mechanism may also play a role in human disease.

Chronic (or Late-Phase) IgE-Associated Allergic Reactions

IgE-mediated mast cell activation in animal models is associated with a late-phase reaction occurring days

after challenge, characterized by infiltration of mononuclear cells into the gastric or intestinal mucosa and mucosal damage. This infiltration of leukocytes is also commonly observed in the mucosa of patients with food allergy. Mast cells release tumor necrosis factor α (TNFα) after activation by IgE cross-linking and this has been shown to contribute to the pathological changes observed in the gastric tissue of mice after allergen challenge. TNFα is up-regulated in the mucosa of patients with food allergy, but its role in clinical symptoms has not been directly investigated in patients.

Eosinophils have also been shown to contribute to gastrointestinal pathology observed after allergen challenge in sensitized mice. Eosinophilia alone is not sufficient for gastrointestinal damage, however, suggesting that an activation step must occur. This, like mast cell activation, probably occurs via allergen/IgE triggering. The eosinophil mediators responsible for gastrointestinal damage have not yet been identified.

Non-IgE-Mediated Food Allergy

The mechanism of non-IgE-mediated reactions to food are the least understood of the gastrointestinal allergic reactions. Peripheral blood mononuclear cells isolated from milk allergic children release TNFα after stimulation with milk proteins and this has been shown to reduce epithelial barrier function *in vitro*. In addition, in children with food protein-induced enterocolitis, increased mucosal TNFα is associated with villous atrophy. TNFα is well known to play a crucial role in other inflammatory diseases such as inflammatory bowel disease or rheumatoid arthritis and it is conceivable that it plays a role in chronic gastrointestinal dysfunction in food allergy, whether from IgE-triggered mast cells or IgE-independent release from monocytes or macrophages.

T lymphocytes are thought to orchestrate allergic diseases through the production of cytokines such as interleukin-4 (IL-4) and IL-13, which contribute to IgE production, or IL-5, which influences eosinophil survival. T cells may also contribute to food allergic reactions by direct modulation of gastrointestinal physiology. Cow's milk-specific T-cell lines grown from the gastrointestinal mucosa of patients with milk-induced enteropathy or eosinophilic gastroenteritis are predominantly of the Th2 subtype, releasing elevated levels of IL-5 and IL-13 but not interferon-γ. In mice, the transfer of T cells from a sensitized mouse to a naive mouse is sufficient to induce diarrhea in response to an allergen challenge. This may be due to the release of IL-4 and IL-13, which have been shown to have direct effects on intestinal epithelial ion secretion and barrier function *in vivo*.

The possibility that IgE-mediated reactions could occur in the absence of detectable serum IgE or positive skin-prick test reactions should also be considered. IgE-bearing cells and activated eosinophils have been shown to be elevated in the lamina propria of patients with active hypersensitivity symptoms in the gastrointestinal tract, despite the lack of allergen-specific IgE or positive skin-prick tests. Therefore, the inability to detect IgE in serum or skin does not rule out the possibility of IgE-mediated reactions occurring in the intestinal mucosa, especially as IgE is produced locally in the mucosal tissue.

See Also the Following Articles

Cow Milk Protein Allergy • Eosinophilic Gastroenteritis • Food Intolerance

Further Reading

Hogan, S. P., Foster, P. S., and Rothenberg, M. E. (2002). Experimental analysis of eosinophil-associated gastrointestinal diseases. *Curr. Opin. Allergy Clin. Immunol.* **2**, 239–248.

Hogan, S. P., Mishra, A., Brandt, E. B., Royalty, M. P., Pope, S. M., Zimmermann, N., Foster, P. S., and Rothenberg, M. E. (2001). A pathological function for eotaxin and eosinophils in eosinophilic gastrointestinal inflammation. *Nat. Immunol.* **2**, 353–360.

Joshi, P., Mofidi, S., and Sicherer, S. H. (2002). Interpretation of commercial food ingredient labels by parents of food-allergic children. *J. Allergy Clin. Immunol.* **109**, 1019–1021.

Kweon, M. N., Yamamoto, M., Kajiki, M., Takahashi, I., and Kiyono, H. (2000). Systemically derived large intestinal CD4(+) Th2 cells play a central role in STAT6-mediated allergic diarrhea. *J. Clin. Invest.* **106**, 199–206.

Li, X. M., and Sampson, H. A. (2002). Novel approaches for the treatment of food allergy. *Curr. Opin. Allergy Clin. Immunol.* **2**, 273–278.

Lin, X. P., Magnusson, J., Ahlstedt, S., Dahlman-Hoglund, A., Hanson, L. L. A., Magnusson, O., Bengtsson, U., and Telemo, E. (2002). Local allergic reaction in food-hypersensitive adults despite a lack of systemic food-specific IgE. *J. Allergy Clin. Immunol.* **109**, 879–887.

Sampson, H. A., Sicherer, S. H., and Birnbaum, A. H. (2001). American Gastroenterological Association technical review on the evaluation of food allergy in gastrointestinal disorders. *Gastroenterology* **120**, 1026–1040.

Sicherer, S. H. (2002). Food allergy. *Lancet* **360**, 701–710.

Yang, P. C., Berin, M. C., Yu, L., and Perdue, M. H. (2001). Mucosal pathophysiology and inflammatory changes in the late phase of the intestinal allergic reaction in the rat. *Am. J. Pathol.* **158**, 681–690.

Yu, L. C., and Perdue, M. H. (2001). Role of mast cells in intestinal mucosal function: Studies in models of hypersensitivity and stress. *Immunol. Rev.* **179**, 61–73.

Shapiro, R. L., Hatheway, C., *et al.* (1998). Botulism in the United States: A clinical and epidemiologic review. *Annu. Intern. Med.* **129**(3), 221–228.
Sirinavin, S., and Garner, P. (2000). Antibiotics for treating *Salmonella* gut infections. *Cochrane Database Syst. Rev.* **2**, CD001167.
Wong, C. S., Jelacic, S., *et al.* (2000). The risk of the hemolytic–uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N. Engl. J. Med.* **342**(26), 1930–1936.

Food Intolerance

M. Cecilia Berin
Mount Sinai School of Medicine, New York

adverse food reaction Abnormal physiologic response to food ingestion, due to either food intolerance or food allergy.
food allergy (hypersensitivity) Immune-mediated reaction to food.
food intolerance Abnormal physiologic response to an ingested food or food additive; intolerance has not been proved to be immunologic in nature and falls under the umbrella term, adverse food reaction.

Food intolerance refers to nonimmunologic adverse reactions to ingested foods; the physiologic reactions resolve after dietary elimination of the food substances and reoccur on subsequent challenges with the foods. This category of adverse food reactions includes well-defined disorders such as enzyme deficiencies (i.e., lactose intolerance) and responses to pharmacological agents in food (i.e., tyramine in cheese or wine). In addition, intolerance includes non-immune-mediated reactions to food additives. Symptoms of food intolerance can appear identical to those arising from immune-mediated food allergy, despite the difference in underlying mechanisms. Therefore, the two categories of disorders are often confused.

ENZYME DEFICIENCY

The most commonly reported food intolerance is due to a deficiency in the brush border enzyme lactase, resulting in intolerance to lactose, the common sugar found in milk (see Table I). The inability to digest and absorb lactose leaves an osmotic load in the small intestine that draws water into the intestinal lumen and causes diarrhea. Patients with lactose intolerance can typically tolerate small amounts of lactose, whereas those with cow's milk allergy (caused by an immune reaction to proteins in milk) must practice complete avoidance of products containing milk proteins. Food intolerance can also occur due to nonspecific loss of brush border enzymes secondary to inflammatory events in the small intestine, where epithelial cells are damaged (such as in celiac or Crohn's disease).

Food-induced symptoms that do not involve the gastrointestinal tract can also occur because of enzyme deficiencies, such as glucose-6-phosphate dehydrogenase (G6PD) deficiency. G6PD deficiency results in hemolytic anemia after ingestion of fava beans ("favism"). Fava beans contain two potent oxidants that damage erythrocyte membranes in the absence of the G6PD enzyme. This genetic enzyme deficiency is not common in the Caucasian North American population, but is common in the Middle East and Africa.

FOOD ADDITIVES AND PRESERVATIVES

Intolerance to a number of food additives has been reported, but only a few cases have been verified by appropriately controlled studies, including an elimination diet and placebo-controlled challenge. The list of causative agents includes flavorings and preservatives (sulfites, monosodium glutamate, benzoic acid) and dyes

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Causes of Food Intolerance

Cause	Food	Symptom
Enzyme deficiency		
Disaccharidase deficiency		
Lactase deficiency	Lactose or milk-containing foods	Diarrhea, abdominal pain, flatulence
Sucrose–isomaltase deficiency	Sucrose-containing foods	Diarrhea
Trehalase deficiency	Mushrooms	Diarrhea, flatulence
Glucose-6-phosphate dehydrogenase	Fava beans	Hemolytic anemia
Pharmacologic agents		
Histamine	Cheese, wine, fermented foods (sauerkraut)	Urticaria, abdominal pain, diarrhea, nausea
Tyramine	Fermented cheeses, chocolate	Migraine
Phenylethylamine	Chocolate, cheese, red wine	Migraine
Additives		
Tartrazine (FD & C Yellow No. 5)	Various foods	Exacerbation of urticaria
Sulfites	Various foods	Bronchospasm, anaphylaxis
Monosodium glutamate	Various foods	Headache, nausea, abdominal pain
Gastrointestinal disease		
Cystic fibrosis (pancreatic insufficiency), gallbladder disease, enteropathy (celiac disease, Crohn's disease)	Various foods	Diarrhea due to malabsorption/maldigestion

(tartrazine). Sulfites, which are antimicrobial and prevent oxidation or browning of food, have been shown to induce bronchospasm in a subset of patients with severe steroid-dependent asthma. The prevalence of sulfite intolerance has been estimated at between 2 and 8% of the adult asthmatic population. Tartrazine, also known as FD & C Yellow No. 5, is one of a family of azo dyes. Tartrazine has been shown to exacerbate skin symptoms after ingestion in a subset of patients with chronic urticaria. Monosodium glutamate, a flavoring agent, has also been reported to induce symptoms, including headache, nausea, flushing, and abdominal pain.

PHARMACOLOGICAL AGENTS

Vasoactive amines compose the largest class of substances responsible for pharmacological reactions to foods. These include histamine, phenylethylamine, and tyramine. This category of food intolerance includes agents that, given in a sufficient dose, induce effects in individuals regardless of host status. Individual variation in sensitivity to histamine may be related to expression levels of the enzyme diamine oxidase, which degrades histamine. Histamine-rich foods include cheese, wine, fish, sausages, and fermented food products such as sauerkraut. These foods can induce food allergy-like symptoms, including nausea and diarrhea, in susceptible individuals. Tyramine (found in chocolate, fermented cheeses, and pickled herring) and phenylethylamine (found in chocolate, cheese, and red wine) can also induce symptoms, including migraines, in susceptible individuals. Susceptibility to tyramine is enhanced in patients taking monoamine oxidase (MAO) inhibitors.

FOOD INTOLERANCE IN INFLAMMATORY BOWEL DISEASE AND IRRITABLE BOWEL SYNDROME

Patients with inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) self-report food intolerance at a much higher rate than do normal controls, and many restrict their diet accordingly. There is currently no conclusive evidence demonstrating that food-induced immunological mechanisms play a role in the pathophysiology of these diseases. Food intolerance may exacerbate symptoms, but there is conflicting evidence on the usefulness of elimination diets in the management of IBD and IBS. Certain intolerances, such as to lactose, may occur secondary to inflammatory damage of the small intestine. The high rate of self-report of food intolerance indicates the awareness of food as a possible trigger of symptoms, and this is an area of research that requires further investigation.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Celiac Disease • Colitis, Ulcerative • Cow Milk Protein Allergy • Crohn's Disease • Food Allergy • Hereditary Fructose Intolerance • Irritable Bowel Syndrome

Further Reading

Alpers, D. H., Stenson, W. F., and Bier, D. M. (2002). "Manual of Nutritional Therapeutics." Lippincott Williams & Wilkins, Philadelphia.

Center for Food Safety and Applied Nutrition, Food and Drug Administration. (2003). Web site: http://www. cfsan. fda.gov.

Wilson, S. H. (2000). Medical nutrition therapy for food allergy and food intolerance. *In* "Krause's Food, Nutrition, and Diet Therapy" (L. K. Mahan and S. Escott-Stump, eds.), 10th Ed., pp. 912–934. W. B. Saunders, Philadelphia.

Food Poisoning

SUZANNE M. MATSUI

VA Palo Alto Health Care System and Stanford University

enteritis necroticans Severe necrotizing disease of the small intestine, associated with high mortality, caused by *Clostridium perfringens* type C; see also pigbel.

enterotoxin Bacterial toxin that exerts its effect by stimulating net fluid secretion by intestinal epithelial cells, without damaging the cells; contrast with other types of bacterial toxins, such as cytoskeletal-altering toxins, cytotoxins, and toxins with immune- or nerve-stimulating activity.

food-borne disease outbreak Two or more cases of a similar illness that occurs after eating a common food.

norovirus Recently adopted name to designate the genus of the viral family *Caliciviridae*; the viruses in this genus were previously called Norwalk-like viruses or small round structured viruses.

pigbel Severe necrotizing disease of the small intestine, associated with high mortality, caused by *Clostridium perfringens* type C, and named after an illness that developed in New Guinea natives after large feasts of inadequately cooked pork.

Food poisoning is an acute illness that develops from less than 1 h to 15 or more hours after the ingestion of food contaminated with bacterial, viral, or parasitic pathogens or toxins (biological or chemical). Although the clinical manifestations of food poisoning are generally gastrointestinal, neurological symptoms may occur with certain types of ingestion. The spectrum of food-borne illness in the United States has changed over time, in association with globalization of food production and broadening of consumer tastes, which have led to greater exposure to food-borne pathogens and chemicals.

INTRODUCTION

In the United States, the morbidity and mortality of food-borne illness have been estimated to be in the range of 76 million illnesses and 5000 deaths, respectively. Food-borne diseases are tracked by the Centers for Disease Control and Prevention through the Food-borne-Disease Outbreak Surveillance System, which has been in place since 1966. During the period 1993–1997, 2751 outbreaks of food-borne illness involving 86,058 persons were documented. For the outbreaks in which a pathogen could be identified (32%), 75% of the outbreaks and 86% of the cases were attributable to bacterial pathogens. In contrast, no etiologic agent could be identified in the majority of outbreaks (68%) and the longer incubation period ($\geq$15 h) associated with these outbreaks suggests a viral cause. These data do not include outbreaks that occurred on cruise ships, those in which the suspect food was consumed outside the United States, and those transmitted through contaminated drinking water (separately tracked). Despite these limitations and the fact that these data represent only a fraction of all outbreaks, the medical and economic impact of food-borne disease cannot be underestimated.

BACTERIAL FOOD POISONING

Bacterial food poisoning is generally attributable to 12 bacteria: *Clostridium perfringens, Staphylococcus aureus,*

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Alpers, D. H., Stenson, W. F., and Bier, D. M. (2002). "Manual of Nutritional Therapeutics." Lippincott Williams & Wilkins, Philadelphia.

Center for Food Safety and Applied Nutrition, Food and Drug Administration. (2003). Web site: http://www. cfsan. fda.gov.

Wilson, S. H. (2000). Medical nutrition therapy for food allergy and food intolerance. *In* "Krause's Food, Nutrition, and Diet Therapy" (L. K. Mahan and S. Escott-Stump, eds.), 10th Ed., pp. 912–934. W. B. Saunders, Philadelphia.

Food Poisoning

SUZANNE M. MATSUI

VA Palo Alto Health Care System and Stanford University

enteritis necroticans Severe necrotizing disease of the small intestine, associated with high mortality, caused by *Clostridium perfringens* type C; see also pigbel.

enterotoxin Bacterial toxin that exerts its effect by stimulating net fluid secretion by intestinal epithelial cells, without damaging the cells; contrast with other types of bacterial toxins, such as cytoskeletal-altering toxins, cytotoxins, and toxins with immune- or nerve-stimulating activity.

food-borne disease outbreak Two or more cases of a similar illness that occurs after eating a common food.

norovirus Recently adopted name to designate the genus of the viral family *Caliciviridae*; the viruses in this genus were previously called Norwalk-like viruses or small round structured viruses.

pigbel Severe necrotizing disease of the small intestine, associated with high mortality, caused by *Clostridium perfringens* type C, and named after an illness that developed in New Guinea natives after large feasts of inadequately cooked pork.

Food poisoning is an acute illness that develops from less than 1 h to 15 or more hours after the ingestion of food contaminated with bacterial, viral, or parasitic pathogens or toxins (biological or chemical). Although the clinical manifestations of food poisoning are generally gastrointestinal, neurological symptoms may occur with certain types of ingestion. The spectrum of food-borne illness in the United States has changed over time, in association with globalization of food production and broadening of consumer tastes, which have led to greater exposure to food-borne pathogens and chemicals.

INTRODUCTION

In the United States, the morbidity and mortality of food-borne illness have been estimated to be in the range of 76 million illnesses and 5000 deaths, respectively. Food-borne diseases are tracked by the Centers for Disease Control and Prevention through the Food-borne-Disease Outbreak Surveillance System, which has been in place since 1966. During the period 1993–1997, 2751 outbreaks of food-borne illness involving 86,058 persons were documented. For the outbreaks in which a pathogen could be identified (32%), 75% of the outbreaks and 86% of the cases were attributable to bacterial pathogens. In contrast, no etiologic agent could be identified in the majority of outbreaks (68%) and the longer incubation period ($\geq$15 h) associated with these outbreaks suggests a viral cause. These data do not include outbreaks that occurred on cruise ships, those in which the suspect food was consumed outside the United States, and those transmitted through contaminated drinking water (separately tracked). Despite these limitations and the fact that these data represent only a fraction of all outbreaks, the medical and economic impact of food-borne disease cannot be underestimated.

BACTERIAL FOOD POISONING

Bacterial food poisoning is generally attributable to 12 bacteria: *Clostridium perfringens, Staphylococcus aureus,*

Copyright 2004, Elsevier (USA). All rights reserved.

Bacillus cereus, Vibrio cholerae /Vibrio parahaemolyticus, Salmonella, Clostridium botulinum, Shigella, toxigenic Escherichia coli, and certain species of Campylobacter, Yersinia, Listeria, and Aeromonas. In addition, group A *Streptococcus* has been implicated in a few outbreaks. Toxin-mediated illness, with short incubation times, as exemplified by the first three pathogens listed, is discussed below.

Clostridium perfringens

C. perfringens is a gram-positive, spore-forming, obligate anaerobe that is normally found in the intestinal flora of humans and animals and in the soil.

The typical episode of food poisoning by *C. perfringens* is characterized by watery diarrhea and severe crampy abdominal pain that develop 8 to 24 h after the suspect meal. Vomiting is usually not a prominent symptom. The illness usually resolves within 24 to 36 h after onset, with no sequelae. Although fatalities are rare, they may occur in debilitated and/or hospitalized patients.

This form of food poisoning is caused by a heat-labile, 35 kDa enterotoxin produced by *C. perfringens* type A. The enterotoxin is a structural protein of the spore coat and is produced during sporulation. Unlike cholera toxin, the *C. perfringens* enterotoxin has greatest activity in the ileum and can inhibit glucose transport, damage intestinal epithelial cells, and induce protein loss.

High attack rates (40 to 50 persons per outbreak) are common. Outbreaks usually are associated with large group gatherings and institutional settings in which precooked meat (beef, chicken, or turkey) that requires reheating is served. When food is cooked in large batches for these gatherings, spores of *C. perfringens* may survive and germinate as the food is cooled. If the food is not reheated to a temperature sufficiently high to kill the organisms, food poisoning may develop among individuals who ingest a large number of organisms and a large amount of toxin.

In contrast, *C. perfringens* type C causes a more severe illness called enteritis necroticans (Darmbrand) or pigbel, which has a high mortality rate ($\sim$40%). Historically, eating rancid meat (post-World War II Germany) and large amounts of poorly cooked pork (New Guinea) has been associated with this illness. Although rare in the United States, outbreaks have been associated with eating chitterlings (hog intestines).

Although the enterotoxin elaborated by *C. perfringens* type C is similar to that produced by type A, it induces a much more severe clinical course that includes bloody diarrhea, necrotizing intestinal damage, and intestinal perforation.

Staphylococcus aureus

S. aureus was the leading cause of food poisoning in the United States until $\sim$1973, but now ranks third. The illness is associated with coagulase-positive strains that elaborate heat-resistant enterotoxins A, B, C1, C2, C3, D, and/or E (28 to 34 kDa). Most outbreaks are caused by strains that produce enterotoxin A alone or together with enterotoxin D. An immunoassay that detects toxin in food and *in vitro* is useful for diagnosis in cases where the organism has been killed, but the preformed, heat-stable toxin remains in the sample. Phage typing has also been useful in outbreak investigations where the organism may be cultured from clinical specimens (emesis or feces). Phage type III alone, or in combination with phage type I, is the most common type found in outbreaks. Matching the phage type between the suspect food and affected individuals, or food handler(s), helps to establish the diagnosis and lines of transmission.

Staphylococci grow well in foods that have a high salt (e.g., canned meats) or sugar/protein (e.g., custards and creams) content and can be passed to prepared food by food handlers who carry toxin-producing *S. aureus*. It is estimated that 20 to 50% of healthy persons carry *S. aureus* on their skin, in their nose, or in their intestinal tract.

After a short incubation period (1 to 6 h), affected individuals develop nausea, vomiting, abdominal cramping, and later, diarrhea. Profuse vomiting may result in dehydration and metabolic alkalosis, which may require medical attention. Despite a stormy acute course, the illness resolves completely in 24 to 48 h and fatalities are rare. Many patients do not seek care because the symptoms resolve relatively quickly.

Bacillus cereus

B. cereus, an aerobic, spore-forming, gram-positive rod is found in soil and water globally, in most raw foods, and in human carriers (10 to 40%). The spores are heat-resistant. *B. cereus* is not a common cause of food poisoning in the United States.

B. cereus is associated with two distinct toxin-mediated types of food poisoning: the emetic syndrome, which has a short incubation period (range 1 to 6 h), and the diarrheal syndrome, which has a longer incubation period (range 8 to 16 h). The type of toxin elaborated depends on the type of food contaminated with *B. cereus*, rather than the strain of the organism.

The emetic toxin is produced in foods such as fried rice left at room temperature, whereas proteinaceous foods are usually associated with the diarrheal syndrome. Typically within 2 h of ingestion of the

preformed emetic toxin (5 to 10 kDa), affected individuals experience vomiting and abdominal cramping. Approximately one-third of them go on to develop diarrhea, but in general, the illness is mild, with prompt resolution in 8 to 10 h. *B. cereus* infection can be proven by isolating 10^5 organisms per gram of suspect food. However, if the food was reheated before consumption, the heat-stable emetic toxin may persist and cause illness in the absence of viable bacteria.

The longer incubation period and large inoculum (10^6 organisms per gram of food) required to induce the diarrheal syndrome suggest intestinal colonization, rather than preformed toxin alone, as the likely mechanism of disease. The diarrheal enterotoxin is a 38 to 46 kDa heat-labile protein. The majority of the patients have diarrhea and cramps. Less than 25% experience vomiting. The illness lasts 20 to 36 h.

DIFFERENTIAL DIAGNOSIS

The onset of gastrointestinal symptoms within 1 to 6 h after eating suggests an illness caused by the ingestion of a preformed toxin or a chemical irritant. If the incubation period is less than 1 h, heavy metal poisoning should be suspected and urine samples collected to screen for toxic chemicals.

Symptoms of zinc poisoning (nausea, abdominal cramps, metallic taste, headache, dizziness, and chills) have been reported to occur from 5 min to 2 h after consuming a beverage containing emetic doses of zinc (225 to 450 mg for adults). The beverage was acidic and had been stored in a galvanized container. The zinc in the galvanized metal was converted to easily absorbable zinc salts by contact with the acidic beverage. The symptoms resolved after vomiting of the zinc salts. The amount of vomiting was generally less than with food poisoning due to a preformed toxin.

Vomiting after drinking accidentally overfluoridated water has been reported to occur within 7 min (median). In the stomach, fluoride combines with hydrogen ions to form the gastric irritant hydrofluoric acid. Other symptoms include nausea, abdominal pain, and diarrhea. Fluoride poisoning also causes significant hyperkalemia and hypocalcemia, which can potentially trigger cardiac dysrhythmias and result in death.

In benign mushroom poisoning, gastrointestinal symptoms (nausea, vomiting, diarrhea, abdominal cramping) begin from a few minutes to 2 to 3 h after consumption. In contrast, the onset of gastrointestinal symptoms is 6 to 12 h from ingestion in the more serious types of mushroom poisoning that involve amatoxins or monomethylhydrazine and the symptoms resolve before progression to life-threatening hepatic or renal disease. Incubation period alone, however, is not a reliable predictor of benign disease, since more than one type of mushroom is usually consumed in these situations.

If a patient develops acute, severe epigastric pain, nausea, and vomiting between 1 and 12 h after eating raw fish, gastric anisakiasis should be suspected. Symptoms usually subside after the nematode larvae are expelled from the stomach by regurgitation. In some cases, the worm may invade the gastric mucosa and cause perforation or cause a local granulomatous response and chronic symptoms. The worm can be seen and removed by upper endoscopy.

Viral gastroenteritis, such as that caused by noroviruses, also should be considered in the differential diagnosis, because it shares a similar constellation of symptoms with food poisoning, including the abrupt and prominent development of acute vomiting and/or diarrhea and it may be related to the ingestion of contaminated seafood. The incubation period for norovirus infection is longer (ranges from 10 to 51 h), the secondary infection rate is significantly higher, and symptoms last longer (1 to 5 days).

Finally, food additives, such as monosodium glutamate (MSG), and natural toxins associated with fish and shellfish consumption should be considered if extragastrointestinal and neurological symptoms are reported. The effect of MSG has a quick onset (<2 h following ingestion) and is characterized by a burning sensation in the chest or chest pressure, diaphoresis, and lacrimation.

Syndromes associated with fish and shellfish include ciguatera fish poisoning (most common; associated with toxins produced by the dinoflagellate *Gambierdiscus toxicus* and other benthic algae), neurotoxic shellfish poisoning (associated with red tides caused by *Gymnodinium breve*, which produces brevotoxin), diarrhetic shellfish poisoning (okadaic acid and dinophysistoxin-1 toxins produced by *Dinophysis fortii* or *D. acuminata*; cases mostly from Japan, some from Europe, and none from the United States), amnesic shellfish poisoning (domoic acid produced by *Pseudonitzschia pungen*, a diatom), and scromboid fish poisoning (high levels of free histamine due to bacterial decomposition of fish tissues after capture; may occur without overt evidence of spoilage).

TREATMENT AND PREVENTION

Most food-borne illnesses resolve without specific therapy. General supportive measures are aimed at maintaining adequate hydration and avoiding the use of anti-motility drugs in cases of inflammatory diarrhea.

Food poisoning is preventable. First, proper preparation of food is essential. For example, a common practice is to cook or steam shellfish until the shells just begin to part (usually less than 1 min of cooking time). If cooking is terminated at that time, the internal temperature of the shellfish meat will likely not have reached the 85 to 90°C maintained for 1.5 min that is required to inactivate hepatitis A virus and noroviruses that may be concentrated in the shellfish.

Second, storage of food at proper holding temperatures will help to prevent many cases of staphylococcal, *B. cereus*, and *C. perfringens* outbreaks. Prompt and adequate refrigeration (<40°F) for cold foods and maintenance of temperature (>140°F) for foods served hot are recommended to prevent the growth of bacterial pathogens.

Third, pathogens such as noroviruses, hepatitis A virus, *Salmonella typhi*, *Shigella* species, *S. aureus*, and *Streptococcus pyogenes* are more likely than other pathogens to be transmitted by an infected food worker who handles the food before it is served. It is, therefore, important for food handlers to practice good handwashing and maintain good personal hygiene, work with food on appropriately disinfected surfaces with clean implements, and abide by workplace policies that prevent symptomatic (or recently symptomatic) employees from working in food preparation.

In conjunction with these practical guidelines, laws have been enacted to promote the safety of domestic and imported food. The application of sound practices at every level of food production will help to prevent incidents such as the contamination of meat products during slaughter or processing or the contamination of fruits and vegetables by polluted water for washing.

See Also the Following Articles

Bacterial Toxins • *Campylobacter* • Diarrhea, Infectious • Foodborne Disease • Food Safety • Gastroenteritis • Necrotizing Enterocolitis • *Salmonella* • *Shigella* • *Yersinia*

Further Reading

De Saxe, M., Coe, A. W., and Wieneke, A. A. (1982). The use of phage typing in the investigation of food poisoning caused by *Staphylococcus aureus* enterotoxins. *Soc. Appl. Bacteriol. Tech. Ser.* **17**, 173.

Mead, P. S., Slutsker, L., Dietz, V., *et al.* (1999). Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5**, 607–625.

Millard, J., Appleton, H., and Parry, J. V. (1987). Studies on heat inactivation of hepatitis A virus with special reference to shellfish. *Epidemiol. Infect.* **98**, 397–414.

Olsen, S. J., MacKinnon, L. C., Goulding, J. S., Bean, N. H., and Slutsker, L. (2000). Surveillance for food-borne-disease outbreaks—United States, 1993–1997. *Morb. Mortal. Wkly. Rep.* **49**, 1–63.

Osterholm, M. T. (1997). Cyclosporiasis and raspberries—Lessons for the future. *N. Engl. J. Med.* **336**, 1597.

Shandera, W. X., Tacket, C. O., and Black, P. A. (1983). Food poisoning due to *Clostridium perfringens* in the United States. *J. Infect. Dis.* **147**, 167.

Terranova, W., and Blake, P. A. (1978). *Bacillus cereus* food poisoning. *N. Engl. J. Med.* **298**, 143.

Food Safety

JUDITH K. SHABERT
Harvard Medical School

food-borne illnesses Diseases, usually either infectious or toxic in nature, caused by agents that enter the body through the ingestion of food.

food-borne pathogens Microorganisms that contaminate food destined for human consumption; *Campylobacter*, *Salmonella*, and *Shigella* are the most frequently implicated bacteria involved in food-borne illnesses.

polychlorinated biphenyls Colorless and odorless chemicals, once widely used in electrical equipment; now banned from use, remain in the environment and contaminate the food chain, primarily through waterways.

Food-borne diseases are a growing public health problem. In industrialized countries, the percentage of people suffering from food-borne diseases each year has been reported to be as high as 30%. In the United States, estimates of 6.5 to 76 million cases of food-borne diseases occur annually, resulting in 325,000 hospitalizations and 5000–9000 deaths. The costs from these diseases and lost productivity are estimated to be $6.5–35 billion annually.

OVERVIEW

The United States has one of the safest food supplies in the world, with an interlocking system for monitoring production, importation, and distribution of food. Food and water safety in the United States is regulated by a number of local, state, and national governmental agencies. Within the Department of Health and Human Services, the Food and Drug Administration (FDA) is the primary regulator for the safety of all domestic and imported food and bottled water sold through interstate commerce. Meat, poultry, and frozen, dried, and liquid eggs are under the authority of the U.S. Department of Agriculture; the labeling and safety of alcoholic beverages (above 7% alcohol) are regulated by the Department of the Treasury's Bureau of Alcohol, Tobacco, and Firearms, and the safety of drinking water, except bottled water, is regulated by the Environmental Protection Agency. In addition, the National Wildlife and Fisheries Agency provides a fee-for-service inspection of fishery production.

Local and state public health departments and the Centers for Disease Control and Prevention (CDC) investigate sources of food-borne disease outbreaks. Food produced and consumed within a state is regulated by state agencies. This becomes particularly important in the regulation and safety of animal foods for human consumption, such as livestock and fish.

The World Health Organization indicates that cases of food-borne diseases may be 300–350 times more frequent than reported cases. Although toxic metals and other environmental hazards play a role in food-borne diseases, pathologic microbes pose the greatest risk to the nation's food supply. Estimates of the incidence of microbial causes of food-borne illness are complicated by newly emerging pathogens, which may not be tested for when suspected outbreaks occur. There is also increased antibiotic resistance in known pathogens.

Persons at greatest risk for food-borne illness are the elderly, infants, and preschool-age children, persons on immunosuppressive medications, those undergoing radiation therapy, and persons infected with the human immunodeficiency virus.

INFECTIOUS CAUSES OF FOOD- AND WATER-BORNE ILLNESS

In the United States, the most common infectious causes of food-borne illness, in order of prevalence, are *Campylobacter*, *Salmonella*, and *Shigella*; poultry is the food most often contaminated with disease-producing organisms. It has been estimated that 60% or more of raw poultry sold in retail stores carries some disease-producing bacteria.

Antimicrobial resistance is emerging as a major public health concern and this is primarily due to the overprescription of antibiotics in clinical medicine. However, the most likely cause of antibiotic resistance in food pathogens is the use of antibiotics in food-producing animals. Antibiotic resistance is present in the microflora of farm animals and of farm animal food products. In the United States, an estimated 50% of all antibiotics manufactured is provided to animals. Antibiotics are prescribed for specific infectious diseases in food animals as they are in humans, but they are also given in subtherapeutic doses as growth promoters and to prevent widespread infection in herds and flocks.

 Copyright 2004, Elsevier (USA). All rights reserved.

In 1995, fluoroquinolones were licensed for therapeutic use in poultry. The use of fluoroquinolones in food animals has led to the emergence of reduced-susceptibility and fluoroquinolone-resistant *Campylobacter* and *Salmonella*. Prior to their use in poultry, no resistant strains of *Campylobacter* were reported in individuals except in those with previous exposure to quinolones. A 1997 study by the Minnesota Health Department documented that 79% of chickens sampled from supermarkets were infected with *Campylobacter* and 20% of those samples carried the antibiotic-resistant strain.

Salmonella spp. colonize and cause disease in both humans and animals. *Salmonella enteritidis* is the most common *Salmonella* infection in human. In 1994, the largest single outbreak of *Salmonella* infection occurred in the United States. An outbreak of *S. enteritidis* affected an estimated 224,000 persons across the country due to contaminated ice cream from one creamery that had nation-wide distribution. The company manufacturing the ice cream was using good manufacturing practices but the vehicles used to transport the pasteurized ice cream premix were transporting unpasteurized eggs as well. Between transports, the vehicles were not adequately sterilized. An estimated six microbes per half-cup serving of ice cream caused symptoms in those affected. This example illustrates how rapidly an infectious agent can penetrate the country because of the great extent to which the United States has consolidated the production of food.

Salmonella typhimurium resistance is another increasing concern. *Salmonella typhimurium* resistance to tetracyclines has increased from 0% in 1948 to 98% in 1998. Since fluoroquinolones have been licensed for use in animals, emerging strains of fluoroquinolone-resistant *S. typhimurium* DT104 have emerged that are also resistant to ampicillin, streptomycin, chloramphenicol, and sulfonamides. Transmission of this strain from animal-derived food (unpasteurized cheese) to humans has now been confirmed.

Escherichia coli O157 produces a potent verotoxin that has been traced to contaminated meat. Subsequent outbreaks have been confirmed in raw sprouts, unpasteurized orange juice, mangoes, and lettuce. Sources of *E. coli* O157 may come from imported produce that is washed in rivers prior to shipment, rather than in potable water. *Listeria monocytogenes*, although not a common cause of food-borne disease, has a fatality rate of up to 30% and is a known cause of miscarriage or of meningitis of the fetus and newborn. The source has been confirmed in soft cheeses.

A number of government programs have been instituted to combat the spread of food-borne diseases from infectious causes. In 1996, the CDC developed a surveillance network to monitor food-borne illnesses. The CDC's Emerging Infections Program Foodborne Diseases Active Surveillance Network (FoodNet) covers 10.8% of the United States population in a number of states throughout the country. This program monitors the outbreak, incidence, and laboratory-confirmed cases of *Campylobacter*, *E. coli* O157, *L. monocytogenes*, *Salmonella*, *Shigella, Vibrio*, *Yersinia enterocolitica, Cryptosporidium*, and *Cyclospora cayetanensis* in an effort to design interventions to reduce infections.

The FDA Center for Veterinary Medicine, together with the CDC and the U.S. Department of Agriculture, has established the National Antimicrobial Resistance Monitoring System (NARMS) to monitor prospectively changes in antimicrobial susceptibilities of enteric pathogens from human and animal clinical specimens, from healthy farm animals, and from carcasses of food-producing animals at slaughter plants. The FDA has also moved to impose regulations to help curb the development of antibiotic-resistant bacteria in animals. The FDA is also enforcing the Hazard Analysis and Critical Control Point (HACCP) system. This system is meant to control microbial hazards as well as chemical and physical hazards. Although HACCP has been in effect for over 20 years for many sectors of the food industry, it now includes seafood processors and importers and meat and poultry producers. Food handlers in the food service industry are an important component of HACCP.

HEAVY METALS AND ENVIRONMENTAL CONTAMINANTS

A variety of heavy metals and environmental contaminants can affect the food supply, but the two substances of greatest concern currently are methylmercury and polychlorinated biphenyls.

Methylmercury

Mercury is released into the atmosphere by degassing from the Earth's crust and from environmental wastes, which are the source of 2700–6000 and 2000–3000 tons, respectively, each year. The first recorded episode of human mercury poisoning occurred in Minimata, Japan, where people who lived downstream from a local industrial plant were exposed to extremely high concentrations of mercury in their drinking water. Their mercury blood levels ranged from 9 to 24 ppm.

Mercury released into the environment undergoes bacterial transformation to methylmercury, and in this

form is absorbed by fish as it crosses their gills. It is then concentrated in large predatory fish as it moves up the food chain. Traces of methylmercury are found in most fish, but concentrations may be higher in areas of industrial mercury pollution. The usual concentration in most fish ranges from < 0.01 to 0.5 ppm. Except for shark, swordfish, large tuna (the type used to make sushi or fresh steaks), tilefish, and king mackerel, few species of fish reach the FDA limit for human consumption of 1 ppm. Freshwater predatory species such as pike and walleye sometimes have methylmercury levels in the 1-ppm range.

In January 2001, the FDA issued a warning for pregnant women and women of childbearing age to limit consumption of shark and swordfish to no more than once a month. For persons other than pregnant women and women of childbearing age, consumption of shark and swordfish should be limited to no more than 7 ounces per week. For fish with concentrations of mercury averaging 0.5 ppm, consumption should be limited to about 14 ounces per week. The types of seafood that make up 80% of the market—canned tuna, shrimp, pollack, salmon, cod, catfish, clams, flatfish, crabs, and scallops—all have methylmercury levels < 0.2 ppm, and restrictions do not apply unless people eat more than 2.2 pounds of these types of seafood per week.

The Canadian government has established guidelines with tighter restrictions than those in the United States. An acceptable level of methylmercury in fish is 0.5 ppm. In addition to the fish that are restricted in the United States, the Canadian government advises that fresh and frozen tuna also be restricted in the diet of pregnant women and women of childbearing age.

Polychlorinated Biphenyls

Polychlorinated biphenyls (PCBs) represent over 200 individual colorless and odorless chemicals that were once widely used in electrical equipment such as transformers and capacitors; the production and use of PCBs were banned in 1976. PCBs are stable and nonflammable, making them ideal for industrial application. These same characteristics have also enabled them to persist in the environment for long periods of time. Of the over 1.2 billion pounds of PCBs produced in the United States prior to 1976, an estimated 0.6 billion pounds still remain in the environment and have primarily been taken up through water systems. PCBs concentrate in fatty flesh of larger fish, such as salmon, lake trout, and carp, as they move up the food chain. Although PCBs can be absorbed through the skin and lungs, they primarily enter the body when ingested in contaminated fatty fish. They readily pass through the placenta and breast milk, so pregnant and nursing women and women of childbearing age should avoid eating fish from water known to be contaminated with PCBs.

FOOD IRRADIATION

The history of food irradiation in the United States is extensive. Patents to use ionizing radiation to kill bacteria in food were issued in 1905, but the first use of food irradiation by the U.S. government was in 1963 to control insects in wheat when the shipment arrived in Russia. To achieve the goal of controlling bacteria and extending shelf life, the U.S. government has approved irradiation of spices, dry vegetable seasonings, fruit and vegetables, poultry, meat, meat by-products (sausage, etc.), fresh shelled eggs, and packaged foods. Food irradiation is widely supported by governmental agencies including the CDC, and industry. However, irradiated food has not met with success in the market place because there is extensive opposition to its use by consumers.

There are three different technologies used to irradiate food: gamma rays (using radioactive cobalt or cesium salt), an electron beam, and X rays. Gamma rays can penetrate food to a depth of several feet; an electron beam penetrates to 3 cm and X rays penetrate to several feet. Doses of irradiation that have been approved vary from 0.05 kGy, to inhibit sprouting in white potatoes, to 30 kGy, to sterilize herbs and spices.

With the exception of thiamin, the nutritional value of food is relatively unaffected by irradiation. Irradiation effectively eliminates pathogenic bacteria, molds, parasites, and insects from food that has been handled in a clean and safe manner. Irradiation does not mitigate the need for good sanitation practices on the farm or in production plants.

Individuals have raised concerns that there is lack of extensive testing on the long-term effects in humans of irradiated food. Food irradiation has been likened to the introduction of a new antibiotic, and opponents to its use suggest that further research is required prior to the large-scale introduction of this technology into food production.

FOOD HANDLING

Although the FDA and other agencies set standards for commercial food handling, the consumer must be aware of basic rules that apply when purchasing and preparing food. Susceptible foods should be refrigerated, meats should be adequately cooked, dishes and kitchen counters should be sanitized periodically

(and especially after preparing raw meat and fish), and only a reliable dealer should supply fresh raw seafood. Other advice is available on the FDA web site (http://www.fda.gov/fdac/features/895_kitchen.html).

MORE INFORMATION

Additional information can be obtained from the following sources:

1. FDA Office of Consumer Affairs, HFE-88, Rockville, Maryland 20857.
2. FDA Consumer Information Line, 1-800-532-4440; 301-827-4420 in the Washington, D.C. area; 10 a.m. to 4 p.m. Eastern time, Monday through Friday.
3. FDA Seafood Hotline, 1-800-FDA-4010; 202-205-4314 in the Washington, D.C. area; 24 hours a day.
4. USDA Meat and Poultry Hotline, 1-800-535-4555; 202-720-3333 in the Washington, D.C. area; recorded messages available 24 hours a day; home economists and registered dietitians available 10 a.m. to 4 p.m. Eastern time, Monday through Friday.

Information is also available from local county extension home economists, local health departments, and food manufacturers.

See Also the Following Articles

Campylobacter • Cholera • *Cryptosporidium* • Foodborne Diseases • *Salmonella* • *Shigella* • *Yersinia*

Further Reading

United States Food and Drug Administration, Office of Consumer Affairs web site. Retreived at http://www.foodsafetyresources.com.

United States Food and Drug Administration, Center for Food Safety and Applied Nutrition web site. "Bad Bug Book." http://vm.cfsan.fda.gov/~mow/intro.html.

Foreign Bodies

CYNTHIA D. DOWNARD, SNOW PENA-PETERSON, AND TOM JAKSIC
Children's Hospital Boston

achalasia Abnormal relaxation of the sphincter at the bottom of the esophagus, causing difficulty swallowing.

bougie A flexible tube used to calibrate or gradually open a narrowed cylindrical passage.

esophagoscopy Inspection of the interior of the digestive tract connecting the mouth and stomach with an endoscope.

mastication The process of chewing food in preparation for swallowing and digestion.

Foreign body ingestion is a common diagnosis in the pediatric emergency department; it is usually due to accidental swallowing of nondigestible objects. Foreign body aspiration is the leading cause of unintentional mortality in children under 1 year of age and accounts for 7% of deaths in children under 4 years of age. Infants and toddlers are at risk for aspiration of food items, especially small hard objects such as peanuts, due to their limited experience with chewing prior to swallowing. Small household objects, such as coins, pins, batteries, and small toys such as Legos and tiddly winks, pose a problem for older children. Although most foreign body ingestions usually result in little morbidity, prompt recognition and treatment are important.

Ingested foreign bodies may move asymptomatically through the esophagus and intestinal tract. If the object does not pass easily, children may develop abdominal pain, vomiting, choking, coughing, cyanosis, refusal to eat or drink, and wheezing. Treatment for foreign body ingestion is not standardized. The size, shape, and nature of the foreign body play an intricate role in the determination of appropriate management.

Copyright 2004, Elsevier (USA). All rights reserved.

Foodborne Diseases

CHRISTOPHER J. GILL* AND DAVIDSON H. HAMER†
*Boston University School of Public Health and †Tufts University School of Medicine

hemolytic uremic syndrome A syndrome defined as the triad of acute renal failure, microangiopathic hemolytic anemia, and thrombocytopenia that follows infection with certain strains of enterohemorrhagic *Escherichia coli*, particularly serotype O157:H7. The syndrome appears to be mediated by bacteriophage-encoded Shiga toxin expressed during infection.

nontyphoidal *Salmonella* (NTS) The most important agents of foodborne illness, second only to campylobacteriosis in incidence. More than 2400 serotypes of NTS have been identified; all are now considered members of a single species, *Salmonella enterica*, and are commonly designated by their capitalized, nonitalicized serotype name. Most NTS are named for the place where they were first isolated. For example, "*Salmonella* Newport" refers to *S. enterica* serotype Newport.

secondary spread Transmission of infection from an index case to close contacts. In foodborne diseases, this usually is via fecal–oral contamination of foodstuffs.

typhoid fever Often called "enteric" fever. This is a syndrome characterized by high fevers, chills, abdominal pain, bacteremia, and metastatic seeding of organs, bones, and other structures with bacteria. Diarrhea may or may not be present. The principal agent is *Salmonella enterica*, serotypes Typhi and Paratyphi; several rarer NTS serotypes have been occasionally implicated as well. Prior to the availability of antibiotics, this disease was frequently fatal.

Foodborne illness includes all diseases that may be infectious or toxin-mediated that are passively introduced into patients by means of a food vehicle. A disease is generally considered foodborne if it meets the following two criteria: (1) two or more persons experience a similar illness, usually gastrointestinal, after ingestion of a common food and (2) epidemiologic analysis implicates food as the source of the illness. The Centers for Disease Control estimate that 76 million new domestic cases of food poisoning occur annually, making them among the most common diseases afflicting modern society. They also incur great human and financial costs, with 4000 to 5000 deaths annually and between 4 and 23 billion dollars per year in direct patient care costs.

CHANGING EPIDEMIOLOGY OF FOODBORNE DISEASES

The epidemiology of foodborne illnesses in the developed world has undergone significant changes during the past half century. Significantly, improvements in sanitation and hygiene have almost eliminated some of the "classical" gastrointestinal diseases, such as cholera and typhoid fever (see Fig. 1). However, these pathogens have been replaced by other pathogens more uniquely adapted to the conditions established by our modern food production and distribution systems. This evolution is typified by the nontyphoidal *Salmonella* (NTS), *Campylobacter jejuni*, and strains of enterohemorrhagic *Escherichia coli* (EHEC), all of which have "emerged" over the past century as significant threats to public health. Remarkably, *Ca. jejuni*, the most common bacterial cause of gastroenteritis in the United States—causing between 2 and 4 million cases of acute diarrhea per year—was not even recognized as a cause of diarrhea until the late 1970s.

The current pattern of infectious foodborne disease outbreaks differs substantially from the pattern 50 years ago. Prior to the Second World War, outbreaks usually occurred in tight clusters and could be more easily attributed to a single locally produced point-source food exposure. Secondary spread from cases to close contacts was relatively common.

In the 21st century, the average U. S. consumer has ~60,000 food items readily available for consumption, most of which can be found in a typical modern supermarket. Very few of these items will have been produced within 500 miles of the consumer, but are instead mass

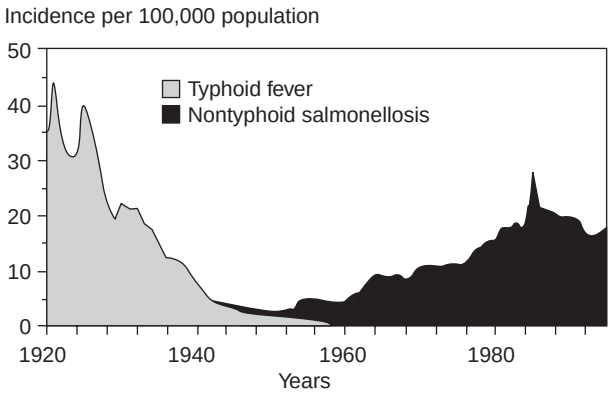


FIGURE 1 Reported incidence of typhoid fever and nontyphoidal salmonellosis, 1920–2000. Adapted from Tauxe *et al.* (1997), *Emerg. Infect. Dis.* **3**(4), 425–434.

 Copyright 2004, Elsevier (USA). All rights reserved.

produced centrally and then widely distributed via an increasingly complex network of distribution channels. Contamination of food items at the point of production thus has the potential for causing widespread outbreaks on an unprecedented scale. This potential was amply demonstrated in 1994 when a contaminated milk truck transporting ice cream premix resulted in over 224,000 cases of intestinal salmonellosis across 41 states.

Another consequence of this widespread distribution of food items is that it presents unique challenges to public health officials since the true scope of an outbreak may easily go unrecognized. This feature also complicates efforts to coordinate strategies to interrupt the chain of transmission. Moreover, for pathogens with low attack rates, it may be difficult to recognize that an outbreak is even occurring. Produce items have increasingly been identified in the context of large-scale outbreaks due to such organisms as NTS, EHEC, and *Cyclospora cayetanensis* (see Table I). Alfalfa and other seed sprouts appear to be particularly efficient vehicles and have caused at least 1500 culture-confirmed cases of disease over the past 10 years (see Table II). Due to the inevitable underreporting of infections, these culture-confirmed cases likely represent closer to 50,000 actual infections in the community.

CAUSES OF FOODBORNE DISEASES

Overview

Numerous infectious and noninfectious agents can cause foodborne illnesses. Table III summarizes the etiologic breakdown and predicted pathogen-specific mortality for the estimated 38 million yearly cases for which a particular agent has been directly implicated; an additional 38 million unexplained, though probably infectious, cases are believed to occur annually. Numerically, viruses are the most frequent agents of diarrheal disease though only approximately 30% of cases of viral gastroenteritis occur via foodborne transmission as opposed to person-to-person transmission.

However, most of the severe morbidity and death is due to bacterial pathogens, particularly NTS, *Ca. jejuni*, and *Listeria monocytogenes*. Unusual sporadic causes of foodborne illness include intoxications due to inorganic poisons, such as heavy metals and agricultural pesticides, and organic toxins, such as pufferfish poisoning and the hepatotoxic death cap mushroom, *Amanita phalloides*. Table IV summarizes some of the less commonly encountered agents of foodborne diseases and their syndromes.

Bacterial Causes of Foodborne Illness

In the developed world, *Ca. jejuni* and NTS account for the majority of foodborne bacterial gastroenteritis, with enteric fever due to *Salmonella typhi* now causing less than 0.1% of all U.S. foodborne disease. Although *Ca. jejuni* is usually associated with small, self-limited point-source outbreaks, usually traced to poultry, raw milk, or contaminated surface water, NTS are hardy and may continue to replicate outside the host animal. These biological features may underlie the propensity of NTS to cause widespread outbreaks involving multiple states or countries.

TABLE I Recent Foodborne Outbreaks Traced to Contaminated Fresh Produce

Year	Pathogen	Vehicle	Cases (No.)	States (No.)	Source
1990	*Salmonella* Chester	Cantaloupe	245	30	Central America
1990	*Salmonella* Javiana	Tomatoes	174	4	U.S.
1990	Hepatitis A	Strawberries	18	2	U.S.
1991	*Salmonella* Poona	Cantaloupe	>400	23	? U.S. vs Central America
1993	*Escherichia coli* O157:H7	Apple cider	23	1	U.S.
1993	*Salmonella* Montevideo	Tomatoes	84	3	U.S.
1994	*Shigella flexneri*	Scallions	72	2	Central America
1995	*Salmonella* Hartford	Orange juice	63	21	U.S.
1995	*E. coli* O157:H7	Leaf lettuce	70	1	U.S.
1996	*E. coli* O157:H7	Leaf lettuce	49	2	U.S.
1996	*Cyclospora cayetanensis*	Raspberries	978	20	Central America
1996	*E. coli* O157:H7	Apple juice	71	3	U.S.
1997	*Cy. cayetanensis*	Raspberries	80	5	Central America
1999	*Cy. cayetanensis*	Basil	64	1	U.S./Mexico
1999	*Salmonella* Thompson	Cilantro	76	1	U.S.
1999	*Salmonella* Baildon	Tomatoes	86	8	U.S.

Note. Adapted from Tauxe *et al*. (1997). *Emerg. Infect. Dis.* 3(4), 425–434.

TABLE II Recent Outbreaks Associated with Seed Sprouts

Year	Pathogen	No. of culture-confirmed cases	Location	Type of sprout	Likely source of contamination
1973	*Bacillus cereus*	4	1 U.S. state	Soy, cress, mustard	Seed
1988	*Salmonella* Saint-Paul	143	United Kingdom	Mung	Seed
1989	*Salmonella* Gold Coast	31	United Kingdom	Cress	Seed and/or sprouter
1994	*Salmonella bovismorbificans*	595	Sweden, Finland	Alfalfa	Seed
1995	*Salmonella* Stanley	242	17 U.S. states, Finland	Alfalfa	Seed
1995–1996	*Salmonella* Newport	133	>7 U.S. states, Canada, Denmark	Alfalfa	Seed
1996	*Salmonella* Montevideo and *S. meleagridis*	~500	2 U.S. states	Alfalfa	Seed and/or sprouter
1996	*Escherichia coli* O157:H7	~6000	Japan	Radish	Seed
1997	*E. coli* O157:H7	126	Japan	Radish	Seed
1997	*S. meleagridis*	78	Canada	Alfalfa	Seed
1997	*Salmonella infantis* and *S. anatum*	109	2 U.S. states	Alfalfa, mung, other	Seed
1997	*E. coli* O157:H7	85	4 U.S. states	Alfalfa	Seed
1997–1998	*Salmonella* Senftenberg	52	2 U.S. states	Clover	Seed and/or sprouter
1998	*E. coli* O157:NM	8	2 U.S. states	Clover, alfalfa	Seed and/or sprouter
1998	*Salmonella* Havana, *S.* Cubana, and *S.* Tennessee	34	5 U.S. states	Alfalfa	Seed and/or sprouter
1999	*Salmonella* Mbandaka	81	4 U.S. states	Alfalfa	Seed

Note. Adapted from Taormina *et al.* (1999). *Emerg. Infect. Dis.* 5(5), 626–634.

Other important bacterial causes of gastroenteritis include *Shigella* spp., highly contagious pathogens that readily spread from person to person and which are the predominant cause of bacillary dysentery worldwide; *Yersinia enterocolitica*, common zoonotic organisms epidemiologically linked to swine herds; and *Vibrio parahaemolyticus*, a sporadic agent of gastroenteritis frequently associated with shellfish consumption in coastal areas of the United States during periods of warmer ocean temperatures.

Recently, Shiga toxin-producing strains of EHEC have emerged as important foodborne pathogens. EHEC share epidemiologic features similar to those of NTS in that they continue to replicate outside of a host animal/patient and thus are capable of causing widely disseminated point-source outbreaks of disease, but share with *Shigella* the ability to cause secondary infections. EHEC are strongly associated with causing the hemolytic uremic syndrome and are now understood to be the most common cause of pediatric end-stage renal disease.

Although *L. monocytogenes* infrequently causes febrile gastroenteritis, controlling foodborne listeriosis remains an important public health priority. This is due to the organism's disproportionate lethality, its tendency to cause permanent neurologic sequelae in its survivors, its innate tolerance to antimicrobials, and its association with both neonatal sepsis and prenatal fetal demise.

Viral Causes of Foodborne Illness

Viral infections are the most common agents of acute, watery, gastroenteritis in adults and children, though mortality is low. Although it is not a cause of gastroenteritis, hepatitis A virus is of particular public health concern. Hepatitis A is highly infectious, is pathogenic, and may cause fatal fulminant hepatitis in 0.5–2% of patients, with risk increasing with age. A highly efficacious vaccine was introduced in 1995. It is used mainly for primary prophylaxis of high-risk populations and international travelers, but has also been used to terminate outbreaks and appears to prevent or attenuate symptomatic disease if given up to 2 weeks following exposure to hepatitis A.

Parasitic Causes of Foodborne Illness

Parasitic infections are comparatively rare but include several important pathogens. *Cryptosporidium parvum* recently achieved notoriety in 1993 after a contaminated municipal water supply caused illness in approximately 400,000 individuals. *Cryptosporidium*

TABLE III Foodborne Illnesses Due to Known Pathogens

Pathogen	Estimated cases[a]	Total cases (%)	Foodborne (%)	Deaths (No.)	Fatality rate
Bacterial					
Campylobacter spp.	2,453,926	14.2	80	124	0.001
Salmonella, nontyphoidal	1,412,498	9.7	95	582	0.0078
Shigella spp.	448,240	0.6	20	70	0.0016
Yersinia enterocolitica	96,368	0.6	90	3	0.0005
Toxigenic *Escherichia coli* (ETEC)	79,420	0.2	70	0	<0.001
E. coli O157:H7 (EHEC)	73,480	0.5	85	61	0.0083
E. coli non-O157:H7 (EHEC)	36,740	0.2	85	7	0.0083
Noncholera *Vibrio* spp.	7,880	<0.1	65	20	0.025
Listeria monocytogenes	2,518	<0.1	99	504	0.2
Brucella spp.	1,554	<0.1	50	11	0.05
Salmonella Typhi[b]	824	<0.1	80	3	0.004
Vibrio vulnificus	94	<0.1	50	37	0.39
Toxin-mediated illnesses					
Clostridium perfringens	248,520	1.8	100	7	<0.001
Staphylococcal food poisoning	185,060	1.3	100	2	<0.001
Streptococcal food poisoning	50,920	0.4	100	0	<0.001
B. cereus	27,360	0.2	100	0	<0.001
Clostridium botulinum	58	<0.1	100	4	0.08
Parasitic					
Giardia lamblia	2,000,000	1.4	10	10	<0.001
Cryptosporidium parvum	300,000	0.2	10	66	0.005
Toxoplasma gondii	225,000	0.8	50	750	<0.001
Cyclospora cayetanensis	16,264	0.1	90	0	<0.001
Trichinella spiralis	52	<0.1	100	0	0.003
Viral					
Caliciviruses (includes Norwalk)	23,000,000	66.6	40	310	<0.001
Rotavirus	3,900,000	0.3	1	30	<0.001
Astrovirus	3,900,000	0.3	1	10	<0.001
Hepatitis A virus	83,391	< 0.1	5	83	0.003
Total cases of diarrheal illness	38,629,641[a]	Bacterial 30%	Parasitic 3%	Viral 67%	

Note. Data from Mead *et al.* (1999). *Emerg. Infect. Dis.* 5(5). EHEC, Enterohemorrhagic *E. coli*; ETEC, enterotoxigenic *E. coli*.

[a] Case numbers are estimates derived from active and passive surveillance data, hospital discharge diagnoses, historical case ratios, and individual outbreaks.

[b] Greater than 70% of typhoid cases were imported from outside of the United States.

infections are frequently chronic and potentially fatal in patients with impaired cell-mediated immunity. At present, no effective antimicrobial agents are available for treatment of cryptosporidiosis; despite showing initial promise in uncontrolled trials, more recent randomized/placebo-controlled trials have shown no benefit to paromoycin for the treatment of autoimmune deficiency syndrome patients with cryptosporidiosis. *Cyclospora cayetanensis* recently achieved notoriety after causing back-to-back outbreaks in 1996 and 1997 linked to contaminated Guatemalan raspberries. *Cryptosporidium parvum, Cy. cayetanenesis, Isospora belli*, and other intestinal coccidian protozoans pathogenic to humans produce a subacute to chronic diarrhea, often complicated by malabsorption and wasting, particularly in immunosuppressed patients.

Giardia lamblia is a common zoonotic protozoan cause of acute self-limited gastroenteritis that generally responds well to antimicrobials. Serious sequelae of giardiasis are rare and usually a consequence of dehydration. Although it is usually a waterborne infection, $\sim$10% of giardiasis may be foodborne.

Bacterial Toxin-Induced Foodborne Illness

Ingestion of preformed bacterial toxins is a frequent cause of foodborne disease outbreaks. The most dangerous of these is the botulinum toxin, for which prompt administration of antitoxin (provided free on request from the Centers for Disease Control and Prevention) may prevent or attenuate the development of a potentially fatal paralysis if administered early in the course of

TABLE IV Uncommon Agents of Food Poisoning and Their Syndromes

Syndrome categories	Nature of disease/comments
Meat-borne	
Gastrointestinal anthrax	Usually caused by consumption of animals killed by anthrax; abdominal pain, nausea, vomiting, hematemesis, bloody diarrhea, and bacteremia progressing to death from septic shock
Trichinosis	Ingestion of larvae of *Trichinella spiralis* in undercooked pork or beef may result in a syndrome of diarrhea, abdominal pain, and vomiting followed by fever, painful myositis, weakness, eosinophilia, and periorbital edema
Fish/shellfish borne	
Anisakiasis	Syndrome of acute epigastric or abdominal pain, nausea, and vomiting after ingesting parasitic larvae in herring, squid, mackerel, salmon, and other white fishes
Ciguatera fish poisoning	Follows ingestion of barracuda, red snapper, amberjack, and grouper; dinoflaggelate neurotoxin concentrated in the fish leads to nausea, vomiting, diarrhea, cramps, hot–cold reversal, myalgias, blurred vision, and paralysis
Scombroid poisoning	A pseudo-allergenic, ingested histamine-mediated syndrome of wheezing, flushing, dizziness, nausea, vomiting, and diarrhea; colonizing gram-negative bacilli convert fish histadine to histamine within the flesh of the fish
Pufferfish poisoning	Sodium channel-blocking tetrodotoxin is concentrated in the fish's liver and ovaries; ingestion leads to respiratory failure and paralysis
Shellfish poisoning (red tide)	Each syndrome results from dinoflaggellate toxins that are concentrated within the tissues of filter-feeding shellfish
Paralytic	Symptoms include paraesthesias, dyspnea, and paralysis
Neurotoxic	Milder than paralytic shellfish poisoning; symptoms include nausea, vomiting, diarrhea, reversal of hot–cold sensation, paraesthesias, and confusion
Diuretic	Dinophysotoxins from the genera *Dinophysis* and *Prorocentrum* mediate this acute, self-limited syndrome of diarrhea and vomiting
Amnestic	Symptoms include gastrointestinal distress, diarrhea, confusion, and temporary or persistent loss of short-term memory formation
Haff syndrome	Epidemic rhabdomyolysis due to an unidentified toxin in bottom-dwelling freshwater fish; originally described near the Black Sea; now reported in California also
Fungal	
Mushroom poisoning	The classic example is the lethal hepatotoxin of *Amanita phalloides*
Mycotoxins	
Ergotism	Epidemic gangrenism and/or psychosis due to rye products contaminated by ergots of the fungus *Claviceps purpurea*
Aflatoxin	*Aspergillus flavus* and *A. parasiticus* produce aflatoxin, a potent carcinogen and putative human carcinogen; presents acutely as fever and fulminant hepatitis
Vegetable	
Neurolathyrism	Chronic consumption of the seeds of the grass pea, *Lathyrus sativus*, causes spasticity and hypermicturation
Hemagglutinin poisoning	Consumption of unsoaked and boiled red kidney beans, *Phaseolus vulgaris*, results in nausea, vomiting, and diarrhea
Ackee fruit (Jamaican vomiting sickness)	Underripe fruits contain high concentrations of hypoglycin, a naturally occurring insulin-like substance; symptoms include nausea, vomiting, confusion, and coma

Note. Reprinted from Gill, C.J. (2001). Foodborne illness. *Curr. Treat. Options Gastroenterol.* 4(1), 23–38, with permission. Copyright Current Medicine Inc.

the disease. Botulinum intoxications most commonly follow ingestion of foods that were contaminated with *Cl. botulinum* spores and then were allowed to incubate under strictly anaerobic conditions, though inoculation of spores into wounds can occasionally cause botulism also. By contrast, the rare syndrome of infant botulism signifies an actual infectious process. This occurs when honey contaminated with *Cl. botulinum* spores is fed to young infants whose intestinal flora has not yet matured to the point where it would naturally suppress the growth of these bacteria.

Strains of *Bacillus cereus*, *Staphylococcus aureus*, *Streptococcus pyogenes*, and *Clostridium perfringens* all can colonize poorly preserved foods or undercooked foods. Although none of these organisms will invade the mammalian gut and lead to a replicative infectious

gastroenteritis, they secrete enteric toxins that do traverse the acid barrier and can lead to an abrupt self-limited syndrome of nausea, vomiting, and diarrhea that is not easily distinguishable from infectious forms of gastroenteritis.

MANAGEMENT OF FOODBORNE DISEASES

The response to foodborne illness has two essential components: the clinical management of acutely ill patients and the accompanying public health investigation as to the etiology, route, vehicle, and extent of community exposure of the agent in question. The public health intervention may be limited to counseling consumers about the risks posed by certain food items or strategies to reduce these risks or issuing health advisories. Other interventions include mandating voluntary or involuntary product recalls and embargos and in rare cases enforcing quarantines. As opposed to failure to prescribe the appropriate antibiotic, complications due to hypovolemia—chiefly metabolic acidosis and electrolyte imbalances—are the major cause of death due to foodborne illness. Oral rehydration salts offer a convenient, cheap, and highly effective mechanism for managing most patients with mild to moderate dehydration, whereas intravenous therapy may be required for seriously ill patients.

The indications for antibiotics are ill-defined and frequently debated by experts in the field. Although most of the bacterial pathogens show *in vitro* sensitivity to commonly used antibiotics, clinical evidence supporting their use for non-travel-associated diarrhea is lacking in most cases, with the notable exceptions of shigellosis and typhoid fever, where therapy is clearly beneficial. Empiric trials of antibiotics for acute gastroenteritis show that antibiotic therapy shortens the average duration of illness by 1−3 days. Caution is raised, however, by other studies indicating that antibiotics may worsen outcomes in the case of infection with certain pathogens. In the case of uncomplicated NTS infections, the indication for antibiotics is hotly disputed, given earlier studies showing that antibiotic therapy prolongs the duration of bacterial shedding, increases relapse rates and has no discernible impact on shortening the duration of illness. Exceptions include bloodstream or disseminated infections or infections in severely immunocompromised patients who are at high risk of developing disseminated disease.

EHEC infections may be of greater concern: epidemiologic and laboratory data indicate that certain DNA-active antibiotics, notably sulfonamides and fluoroquinolones, may increase the risk of developing hemolytic uremic syndrome (HUS) for patients with EHEC infections. Despite several reports supporting the use of fosfomycin in this disease, the evidence proving that this drug reduces the risk of HUS is inconclusive. For these reasons, most authorities recommend that antibiotics be used with great caution when EHEC infections are suspected. As with NTS infections, antibiotics are indicated if patients appear to have disseminated disease or bacteremia.

Acknowledgments

Support for this work came from National Institutes of Health Training Grant 5T32AI0738 and National Research Services Award Training Grant T32 HS00060-09, from the agency for Healthcare Research and Quality.

See Also the Following Articles

Campylobacter • Cholera • _Cryptosporidium_ • Food Safety • Hepatitis A • _Salmonella_ • _Shigella_ • _Yersinia_

Further Reading

Beuchat, L. R., and Ryu, J. H. (1997). Produce handling and processing practices. *Emerg. Infect. Dis.* **3**(4), 459−465.

Bottone, E. J. (1997). *Yersinia enterocolitica:* The charisma continues. *Clin. Microbiol. Rev.* **10**(2), 257−276.

Boyce, T. G., Swerdlow, D. L., *et al.* (1995). *Escherichia coli* O157:H7 and the hemolytic−uremic syndrome. *N. Engl. J. Med.* **333**(6), 364−368.

Centers for Disease Control and Prevention Prevention of hepatitis A through active or passive immunization: Recommendations of the Advisory Committee on Immunizations Practices (ACIP). **45**, 1−30.

Crane, J. K. (1999). Preformed bacterial toxins. *Clin. Lab. Med.* **19**(3), 583−599.

Hardy, M. E. (1999). Norwalk and "Norwalk-like viruses" in epidemic gastroenteritis. *Clin. Lab. Med.* **19**(3), 675−690.

Hennessy, T. W., Hedberg, C. W., *et al.* (1996). A national outbreak of *Salmonella enteritidis* infections from ice cream. The Investigation Team. *N. Engl. J. Med.* **334**(20), 1281−1286.

MacKenzie, W. R., Schell, W. L., *et al.* (1995). Massive outbreak of waterborne *Cryptosporidium* infection in Milwaukee, Wisconsin: Recurrence of illness and risk of secondary transmission. *Clin. Infect. Dis* **21**(1), 57−62.

Mannheimer, S. B., and Soave, R. (1994). Protozoal infections in patients with AIDS: Cryptosporidiosis, isosporiasis, cyclosporiasis, and microsporidiosis. *Infect. Dis. Clin. North Am.* **8**(2), 483−498.

McLauchlin, J., Audurier, A., *et al.* (1991). Treatment failure and recurrent human listeriosis. *J. Antimicrob. Chemother.* **27**(6), 851−857.

Mead, P. S., Slutsker, L., *et al.* (1999). Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5**(5), 607−625.

Ortega, Y. R., and Adam, R. D. (1997). *Giardia*: Overview and update. *Clin. Infect. Dis.* **25**(3), 545−559.

Sack, D. A. (1991). Use of oral rehydration therapy in acute watery diarrhoea: A practical guide. *Drugs* **41**(4), 566−573.

Salam, M. A., and Bennish, M. L. (1991). Antimicrobial therapy for shigellosis. *Rev. Infect. Dis.* **13**(Suppl. 4), S332−S341.

Shapiro, R. L., Hatheway, C., *et al.* (1998). Botulism in the United States: A clinical and epidemiologic review. *Annu. Intern. Med.* **129**(3), 221–228.

Sirinavin, S., and Garner, P. (2000). Antibiotics for treating *Salmonella* gut infections. *Cochrane Database Syst. Rev.* **2**, CD001167.

Wong, C. S., Jelacic, S., *et al.* (2000). The risk of the hemolytic–uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N. Engl. J. Med.* **342**(26), 1930–1936.

Food Intolerance

M. Cecilia Berin

Mount Sinai School of Medicine, New York

adverse food reaction Abnormal physiologic response to food ingestion, due to either food intolerance or food allergy.

food allergy (hypersensitivity) Immune-mediated reaction to food.

food intolerance Abnormal physiologic response to an ingested food or food additive; intolerance has not been proved to be immunologic in nature and falls under the umbrella term, adverse food reaction.

Food intolerance refers to nonimmunologic adverse reactions to ingested foods; the physiologic reactions resolve after dietary elimination of the food substances and reoccur on subsequent challenges with the foods. This category of adverse food reactions includes well-defined disorders such as enzyme deficiencies (i.e., lactose intolerance) and responses to pharmacological agents in food (i.e., tyramine in cheese or wine). In addition, intolerance includes non-immune-mediated reactions to food additives. Symptoms of food intolerance can appear identical to those arising from immune-mediated food allergy, despite the difference in underlying mechanisms. Therefore, the two categories of disorders are often confused.

ENZYME DEFICIENCY

The most commonly reported food intolerance is due to a deficiency in the brush border enzyme lactase, resulting in intolerance to lactose, the common sugar found in milk (see Table I). The inability to digest and absorb lactose leaves an osmotic load in the small intestine that draws water into the intestinal lumen and causes diarrhea. Patients with lactose intolerance can typically tolerate small amounts of lactose, whereas those with cow's milk allergy (caused by an immune reaction to proteins in milk) must practice complete avoidance of products containing milk proteins. Food intolerance can also occur due to nonspecific loss of brush border enzymes secondary to inflammatory events in the small intestine, where epithelial cells are damaged (such as in celiac or Crohn's disease).

Food-induced symptoms that do not involve the gastrointestinal tract can also occur because of enzyme deficiencies, such as glucose-6-phosphate dehydrogenase (G6PD) deficiency. G6PD deficiency results in hemolytic anemia after ingestion of fava beans ("favism"). Fava beans contain two potent oxidants that damage erythrocyte membranes in the absence of the G6PD enzyme. This genetic enzyme deficiency is not common in the Caucasian North American population, but is common in the Middle East and Africa.

FOOD ADDITIVES AND PRESERVATIVES

Intolerance to a number of food additives has been reported, but only a few cases have been verified by appropriately controlled studies, including an elimination diet and placebo-controlled challenge. The list of causative agents includes flavorings and preservatives (sulfites, monosodium glutamate, benzoic acid) and dyes

 Copyright 2004, Elsevier (USA). All rights reserved.

(and especially after preparing raw meat and fish), and only a reliable dealer should supply fresh raw seafood. Other advice is available on the FDA web site (http://www.fda.gov/fdac/features/895_kitchen.html).

MORE INFORMATION

Additional information can be obtained from the following sources:

1. FDA Office of Consumer Affairs, HFE-88, Rockville, Maryland 20857.
2. FDA Consumer Information Line, 1-800-532-4440; 301-827-4420 in the Washington, D.C. area; 10 a.m. to 4 p.m. Eastern time, Monday through Friday.
3. FDA Seafood Hotline, 1-800-FDA-4010; 202-205-4314 in the Washington, D.C. area; 24 hours a day.
4. USDA Meat and Poultry Hotline, 1-800-535-4555; 202-720-3333 in the Washington, D.C. area; recorded messages available 24 hours a day; home economists and registered dietitians available 10 a.m. to 4 p.m. Eastern time, Monday through Friday.

Information is also available from local county extension home economists, local health departments, and food manufacturers.

See Also the Following Articles

Campylobacter • Cholera • *Cryptosporidium* • Foodborne Diseases • *Salmonella* • *Shigella* • *Yersinia*

Further Reading

United States Food and Drug Administration, Office of Consumer Affairs web site. Retreived at http://www.foodsafetyresources.com.

United States Food and Drug Administration, Center for Food Safety and Applied Nutrition web site. "Bad Bug Book." http://vm.cfsan.fda.gov/~mow/intro.html.

Foreign Bodies

CYNTHIA D. DOWNARD, SNOW PENA-PETERSON, AND TOM JAKSIC
Children's Hospital Boston

achalasia Abnormal relaxation of the sphincter at the bottom of the esophagus, causing difficulty swallowing.

bougie A flexible tube used to calibrate or gradually open a narrowed cylindrical passage.

esophagoscopy Inspection of the interior of the digestive tract connecting the mouth and stomach with an endoscope.

mastication The process of chewing food in preparation for swallowing and digestion.

Foreign body ingestion is a common diagnosis in the pediatric emergency department; it is usually due to accidental swallowing of nondigestible objects. Foreign body aspiration is the leading cause of unintentional mortality in children under 1 year of age and accounts for 7% of deaths in children under 4 years of age. Infants and toddlers are at risk for aspiration of food items, especially small hard objects such as peanuts, due to their limited experience with chewing prior to swallowing. Small household objects, such as coins, pins, batteries, and small toys such as Legos and tiddly winks, pose a problem for older children. Although most foreign body ingestions usually result in little morbidity, prompt recognition and treatment are important.

Ingested foreign bodies may move asymptomatically through the esophagus and intestinal tract. If the object does not pass easily, children may develop abdominal pain, vomiting, choking, coughing, cyanosis, refusal to eat or drink, and wheezing. Treatment for foreign body ingestion is not standardized. The size, shape, and nature of the foreign body play an intricate role in the determination of appropriate management.

Copyright 2004, Elsevier (USA). All rights reserved.

COINS

Coins are the most frequently ingested foreign body in the pediatric population. Treatment following coin ingestion depends on the position of coin at the time of diagnosis. A coin that has passed unaided through the esophagus to the stomach will most likely pass uneventfully through the remaining gastrointestinal tract. However, coins frequently become lodged in the esophagus and may require medical intervention for removal.

Diagnosis

Once foreign body ingestion is suspected, plain radiographs are useful in localizing radioopaque substances such as coins (Fig. 1). The most common site of obstruction is relatively high in the esophagus, at the level of the cricopharyngeus. Contrast studies can be used to identify nonradioopaque foreign bodies. Ultrasound may also demonstrate the location of foreign body with visualization of a hyperechoic substance within the alimentary tract.

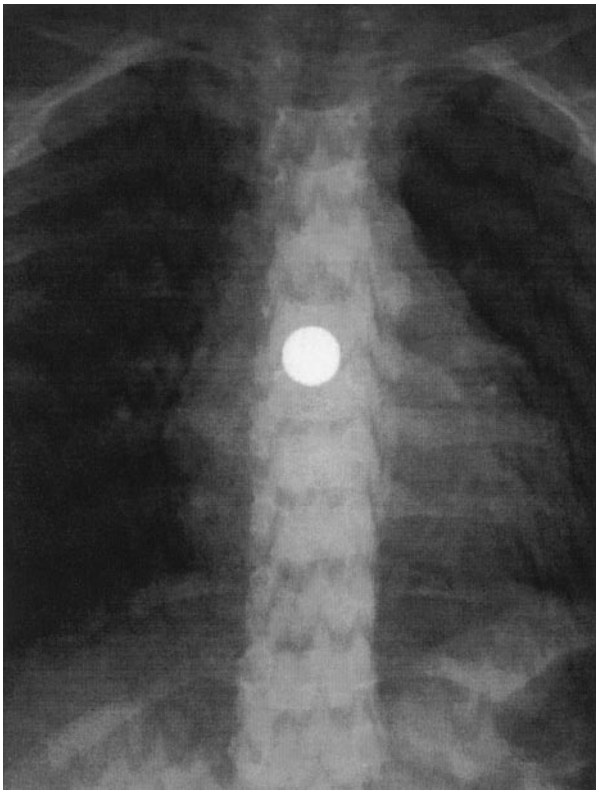


FIGURE 1 Plain chest radiograph showing a coin lodged in the esophagus. This object was removed by esophagoscopy.

Treatment

Coins that become lodged in the esophagus may pass to the stomach spontaneously. If a coin has been in the esophagus for less than 24 h, it is acceptable to monitor a child for an additional 12–24 h for spontaneous passage prior to removing the coin in an invasive manner.

Coins lodged in the esophagus are reliably removed by esophagoscopy. Perforation is rare when removing a rounded object, such as a coin, in an operating room under general endotracheal anesthesia.

Alternatively, a Foley catheter, passed distal to the coin and then inflated, may be used to withdraw esophageal foreign bodies under radiographic guidance. This technique does not require general anesthesia, but does require a cooperative patient. The risk of using this method of foreign body extraction is that the coin, when being withdrawn from the esophagus, could easily pass into the trachea from the oropharynx and cause airway compromise. Series have reported up to 96% success with removal of coins using this technique, but the possibility of aspiration has resulted in relatively few centers adopting this approach.

A third method being utilized for the removal of esophageal coins is the penny pincher technique, which involves the use of grasping forceps covered by a soft rubber catheter. After this device is directed into the esophagus fluoroscopically, the forceps are advanced from the shield and aligned with the foreign body. The forceps are used to grasp the coin, which is then withdrawn under radiographic guidance. Like the Foley catheter technique, no anesthesia or sedation is needed; however, with forceps control of the foreign body, the risk of aspiration during removal is theoretically decreased. Experience with this technique is limited and this special catheter is not widely available.

Outcome

Both esophagoscopy and the Foley catheter techniques have approximately 90% success rates as methods for foreign body removal. When a coin is unable to be removed by either method, it can simply be pushed through to the stomach, with the expectation that it will pass through the remaining gastrointestinal tract. The expulsion of the foreign body should be ensured with careful examination of the stool. If obstructive symptoms develop, repeat radiographs can visualize the position of the foreign body and treatment can be appropriately determined. In the rare event that a coin becomes lodged at the ileocecal valve, operative removal may be necessary.

PINS

Diagnosis

The ingestion of pins (and other sharp metal objects such as screws) is dangerous due to the potential for perforation of the alimentary tract. When pin ingestion is suspected, plain radiographs can usually identify the position of the foreign body.

Treatment

Ingested pins should be removed immediately after failure of advancement is established because of the risk of perforation. In addition to direct perforation, longer foreign bodies, such as hair pins, may become lodged transversely in the esophagus and produce pressure necrosis and fistula formation. Once pin ingestion is diagnosed, endoscopic removal of the foreign body is undertaken immediately to minimize the potential for perforation or fistula formation. Pins should be endoscopically removed even if they pass spontaneously to the stomach. A magnet may also be used to remove metallic foreign bodies such as pins, provided that precautions are taken to minimize the likelihood of perforation on withdrawal.

Outcome

Pin ingestion is worrisome due to the risk of alimentary tract perforation. Every effort should be directed toward removing the object while it is within the upper gastrointestinal tract. Should the child present after an ingested pin has passed through the stomach, the stool must be carefully examined to ensure enteral elimination of the object. If symptoms suggesting perforation or fistula formation develop, repeat radiographs should be obtained. These complications require operative intervention for correction.

BATTERIES

Diagnosis

Button-style batteries are frequently ingested because of their small size and household availability in daily items such as hearing aids, hand-held games, and cameras. Ingested batteries require immediate attention due to the possible leakage of alkaline fluid and discharge of electric current, which can cause severe necrosis and perforation when lodged in the esophagus. An initial radiograph should be performed on patients who have ingested a battery in order to confirm the diagnosis and determine the location of the battery.

Treatment

Immediate endoscopic removal is mandatory when a battery is lodged in the esophagus. Batteries lodged in the esophagus for as little as 4 h have been associated with significant mucosal burns. Esophagoscopy, in preference to other extraction methods, should be used to remove impacted esophageal batteries to allow for direct visualization of the adjacent mucosa. If mucosal damage is present, scarring and subsequent narrowing of the esophagus may occur.

Batteries that have passed the esophagus can be left to navigate the remaining gastrointestinal tract, although intestinal damage from the leakage of contained fluid may still occur. The diameter and chemical type of battery should be determined by examining a similar battery or the device with which the ingested battery is associated. Larger batteries (>15 mm) may not readily pass through the pylorus and ileocecal valves in young children. If the battery is retained in the stomach, it can be removed endoscopically. If it is lodged further in the intestinal tract, operative removal may be necessary.

If a battery is determined to contain mercuric oxide (approximately 25% of ingestions) and the battery splits in the intestinal tract, serum and urine mercury levels should be monitored for toxicity. Serial radiographs are not necessary unless the patient develops obstructive symptoms or passage of the battery in the stool has not been noted. Emetics are not useful for battery expulsion after ingestion.

Outcome

Over 90% of batteries located in the esophagus at the time of diagnosis are successfully removed. However, in one large series, 40% of patients with batteries removed from the esophagus experienced serious adverse outcomes, including esophageal perforation, scarring requiring dilations, and death. If batteries pass unaided through the esophagus, there is a very small likelihood that operative intervention will be required for removal and outcome is generally excellent. The stool should be carefully inspected to document successful evacuation of ingested batteries.

Food

As solid foods are introduced into the diet, inadequate mastication can lead to food becoming lodged in the alimentary tract. This is a particular problem for children who have undergone prior repair of esophageal atresia early in life, with the food becoming lodged in the hypomotile distal esophagus. Severe acid reflux causing esophageal stricture may also cause food to lodge in the

esophagus. Additionally, esophageal dysmotility, as seen in achalasia, may come to clinical attention after food becomes stuck in the esophagus. Small bones in foods such as chicken and fish are also easily swallowed by children.

Diagnosis

Detection of radiolucent foreign bodies such as food products is difficult. A negative plain radiograph in this situation does not indicate the absence of a foreign body. Gastrointestinal contrast studies may be used in this situation to identify radiolucent foreign bodies. A history of foreign body ingestion and symptomatic presentation with difficulty swallowing and excess saliva production should guide treatment.

Treatment

Food that becomes lodged in the esophagus can be removed by esophagoscopy. Alternatively, digestible food products can be pushed through to the stomach using a bougie and allowed to pass through the remaining alimentary tract. Bones from chicken and fish should be treated as sharp objects and removed if at all possible due to the risk of intestinal perforation. Removal of a lodged food product should be followed with contrast evaluation to search for underlying esophageal stricture or dysmotility.

Outcome

Unless the ingestion involves sharp bones, obstruction of the alimentary tract with inadequately chewed food is usually well tolerated. Eventual outcome depends on the underlying problem predisposing the patient to swallowing difficulty. All children should be reminded to chew food well before swallowing.

BEZOARS

Bezoars are caused by repeated swallowing of an indigestible foreign body. Most frequently, this material is from an excessively fiber-rich diet (phytobezoar) or hair (trichobezoar). Bezoars present with symptoms of intestinal obstruction and are sometimes diagnosed by plain radiographs or endoscopy. If the bezoar is in the stomach, endoscopic removal of the material may be attempted. Operative intervention with a possible gastrotomy or enterotomy is often required to relieve the intestinal obstruction. Psychological evaluation, as well as an evaluation for anemia, should accompany surgical therapy for this condition.

CONCLUSION

A variety of household objects and foods can be accidentally ingested in the pediatric population. Treatment following foreign body ingestion is determined by the size, shape, and nature of the swallowed material, with sharp objects and batteries lodged in the esophagus requiring urgent removal. If a foreign body has passed through the upper intestinal tract, the stool should be monitored to ensure successful transit of the foreign body. Few serious problems are associated with most foreign body ingestions, but timely recognition and treatment are important in minimizing adverse outcomes.

See Also the Following Articles

Achalasia • Bezoars • Computed Tomography (CT) • Endoscopic Ultrasonography • Ultrasonography

Further Reading

Arana, A., Houser, B., Hachimi-Idrissi, S., and Vandenplas, Y. (2001). Management of ingested foreign bodies in childhood and review of the literature. *Eur. J. Pediatr.* **160**, 468–472.

Litovitz, T., and Schmitz, B. F. (1992). Ingestion of cylindrical and button batteries: An analysis of 2382 cases. *Pediatrics* **89**, 747–757.

Macpherson, R. I., Hill, J. G., Othersen, H. B., Tagge, E. P., and Smith, C. D. (1996). Esophageal foreign bodies in children: Diagnosis, treatment, and complications. *Am. J. Roentgenol.* **166**, 919–924.

Soprano, J. V., and Mandl, K. D. (2000). Four strategies for the management of esophageal coins in children. *Pediatrics* **105**, e5. Available at http://www.pediatrics.org/cgi/content/full/105/1/e5.

Stricker, T., Kellenberger, C. J., Neuhaus, T. J., Schwoebel, M., and Braegger, C. P. (2001). Ingested pins causing perforation. *Arch. Dis. Childhood* **84**, 165–166.

Fulminant Hepatic Failure

JOHN J. POTERUCHA
Mayo Clinic

aminotransferases Usually two types, alanine aminotransferase and aspartate aminotransferase; these hepatocellular enzymes are released into the bloodstream with injury of hepatocytes, thus elevated levels serve as markers of liver cell injury.
cerebral edema Increased extracellular water in the brain, leading to increased intracranial pressure.
coagulopathy Alteration in the normal clotting mechanisms of blood.
hepatic encephalopathy Changes in mental status due to liver dysfunction.

Fulminant hepatic failure is a clinical syndrome of severe acute liver failure with hepatic encephalopathy in a patient without a prior history of liver disease. The most common causes are medications, especially acetaminophen, and viral hepatitis A and B. Clinical manifestations include the cardinal features of liver failure and encephalopathy but also coagulopathy, hypoglycemia, infections, and various abnormalities of other organ systems. Models have been developed to help predict patient outcome to assist in management, including the timing of liver transplantation.

INTRODUCTION

Fulminant hepatic failure (FHF) has traditionally been defined as the presence of acute liver failure including the development of hepatic encephalopathy within 8 weeks after the onset of jaundice in a patient without a prior history of liver disease. Because not all patients with severe acute liver disease meet this definition, some have proposed to use the term "acute liver failure," which encompasses other clinical scenarios, including FHF. Various classifications are shown in Table I. Different classification schemes are proposed because of diagnostic and prognostic differences between groups. For example, hyperacute liver failure is most often due to ischemia or acetaminophen and may often be reversible, assuming the offending agent is removed. On the other hand, late-onset hepatic failure, often due to viruses or idiopathic causes, may lead to manifestations of portal hypertension, is seldom complicated by cerebral edema, and yet has a poor prognosis. Because most studies have used the traditional definition of FHF, that will be the nomenclature used herein.

There are about 2000 cases of FHF in the United States each year. The overall mortality of FHF without liver transplantation is up to 90%. Because many of these patients are young and previously healthy, the outcomes of this relatively unusual condition are particularly tragic. Specific management, including liver transplantation, is available for this group of patients and therefore knowledge of management strategies is important.

ETIOLOGY

Determining the etiology of FHF is important for two reasons: first, specific therapy may be available, such as with acetaminophen hepatotoxicity or herpes hepatitis, and second, prognosis will differ depending on etiology. For instance, spontaneous recovery rate with FHF due to acetaminophen or hepatitis A is $>50\%$ and therefore a more cautious approach before proceeding with liver transplantation would be advised. On the other hand, spontaneous recovery from FHF due to Wilson's disease is very unusual and therefore early liver transplantation would be recommended. In addition, determination of a specific etiology may have implications for other patients. Identification of a hepatotoxic agent is certainly helpful in monitoring other patients on the same drug. Identification of a viral cause of FHF has implications for other patients that have been exposed to the transmissible agent. Patients and family members need to be asked about risk factors for liver disease, medications, and a family history of liver disease that may give clues to the etiology of the acute hepatitis.

The etiology of FHF varies according to region, with hepatitis B being a more common cause in areas such as Asia, where hepatitis B is endemic, whereas studies from Great Britain generally report a high number of acetaminophen hepatotoxicity cases. The U.S. Acute Liver Failure Study Group has put together a coordinated effort from a number of centers attempting to better define the causes and outcome of acute liver failure in the United States. The most common identifiable

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Acute Liver Failure

Classification	Source
Acute hepatic failure (rapidly developing impairment of liver function)	Bernuau *et al.* (1986)
Severe acute hepatic failure (prothrombin time or factor V concentration <50% of normal, with or without encephalopathy)	
Fulminant hepatic failure (encephalopathy within 2 weeks of onset of jaundice)	
Subfulminant hepatic failure (encephalopathy between 3 and 12 weeks after onset of jaundice)	
Acute liver failure (requires encephalopathy)	O'Grady *et al.* (1993)
Hyperacute liver failure (0–7 days between onset of jaundice and encephalopathy)	
Acute liver failure (8–28 days)	
Subacute liver failure (29–72 days)	
Late-onset acute liver failure (56–182 days)	
Acute liver failure (encephalopathy within 4 weeks after symptom onset)	Tandon *et al.* (1999)
Hyperacute liver failure (within 10 days)	
Fulminant liver failure (10–30 days)	
Acute liver failure not otherwise specified	
Subacute liver failure (development of ascites and/or encephalopathy from 5 to 24 weeks after symptom onset; may be subclassified based on etiology)	

cause is acetaminophen hepatotoxicity followed by hepatitis B and hepatitis A (see Fig. 1).

FHF is usually evident from the time of presentation based on the cardinal features of liver failure and encephalopathy. It can sometimes be difficult to distinguish FHF from an acute presentation of chronic liver disease. Clues to the presence of a more chronic liver condition include spider angiomata and manifestations of portal hypertension, although portal hypertension may be seen in patients with late-onset acute hepatic failure. Also in the differential diagnosis of FHF is multisystem organ failure associated with sepsis, which can occasionally be accompanied by nonspecific mental status changes that mimic hepatic encephalopathy.

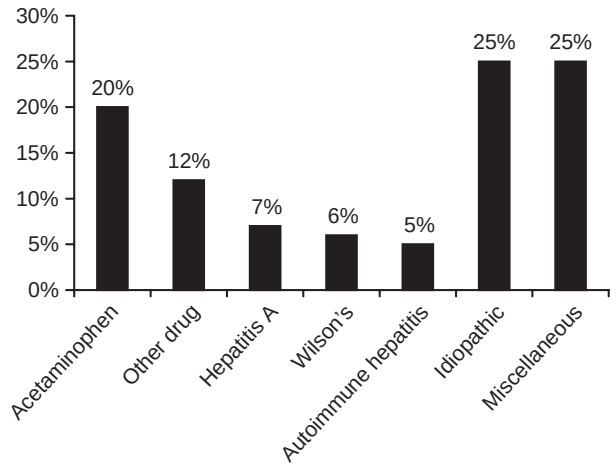


FIGURE 1 Etiology of acute liver failure in the United States Miscellaneous causes include Budd–Chiari syndrome, herpes, Epstein–Barr virus, paramyoxovirus, *Amanita* poisoning, ischemia, malignancy, acute fatty liver of pregnancy, and others. Data from Schiodt *et al.* (2003).

CLINICAL MANIFESTATIONS

The presenting symptoms of FHF are usually those of acute hepatitis, including malaise, nausea, and jaundice. Portal systemic encephalopathy is a required feature of the syndrome and can be staged as noted in Table II. Manifestations may range from subtle mental status changes, such as difficulty with concentration, to coma. The presence of encephalopathy in a patient with acute liver disease is an ominous sign and therefore mental status of patients with acute hepatitis should be frequently assessed.

Laboratory features of FHF are consistent with severe liver dysfunction. Aminotransferases are variably elevated, although usually quite high. The highest levels of aminotransferases, occasionally as high as 10,000 U/liter, are seen with acetaminophen hepatotoxicity or unusual viruses such as herpes. Viral hepatitis and other drug-induced liver diseases usually result in aminotransferase levels of 1000–5000 U/liter. Fulminant Wilson's disease is characterized by only modest aminotransferase elevations and a normal or only minimally elevated alkaline phosphatase, despite other more typical laboratory evidence of liver failure. Evidence of hepatocellular dysfunction includes a prolonged prothrombin time due to poor hepatic synthesis of clotting factors, high bilirubin secondary to decreased ability of the liver to excrete bilirubin, and sometimes a low albumin level, which may be due to cytokine-induced

TABLE II Stages of Hepatic Encephalopathy

Stage	Feature
I	Changes in behavior with minimal change in level of consciousness
II	Gross disorientation, gross slowness of mentation, drowsiness, asterixis, inappropriate behavior, able to maintain sphincter tone
III	Sleeping most of the time, arousable to vocal stimuli, marked confusion, incoherent speech
IV	Comatose, unresponsive to pain, includes decorticate or decerebrate posturing

increases in degradation or poor liver synthetic function.

Encephalopathy

The encephalopathy associated with FHF is likely different than that seen in chronic liver disease. The major difference is the propensity of encephalopathy of acute liver disease to progress to cerebral edema. Clinically, the encephalopathy is often associated with elevations in serum ammonia, although it is likely that alterations in other neurotransmitters are involved in causing the mental status changes. Marked elevations in serum ammonia are associated with a higher likelihood of subsequent cerebral edema and a poorer outcome of FHF. The exact mechanisms for the development of cerebral edema have not been clarified but may involve disruptions in blood–brain barrier and interference with mechanisms of cellular osmolarity.

Cerebral edema is estimated to cause of about 20% of deaths in patients with FHF; however, in autopsy studies of patients having died from FHF, cerebral edema is present in 50% of cases. Cerebral edema leads to death by causing brain ischemia and cerebral herniation. Symptoms and signs of cerebral edema are not specific and therefore intracranial pressure monitoring is often necessary. Computed tomography (CT) scans of the brain are relatively insensitive for detecting cerebral edema, although they are useful to exclude other causes of mental status changes, particularly intracerebral bleeding.

Coagulopathy

The liver is the site of synthesis of many clotting factors, including factors II, V, VII, IX, and X. Poor hepatocellular function leads to decreased synthesis of these factors, which can be measured directly or by determination of prothrombin time. Consumption of clotting factors due to intravascular coagulation and fibrinolysis (disseminated intravascular coagulation) can also contribute to the coagulopathy of FHF. Because factor V has the shortest half-life of the clotting factors synthesized in the liver, it is often used as a marker of residual liver function and has been utilized to help determine prognosis of FHF.

Hypoglycemia

Hypoglycemia is a frequent manifestation of FHF and glucose should be carefully monitored in all patients. The hypoglycemia is likely due to both inadequate degradation of insulin and diminished production of glucose by the diseased liver.

Infections

Infections are another common cause of death in patients with FHF. Reasons for the infections are multiple, but likely reflect severe illness and need for multiple interventions and monitoring. Typical clinical features of infection such as fever and leukocytosis are not reliable in patients with FHF. A high index of suspicion needs to be maintained and any clinical deterioration should mandate a search for infection. The presence of infection and progression of encephalopathy have been correlated in patients with FHF.

Cardiovascular System

A hyperdynamic circulation and reduction in systemic vascular resistance frequently accompany FHF. Although these features may be well tolerated, occasionally patients can develop hemodynamic compromise. Monitoring parameters may mimic septic shock. Fluid resuscitation is often necessary, although caution is advised because excessive fluid administration may worsen intracranial pressure; vasopressors are often necessary.

Other Organ Systems

Hypoxemia is due to multiple factors, including pneumonia or noncardiogenic pulmonary edema associated with adult respiratory distress syndrome. Renal and electrolyte abnormalities occur due to underlying disease such as Wilson's disease, functional renal failure due to sepsis or hepatorenal syndrome, or acute tubular necrosis. Renal dysfunction may be more common when FHF is due to acetaminophen hepatotoxicity. Monitoring of electrolytes, including sodium, potassium, bicarbonate, magnesium, and phosphorus, is important. Lactic acidosis is also common in FHF, likely due to hypoperfusion and inability of the diseased liver to clear lactate. The presence of acidosis is a risk factor

TABLE III Criteria for Liver Transplantation in Fulminant Hepatic Failure

King's College Hospital recommendations (any one of the three)
1. Fulminant hepatic failure due to Wilson's disease or Budd–Chiari syndrome
2. Acetaminophen-induced if either of the following criteria are met
a. pH < 7.3 24 hours after overdose
b. Creatinine >3.4 mg/dl and prothrombin time <100 seconds and grades 3 and 4 encephalopathy
3. Nonacetaminophen if either
a. International normalized ratio (INR) >6.5 or
b. Any three of the following: INR >3.5, more than 7 days from jaundice to encephalopathy, indeterminate or drug-induced cause, age <10 years, age >40 years, bilirubin >17.5 mg/dl
France recommendations
1. Grades 3 and 4 encephalopathy and
2. Factor V >20% of normal if age <30 years or <30% of normal if age >30 years

for poor outcome in FHF and has been incorporated into prognostic models.

PREDICTIVE MODELS AND MANAGEMENT

There have been several proposed models attempting to predict outcome of FHF. These have been developed to facilitate optimal timing of liver transplantation before the patient becomes so ill that there are contraindications to transplantation. Some of these models are summarized in Table III. The most well known and widely used are the King's College criteria. Liver transplantation likely improves mortality, although this has never been assessed except by comparison with historical controls.

All patients with FHF should be hospitalized. Patients who have no contraindications to liver transplantation should be transferred to a center where transplantation is available. Tests to establish an etiology are advised; history from the patients or family members should include a careful medication history, exam, serologic tests for viral hepatitis, autoantibodies, ceruloplasmin in appropriate patients, and liver ultrasound. Sedatives are avoided unless intubation or intracranial pressure monitoring is necessary.

Encephalopathy and Cerebral Edema

The appearance of encephalopathy precedes cerebral edema and therefore patients with acute hepatitis and evidence of liver failure need to be carefully monitored for mental status changes. Patients with encephalopathy should receive lactulose, although this agent is not as effective in acute liver failure as it is in chronic liver disease and may not prevent later cerebral edema. Patients with encephalopathy should be carefully monitored because rapid deterioration can occur. Patients with stage II encephalopathy are generally admitted to the intensive care unit for close monitoring of mental status and vital signs. It is especially important to avoid sedatives at this point to allow close monitoring of mental status. Most centers will also perform CT of the head to exclude an alternative cause for mental status changes.

Patients who reach stage III encephalopathy are at significant risk of progression to cerebral edema. Because clinical signs and head CT scans are insensitive for detecting increased intracranial pressure (ICP), many centers will institute ICP monitoring when patients reach stage III encephalopathy. This is usually preceded by endotracheal intubation and mechanical ventilation. A variety of ICP monitors are used, all of which can be complicated by infection and bleeding. The goal of ICP monitoring is to allow treatment of high pressures but also to identify which patients become too ill for liver transplantation because of excessively high ICP for a prolonged period of time. In general, the goal is to keep ICP less than 40 mmHg and cerebral perfusion pressure (the difference between mean arterial pressure and ICP) between 60 and 100 mmHg. Excessively high cerebral perfusion pressures (above 120 mmHg) can result in worsening cerebral edema.

Maneuvers that cause straining, including tracheal suctioning, should be avoided or limited. Paralyzing agents and sedatives may be necessary, although they limit further assessment of neurologic status. For ICP >20 mmHg or cerebral perfusion pressure below 60 mmHg, head elevation to 20°, hyperventilation to a P_aCO_2 of 25 mmHg, and mannitol (if renal function is intact) are advised. Barbiturate-induced coma or hypothermia can be used for refractory cases. A prolonged increase in ICP above mean arterial pressure may signify brain death and generally is a contraindication to liver transplantation. Sudden decreases in ICP may indicate brain herniation.

Management of Other Specific Complications

The prolonged prothrombin time seen in patients with FHF is a simple noninvasive parameter to follow and coagulopathy is generally not corrected unless there is bleeding or planned interventions such as monitoring devices. If bleeding occurs or invasive procedures are necessary, than fresh frozen plasma is generally used

first. Administration of platelets and fibrinogen may be necessary in certain circumstances. Continuous infusion of 5 or 10% dextrose is administered to keep the plasma glucose between 100 and 200 mg/dl. Plasma glucose should be carefully monitored, at least twice daily. Both bacteremia and fungemia are frequent enough that periodic blood cultures are advised and prophylaxis with antimicrobials may be used, although this has not been shown to impact survival.

Liver Transplantation

Liver transplantation has revolutionized the management of FHF. FHF is the indication for 6% of liver transplants in the United States. Even though survival with transplantation for FHF is lower than that for transplantation for other indications, outcomes are an improvement over the dismal survival rates seen in FHF when poor prognostic criteria (such as the King's College criteria) are met. Transplantation should be performed when a poor outcome is anticipated yet before the patient has uncontrolled sepsis or prolonged periods of increased ICP, such that recovery even with a functioning transplanted liver is not possible.

Contraindications to liver transplantation are advanced age, multisystem organ failure, uncontrolled sepsis, and prolonged increases in ICP. In the United States, patients with FHF are given highest priority for organs and therefore are often transplanted within 1 week. The 1-year survival after transplantation is about 60–75%.

Liver-Assist Devices

Artificial liver-assist devices have been in development for over a decade. Charcoal hemoperfusion and plasmapheresis have not been demonstrated to be beneficial. The newest system is the molecular adsorbent recycling system, which is based on dialysis of blood with a membrane that is coated with albumin to allow removal of both water-soluble and albumin-bound toxins. This promising system has yet to be formally studied.

Bioartificial livers use cultured porcine or hepatoma cell line hepatocytes in hollow fiber cartridges to provide metabolic functions. Such devices are able to achieve certain metabolic functions of the liver but have not yet been shown to improve survival compared to standard care, including liver transplantation. The ideal device would allow time for recovery of the diseased liver. Currently, the devices are used as a bridge to transplantation, keeping the patient stable until a liver becomes available.

See Also the Following Articles

Budd–Chiari Syndrome • Cirrhosis • Hepatic Encephalopathy • Hepatitis A • Hepatitis B • Hepatorenal Syndrome • Liver Transplantation

Further Reading

Bernuau, J., Rueff, B., and Benhamou, J. P. (1986). Fulminant and subfulminant liver failure: Definitions and causes. *Semin. Liver Dis.* **6**, 97–106.

Clemmesen, J. O., Larsen, F. S., Konsrug, J., *et al.* (1999). Cerebral herniation in patients with acute liver failure is correlated with arterial ammonia concentration. *Hepatology* **29**, 648–653.

O'Grady, J. G., Alexander G. J., Hayllar, K. M., and Williams, R. (1989). Early indicators of prognosis in fulminant hepatic failure. *Gastroenterology* **97**, 439–445.

O'Grady, J. G., Schalm, S. W., and Williams, R. (1993). Acute liver failure: Redefining the syndromes. *Lancet* **342**, 273–275.

Samuel, D. (2002). Treatment of patients with hepatic failure. *J. Gastroenterol. Hepatol.* **17**, S274–S279.

Schiodt, F. V., Atillasoy, E., Shakil, A. O., *et al.* (1999). Etiology and outcome for 295 patients with acute liver failure in the United States. *Liver Transplant. Surg.* **5**, 29–34.

Schiodt, F. V., Davern, T. J., Shakil, A. O., *et al.* (2003). Viral hepatitis-related acute liver failure. *Am. J. Gastroenterol.* **98**, 448–453.

Stevens, A. C., Busuttil, R., Han, S., *et al.* (2001). An interim analysis of a phase II/III prospective randomized, multicenter, controlled trial of the HepatAssist Bioartificial Liver Support System for the treatment of fulminant hepatic failure (abstract). *Hepatology* **34**, A299.

Trey, C., and Davidson, C. S. (1970). The management of fulminant hepatic failure. *Prog. Liver Dis.* **3**, 282–298.

Vaquero, J., and Blei, A. (2003). Etiology and management of fulminant hepatic failure. *Curr. Gastroenterol. Rep.* **5**, 39–47.

Functional (Non-Ulcer) Dyspepsia

HENRY P. PARKMAN
Temple University School of Medicine

dyspepsia Refers to symptoms originating in the upper gastrointestinal tract; symptoms include upper abdominal pain/discomfort, early satiety, postprandial abdominal bloating/distension, and nausea without vomiting.
functional dyspepsia Refers to dyspeptic symptoms with no definable organic cause; diagnosis is based on negative findings after evaluation of medical history, physical examination, blood tests, and upper endoscopy.
gastroparesis Delayed gastric emptying in the absence of obstruction.

Dyspepsia refers to symptoms originating in the upper gastrointestinal tract, including upper abdominal pain/discomfort, early satiety, postprandial abdominal bloating/distension, and nausea with or without vomiting. Dyspepsia is a common condition with an estimated prevalence of approximately 20% of the population. Only about 25% of people with dyspeptic symptoms actually seek medical care. Dyspepsia accounts for 2–5% of patient visits to primary care physicians and 8% of patient visits to gastroenterologists. In approximately 60% of patients with dyspepsia, there is no apparent cause for the condition, and the diagnosis is "functional (nonulcer)" dyspepsia.

INTRODUCTION

Patients with functional (nonulcer) dyspepsia present with upper abdominal pain or discomfort with or without symptoms of early satiety, nausea, and/or vomiting with no definable organic cause. The Rome II criteria (Table I) help to diagnose functional dyspepsia. An upper endoscopy is suggested for evaluation of persistent or alarming dyspeptic symptoms. In young patients with uninvestigated dyspepsia, testing for *Helicobacter pylori* and treating if tests are positive will improve symptoms from underlying ulcer disease and will help prevent future development of ulcers. However, in patients with functional dyspepsia who have a negative endoscopy but a positive *H. pylori* test, the symptom response to *H. pylori* therapy is marginal. Lifestyle and dietary modifications are often suggested in the initial treatment. Antacids and over-the-counter histamine-2 (H2) receptor antagonists may be helpful as an "on-demand" therapy for intermittent symptoms. If the predominant symptom is epigastric pain (ulcerlike functional dyspepsia), H2 receptor antagonists or proton pump inhibitors are the initial treatment of choice. If fullness, bloating, early satiety, or nausea is the predominant complaint (dysmotility-like functional dyspepsia), a prokinetic agent may help. Metoclopramide is the only available effective prokinetic agent at present. If these treatments fail, patient reevaluation for other disorders (including other functional bowel disorders) is advised. Low-dose tricyclic antidepressant agents may be helpful for treatment of refractory dyspeptic symptoms.

TABLE I Rome II Criteria (1999) for Functional Dyspepsia

Persistent or recurrent abdominal pain or discomfort centered in the upper abdomen; discomfort is defined as an unpleasant sensation and may include fullness, bloating, early satiety, and nausea
Symptom duration of at least 12 weeks, which need not be consecutive, within the preceding 12 months
No evidence of organic disease (including at upper endoscopy) that is likely to explain the symptoms
No evidence that dyspepsia is exclusively relieved by defecation or is associated with the onset of a change in stool frequency or stool form (i.e., not irritable bowel syndrome)

The 1999 Rome II criteria for the diagnosis of functional dyspepsia listed in Table I have been useful in standardizing the diagnosis of functional dyspepsia in clinical studies. The Rome group has also suggested subcategorizing patient diagnoses, based on the predominant symptom, as ulcerlike, dysmotility-like, and nonspecific dyspepsias (Table II). Refluxlike dyspepsia in the Rome I criteria, with emphasis on predominant heartburn/regurgitation, is now categorized as primary gastroesophageal reflux disease (GERD). Specific symptoms, however, do not always correlate with pathophysiological abnormalities and response to treatment.

A variety of abnormalities have been implicated in functional dyspepsia (Table III). These include

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Rome II Subgroups for Functional Dyspepsia (Based on Predominant or Most Bothersome Symptom)[a]

Subgroup	Symptom
Ulcer like	Pain in upper abdomen
Dysmotility-like	Discomfort, fullness, early satiety, bloating, nausea
Nonspecific	No predominant symptom

[a] Overlap syndromes (not primarily dyspepsia, if main symptom): gastroesophageal reflux disease, with heartburn and acid regurgitation; irritable bowel syndrome, with pain relief following defecation; and biliary tract disease, with right upper quadrant abdominal pain.

inflammation, motility disturbances, visceral hypersensitivity, and psychological factors. Delayed gastric emptying, impaired gastric accommodation to a meal, gastric dysrhythmias, and visceral hypersensitivity are important pathophysiological factors. Delayed gastric emptying is present in approximately 35% of patients. As measured by electrogastrography (EGG), gastric myoelectrical abnormalities such as tachygastria, bradygastria, and a decreased postprandial : fasting power ratio are found in about 40% of patients with functional dyspepsia. Specific dyspeptic symptoms and their severity have generally correlated poorly with the degree of gastric stasis. However, in a large series of patients with functional dyspepsia, the severity of postprandial fullness and vomiting correlated with delayed gastric emptying. Regional gastric function abnormalities may be present in many dyspeptic patients and appear to correlate with symptoms. Reduced postprandial fundic relaxation and impaired accommodation are found in 40% of patients with functional dyspepsia. Transit from the proximal stomach into the antrum is rapid, leading to early antral distension, which correlates with symptoms. Hypersensitivity to gastric distension occurs in 35% of patients. Specific symptoms in dyspeptic patients may be related to the underlying pathophysiological abnormality. Early satiety is associated with impaired fundic accommodation, pain and belching are associated with hypersensitivity to gastric distension, and vomiting appears to be associated with delayed gastric emptying.

TABLE III Pathophysiological Mechanisms Linked to Functional Dyspepsia

Inflammation	*Visceral hypersensitivity*
Gastric acid sensitivity	Gastric
Helicobacter pylori gastritis	Duodenal
Enterogastric (bile) reflux	
Motility disturbances	*Psychologic factors*
Delayed gastric emptying/antral hypomotility	
Impaired fundic relaxation/antral distension to meal	

EVALUATION

Many patients seeking medical attention for dyspeptic symptoms are concerned about the possibility of other serious diseases, rather than the symptoms of dyspepsia. Definitive diagnosis of functional dyspepsia requires a normal esophagogastroduodenoscopy (upper endoscopy) during a symptomatic period when patients are not taking agents that suppress gastric acid. Upper endoscopy is not usually required in young patients (<50 years old) unless there are "alarm symptoms" such as dysphagia, weight loss, or anemia. Endoscopy rules out the presence of serious disease and helps to reassure the patient. This reassurance may be a useful initial step in management; some studies have shown that the endoscopic procedure alone may reduce symptoms.

Further noninvasive tests, such as gastric emptying scintigraphy and electrogastrography, can be performed to identify delayed gastric emptying and gastric dysrhythmias. These tests can demonstrate abnormalities in approximately 30–45% of patients with functional dyspepsia. As compared to patients with normal emptying, patients with delayed gastric emptying or an abnormal EGG may respond better to prokinetic agents, thus these tests may have prognostic importance. However, the link between prokinetic agent treatment and improved gastric emptying and relief from symptoms has not been proved.

Satiety testing with water or a nutrient drink may be used to evaluate impaired accommodation and sensation. Satiety tests offer the potential to evaluate gastric accommodation in a noninvasive way.

TREATMENT

Treatment of functional dyspepsia is influenced by several factors, including the episodic nature of symptoms, the occurrence of spontaneous remission, the heterogeneity of abnormalities, and the high placebo response rate, all of which have made it difficult to evaluate the actual response to drugs in many clinical studies. At the present time, there are several approaches to treatment, due in part to the proposed multiple causes of functional dyspepsia (Table IV). The treatment modalities that have been tested extensively are gastric acid suppressants, promotility agents, and *H. pylori* eradication. The aim of therapy in functional dyspepsia is to provide adequate symptom control and, if possible, to abolish symptoms. Another important aspect of drug therapy is

TABLE IV Multiple Causes and Therapies Proposed for Functional Dyspepsia

Multiple causes	Multiple therapies
Gastric acid	Antisecretory agents
Helicobacter pylori	Antibiotics
Dysmotility	Prokinetic agents
Hypersensitivity	Sensory modulators
Psychogenic	Psychotropic agents

to improve patient functioning and quality of life while minimizing the side effects.

Lifestyle and Dietary Changes

At present there is little evidence to implicate dietary habits in causing dyspeptic symptoms or to suggest that dietary modifications help relieve symptoms. Small, frequent meals rather than large meals should help relieve symptoms, assuming that the etiology involves receptive fundic relaxation and antral distension. Eating a low-fat diet, decreasing coffee intake, cessation of smoking, and limiting alcohol intake might be helpful in some patients. Patients should be advised to stop taking nonsteroidal antiinflammatory drugs (NSAIDs) and any other medication likely to cause dyspeptic symptoms, including digoxin, potassium supplements, and erythromycin. A lactose-free diet may be helpful in some patients. A high-fiber diet may delay gastric emptying and make the symptoms, especially bloating, worse.

Gastric Acid Suppressants

Gastric acid suppressants have been tested for several decades for their potential to treat functional dyspepsia. In general, improvement in dyspeptic symptoms with acid-suppressing agents has been moderate, averaging only 25% greater than with placebo. Antacids provide little help for chronic symptoms. Fourteen of the 24 studies that used prescription doses of acid-suppressing agents to treat nonulcer dyspepsia have reported a positive effect on symptoms in 35–80% of patients, compared to improvement in 30–60% of patients using placebo.

Histamine type 2 receptor antagonists (H2RAs) have a therapeutic effect of approximately 20% over placebo as reported in meta-analyses. However, a number of well-designed trials showed no benefit. The best response is seen in patients with endoscopy-negative GERD overlapping with functional dyspepsia.

Proton pump inhibitors (PPIs) are superior to placebo in promoting the complete relief of symptoms in ulcerlike dyspepsia with predominant epigastric pain. Complete relief of symptoms was obtained in 38% of patients treated with omeprazole (20 mg/day), compared to 28% with placebo. Omeprazole was not effective in dysmotility-like dyspepsia. In the near future, omeprazole will become generic as well as an over-the-counter medication, and the costs of PPI therapy will be dramatically reduced.

Promotility Agents

Promotility compounds, including metoclopramide, cisapride, and domperidone, have improved dyspeptic symptoms in patients with functional dyspepsia more effectively than placebo in the majority of reported studies. On average, the improvement was 40–50% greater than with placebo. The few studies comparing the effects of acid-suppressing and promotility agents on dyspeptic symptoms have favored promotility agents. In a recent meta-analysis, both cisapride and domperidone seemed to be effective in treating functional dyspepsia. One meta-analysis, however, suggested the possibility of publication bias with prokinetic agents. Cisapride was taken off the market due to its adverse cardiac effects and domperidone is not available in the United States. This limits the current choice of available agents to metoclopramide.

Metoclopramide is presently approved for short-term treatment of gastroparesis. Although a limited number of studies have directly tested use of this agent in functional dyspepsia, metoclopramide appears to be superior to placebo in the treatment of dysmotility-like dyspepsia. The main side effects are antidopaminergic and include sedation, diarrhea, and extrapyramidal symptoms, which occur in up to 25% of patients. These adverse reactions are most common in younger patients. Tardive dyskinesia, which may be irreversible even after stopping the medication, occurs rarely in elderly patients.

Erythromycin and other motilin receptor agonists have not demonstrated efficacy in treating functional dyspepsia. The motilin receptor agonist ABT-229 is ineffective in dyspeptic patients with delayed gastric emptying. Erythromycin and other motilin receptor agonists increase proximal gastric tone, which possibly aggravates the impaired fundic accommodation and worsens symptoms in functional dyspepsia.

Newer prokinetic agents under evaluation for treatment of functional dyspepsia include 5-hydroxytryptamine isotype 4 (5-HT4) agonists (tegaserod), dopamine antagonists (levosulpiride analogues), cholecystokinin antagonists (loxiglumide), and several motilin agonists. Tegaserod has been shown to improve symptoms and

gastric emptying in some studies. Fundic relaxing agents may improve the fundic accommodation response and may be helpful in patients with early satiety. 5-HT1 agonists (sumatriptan, buspirone), α-adrenergic receptor agonists (clonidine), and nitric oxide donors (glyceryl trinitrate) have been tried in a small number of series. Sumatriptan induces fundic relaxation through a nitric oxide-mediated pathway and improves meal-induced satiety in patients with functional dyspepsia. It has to be administered subcutaneously and can cause headaches. The α-adrenergic agonist clonidine reduces proximal gastric tone and pain perception during gastric distension in normal subjects. Thus, clonidine has the potential to reduce gastric sensation and increase gastric compliance. Clonidine, however, can delay gastric emptying and cause significant hypotension. Sildenafil (Viagra) has been shown to increase gastric accommodation. In animal models, it also relaxes the pylorus and improves gastric emptying, but in humans it can slow gastric emptying.

Eradication of *H. pylori*

Helicobacter pylori eradication is often suggested as first-line therapy for uninvestigated dyspepsia; patients with dyspeptic symptoms caused by ulcers will often be cured with this approach. The overall benefit of *H. pylori* eradication in uninvestigated dyspepsia depends on the underlying ulcer rate and the proportion of cases of ulcers that are associated with *H. pylori*. The latter has declined from a 90% association of *H. pylori* with duodenal ulcers in the 1980s to 60–65% at the present time. Although *H. pylori* is detected slightly more frequently in functional dyspeptic patients than in normal controls, only three of eight studies found that eradication of the bacteria improved patient symptoms. The 1994 National Institutes of Health (NIH) consensus conference on *H. pylori* did not support *H. pylori* eradication for treatment of functional dyspepsia: only 1 patient in 15 is likely to have improved symptoms. However, *H. pylori* eradication may decrease ulcer diathesis and may reduce the need for endoscopy, potentially reducing treatment costs. On a global scale, it may also reduce the eventual development of intestinal mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. The potential drawbacks are treatment costs, side effects of antibiotic therapy, and antibiotic resistance. Moreover, *H. pylori* eradication may increase the chance of development of GERD/reflux esophagitis. Two recent large randomized studies on the effect *H. pylori* eradication in functional dyspepsia showed different results. One study suggested that symptomatic benefit from eradication may occur in a few patients whereas the other study suggested no benefit at all. The 1997 American Digestive Health Foundation (ADHF) panel on *H. pylori* and functional dyspepsia suggested that testing for *H. pylori* should be undertaken only if treatment is intended for a positive test. Testing should be considered in patients younger than age 45 years presenting with dyspepsia for the first time and in whom an ulcer is likely.

Psychotropic Medications

Psychotropic medications, including tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs), are being used in the treatment of functional dyspepsia based on the evidence that augmented visceral sensitivity may be an important pathophysiological factor. Low-dose tricyclic antidepressants have analgesic effects that are independent of their psychological effects. The mechanism of action in treatment of functional dyspepsia is not clear, but may include reduction in visceral sensitivity. Two studies using tricyclic antidepressants reported a decrease in dyspeptic symptoms. In clinical practice, psychotropic medications have been recommended for patients with severe or intractable symptoms that interfere with their daily activities and are unresponsive to other treatment regimes. Major side effects are related to the anticholinergic action, causing dry mouth, blurred vision, and urinary retention. Orthostatic hypotension and reflex tachycardia can occur in the elderly. Sedation may be prominent, especially during the first few weeks of therapy; nighttime dosing is used to prevent this. In patients with poor sleep habits, the sedation effect can sometimes be advantageous. Agents with strong anticholinergic properties, such as amitriptyline, may adversely affect gastric motility. However, compared to amitriptyline, nortriptyline (Pamelor) and desipramine (Norpramin) have fewer side effects. Amitriptyline is the most commonly used and most thoroughly studied drug in this class for the treatment of functional dyspepsia. Dosages required for symptomatic improvement and complete remission are usually much lower than those conventionally used to treat depression. SSRIs may also be helpful but have not been studied in large-scale clinical trials.

Alternative Medicine

Complementary and alternative medicines are used frequently by patients with functional dyspepsia. Alternative medicine regimes, including special diets with herbal teas, yoga, and acupuncture, are used in Germany. Some commercially available herbal preparations, including Iberogast, which has extracts from

bitter candy tuft, and Enteroplant, which contains peppermint oil and caraway oil, have been shown in placebo-controlled trials to improve symptoms. Psychodynamic interpersonal psychotherapy and hypnotherapy have been reported effective in treating functional dyspepsia.

See Also the Following Articles

Antacids • Budd–Chiari Syndrome • Complementary and Alternative Medicine • Electrogastrography • Gastric Emptying • Gastric Motility • Gastroesophageal Reflux Disease (GERD) • H2-Receptor Antagonists • *Helicobacter pylori* • Proton Pump Inhibitors • Upper Gastrointestinal Endoscopy

Further Reading

Blum, A. L., Talley, N. J., O' Morain, C., *et al.* (1998). Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1875–1881.

Clouse, R. E. (1994). Antidepressants for functional gastrointestinal syndromes. *Dig. Dis. Sci.* **39**, 2352–2363.

Finney, J. S., Kinnersley, N., Hughes, M., *et al.* (1998). Meta-analysis of antisecretory and gastrokinetic compounds in functional dyspepsia. *J. Clin. Gastroenterol* **26**, 312–320.

Fisher, R. S., and Parkman, H. P. (1998). Management of nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1376–1381.

Friedman, L. S. (1998). *Helicobacter pylori* and nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1928–1930.

McColl, K., Murray, L., El-Omar, E., *et al.* (1998). Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with non-ulcer dyspepsia. *N. Engl. J. Med.* **339**, 1869–1874.

Tack, J., Coulie, B., Andrioli, A., *et al.* (1998). Influence of sumatriptan on gastric fundus tone and of the perception of gastric distension in man. *Gut* **46**, 468–473.

Talley, N. J., Silverstein, M. D., Agreus, L., *et al.* (1998). AGA technical review: Evaluation of dyspepsia. *Gastroenterology* **114**, 582–595.

Talley, N. J., Stanghellini, V., Heading, R. C., *et al.* (1998). Functional gastrointestinal disorders. *Gut* **45**(Suppl. II), 1137–1142.

Veldhuyzen van Zanten, S. J. O., Cleary, C., Talley, N. J., *et al.* (1996). Drug treatment of functional dyspepsia. A systematic analysis of trial methodology with recommendations for design of future trials. *Am. J. Gastroenterol.* **91**, 660–673.

Fungal Infections

MICHAEL ELLIS

United Arab Emirates University Medical School, Al-Ain

highly active antiretroviral therapy Combination of different anti-human immunodeficiency virus (HIV) drugs that have markedly increased the prognosis for patients with HIV.

leukoplakia Thick white patch on mucous membrane.

peritonitis Inflammation of the peritoneum.

pseudomembrane False membrane. A layer of exudate on surface of skin or mucous membrane.

The increasing tide of fungal infections reflects the upsurge in the immunocompromised population. *Candida* spp. now rank fourth among bloodstream pathogens causing fungemia. The pulmonary tract, sinus tract, and skin are common target organs whereas the gastrointestinal tract rarely is involved apart from the oropharyngo-esophageal tract, the peritoneal cavity, and the hepatobiliary system. Compared to bacteria, fungi are distinctly unusual pathogens in the etiology of gastrointestinal mucosal–luminal infectious disease. Nevertheless, they cause substantial morbidity and mortality, pose problems for precision diagnosis, and can be challenging to treat. Fungal colonization in the gut serves as an important reservoir for bloodstream dissemination when the mucosal barrier is compromised for example, during chemotherapy-induced mucositis.

CANDIDAL OROPHARYNGO-ESOPHAGITIS AND NONESOPHAGEAL GUT SITES

Candida spp., mainly *Candida albicans*, colonize the gastrointestinal tract in up to 55% of normal individuals, depending on diet, age, and other factors. Hospitalization, broad-spectrum antibiotic use, steroids, and other immunosuppressive agents, e.g., in

 Copyright 2004, Elsevier (USA). All rights reserved.

bitter candy tuft, and Enteroplant, which contains peppermint oil and caraway oil, have been shown in placebo-controlled trials to improve symptoms. Psychodynamic interpersonal psychotherapy and hypnotherapy have been reported effective in treating functional dyspepsia.

See Also the Following Articles

Antacids • Budd–Chiari Syndrome • Complementary and Alternative Medicine • Electrogastrography • Gastric Emptying • Gastric Motility • Gastroesophageal Reflux Disease (GERD) • H2-Receptor Antagonists • *Helicobacter pylori* • Proton Pump Inhibitors • Upper Gastrointestinal Endoscopy

Further Reading

Blum, A. L., Talley, N. J., O' Morain, C., *et al.* (1998). Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1875–1881.

Clouse, R. E. (1994). Antidepressants for functional gastrointestinal syndromes. *Dig. Dis. Sci.* **39**, 2352–2363.

Finney, J. S., Kinnersley, N., Hughes, M., *et al.* (1998). Meta-analysis of antisecretory and gastrokinetic compounds in functional dyspepsia. *J. Clin. Gastroenterol* **26**, 312–320.

Fisher, R. S., and Parkman, H. P. (1998). Management of nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1376–1381.

Friedman, L. S. (1998). *Helicobacter pylori* and nonulcer dyspepsia. *N. Engl. J. Med.* **339**, 1928–1930.

McColl, K., Murray, L., El-Omar, E., *et al.* (1998). Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with non-ulcer dyspepsia. *N. Engl. J. Med.* **339**, 1869–1874.

Tack, J., Coulie, B., Andrioli, A., *et al.* (1998). Influence of sumatriptan on gastric fundus tone and of the perception of gastric distension in man. *Gut* **46**, 468–473.

Talley, N. J., Silverstein, M. D., Agreus, L., *et al.* (1998). AGA technical review: Evaluation of dyspepsia. *Gastroenterology* **114**, 582–595.

Talley, N. J., Stanghellini, V., Heading, R. C., *et al.* (1998). Functional gastrointestinal disorders. *Gut* **45**(Suppl. II), 1137–1142.

Veldhuyzen van Zanten, S. J. O., Cleary, C., Talley, N. J., *et al.* (1996). Drug treatment of functional dyspepsia. A systematic analysis of trial methodology with recommendations for design of future trials. *Am. J. Gastroenterol.* **91**, 660–673.

Fungal Infections

MICHAEL ELLIS

United Arab Emirates University Medical School, Al-Ain

highly active antiretroviral therapy Combination of different anti-human immunodeficiency virus (HIV) drugs that have markedly increased the prognosis for patients with HIV.

leukoplakia Thick white patch on mucous membrane.

peritonitis Inflammation of the peritoneum.

pseudomembrane False membrane. A layer of exudate on surface of skin or mucous membrane.

The increasing tide of fungal infections reflects the upsurge in the immunocompromised population. *Candida* spp. now rank fourth among bloodstream pathogens causing fungemia. The pulmonary tract, sinus tract, and skin are common target organs whereas the gastrointestinal tract rarely is involved apart from the oropharyngo-esophageal tract, the peritoneal cavity, and the hepatobiliary system. Compared to bacteria, fungi are distinctly unusual pathogens in the etiology of gastrointestinal mucosal–luminal infectious disease. Nevertheless, they cause substantial morbidity and mortality, pose problems for precision diagnosis, and can be challenging to treat. Fungal colonization in the gut serves as an important reservoir for bloodstream dissemination when the mucosal barrier is compromised for example, during chemotherapy-induced mucositis.

CANDIDAL OROPHARYNGO-ESOPHAGITIS AND NONESOPHAGEAL GUT SITES

Candida spp., mainly *Candida albicans*, colonize the gastrointestinal tract in up to 55% of normal individuals, depending on diet, age, and other factors. Hospitalization, broad-spectrum antibiotic use, steroids, and other immunosuppressive agents, e.g., in

 Copyright 2004, Elsevier (USA). All rights reserved.

organ transplantation, increase fecal carriage rate and intensity. Widespread azole antifungal usage has been significantly associated with increased colonization of non-*albicans* spp., which may be resistant to azoles and other antifungal drugs. However, colonization does not equate with clinical infection and claims that the presence of gut *Candida* per se in an otherwise normal individual produces a symptom complex of diarrhea, debility, and weight loss has not been substantiated by vigorous clinical studies. On the other hand, there is little dispute that mucosal plaque formation and ulceration occur in patients with cancer and primary or acquired immunodeficiency states, particularly HIV infection. Oropharyngeal candidiasis presents clinically in four aspects: (1) pseudomembrane formation (Fig. 1), in which the plaques that constitute the pseudomembranes are composed of superficial hyperplastic, hyperkeratotic squamous epithelium with inflammatory cells, infiltrating fungal elements, and superinfecting bacteria; (2) acute erythematous epithelial atrophy, thinning, and papillary loss, which may be secondary to the pseudomembranous form; (3) chronic atrophy typified by mucosal edema and erythema (particularly in the elderly patients who have dentures), which may lead to hyperplasia (*Candida* leukoplakia), possibly premalignant; and (4) angular fissuring cheilitis.

Most cases present in the setting of HIV infection in which the key immunologic defect is a low CD4 T cell count. In non-HIV situations, steroid use, poor dental care, smoking, and antibiotic use are important predisposing factors. Increasingly recognized are other severe immunosuppressive states, including renal transplant patients receiving immunosuppressive therapy. Apart from the chronic atrophic form, oral discomfort is the commonest symptom and may result in poor food intake. Extension of the infection to the esophagus

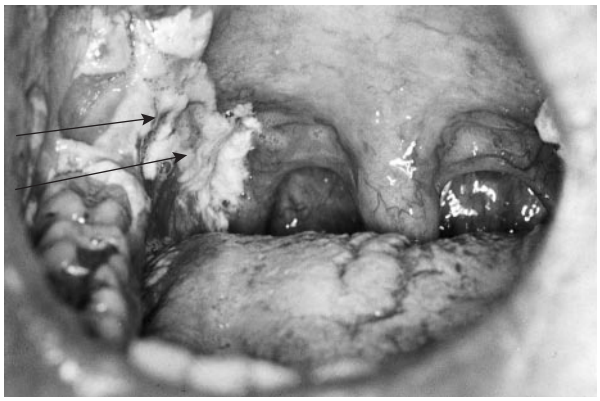


FIGURE 1 Pseudomembranous candidiasis of the buccal cavity in an HIV-infected individual.

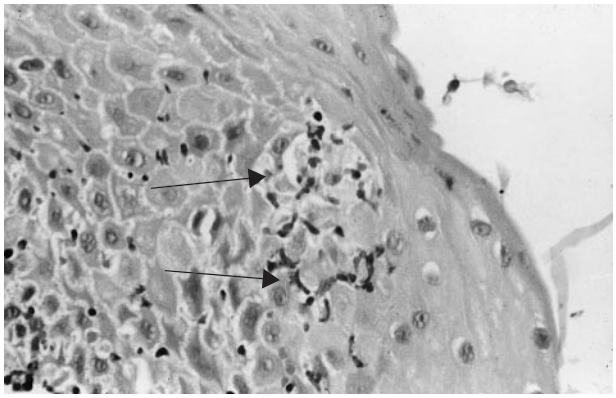


FIGURE 2 Esophageal biopsy showing infiltrating fungal elements from an HIV-infected patient with esophageal candidiasis.

(Fig. 2) results in painful and difficult swallowing due to the erosions and ulcerations (Fig. 3a). More severe esophageal disease is associated with coalescence of the plaques, leading to circumferential disease and luminal impingement. The combination of oral candidiasis and retrosternal pain on swallowing is reliably predictive of esophageal candidiasis, without recourse to endoscopy or barium swallow imaging examinations. Systemic candidiasis arising from this source is rare.

Oroesophageal candidiasis is best treated with systemically administered antifungals. Oral fluconazole (200–400 mg/day) or itraconazole, with the cyclodextrin vehicle to optimize absorption (100–200 mg/day), given for 2 weeks, is effective, with a rapid response seen in many patients (Fig. 3b). The relapses that occur in more than 30% of HIV patients are preferably managed by treating each episode rather than placing the patient on prophylactic antifungals, a course that fuels the emergence of drug resistance. Nevertheless, around 10% of all patients with advanced HIV infection (CD4 T cell count $< 100/\mu l$) have fluconazole-resistant infection associated with high minimum inhibitory concentrations (MICs) to the drug. Such cases can be managed either with high dosages of fluconazole (up to 1600 mg/day) or with liposomal amphotericin B or caspofungin. An integral part of the management of these patients is correction of the underlying immunodeficiency; for example, treating with highly active antiretroviral therapy (HAART) in HIV patients raises the CD4 T cell count.

After the esophagus, the stomach is the next most frequently involved site, followed by the small and large bowels. Underlying malignancy is often the setting in these patients. The etiologic role of histamine-2 (H2) blockers in increasing fungal disease susceptability remains a moot point. Plaques, fold thickening, superficial

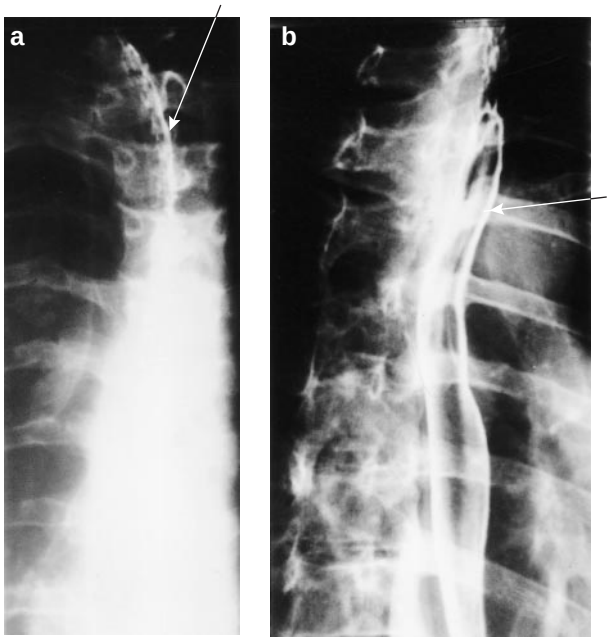


FIGURE 3 (a) Barium swallow, showing extensive ulceration and narrowing of the lumen. (b) The same patient after 3 days or oral fluconazole: normal swallow.

and penetrating ulcers, and perforation have all been described.

HEPATOSPLENIC CANDIDIASIS

Hepatosplenic candidiasis is a distinct form of disseminated candidiasis that occurs in the setting of leukemia. The classic scenario is of a patient with neutropenic fever and negative bacterial cultures, absent clinical foci, who subsequently resolves the neutropenia but has persistent fever for up to several weeks. Other cardinal features include weight loss, vague general unwell being and upper abdominal pain. There is hepatomegaly or splenomegaly or both in one-half of the patients. The hallmark biochemical abnormality is an elevated serum alkaline phosphatase. Upper abdominal imaging reveals multiple translucent lesions in the liver and spleen (Fig. 4). Liver biopsy may show fungal elements but tissue culture and blood culture are usually negative. Serological examination with *Candida* mannan antigen and antibody can be helpful in that situation (Fig. 5). Management with either liposomal amphotericin B (3–5 mg/kg/day) or fluconazole ($\geq$400 mg/day) continuing for 4 weeks after symptoms settle achieves success in 80% of cases. Widespread fluconazole prophylaxis use in leukemia patients has decreased the incidence of hepatosplenic candidiasis dramatically in recent years.

FUNGAL PERITONITIS

Fungal infections of the peritoneal cavity are usually due to *Candida* spp. They have become increasingly recognized, paralleling the increased numbers of patients undergoing continuous ambulatory peritoneal dialysis (infection incidence up to 20%) and gastrointestinal surgery (often associated with bacterial peritonitis and insidious in onset, but especially with anastomotic breakdown situations). Broad-spectrum antibiotic usage, parenteral nutrition, and immunosuppression are other important predisposing factors. Dissemination in patients who have candidal peritonitis in general is uncommon but occurs in up to 25% of cases of peritonitis secondary to intestinal perforation and following gastrointestinal surgery, with mortality as high as 80%. The presence of *Candida* in intraoperative specimens recovered from patients with intraabdominal perforations significantly relates to the risk of death. Fungal peritonitis is best managed with systemic antifungals and removal of the dialysis catheter if appropriate.

EMERGING AND UNUSUAL FUNGAL INFECTIONS

Several new fungal pathogens have emerged in recent years as important but rare causes of gastrointestinal disease. Increased treatment involving immunosuppression, selective antifungal drug pressure, and changing hospital environmental conditions drive this emergence. Dematiaceous fungi, including *Cladophialophora*, which causes esophagitis in small intestinal transplantation, and *Penicillium*, which causes peritonitis in chronic ambulatory peritoneal dialysis, have been described. Basidiobolomycosis, caused by a member of the Entomophthoraceae family, can mimic

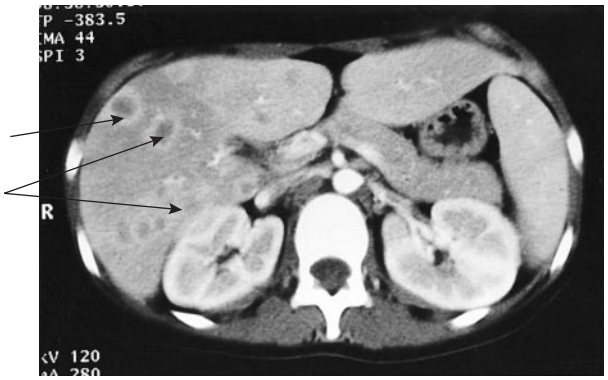


FIGURE 4 Computed tomography scan of a patient with acute myeloid leukemia and fever unresponsive to antibiotics, showing multiple hypolucent liver densities with enhanced rims.

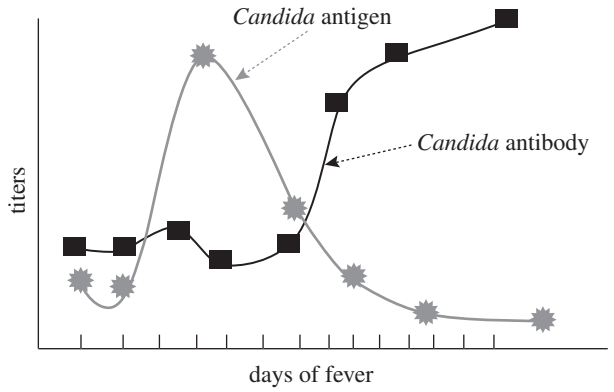


FIGURE 5 Serial *Candida* antigen (mannan) and antibody (antimannan) measurements in a patient with hepatosplenic candidiasis.

inflammatory bowel disease. Other members of the Zygomycetes, known gut wall invaders, include *Absidia*, *Rhizopus*, and *Mucor*, which producing nonspecific abdominal pain and hematemesis. Perforation can occur and is usually fatal. Another feature of the changing fungal profile is the increasing incidence of non-*albicans Candida* spp.

Several well-known invasive fungi have expanded their target organs to include the gastro-hepatobiliary system. Among these are *Aspergillus* spp., now documented in esophagitis and cecal deep wall involvement with ulceration and hemorrhage. Recent autopsy studies have indicated that in patients with hematological malignancy, *Aspergillus* infection occurred in 11% and the gastrointestinal tract was involved in 20%, which is an extraordinarily high figure. Most of these infections are not diagnosed postmortem.

Candida spp. are also known to involve the gallbladder, where they can present as a fungal ball, and bile ducts. *Cryptococcus* spp. are known to produce esophagitis and peritonitis. All of these new and unusual fungal infections, although uncommon, present considerable challenges for diagnosis and successful treatment.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Candidiasis • Gastric Infection (non-*H. pylori*) • Microflora, Overview

Further Reading

Blinzler, L., Fischer, K., Just, H.-M., and Heuser, D. (1997). Role of fungal infections in peritonitis. *Langenbecks Archiv. Chir.* **383**, S5–S8.

Collier, L., Balows, A., and Sussmann, M. (1998). "Topley and Wilson's Microbiology and Microbial Infections. Volume 4, Medical Mycology," 9th Ed. Arnold, London, Sydney, Auckland.

Edwards, J. E., Jr., Bodey, G. P., Bowden, R. A., *et al.* (1997). International conference for the development of a consensus on the management and prevention of severe candidal infections. *Clin. Infect. Dis.* **25**, 43–59.

Fridkin, S. K., and Jarvis, W. R. (1996). Epidemiology of nosocomial fungal infections. *Clin. Microbiol. Rev.* **9**, 499–511.

Kami, M., Hori, A., Takaue, Y., and Mutou, Y. (2002). The gastrointestinal tract is a common target of invasive aspergillosis in patients receiving cytotoxic chemotherapy for hematological malignancy. *Clin. Infect Dis.* **35**, 105–106.

Kontoyiannis, D. P., Luna, M. A., Samuels, B. I., and Bodey, G. P. (2000). Hepatosplenic candidiasis. A manifestation of chronic disseminated candidiasis. *Infect. Dis. Clin. North Am.* **14**, 721–739.

Vazquez, J. A. (2000). Therapeutic options for the management of oropharyngeal and esophageal candidiasis in HIV/AIDS patients. *HIV Clin. Trials* **1**, 47–59.

Galactosemia

DOROTHY K. GRANGE
Washington University, St. Louis, Missouri

allele Alternate form of the DNA sequence of a gene at a given locus (position of a gene on a chromosome).
carrier A person who is a heterozygote for a mutant allele.
heterozygote An individual with two different alleles at a specific locus.
hypergonadotropic hypogonadism Reduced function of the gonads (ovaries or testes) due to congenital failure of development or postnatal damage or destruction, associated with elevation of gonadotropin (pituitary hormone) levels.
missense mutation A mutation that changes a codon for a specific amino acid to a different amino acid and results in a change in the amino acid sequence of a protein.
point mutation A mutation that involves a single nucleotide base-pair change in the DNA sequence of a gene.
verbal dyspraxia Expressive speech problem that results from the inability to properly program the muscles needed for speech, manifested by articulation problems, incorrect usage of words, or errors in word ordering.

Galactosemia is an autosomal recessive inborn error of carbohydrate metabolism that results in the inability to metabolize galactose to glucose. It occurs in approximately 1 in 60,000 newborns. Mason and Turner originally described the condition in 1935, when galactose was found in the urine of an affected patient. Individuals with galactosemia are intolerant of dietary lactose and galactose, primarily found in milk and milk products. If untreated, the disorder can cause liver failure, kidney dysfunction, sepsis, and death. If it is diagnosed soon after birth and treated by removal of lactose and galactose from the diet, the symptoms will resolve and many of the long-term complications, including cataracts and mental retardation, can be prevented.

BIOCHEMICAL PATHWAY

Dietary lactose is hydrolyzed to glucose and galactose by lactase in the intestine. Galactose is then further metabolized to uridyl diphosphate (UDP)-glucose by a series of reactions involving three enzymes, galactokinase (GALK), galactose-1-phosphate (PO_4) uridyl transferase (GALT), and UDP galactose-4-epimerase (GALE). Although galactosemia and galactosuria can be caused by a deficiency of any one of these enzymes, classic galactosemia, the condition of greatest clinical importance, is caused by a defect in GALT. Galactokinase catalyzes the conversion of galactose to galactose-1-PO_4 (Fig. 1). Galactose-1-PO_4 is then converted to glucose-1-PO_4 by a two-step reaction involving GALT and UDP galactose-4-epimerase. GALT catalyzes the transfer of UDP to galactose-1-PO_4 to form UDP galactose. UDP-galactose is then converted to UDP-glucose by UDP galactose-4-epimerase. Galactose is incorporated into glycoproteins and glycolipids, using UDP-galactose as the substrate for galactosylation reactions. In individuals with GALT deficiency, galactose and galactose-1-PO_4 accumulate in tissues, as well as the reduction product, galactitol, and the oxidation product, galactonate, which are derived from alternate metabolic pathways for galactose.

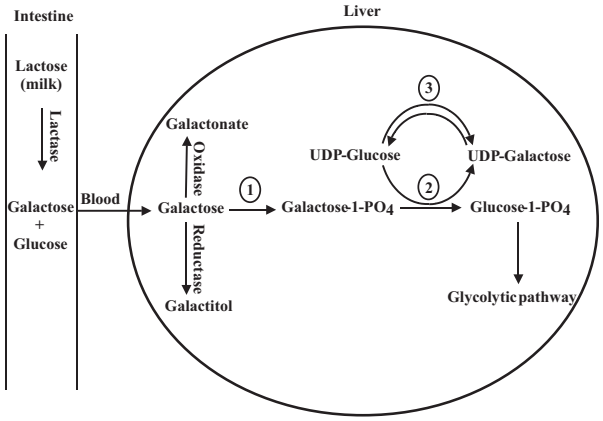


FIGURE 1 Lactose is metabolized by lactase in the intestine to galactose and glucose. Galactose is metabolized in the liver via the pathway shown in the figure. The enzymes required include (1) galactokinase, (2) galactose-1-PO_4 uridyltransferase (GALT), and (3) UDP galactose-4-epimerase. The defective enzyme in classic galactosemia is GALT (2). Galactose-1-PO_4 and galactose, as well as galactitol and galactonate, are elevated in patients with classic galactosemia.

CLINICAL PRESENTATION

A newborn infant with GALT deficiency appears normal at birth. However, if the baby is fed breast milk or a

 Copyright 2004, Elsevier (USA). All rights reserved.

lactose-based infant formula, symptoms will usually begin to develop within the first few days of life. Symptoms include vomiting, diarrhea, poor weight gain or weight loss, jaundice, and lethargy (see Table I). The baby will develop hepatomegaly and sometimes encephalopathy. If lactose-containing feedings are continued, the condition can be rapidly fatal.

Laboratory testing reveals liver dysfunction with elevated liver transaminases (alanine aminotransferase and aspartate aminotransferase), prolonged prothrombin time (PT) and partial thromboplastin time (PTT), and elevated bilirubin and ammonia. Hypoglycemia occasionally occurs during acute neonatal illness. The hyperbilirubinemia is primarily unconjugated initially, but there may be a conjugated component as the liver disease worsens over several days. The coagulopathy may be severe and can lead to significant bleeding. Vitreal hemorrhage and intracranial hemorrhage have been reported as complications. Kidney dysfunction, with development of renal Fanconi syndrome, is also a frequent complication. There is evidence of renal tubular disease, with hyperchloremic metabolic acidosis, glycosuria, aminoaciduria, and albuminuria. Urinalysis may reveal a positive test for reducing substances, indicating galactosuria. It is important to note that galactosuria can clear rapidly if the infant is placed on intravenous fluids, and thus, the absence of reducing substances in the urine does not rule out galactosemia. Other sugars in the urine besides galactose, such as fructose or pentose, can cause the test to be positive. Laboratory testing in suspected cases of galactosemia should include quantitative red blood cell (RBC) GALT activity and galactose-1-PO_4 measurements. RBC levels of galactose-1-PO_4 are markedly elevated in affected infants. RBC GALT activity is extremely low or absent in classic galactosemia.

In the untreated infant, cataracts can develop shortly after birth, although they most commonly develop after 3–4 weeks of lactose feedings. All infants with galactosemia should have an ophthalmology evaluation, which should include a slit lamp examination, since the cataracts may be missed on routine ophthalmoscopy. The cataracts in galactosemia are thought to be caused by injury to the lens from osmotic swelling due to the accumulation of galactitol. Cataracts usually regress spontaneously after treatment is initiated. The presence of such cataracts may aid in diagnosis while awaiting the results of laboratory testing.

Encephalopathy may occur in affected newborns, resulting in hypotonia, lethargy, and a reduced level of consciousness. Occasionally, pseudotumor cerebri may develop due to increased intracranial pressure; there may be increasing head circumference with bulging of the anterior fontanel and evidence of brain edema on computed tomography or magnetic resonance imaging scan. The encephalopathy is believed to be related to elevated galactitol in the brain, in some cases in combination with hyperammonemia. If galactosemia is untreated or diagnosed late, patients may have irreversible neurologic injury, with variable degrees of mental retardation.

Infants with galactosemia have an increased susceptibility to sepsis, particularly due to *Escherichia coli* and other gram-negative bacteria. Granulocyte function has been shown to be impaired by galactose and its metabolites. Thus, all ill newborns with galactosemia

TABLE I Clinical Findings in Symptomatic Neonates with Galactosemia

Abnormality	Percentage of patients affected	Additional findings
Hepatocellular damage	89	Jaundice (74%) Hepatomegaly (43%) Abnormal liver function tests (10%) Coagulation disorders (9%) Ascites (4%)
Feeding intolerance	76	Vomiting (47%) Diarrhea (12%) Poor feeding (23%)
Failure to thrive	29	
Lethargy	16	
Seizures	1	
Sepsis	10	*Escherichia coli* (26 cases), *Klebsiella* (3 cases), *Enterobacter* (2 cases), *Staphylococcus* (1 case), Beta *Streptococcus* (1 case), *Streptococcus faecalis* (1 case)

Reprinted from Waggoner *et al.* (1990). Long-term prognosis in galactosemia: Results of a survey of 350 cases. *J. Inherit. Metab. Dis.* **13**, p. 804. With kind permission of Kluwer Academic Publishers.

should be evaluated for possible sepsis and, until culture results are assessed, should receive broad-spectrum antibiotic coverage.

Because of widespread newborn screening for galactosemia, many newborns with mild variant forms of galactosemia are now being detected (see below). The most common of these is Duarte-classic galactosemia (DG) heterozygosity. In the vast majority of cases, there are no serious signs or symptoms of illness. Liver dysfunction and kidney dysfunction are not seen in affected newborns. Some infants with DG galactosemia have a mild increase in RBC galactose-1-PO_4 levels and GALT activity is approximately 20% of normal. There may be an increased risk for presenile cataracts in adulthood.

MOLECULAR GENETICS

All forms of galactosemia are inherited in an autosomal recessive fashion. Thus, parents who have had one affected child are both obligate carriers (heterozygotes) and have a 25% risk of having an affected child in each subsequent pregnancy. Approximately 1 in 120 individuals is a heterozygote for classic galactosemia. Over 170 different mutations have been identified in the GALT gene, which is located on chromosome 9p13. The most common mutation in classic galactosemia is the Q188R point mutation, found primarily in northern European populations and in patients of European descent. This mutation accounts for approximately 55% of mutant alleles in the Caucasian population. Another common mutation is the K285N point mutation, originating in eastern and central Europe. In some parts of Europe, this mutation is found on 25–40% of mutant alleles, but in North America, it accounts for approximately 5% of mutant alleles. Both of these mutations result in essentially total loss of GALT enzyme activity. The S135L mutation is almost always found in patients of African descent. In African Americans, it accounts for approximately 50% of the mutant alleles. Although the red blood cell GALT activity associated with the S135L mutation is very low, blood levels of galactose metabolites are lower and are easier to control than with other classic galactosemia mutations, reflecting higher GALT activity in other cells, with up to 5% of normal activity in the leukocytes and 10% activity in other tissues. Other common classic galactosemia alleles, also missense mutations, include L195P, Y209C, and F171S. Approximately 70% of affected individuals have two of these six common alleles, whereas 10% have one common allele and one "private" or rare allele, and approximately 20% have two rare alleles.

The Duarte allele (N314D point mutation) is very common and is present in 6–13% of the general population. Thus, approximately 1 in 3000 individuals is a compound heterozygote for the Duarte allele and a classic galactosemia allele. Compound heterozygotes have GALT activity between 5 and 20% of control values. Other mutations are sometimes detected via newborn screening, such as the Los Angeles allele, which is a gain-of-function mutation and results in increased GALT activity to approximately 140% of normal. Individuals who are compound heterozygotes for the classic galactosemia allele and the Los Angeles allele (LA/G) may be detected by newborn screening because the overall GALT enzyme activity is reduced. However, this genotype does not cause clinical disease.

NEWBORN SCREENING

Most states in the United States (48 of 50) and many developed countries of the world perform newborn screening for galactosemia. Although in many cases a newborn with classic galactosemia is beginning to show signs of illness before the results of the newborn screen are available, the goal of screening is early detection and prevention of both acute and long-term complications. Newborn screening methods vary among states, but usually involve fluorescent GALT enzyme assay and/or microbiologic or fluorescence measurement of galactose metabolites on red blood cells obtained from a filter paper blood spot. States that screen only by using GALT enzyme activity measurement would not detect cases of GALK and GALE deficiency, whereas those disorders would be detected by elevated galactose metabolites if both tests are performed. The enzyme activity on the screen can be falsely low if the filter paper is exposed to heat, a problem often encountered during transport of the filter papers to a centralized state laboratory in the summer months. The galactose metabolite measurements could be falsely low in an affected infant who has not yet consumed adequate lactose-containing milk at the time the blood sample is obtained, because the sample is collected too soon after birth or because of illness, poor feeding, or use of a soy-based or elemental formula. Therefore, it is critical that the laboratory be notified of the infant's feeding status to ensure accurate screening results.

Infants who are compound heterozygotes for the Duarte allele and a classic galactosemia allele (DG genotype) are often detected by newborn screening due to low GALT activity and, sometimes, mildly elevated galactose metabolites. GALT activity in these cases is approximately 20% of normal. Other GALT variants,

such as the Los Angeles allele, may be detected as a result of newborn screening.

TREATMENT

Dietary therapy for galactosemia has been used for nearly 50 years and involves removal of lactose- and galactose-containing foods and beverages from the diet. This means that breast-feeding or feeding with a cow's milk-based infant formula must be discontinued as soon as the diagnosis is made. Soy formula is used in infancy. As the baby gets older, galactose-free foods or foods that are extremely low in galactose, such as cereals, fruits, vegetables, and meats, are introduced. Some fruits and vegetables that contain galactose must be avoided and all milk and milk products (cheese, ice cream, etc.) must be removed from the diet. Parents must learn to read food labels and are provided with instructional materials about foods and medications that contain lactose. Treatment centers for children with galactosemia employ dieticians trained in metabolic disease treatment to assist patients and families and to monitor their progress. The diet must be maintained for the lifetime of the individual. Because of the lack of milk products in a galactose-free diet, calcium supplements must be provided to older individuals who are no longer consuming infant formula.

Long-term monitoring of compliance with dietary therapy is required for galactosemia patients. Growth and development are followed closely. After initiation of treatment, the galactose-1-PO_4 level will decrease, usually over a period of weeks to months. Regular clinic visits and galactose-1-PO_4 measurements on red blood cells can determine whether the patient is adhering to the prescribed diet. Urinary galactitol measurement is another way to monitor treatment. Patients with classic galactosemia are usually unable to lower their galactose-1-PO_4 levels to the normal range, despite good compliance with the diet, presumably because of the endogenous conversion of glucose to galactose via GALE, which has bidirectional activity. Galactose-1-PO_4 levels less than 3.5 mg/dl are considered adequate for galactosemics on diet, whereas the normal level is less than 0.5 mg/dl.

Although long-term complications are thought to be unlikely in DG galactosemia, many infants are placed on a galactose-restricted diet for some period of time after birth, ranging from 6 months to 2 years, depending on the treatment center. Some centers do not treat DG galactosemia. There have been no well-controlled studies to determine whether there are significant developmental or other long-term health risks in patients with DG galactosemia.

LONG-TERM COMPLICATIONS

Ovarian Dysfunction

Among women with galactosemia, there is a high risk for ovarian dysfunction and infertility. Hypergonadotropic hypogonadism occurs in 75% of affected females, manifested as primary amenorrhea and lack of secondary sexual development, delayed or abnormal puberty, or secondary amenorrhea due to premature ovarian failure in the 20s or 30s. Streak ovaries have been described pathologically. Risk factors for the development of ovarian dysfunction include homozygosity for the Q188R mutation, poor control of galactose-1-PO_4 levels (>3.5 mg/dl), and little ability to metabolize galactose (<5% galactose oxidation). These problems occur despite early diagnosis and treatment. The height of the galactose level at birth does not correlate with the degree of ovarian dysfunction, but there is some evidence that tight control of galactose-1-PO_4 levels throughout childhood may decrease the risk for ovarian failure. The cause of the ovarian dysfunction is poorly understood. It may be related to the effects of potentially toxic metabolites of galactose, such as galactitol, on the oocytes. Some women with galactosemia have been able to become pregnant and some have had children via ovum donation and *in vitro* fertilization techniques. Preteen girls with galactosemia should be evaluated by an endocrinologist and supplemental hormone therapy should be provided as needed to ensure pubertal development. Males with galactosemia do not have testicular dysfunction and have normal pubertal development and fertility.

Speech and Language Disorders

Despite good dietary control of the disorder, many children with classic galactosemia will exhibit speech and language problems, specifically verbal dyspraxia. Sixty-five percent of patients have been found to have these deficits. Most problems are in expressive language skills, speech articulation, word retrieval, and short-term memory. Receptive language skills are normal. Sibling pairs are similarly affected, even if the second child never received lactose or became ill as a newborn, suggesting a prenatal effect during brain morphogenesis.

Neurological Abnormalities

Some individuals with galactosemia may have later onset neurological abnormalities, again despite good dietary control. Patients have developed mild to severe

ataxia and a coarse intention tremor, beginning at 9–14 years of age. One study has shown a decline in IQ with age. The cause of these problems is not well understood. It has been postulated that there may be a deficiency of galactose-containing glycolipids or glycoproteins in the brain.

In summary, long-term follow-up of patients with galactosemia indicates that, although the outcomes have greatly improved with early diagnosis and treatment, certain health problems occur despite compliance with therapy. The cause of these problems is unclear, but may be related to prenatal elevation of galactose, galactose-1-PO_4, and galactitol, chronic postnatal intoxication by galactose metabolites as well as endogenous production of galactose, deficiency of UDP-galactose, or deficiency of galactose-containing glycolipids or glycoproteins in the brain.

OTHER FORMS OF GALACTOSEMIA: GALACTOKINASE AND UDP GALACTOSE 4-EPIMERASE DEFICIENCIES

GALK deficiency results in the development of cataracts. It does not cause systemic illness as seen in GALT deficiency, although pseudotumor cerebri has been reported occasionally as a complication in untreated individuals. This enzyme defect can be detected by newborn screening in those states that measure RBC galactose and/or galactose-1-PO_4 levels, which are elevated in affected individuals. It is treated with dietary restriction of galactose. The cataracts resolve with treatment.

GALE deficiency can also be detected through newborn screening, again if the state laboratory measures galactose metabolites. In North America, the most common form of GALE deficiency is limited to the red blood cells only and is most often seen in African Americans. Although RBC galactose-1-PO_4 levels are elevated and RBC GALE activity is absent or very low, GALE activity is 60–70% of normal in other tissues, including liver and fibroblasts. No treatment is required and patients are clinically well with normal growth and development. There have been rare patients reported in inbred populations with generalized GALE deficiency and, in those cases, the disorder presents in a manner similar to classic GALT deficiency.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Cow Milk Protein Allergy • Glycogen Storage Disease • Hereditary Fructose Intolerance • Malabsorption • Tyrosinemia

Further Reading

Elsas, L. J., and Lai, K. (1998). The molecular biology of galactosemia. *Genet. Med.* **1**, 40–48.

Gene Clinics Web Site. Available at http://www.geneclinics.org/servlet/access?db=geneclinics&site=gt&id=8888891&key=i9ku6oVbzxtiZ&gry=&fcn=y&fw=I-tG&filename=/profiles/galactosemia/index.html.

Guerrero, N. V., Singh, R. H., Manatunga, A., Berry, G. T., Steiner, R. D., and Elsas, L. J. (2000). Risk factors for premature ovarian failure in females with galactosemia. *J. Pediatr.* **137**, 833–841.

Holton, J. B., Walter, J. H., and Tyfield, L. A. (2001). Galactosemia. *In* "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th Ed., Vol. 1, pp. 1553–1587. McGraw Hill, New York.

Nyhan, W. L., and Ozand, P. T. (1998). Galactosemia. *In* "Atlas of Metabolic Diseases," pp. 322–329. Chapman & Hall Medical, London.

Parents of Galactosemic Children Web Site. Available at http://www.galactosemia.org and http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?230400.

Petry, K. G., and Reichardt, J. K. V. (1998). The fundamental importance of human galactose metabolism: Lessons from genetics and biochemistry. *Trends Genet.* **14**, 98–102.

Schweizer, S., Shin, Y., Jakobs, C., and Brodehl, J. (1993). Long-term outcome in 134 patients with galactosemia. *Eur. J. Pediatr.* **152**, 36–43.

Tyfield, L. A. (2000). Galactosemia and allelic variation at the galactose-1-phosphate uridyltransferase gene: A complex relationship between genotype and phenotype. *Eur. J. Pediatr.* **159**, S204–S207.

Waggoner, D. D., Buist, N. R., and Donnell, G. N. (1990). Long-term prognosis in galactosaemia: Results of a survey of 350 cases. *J. Inherit. Metab. Dis.* **13**, 802–818.

Webb, A. L., Singh, R. H., Kennedy, M. J., and Elsas, L. J. (2003). Verbal dyspraxia and galactosemia. *Pediatr. Res.* **53**, 396–402.

Gallbladder Cancer

DAVID J. POMPA AND JOHN M. CARETHERS
University of California, San Diego

anomalous pancreaticobiliary ductal junction An anatomical variant of the juncture of the common bile duct and pancreatic duct found in association with gallbladder carcinoma.
carcinogenic Of or pertaining to the ability to cause the development of cancer.
choledochal cyst An often saccular dilation or other abnormal widening of a portion of the biliary tree.
cholelithiasis The presence of gallstones in the gallbladder.
common bile duct The duct formed by the juncture of the cystic duct and common hepatic duct, permitting the flow of bile from the liver and gallbladder to the small intestine.
mutagenic Of or pertaining to any chemical or physical environmental agent capable of inducing genetic mutations or increasing the mutation rate.
pancreatic duct (duct of Wirsung) The primary secretory channel of the pancreas.

Gallbladder carcinoma is an uncommon cancer with two distinct risk factors: cholesterol gallstones and anomalous pancreaticobiliary ductal junction. Few symptoms herald its presence, as late clinical presentations are the rule. Although surgery offers the opportunity for cure, other treatment options are limited in efficacy.

EPIDEMIOLOGY

Gallbladder carcinoma (GBCa) is an infrequent cancer, with an incidence ranging from 2.5 cases/100,000 population in the United States to 7.5 cases/100,000 population in parts of South America. A genetic predisposition to GBCa is undoubtedly important as women are affected three times as often as men, whereas people of Native American, Mexican, Japanese, Chilean, and Bolivian descent have the highest incidence worldwide. Most patients with this disease are age 65 or older.

ETIOLOGY/RISK FACTORS

Several disease processes are strongly associated with the development of GBCa. Cholesterol gallstones are the most commonly associated condition. Present in 70–90% of GBCa cases, they yield a threefold increase in risk for GBCa. The presence of cholelithiasis for a duration >40 years seems to increase the risk of GBCa as well. In its most chronic form, cholelithiasis leads to a "porcelain" gallbladder with partial or complete calcification of the gallbladder wall and imparts a significant risk of cancer.

Anomalous pancreaticobiliary ductal junction (APBDJ) occurs when the pancreatic duct and common bile duct join abnormally outside of the duodenal wall, yielding a longer common channel. Often associated with choledochal cysts, it is estimated that 11–18% of GBCa cases have an associated APBDJ. It is more common for GBCa to occur at younger ages in patients with this anomaly.

Gallbladder polyps are known to harbor cancer and represent another risk for GBCa. Gallbladder polyps in excess of 10 mm require that the gallbladder be removed surgically as the prevalence of GBCa increases significantly beyond this size (Table I).

Other etiologic factors with debatable roles in the development of GBCa include chronic *Salmonella typhi* carriage, various environmentally toxic compounds used in the automobile, textile, rubber, and metal industries, and previous gastrointestinal surgery.

PATHOGENESIS

The pathogenesis of GBCa depends on the etiologic factor implicated in its genesis. Cholelithiasis may lead to a vicious cycle of chronic gallbladder mucosal irritation with perpetual epithelial repair, resulting in carcinogenic mutations. Bile contains a multiplicity of

TABLE I Prevalence of Gallbladder Carcinoma According to Polyp Size

Polyp size	Prevalence of cancer
<10 mm	≤ 5%
10–15 mm	11–23%
>15 mm	50–70%

 Copyright 2004, Elsevier (USA). All rights reserved.

compounds that are potentially mutagenic, but the exact constituent(s) that may lead to cancer is yet to be defined.

Anomalous pancreaticobiliary ductal junction promotes free reflux of pancreatic juice into the biliary tree and, by altering the gallbladder milieu, promotes inflammation and proliferative activity of the gallbladder epithelium. This may facilitate mutagenesis at the cellular and nuclear levels. It is well known that up-regulation of the K-*ras* oncogene and mutations in the *p53* tumor suppressor gene exist more frequently in patients with APBDJ.

CLINICAL CONSIDERATIONS

Gallbladder carcinoma is a slowly progressive tumor, taking at least 15 years to develop. At presentation, symptoms are very nonspecific and their presence usually heralds a late presentation of this cancer. Nausea, vomiting, anorexia, and weight loss are typical. A fortunate patient will have the tumor located in the neck of the gallbladder intermittently occluding the cystic duct, causing symptoms mimicking cholecystitis. This may lead to an earlier diagnosis and increased chance for curative resection. Jaundice yields a particularly poor prognosis in that 85% of these tumors are surgically unresectable due to extensive local invasion into the adjacent hepatic parenchyma.

DIAGNOSIS

Most early-stage cancers are diagnosed incidentally at the time of cholecystectomy for unrelated reasons. If constitutional symptoms predominate, workup usually includes abdominal ultrasonography. Findings at ultrasound suggestive of GBCa include a thick gallbladder wall and/or lumenal mass, gallbladder calcifications, or obvious hepatic infiltration. Abdominal computed tomography defines more clearly the extent of disease, whereas ERCP with brushings may ultimately be diagnostic. Endoscopic ultrasound's role in staging these tumors is evolving.

STAGING/PROGNOSIS

The staging of GBCa follows that set forth by the American Joint Commission on Cancer and groups cancers into stages 0 through IV, with stage IV being the most advanced. For the less advanced stages 0 through II, prognosis is encouraging, as curative resection is possible. Stage III and IV cancers, however, have 1-year survivals of 25 and 10%, respectively, which reflect significant lymph node involvement and metastatic disease.

TREATMENT

Surgical resection of GBCa is the only potentially curative treatment modality. Chemotherapy using 5-FU (5-fluorouracil) based regimens has been explored, but the small number of patients treated limits interpretability. Combination therapy with surgery and either neoadjuvant or adjuvant chemotherapy may impart a trend toward improved survival in a heterogeneous group of case series and reports. Most centers treating GBCa would offer chemotherapy as an option. Radiation therapy has yet to be shown as a significant adjunct in the treatment of GBCa.

See Also the Following Articles

Cancer, Overview • Cholangiocarcinoma • Cholecystectomy • Cholelithiasis, Complications of • Gallstones, Pathophysiology of • Porcelain Gallbladder

Further Reading

Donohue, J. H., Stewart, A. K., and Menck, H. R. (1998). The national cancer data base report on carcinoma of the gallbladder, 1989–1995. *Cancer* **83**, 2618–2628.

Sheth, S., Bedford, A., and Chopra, S. (2000). Primary gallbladder cancer: Recognition of risk factors and the role of prophylactic cholecystectomy. *Am. J. Gastroenterol.* **95**, 1402–1410.

Strasberg, S. M., and Drebin, J. A. (1999). Tumors of the biliary tree. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., pp. 2235–2340. Lippincott Williams & Wilkins, Philadelphia, PA.

Gallbladder, Pediatric

M. JAMES LOPEZ
University of Michigan Medical Center, Ann Arbor

acalculous Lacking a gallstone.
adenoma (adenomatous) Benign epithelial tumor in which there are recognizable glands or in which cells are clearly derived from glandular epithelium.
agenesis Absence of an organ.
calculous Abnormal concretion, usually containing mineral salts; here, specifically a gallstone.
cholecystitis Inflammation of the gallbladder.
choledocholithiasis Stone in the common bile duct.
cholelithiasis Presence or formation of gallstones.
cystic duct Gallbladder duct that connects to the common bile duct.
empyema Accumulation of pus in a cavity of the body; can refer to accumulation external to or in the gallbladder.
functional Affecting function but not structure; in some cases, refers to absence of a specific disease entity although there is a symptom.
neoplasm Abnormal tissue growth that is uncontrolled and progressive but can be either benign or malignant.
peritonitis Inflammation of the peritoneum; in the case of cholecystitis, secondary to perforation, and extension of infection to involve the peritoneum and abdominal cavity.
Peutz–Jeghers Hereditary syndrome characterized primarily by gastrointestinal polyps (usually hamartomas).
polyp Protruding growth from a mucous membrane.

Gallbladder disease in children runs the gamut, from gallstones to neoplasms. In children, all gallbladder disorders appear to occur much less commonly than they do in adults. However, any estimate of prevalence or incidence is very difficult in children. In certain populations, there may be an increased incidence of diseases such as gallstones, but prospective data for these disorders are almost uniformly lacking.

ACALCULOUS DISORDERS OF THE GALLBLADDER

Acalculous Cholecystitis

Acalculous cholecystitis results from acute inflammation of the gallbladder. Reports of children with cholecystitis suggest that up to half may have acalculous cholecystitis. This inflammatory condition is generally categorized as either primary or secondary.

Primary inflammation has no identifiable cause, whereas secondary inflammation is seen in association with a variety of conditions, some similar to those associated with hydrops of the gallbladder. Conditions reported in association with acute inflammation of the gallbladder include septic infections with a variety of bacterial organisms, bacterial enteric infections, and streptococcal infections; pneumonias; viral infections such as gastroenteritis; opportunistic infections (e.g., cryptosporidium, cytomegalovirus, and fungal organisms) in immunocompromised patients, in patients following surgery, in severe illnesses such as burns, and in chronic inflammatory conditions such as Crohn's disease; congenital anomalies of the biliary tree causing cystic duct obstruction; and infections such as leptospirosis, Rocky Mountain spotted fever, and typhoid fever.

Patients usually present with biliary complaints, which include right upper quadrant pain, nausea, vomiting, and fever. On examination, they can have a mass, although this is unusual, and can have either localized right upper quadrant discomfort or generalized abdominal pain. Findings can be obscured by the abnormalities of the underlying associated disease process or the difficulty of examining some patients, e.g., patients with severe burns.

Recognition of the condition is important because of the complications of perforation, gangrene, and empyema that can result when the condition remains unrecognized. Because the symptoms can be consistent with a variety of other entities, such as peritonitis, intussusception, hepatitis, appendicitis, and choledochal cyst with cholangitis, distinguishing on clinical grounds alone can be difficult. Ultrasonography can show thickening of the gallbladder wall, absence of gallstones, edema around the gallbladder, and distension, but provides only supportive evidence of the clinical suspicion. Failure to fill the gallbladder on cholescintigraphy in the presence of good hepatic uptake is consistent with cholecystitis. However, a strong clinical suspicion is often the best clue to

 Copyright 2004, Elsevier (USA). All rights reserved.

the diagnosis. In most cases, cholecystectomy is recommended to avoid complications in a patient who is often already ill.

Hydrops of the Gallbladder

Acute hydrops is defined as gallbladder distension in the absence of evidence of obstruction (e.g., gallstones), inflammation, or congenital anomalies. The precise cause is unknown, but hydrops is seen only in association with other conditions, including Kawasaki syndrome, mesenteric adenitis, viral hepatitis, streptococcal or staphylococcal infection, Henoch–Schönlein purpura, hypokalemia, Sjogren syndrome, nephrotic syndrome, sepsis, parenteral nutrition, fasting, α1-antitrypsin deficiency, leukemia, familial Mediterranean fever, and systemic sclerosis. The unifying factor in all of these illness is unclear, but may relate to inflammation of the cystic duct secondary to a systemic illness, adenopathy compressing on the cystic duct leading to extrinsic obstruction, or possibly circulating factors (e.g., cytokines) that affect function of the gallbladder.

In general, patients with this disorder present with acute, crampy abdominal pain, nausea, and vomiting. Because the symptoms are nonspecific and not specifically biliary, a careful examination, looking for upper abdominal tenderness or mass in the right upper quadrant, is essential. Generally, there is neither fever nor jaundice. Ultrasonography is the most useful diagnostic test for verifying the diagnosis and will reveal a markedly distended gallbladder.

Once the diagnosis has been established, treatment is symptomatic and conservative, with spontaneous resolution occurring within days to weeks in the vast majority of patients. There is no evidence in these patients of chronic biliary dysfunction following resolution.

Neoplasms

Neoplasms of the gallbladder are extremely uncommon in childhood. Benign tumors (adenomatous polyps) have been described in patients with Peutz–Jeghers syndrome. Isolated adenomatous polyps have also been described in the absence of systemic illness. In each case, surgical removal is recommended because there is at least theoretical risk of progression to malignancy. Primary malignancies of the gallbladder are exceedingly rare in children. However, embryonal rhabdomyosarcoma has been described in the gallbladder and biliary tract in childhood. Thankfully, these are rarely found, because prognosis is extremely poor.

Congenital Abnormalities

Isolated congenital anomalies of the gallbladder are uncommon. Both gallbladder agenesis and torsion of the gallbladder secondary to a congenitally long mesentery have been described. This latter problem presents with clinical features essentially indistinguishable from cholecystitis (see later). Cholecystectomy is the treatment of choice.

Functional Disorders

Biliary dyskinesia is thought to be a motor dysfunction of the gallbladder and the sphincter of Oddi. Patient complaints are generally described as a recurrent, continuous biliary type of pain. Decreased gallbladder emptying on a quantitative cholescintigraphy is used as the standard for assessing this dysfunction in adults. However, there are no data to support the presence of such dysfunction in children or what would be normal values for such a functional study in children. At present, neither recommendations for identification nor treatment of such dysfunction can be made. It is more likely that many children thought to have this disorder have functional recurrent abdominal pain.

CHOLELITHIASIS

Gallstones

Gallstones in children can form *in utero* and be diagnosed in the fetus with prenatal ultrasounds or can be found with abdominal ultrasonography in otherwise asymptomatic infants or children. It is not clear what factors contribute to the formation of gallstones or what the incidence or prevalence is in children. In younger children, genetic predisposition is a likely important contributor to their formation.

Recent work by Paigen and colleagues in mouse models has suggested that genes that may contribute to the formation of these gallstones fall into several general categories. The processes suggested to be involved include cholesterol secretory rate, increased cholesterol delivery to the liver, up-regulation of cholesterol synthesis, development of type 2 diabetes and obesity, biliary hypersecretion of cholesterol, inhibition of bile salt synthesis, and coupling of biliary cholesterol to other biliary lipids. Further work in both mouse models and high-risk human populations should help elucidate how various genetic mutations or polymorphisms influence the process of gallstone formation (cholesterol stones).

Such genetic predisposition appears to underlie the increased incidence or prevalence observed in certain human populations. Some populations appear to be at much higher risk than others are, including some Native Americans (Amerindians), Swedes, and Czechs, whereas very low incidences are reported in Canadian Eskimos and in people in certain areas of Africa. In the few studies available in (Italian and Swedish) children, there is an increased incidence and prevalence in girls, with a rapid divergence of the incidence curves for boys and girls after 11 years of age. From these studies, Italian children between 6 and 19 years of age had an overall prevalence of 0.13%, with girls having a prevalence of 0.27%. In Swedish children, gallstones were very infrequent prior to 11 years of age. In studies of Amerindians, girls had an overall prevalence of around 10% during their teenage years, but this rose to 70% by their fourth decade. Amerindian men are also at increased risk, with reported prevalences of less than 2% during their teen years but rising to 25% during their forties and increasing to 70% by age 60. Variations noted in various human populations suggest that genetic differences may underlie the noted differences in prevalence. The fact that in the United States there can be vastly different incidences on diets that are roughly equivalent (Amerindians $\gg$ Caucasians $\gg$ African Americans) again points to the influence of genetic susceptibilities as important contributors to the burden of gallstone disease. Further support for genetic susceptibility to gallstone formation is found in the differences among Hispanics in the United States. Mexican Americans are at a much higher risk for gallstone formation and gallbladder cancer when compared to Puerto Ricans and Cubans. Such differences may relate to the higher prevalence of Amerindian genes in the Mexican American population. However, even among North American Amerindians, there is considerable variation in the prevalence from population to population. In individuals with 100% Amerindian heritage, those in the Dakotas had a prevalence of around 45%, those in Oklahoma had a prevalence of around 73%, and those in Arizona had a prevalence of approximately 93%. Prevalences observed were similar in both males and females. Thus, the contribution of genetic background can be inferred from such data, but remains to be better established on an experimental basis before specific recommendations can be made for screening or treatment based on genetic susceptibility.

In children, most gallstones are either pigment stones or cholesterol stones. Pigment stones are categorized as brown or black. Black pigment stones are primarily made of calcium carbonate and phosphate, whereas brown stones have large amounts of calcium soaps (fatty acid salts). In both, the underlying framework consists of mucin glycoproteins secreted by the biliary epithelium. Brown pigment stones form in the presence of both stasis and infection. Black pigment stones form most commonly in the presence of chronic hemolytic disease (e.g., hereditary spherocytosis, sickle cell anemia, or glucose-6-phosphate dehydrogenase deficiency), during fasting while on total parenteral nutrition, and because of cirrhosis or chronic cholestasis [including conditions such as Wilson's disease and progressive familial intrahepatic cholestasis type 1 (Byler's disease)]. The use of certain drugs, such as ceftriaxone, causes development of sludge (called pseudolithiasis) and may also contribute to stone formation.

There are some differences in etiology at different ages. Gallstones in infants are seen in association with total parenteral nutrition, sepsis, chronic lung disease, hemolytic anemia, malabsorption, abdominal surgery, hepatobiliary problems, and necrotizing enterocolitis. However, in many of these disorders, the children will be receiving intravenous nutrition or will be very limited in their oral intake, which may in and of itself be a major contributor to stone formation. In older children and adolescents, there is an increased association with both obesity and pregnancy.

Although it is estimated that up to 80% of gallstones in adults are silent, they can cause significant clinical symptoms. The most consistent symptoms relate to obstruction of the cystic or biliary duct. It is not known how many children without symptoms actually have gallstones, nor if those who do have "silent" gallstones ever develop symptoms during the course of their life. Those who develop symptomatic cholelithiasis have been described at all ages in childhood. The frequency of symptomatic patients at any age appears to be related to the risk for developing gallstones. For all symptomatic cases of cholelithiasis, cholecystectomy is the treatment of choice. With asymptomatic gallstones, most clinicians would recommend no treatment unless symptoms arise.

Complications of Cholelithiasis

Both chronic and acute cholecystitis can result from cholelithiasis. Gallstone obstruction of the cystic duct of the gallbladder appears to be the cause of acute cholecystitis. The symptoms that characterize acute cholecystitis are similar to those of symptomatic cholelithiasis without cholecystitis. They include "biliary colic," which is severe persistent abdominal pain that may localize to the right upper quadrant, with anorexia and vomiting. High fever is unusual. Physical examination may reveal right upper quadrant tenderness and a Murphy's sign

(inspiratory arrest secondary to pain in this area). In adults, a palpably enlarged gallbladder is found in up to a third of patients.

Laboratory abnormalities can include hyperbilirubinemia and elevated liver enzymes (alkaline phosphatase and aminotransferases). Leukocytosis is commonly observed. If there is choledocholithiasis, then amlyase and lipase may also be elevated. A suspected diagnosis of acute cholecystitis is best confirmed by ultrasonography, with the visualization of gallstones in the gallbladder. Failure to visualize the gallbladder following cholescintigraphy is also consistent with the diagnosis of cholecystitis. Management is initially conservative, with supportive care as needed and antibiotics. Cholecystectomy is the treatment of choice. All patients with acute cholecystitis should be carefully monitored for signs of potential complications, which include perforation and empyema.

Chronic cholecystitis undoubtedly occurs in children but is poorly characterized as an entity. This may in part be related to the standard of surgical removal of the gallbladder following an episode of acute cholecystitis in children. Because many patients with chronic cholecystitis have an antecedent history of an episode of acute cholecystitis, children may be less likely to present with this particular entity. Those who do have the disorder can have essentially asymptomatic disease or symptoms consistent with biliary colic. Treatment is a cholecystectomy.

Choledocholithiasis is not thought to be a common problem in children but is usually symptomatic, in marked contrast to studies in adults, who can be symptom free in this condition. In children, choledocholithiasis can cause pancreatitis, which can be the presenting problem for these children. Almost all children with duct obstruction secondary to gallstones have biliary colic and severe abdominal pain. If they have cholangitis as well, they will have fever and tenderness on exam. Most commonly, patients with obstruction will have a leukocytosis and elevated liver enzymes. An elevated conjugated bilirubin is also consistent with the diagnosis.

Initial diagnostic studies include ultrasonography. Dilatation of the common bile duct suggests an impacted stone or a congenital anomaly of the bile duct. Because biliary stones, especially those near the ampulla, can be missed on ultrasonography, the absence of a visualized stone does not exclude the diagnosis. Magnetic resonance cholangiopancreatography or computed tomography are both useful tools in the visualization of the common bile duct and pancreatic ducts and can detect stones in both the common bile duct and in some cases stones near the ampulla. In most cases, endoscopic retrograde cholangiopancreatography is necessary both for diagnosis and for treatment. Following treatment, if there are still gallstones in the gallbladder, a cholecystectomy is recommended.

See Also the Following Articles

Cholecystectomy • Cholelithiasis, Complications of • Gallstones, Pathophysiology of • Ultrasonography

Further Reading

Carey, M. C., and Paigen, B. (2002). Epidemiology of the American Indians' burden and its likely genetic origins. *Hepatology* **36**(4, Pt. 1), 781–791.

Friedman, G. D., Kannel, W. B., *et al.* (1966). The epidemiology of gallbladder disease: Observations in the Framingham Study. *J. Chronic. Dis.* **19**(3), 273–292.

Friesen, C. A., and Roberts, C. C. (1989). Cholelithiasis. Clinical characteristics in children. Case analysis and literature review. *Clin. Pediatr. (Phila.)* **28**(7), 294–298.

Heubi, J. E., Lewis, L. G., *et al.* (2001). Diseases of the gallbladder in infancy, childhood, and adolescence. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), pp. 343–362. Lippincott Williams & Wilkins, Philadelphia.

Holcomb, G. W., Jr., and Holcomb, G. W., 3rd (1990). Cholelithiasis in infants, children, and adolescents. *Pediatr. Rev.* **11**(9), 268–274.

Jorgensen, T. (1987). Prevalence of gallstones in a Danish population. *Am. J. Epidemiol.* **126**(5), 912–921.

Lammert, F., Carey, M. C., *et al.* (2001). Chromosomal organization of candidate genes involved in cholesterol gallstone formation: a murine gallstone map. *Gastroenterology* **120**(1), 221–238.

Paigen, B., and Carey, M. C. (2002). Gallstones. *In* "The Genetic Basis of Common Diseases" (R. A. King, J. I. Rotter, and A. G. Motulsky, eds.), pp. 298–335. Oxford University Press, Oxford and New York.

Palasciano, G., Portincasa, P., *et al.* (1989). Gallstone prevalence and gallbladder volume in children and adolescents: An epidemiological ultrasonographic survey and relationship to body mass index. *Am. J. Gastroenterol.* **84**(11), 1378–1382.

Reif, S., Sloven, D. G., *et al.* (1991). Gallstones in children. Characterization by age, etiology, and outcome. *Am. J. Dis. Child.* **145**(1), 105–108.

Sears, H. F., Golden, G. T., *et al.* (1973). Cholecystitis in childhood and adolescence. *Arch. Surg.* **106**(5), 651–653.

Shaffer, E. A. (2000). Gallbladder disease. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (W. A. Walker, P. R. Durie, J. R. Hamilton, J. A. Walker-Smith, and J. B. Watkins, eds.), pp. 1291–1311. Decker, Ontario, B. C.

Suma, V., Marini, A., *et al.* (1998). Fetal gallstones: Sonographic and clinical observations. *Ultrasound Obstet. Gynecol.* **12**(6), 439–441.

Trotman, B. W., and Soloway, R. D. (1975). Pigment vs cholesterol cholelithiasis: Clinical and epidemiological aspects. *Am. J. Dig. Dis.* **20**(8), 735–740.

Gallstones, Pathophysiology of

CYNTHIA W. KO AND SUM P. LEE
University of Washington, Seattle

antinucleating proteins Biliary proteins that can inhibit nucleation of cholesterol crystals *in vitro*. Examples include apolipoprotein A-I, apolipoprotein A-II, and immunoglobulin A.

biliary sludge Microscopic precipitates in bile that can be visualized with ultrasonography or bile microscopy. Can produce symptoms and complications similar to those of gallstones. Also known as microlithiasis or pseudolithiasis.

cholesterol saturation index In a given bile sample, the ratio of the actual amount of cholesterol to the maximum cholesterol-carrying capacity of that sample, determined *in vitro*. Bile with a cholesterol saturation index greater than 1 is considered supersaturated.

cholesterol supersaturation State in which the amount of cholesterol in bile exceeds the cholesterol-carrying capacity of the biliary lipids. A prerequisite for cholesterol gallstone formation.

nucleation Initial precipitation of solid cholesterol crystals from supersaturated bile. An initial step in gallstone formation.

pronucleating proteins Biliary proteins that can promote nucleation of cholesterol crystals. Examples include mucin, anionic peptide fraction, and phospholipase C.

Gallstones are extremely common in Western societies. Over 20 million Americans are believed to have gallstones, and approximately 700,000 cholecystectomies are performed every year. Symptoms and complications related to gallstones are among the most costly gastroenterologic disorders, at an estimated annual cost of almost $6.5 billion. In Western societies, gallstones are composed primarily of cholesterol. The formation of cholesterol gallstones requires three fundamental pathophysiological factors: hepatic secretion of cholesterol-supersaturated bile, nucleation of biliary cholesterol to form crystals, and gallbladder stasis.

SYNTHESIS OF BILIARY LIPIDS

Bile is a complex substance composed of lipids, proteins, electrolytes, and water. There are three principal biliary lipids: cholesterol, bile acids, and phospholipids (primarily phosphatidylcholine). Many species of proteins are present in bile, and they are derived from serum proteins as well as from hepatocyte, bile duct, and gallbladder epithelial secretion. Gallstone formation is determined by the physical–chemical interactions of biliary lipids, as well as interactions with biliary proteins.

Cholesterol

The total body pool of cholesterol is derived from dietary absorption and *de novo* synthesis. The liver is the primary site of cholesterol synthesis in the body. Free cholesterol is synthesized in the hepatocyte endoplasmic reticulum via the rate-limiting enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase. Cholesterol derived from lipoproteins is also taken up by the liver. Cholesterol enters the liver in esterified form and is transported to the lysosomes, where it is converted into its free form. It is then transported to the endoplasmic reticulum and is reesterified for storage. Cholesterol esters in the endoplasmic reticulum are continually undergoing hydrolysis, providing a constant supply of free cholesterol. This free pool in the endoplasmic reticulum is the source of cholesterol for bile acid and lipoprotein synthesis and for secretion into bile. Approximately 80% of biliary cholesterol originates from an existing pool in the liver. The remaining 20% of biliary cholesterol is derived from new hepatic synthesis. Total hepatic cholesterol is tightly regulated within a narrow range.

Bile Acids

Bile acids are synthesized from cholesterol through the rate-limiting microsomal enzyme cholesterol 7α-hydroxylase. Bile acids, which are amphiphilic, can solubilize and transport hydrophobic cholesterol in aqueous solutions such as bile. Two primary bile acids, chenodeoxycholic acid (CDCA) and cholic acid (CA), are synthesized *in vivo*. The relative proportions of these are regulated by the activity of the enzyme 12α-hydroxylase, which converts CDCA to CA. The cholesterol-carrying capacity of bile is influenced by the relative proportions of these two bile acids, because less cholesterol is solubilized by CA than by CDCA. 12α-Hydroxylase can be induced by estrogens, resulting

 Copyright 2004, Elsevier (USA). All rights reserved.

in higher proportions of CA and more lithogenic bile. This effect of estrogen may explain some of the female propensity for gallstone formation.

After synthesis, carrier proteins bind bile acids to transport them through the hepatocyte. Bile acids are then conjugated with glycine and taurine, which increases their water solubility. They are then actively secreted by bile acid transporters located in the bile canalicular membrane of the hepatocyte. Bile acids are initially excreted as individual molecules. However, at a critical micellar concentration, they can aggregate to form micelles that are capable of solubilizing cholesterol and phospholipids. When bile acids enter the intestine, they are metabolized by intestinal bacteria, deconjugated, and dehydroxylated. In this process, CDCA is converted to lithocholic acid and CA is converted to deoxycholic acid. Bile acids are reabsorbed in the terminal ileum and circulate via the portal vein to the liver. In the liver, they are reabsorbed, reconjugated, and resecreted into bile via the enterohepatic circulation.

Phospholipids

It is believed that phospholipids in bile are derived from the hepatocyte canalicular membrane or from membranous structures within the hepatocyte, such as the endoplasmic reticulum or lysosomes. Phosphatidylcholines comprise over 95% of the phospholipids in bile. Phospholipids are amphiphilic molecules, similar to bile acids, and can solubilize and transport cholesterol. Phospholipids are believed to be translocated across the canalicular lipid bilayer by a protein similar to the multidrug resistance gene protein. Cholesterol and phospholipids then aggregate and are secreted from the outer leaflet of the canalicular membrane into bile as vesicles.

CHOLESTEROL SECRETION AND SOLUBILIZATION

The primary route for elimination of cholesterol from the body is biliary secretion. As a hydrophobic molecule, cholesterol is relatively insoluble in aqueous solutions such as bile. Thus, in bile, cholesterol must be solubilized by amphipathic biliary lipids such as bile salts and phospholipids. Free cholesterol is virtually insoluble in aqueous solution. Bile acids, because of their unique amphipathic properties, are able to solubilize cholesterol and phospholipids in mixed micelles.

The major driving force behind the biliary secretion of cholesterol and phospholipid is bile acid secretion. Bile acids are actively secreted against a gradient into the bile canaliculus, inducing subsequent secretion of cholesterol–phospholipid vesicles. Hydrophobic bile acids are more effective than other bile acids as stimulators of lipid secretion, although the maximum rate of biliary lipid secretion is similar for all bile acids. In the bile canaliculus, cholesterol can be solubilized with phospholipids (such as phosphatidylcholine) as unilamellar vesicles ranging in size from 40 to 100 nm. These vesicles consist of a bilayer of phospholipid interdigitated with cholesterol, without associated bile acids. These vesicles are believed to be derived from the outer membrane leaflet of the canalicular membrane. Vesicles comprise a separate and distinct carriage system that probably represents the major mode of cholesterol transport from the hepatocyte.

Within the bile canaliculus, the bile acid concentration gradually increases, eventually reaching a critical micellar concentration. At this concentration, cholesterol–phospholipid vesicles can be converted into mixed micellar carriers containing cholesterol, phospholipids, and bile salts. This dynamic interchange between vesicular and micellar forms of cholesterol carriage continues within the bile ductules, ducts, and gallbladder. Bile salts are more efficient at solubilizing phospholipids than cholesterol. As lipid carriers are converted from vesicles into micelles, the remaining vesicles become progressively enriched with cholesterol.

The proportion of cholesterol transported in the micellar form can vary dramatically, depending on bile acid concentrations. For example, at low bile acid secretion rates, as with fasting, cholesterol hepatic bile is carried primarily in vesicles. Conversely, in the gallbladder, the higher bile acid concentrations favor a shift of cholesterol into the micellar forms. Thus, the degree of bile cholesterol saturation varies depending on physiologic conditions and the relative concentrations of the various biliary lipids.

Phase equilibrium diagrams have been constructed to characterize cholesterol solubility at various concentrations of biliary lipids. Using these diagrams, the cholesterol saturation index (CSI) can be calculated. The CSI is the ratio of the actual amount of cholesterol in a given bile sample to the maximum cholesterol carrying capacity of that sample determined *in vitro*. Bile with a CSI greater than 1 is considered supersaturated. Supersaturated bile that does not form cholesterol crystals is called "metastable."

CHOLESTEROL SUPERSATURATION

The amount of cholesterol secreted by the liver can sometimes exceed the carrying capacity of the biliary lipids, resulting in bile that is supersaturated and

metastable. Various pathogenic conditions can disrupt the tight balance of biliary lipid secretion, resulting in absolute biliary cholesterol hypersecretion or relative bile acid hyposecretion.

Biliary cholesterol hypersecretion may occur by several different mechanisms. For example, nonobese gallstone patients may have higher intracellular levels of total and free cholesterol, suggesting enhanced intracellular cholesterol transport. Clinical situations associated with excessive secretion of biliary cholesterol include obesity, rapid weight loss, aging, drug effects, and hormonal therapy. In obesity, it is believed that activity of HMG-CoA reductase, the rate-limiting enzyme in cholesterol synthesis, is increased, leading to higher total body cholesterol. Paradoxically, with rapid weight loss, cholesterol is rapidly mobilized from adipose tissue, with resultant increases in biliary secretion. With aging, there may be a decrease in activity of the 7α-hydroxylase enzyme, a key enzyme in the metabolism of cholesterol to bile salts. Various medications, including clofibrate, estrogens, and progesterone, affect cholesterol secretion by altering hepatic cholesterol uptake or metabolism.

Alternatively, biliary cholesterol supersaturation may occur with relative bile acid hyposecretion. In patients with ileal disease or who have undergone ileal bypass or resection, bile acid absorption is decreased, leading to a diminished bile acid pool. In patients with primary biliary cirrhosis or other chronic cholestatic diseases, bile acid secretion may be decreased. Recent attention has focused on the role of deoxycholic acid. In nonobese patients with cholesterol gallstones, cholic acid and chenodeoxycholic acid pools are reduced, and deoxycholic acid is often increased in bile. Deoxycholic acid may influence the amount of cholesterol and mucin secreted into bile. Prolonged intestinal transit seems to be the underlying reason for the increased percentage of deoxycholic acid in gallstone patients. Associated with these differences in intestinal transit, gallstone patients have higher concentrations of anaerobic and gram-positive aerobic bacteria in the colon, as well as significantly higher activity of the 7α-dehydroxylase enzyme, which is critical in the synthesis of deoxycholic acid formation. In the future, altering the metabolism of deoxycholic acid or accelerating intestinal transit may prove to be a mechanism to prevent gallstone formation.

Recent animal studies have indicated that specific genes determine the incidence of cholesterol stone formation in mice fed a lithogenic diet. Some of these genes may act through their role in regulation of biliary cholesterol or bile acid secretion. For example, apolipoprotein E (apoE), an important component of very-low-density and high-density lipoproteins and chylomicrons, has been the subject of many studies. It has a central role in hepatic lipoprotein catabolism and regulates the plasma cholesterol response to dietary cholesterol. Some, but not all, previous studies have associated the E4 allele of apoE with a higher risk of cholesterol gallstones. The E4 genotype of apolipoprotein E is reported to be associated with gallstone formation, and patients with the E4 genotype have higher gallstone cholesterol content. In apoE-deficient mice, the increase in plasma cholesterol in response to dietary cholesterol is attenuated. The apoE-deficient mice also have an attenuated increase in biliary or hepatic cholesterol concentration and cholesterol saturation index in response to a lithogenic diet. These mice form markedly fewer cholesterol crystals or stones compared to wild-type mice when fed a lithogenic diet. Thus, apoE may play a role in regulating biliary cholesterol secretion, while also playing a role in gene–diet interactions in gallstone formation.

Other ongoing work is characterizing the roles of various genes in mouse models of gallstone formation. For example, the murine *Lith1* gene is associated with gallstone disease in a dominant fashion. The bile salt export pump has been identified as a candidate protein for this *Lith1* gene. Other candidate genes for gallstones in mice include those for HMG CoA reductase, for the canalicular multispecific organic anion transporter (CMOAT), for the intracellular lipid transporters (Pctp and Fabp6), for lipoprotein lipase, and for lecithin–cholesterol acyltransferase. Further work will characterize these genes and hopefully identify their human homologues.

NUCLEATION OF CHOLESTEROL CRYSTALS

An early and indispensable step in cholesterol gallstone formation (Fig. 1) is nucleation, the precipitation of solid cholesterol crystals in saturated bile. Aggregation and fusion of cholesterol–phospholipid vesicles is a crucial step in this process, and cholesterol in stones appears to be derived preferentially from these vesicles. Thus, vesicular cholesterol may be more unstable and prone to precipitate than that in mixed micelles.

Vesicles in bile do not behave homogeneously. Vesicles in gallbladder bile are likely to participate in nucleation; vesicles in hepatic bile are much more resistant to nucleation. These differences are caused partly by the difference in their relative cholesterol-to-phospholipid ratio. Vesicles with increased ratios of cholesterol to phospholipid are more prone to aggregate, fuse, and nucleate. Vesicles with lower

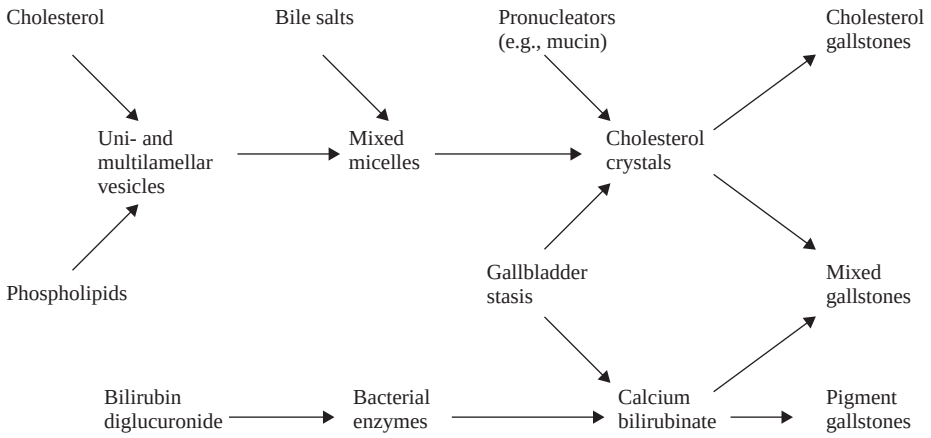


FIGURE 1 Cholesterol gallstone formation. Cholesterol and phospholipids are initially secreted into bile as uni- and multilamellar vesicles. With the addition of bile salts, mixed micelles can form. Further addition of pronucleating agents, such as mucin, in combination with gallbladder stasis can lead to nucleation of cholesterol crystals from supersaturated bile. These crystals may aggregate and grow, eventually forming gallstones.

cholesterol-to-phospholipid ratios, as seen in hepatic bile, are more stable and less prone to nucleation. Conversely, higher concentrations of bile acids, as seen in the gallbladder, favor preferential removal of phospholipids over cholesterol from vesicles during the transition to micellar carriage. The residual vesicles are relatively cholesterol enriched and prone to nucleate.

In humans, secretion of cholesterol-supersaturated bile is common, but not all people with supersaturated bile develop stones. Thus, factors in addition to supersaturated bile must be present for gallstones to form. Factors influencing the rate of crystal nucleation have been a subject of intense study. Biliary proteins have generated the most attention, because total biliary protein content is increased in bile with cholesterol crystals compared with samples without crystals. Several studies have suggested that protein–lipid interactions in bile might be important in promoting or inhibiting crystal formation.

Several pronucleating and antinucleating proteins have been identified in bile. Some putative pronucleating and antinucleating proteins are listed in Table I. All of the postulated pronucleating proteins can bind lipid microaggregates in bile. It is postulated that specific hydrophobic peptide segments of these proteins bind to cholesterol-enriched microdomains in biliary vesicles, eventually leading to their aggregation and fusion. This, in turn, facilitates precipitation of microscopic cholesterol crystals. Nevertheless, the experiments on pro- and antinucleators have often been performed *in vitro* or in animal models, and the significance of many of these proteins *in vivo* or in humans has not been fully established.

The most widely studied pronucleating agent is mucin, a high-molecular-weight glycoprotein secreted by the gallbladder epithelium that is the major organic constituent of gallbladder mucus. Mucin consists of a polypeptide core with multiple oligosaccharide side chains. The polypeptide cores contain hydrophobic domains that can bind cholesterol and phospholipid, thereby bringing vesicles and micelles into close contact. After this binding has occurred, the lipid membranes can fuse and mix, forming multilamellar vesicles. In this way, mucin can act as a pronucleator. Addition of mucin to supersaturated bile will greatly

TABLE I Postulated Pronucleating and Antinucleating Proteins

Pronucleating proteins	Antinucleating proteins
Mucin	Apolipoproteins A-I and A-II
Concanavilin A binding protein	Immunoglobulin
Phospholipase C	Biliary protein fractions with lectin-binding properties
Anionic fraction peptide (calcium-binding protein)	
α1-Acid glycoprotein	
Aminopeptidase N	
Immunoglobulins G and M	
Haptoglobin	
Fibronectin	
Phospholipase A_2	
α1-Antichymotrypsin	

facilitate cholesterol crystal nucleation. Mucin may also be incorporated into the core of growing crystals, providing an architectural nidus for further crystal growth into stones. Mucin hypersecretion is a necessary antecedent to stone formation in a number of animal models of lithogenesis. The stimulus for mucin hypersecretion is not clear, although prostaglandins may play a role. For example, the hypersecretory response can be inhibited by aspirin. In some animal models, aspirin ingestion can inhibit formation of cholesterol crystals and gallstones. Although prostaglandins may play a role in mucin hypersecretion, the role of other factors, such as hydrophobic bile salts or lithogenic bile, has not been clarified. More recently, oxysterols have been identified in bile of patients with pigment gallstones, and their presence may also affect mucin secretion.

Calcium is another important factor influencing crystal nucleation. Calcium salts of bilirubin, phosphate, or carbonate are present in the center of many, if not all, cholesterol gallstones, and may serve as nidi for cholesterol crystallization. Calcium has been shown to bind to mixed micelles, simple micelles, and simple vesicles *in vitro*, potentially explaining its role in cholesterol crystal nucleation. Calcium may also regulate gallbladder secretion of electrolytes and glycoproteins such as mucin. However, studies of biliary calcium levels in gallstone patients are conflicting.

ROLE OF THE GALLBLADDER

The gallbladder is an essential organ in the pathogenesis of cholesterol gallstones. Cholecystectomy essentially prevents *de novo* formation of cholesterol stones. Both gallbladder motility and mucosal function are believed to be important. The gallbladder serves as more than a storage organ. The gallbladder can concentrate and acidify bile, substantially altering the solubility of cholesterol, calcium, and bilirubin in the process. Compared to hepatic ductal bile, gallbladder bile is more concentrated and prone to nucleation. Mucin gels accumulate with prolonged bile storage in gallbladder stasis. The viscosity of these gels may interfere with mechanical emptying. Increased cholesterol nucleation and crystal growth as well as enzymatic and nonenzymatic hydrolysis of bilirubin conjugates could occur with prolonged stasis. The resulting unconjugated bilirubin may precipitate with calcium. Thus, in the presence of gallbladder stasis, bile becomes more concentrated and lithogenic. In this setting, cholesterol crystals are more likely to form. Furthermore, after the initial nucleation step, cholesterol crystals will remain in a static gallbladder. Thus, with gallbladder stasis, cholesterol crystals are more likely to form, remain in the biliary system, and grow into stones. If gallbladder motility is adequate, nascent cholesterol crystals can be emptied, and further growth of these crystals into stones cannot occur.

Altered gallbladder contractility before cholesterol stone formation has been shown in a lithogenic animal model. Many clinical situations associated with gallbladder stasis are also associated with higher risk of gallstones. For example, prolonged total parenteral nutrition (TPN) induces profound gallbladder hypomotility and stasis, predisposing to subsequent biliary sludge and possible stone development. High spinal cord injury is linked to gallbladder stasis and gallstone formation. Pregnancy, oral contraceptives, obesity, diabetes mellitus, and octreotide therapy are also associated with impaired emptying and cholelithiasis.

Gallbladder mucosal function plays a vital role in lithogenesis, but the role of mucosal function is not as well studied compared to gallbladder motility. The gallbladder mucosa absorbs excess water and electrolytes, concentrating bile. The mucosa also normally secretes hydrogen ions and mucin. Transport of sodium and chloride ions across the mucosa is altered in animal models of lithogenesis. The role of the gallbladder in concentrating bile is magnified with gallbladder stasis.

In addition to concentrating bile, the gallbladder mucosa can secrete proteins, such as mucin, which might serve as sites for cholesterol crystal precipitation. Alterations of lipid metabolism in the gallbladder epithelium that might increase cholesterol content have also been studied. Hydroxymethylglutaryl-CoA reductase activity, governing cholesterol synthesis, is lower in the gallbladder mucosa than in hepatic tissue, and acylcoenzyme A : cholesterol acyltransferase (ACAT) activity catalyzing the esterification of cholesterol is several times higher than in the liver. A recent study also suggests that normal human gallbladders can absorb phospholipids and cholesterol from bile *in vitro*, resulting in bile becoming less lithogenic. In contrast, gallbladders from patients with cholesterol stones were less able to absorb cholesterol and phospholipids, resulting in potentially more lithogenic bile. This impaired ability to absorb biliary lipids may be an additional pathogenetic factor for gallstones.

MIXED CHOLESTEROL STONES, PIGMENT STONES, AND THE ROLE OF BACTERIA

Cholesterol gallstones commonly contain other constituents, including proteins and calcium salts of bilirubin or carbonate. Although sometimes called mixed stones,

they are still classified as cholesterol stones, because cholesterol compromises more than 70% of the stone by weight. Pigment gallstones can be subclassified into brown and black types, which differ in morphology, pathogenesis, and clinical associations. In contrast to cholesterol gallstones, ethnic origin is not an important risk factor for pigment gallstones. Black pigment stones can occur in persons with no predisposing conditions, but important risk factors associated with the formation of this type of stone include chronic hemolysis (e.g., sickle cell disease), thalassemia, prosthetic cardiac valves, advancing age, and cirrhosis. Black pigment stones seldom coexist with cholesterol stones in the same gallbladder.

In Asia, brown pigment stones are often found in the bile ducts and are strongly associated with bacterial infection. There is a high incidence of infection with *Escherichia coli* and other intestinal organisms such as *Enterococcus, Enterobacter, Pseudomonas,* and *Proteus.* Infection with anaerobic organisms such as *Clostridium* and *Bacteroides* species can also be documented. Infected bile exhibits high bacterial β-glucuronidase activity, believed to be an important enzymatic process in bilirubin deconjugation. Bilirubin, like cholesterol, is insoluble in water. After glucuronidation in the liver, bilirubin is secreted into bile, mostly as the diglucuronide (75–80%) or the monoglucuronide (20%), but with a small amount of the unconjugated form (3%). Because mixed cholesterol and pigment stones contain salts of calcium and bilirubin, deconjugation and precipitation of bilirubin is essential in the pathogenesis of these stones.

The role of biliary β-glucuronidase in the pathogenesis of pigment gallstones has been controversial. It has been hypothesized that bacterial β-glucuronidase facilitates the hydrolysis of conjugated bilirubin into insoluble bilirubin, providing the substrate for formation of infection-related brown pigment stones. However, this hypothesis does not seem to explain the pathogenesis of black pigment gallstones. β-Glucuronidase activity in uninfected bile has been demonstrated, suggesting that this enzyme may also originate in the biliary tract epithelium. In support of this, gallbladder bile from patients with pigment, cholesterol, or mixed gallstones has been shown to have a higher proportion of bilirubin monoglucuronide compared to bile from controls. However, another pathogenic factor, gallbladder stasis, must be present for precipitates of calcium bilirubinate to form. In a scenario similar to cholesterol gallstone formation, gallbladder stasis can provide an opportunity for nonenzymatic hydrolysis of conjugated bilirubin and its subsequent precipitation. For example, prolonged total parenteral nutrition, with the subsequent gallbladder stasis, can result in the formation of gallbladder sludge and black pigment stones.

The gallbladder may promote pigment gallstone formation in other ways. Calcium solubility is vitally linked to black and brown pigment gallstones, because both kinds of stones contain predominantly calcium bilirubinate. Hepatic bile is supersaturated with calcium carbonate. Normal gallbladder epithelium can acidify bile, which increases the solubility of calcium carbonate. An inflamed gallbladder may be unable to acidify bile, promoting pigment gallstone formation. The gallbladder epithelium secretes mucin into bile, which can bind to bilirubin and other hydrophobic lipids and is contained in black and brown pigment gallstones.

The role of biliary tract infection in gallstone pathogenesis is receiving renewed interest. It has traditionally been believed that bacterial infection is involved only in the pathogenesis of mixed pigment ("brown") stones, in which bacteria are believed to facilitate hydrolysis of conjugated bilirubin to its less soluble, unconjugated form. More recently, using molecular genetic techniques, it has been possible to identify bacteria in both mixed and pure (>90%) cholesterol stones. It has recently been hypothesized that the inciting factor in brown pigment stone formation is bacterial infection and bilirubin deconjugation. However, after initiation of the stone-forming process, bile may evolve in its composition, leading to formation of mixed or even pure cholesterol stones. Similarly, stones may act as foreign bodies and enhance bacterial colonization, resulting in precipitation of bilirubinate salts or remodeling of the existing stones. Thus, it is likely that significant overlaps exist in the pathogenesis of cholesterol and pigment stones. Study in this area has been hampered by the fact that it is often difficult, if not impossible, to determine whether bacterial infection or stone formation was the initial event. The development of gallstones likely includes the interaction of bacterial and physical–chemical mechanisms over a period of years, leading to diverse types, shapes, and numbers of stones.

BILIARY SLUDGE

Once sufficient numbers of crystals have formed, they may be visible by ultrasonography and produce the same range of clinical symptoms as gallstones. These precipitates in bile have been called by many names, including biliary sludge, microlithiasis, and pseudolithiasis. Sludge is best diagnosed by microscopic examination of a fresh sample of gallbladder bile. Sludge also appears ultrasonographically as low-amplitude echoes (without postacoustic shadow) that layer with gravity.

Biochemically, sludge is composed of cholesterol monohydrate crystals, calcium bilirubinate granules, or other calcium salts embedded in a mucus gel. Calcium bilirubinate granules are present in almost all cases, with bilirubin usually in its unconjugated and least soluble form. Proteins and other substances such as ceftriaxone are also important components. Sludge also contains a large proportion of mucin, protein–lipid complexes, and undefined residue. The calcium precipitates, with cholesterol crystals 50 μm or more in diameter, produce the characteristic ultrasonographic echoes in sludge. The deformable mucin gel accounts for its unique layering and flow characteristics.

In certain clinical conditions, sludge may evolve into cholelithiasis, and it has been hypothesized that the crystals in sludge aggregate and grow to form gallstones. Thus, sludge is believed by many to be a necessary precursor to stones. For example, in a prospective study of patients receiving prolonged TPN, 6% of the patients developed sludge during the initial 3 weeks. By 4–6 weeks, 50% of the patients had sludge, and after 6 weeks, sludge was seen universally. Stones developed in 43% of patients during followup. At this time, the risk factors for evolution of sludge into stones are unknown.

Gallbladder sludge may also appear spontaneously in selected individuals. For example, in the previous study of patients receiving prolonged TPN, sludge resolved with reinstitution of oral feedings in all the patients by the end of 4 weeks. Another study documented the clinical outcome of a group of patients presenting with upper abdominal pain and sludge documented on initial ultrasound examination. In approximately 18% of patients, the sludge resolved over a 2-year period. It had a disappearing and reappearing course in about 60% of patients. Gallstones also developed in 14% of patients. Other retrospective studies have demonstrated similar findings.

SUMMARY

The pathogenesis of cholesterol gallstones is complex and not completely understood. It is believed that a primary factor is altered secretion of biliary cholesterol, phospholipid, and bile acids from the liver, resulting in cholesterol supersaturation of bile. As the bile is progressively concentrated during its passage through the bile ducts and during storage in the gallbladder, the vesicular fraction of biliary lipids becomes relatively cholesterol enriched. These lipid vesicles aggregate and fuse. Gallbladder mucin hypersecretion and stasis develop, perhaps through the action of chemical mediators such as prostaglandins. With the addition of pronucleating proteins such as mucin, nucleation of cholesterol crystals can occur. The cholesterol crystals formed can be retained within the gallbladder, where they grow and conglomerate with mucin glycoprotein and other constituents, such as calcium and bilirubin. As these crystals aggregate and enlarge, they may become visible on ultrasonography as biliary sludge. Continued gallbladder stasis and impaired emptying of sludge eventually lead to macroscopic stone formation.

The pathogenesis of pigment stones has recently been revisited with the documentation of bacteria and bacterial biofilms in gallstones. Bacterial enzymatic deconjugation of bilirubin is a key factor in formation of bilirubinate stones, and may be an initiating factor for cholesterol gallstone formation as well. Ongoing work is focused on the role of pronucleating and antinucleating proteins, as well as further clarification of the role and timing of bacterial infection in gallstone pathogenesis.

See Also the Following Articles

Bile Composition • Bile Flow • Bile Formation • Cholecystectomy • Cholelithiasis, Complications of • Cholesterol Absorption

Further Reading

Amigo, L., Quinones, V., Mardones, P., Zanlungo, S., Miquel, J. F, Nervi, F., and Rigotti, A. (2000). Impaired biliary cholesterol secretion and decreased gallstone formation in apolipoprotein E-deficient mice fed a high-cholesterol diet. *Gastroenterology* **118**, 772–779.

Donovan, J. M. (1999). Physical and metabolic factors in gallstone pathogenesis. *Gastroenterol. Clin. North Am.* **28**, 75–97.

Haigh, W. G, and Lee, S. P. (2001). Identification of oxysterols in human bile and pigment gallstones. *Gastroenterology* **121**, 118–123.

Khanuja, B., Cheah, Y. C, Nishina, P. M., Wang, D. Q. H., Chen, H. W., Billheimer, J. T., Carey, M. C., and Paigen, B. (1995). *Lith1*, a major gene affecting cholesterol gallstone formation among inbred strains of mice. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7729–7733.

Ko, C. W., Sekijima, J. H., and Lee, S. P. (1999). Biliary sludge. *Ann. Intern. Med.* **130**, 301–311.

Lammert, F., Carey, M. C., and Paigen, B. (2001). Chromosomal organization of candidate genes involved in cholesterol gallstone formation: A murine gallstone map. *Gastroenterology* **120**, 221–238.

Lee, S. P., Maher, K., and Nicholls, J. F. (1988). Origin and fate of biliary sludge. *Gastroenterology* **94**, 170–176.

Smit, J. J., Schinkel, A. H, Oude Elferink, R. P., Groen, A. K., Wagennar, E., van Deemter, L., Mol, C. A., Ottenhoff, R., van der Lugt, N. M., and van Roon, M. A. (1993). Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. *Cell* **75**, 451–462.

Swidsinski, A., and Lee, S. P. (2001). The role of bacteria in gallstone pathogenesis. *Front. Biosci.* **6**, E93–E103.
Swidsinski, A., Ludwig, W., Pahlig, H., and Priem, F. (1995). Molecular genetic evidence of bacterial colonization of cholesterol gallstones. *Gastroenterology* **108**, 860–864.
Van Erpecum, K. J., Wang, D. Q. H., Lammert, F., Paigen, B., Groen, A. K., and Carey, M. C. (2001). Phenotypic characterization of *Lith* genes that determine susceptibility to cholesterol cholelithiasis in inbred mice: Soluble pronucleating proteins in gallbladder and hepatic biles. *J. Hepatol.* **35**, 444–451.

Gastrectomy

ALEXANDER PEREZ AND MALCOLM K. ROBINSON
Harvard Medical School

antrectomy Type of partial gastrectomy in which the stomach's antrum is removed.
Billroth I and II Reconstructive surgical procedures used after partial gastrectomy in which the remaining part of the stomach is connected to the duodenum or jejunum, respectively.
endoscope Medical instrument used to examine the inside of the stomach and duodenum; consists of a thin, long, flexible tube that contains a light and camera, which can be easily passed through the mouth and into the stomach and duodenum. A small, flexible wirelike instrument can be inserted through the endoscope and out its end to sample or remove stomach masses such as polyps.
esophagojejunostomy Surgical creation of an artificial passage between the esophagus and jejunum; the passage is used to reconstruct the intestinal tract following a total gastrectomy.
gastrectomy Surgical procedure in which part (i.e., a "partial" gastrectomy) or all (i.e., a "total" gastrectomy) of the stomach is removed.
gastroduodenostomy Type of gastroenterostomy in which a portion of the stomach is connected to the duodenum.
gastroenterostomy Surgical creation of an artificial passage between the stomach and any part of the small intestine (e.g., the duodenum or jejunum).
gastrojejunostomy Type of gastroenterostomy in which a portion of the stomach is connected to the jejunum.
gastrostomy Artificial opening in the stomach wall, usually created by surgical means. Not to be confused with gastrectomy.
palliation Treatment designed to improve a patient's symptoms without curing the disease causing the symptoms.
sphincter Ringlike band of muscle fibers that can intermittently contract to constrict an opening, thereby controlling flow of substances out of a structure. For example, the pyloric sphincter controls flow of contents between the stomach and duodenum.
stomach polyp Raised mass of the inner lining of the stomach; protrudes into the lumen of the stomach.
tumor Abnormal collection of cells that form a mass or polyp; in the lay community, at times mistakenly assumed to be equivalent to "cancer," although tumors can be either malignant (i.e., cancerous) or benign (noncancerous). Malignant and benign tumors are distinguished by their appearance under the microscope and/or their propensity to spread throughout the body.
ulcer (stomach or duodenum) Local defect or erosion of the inner lining of the stomach or duodenum.

The word "gastrectomy" comes from the Greek *gaster*, meaning stomach, and the root word *ectomia*, meaning to surgically excise or remove. Christian Theodur Billroth, a well-known surgeon from Vienna, first described partial gastrectomy procedures in 1881. In recognition of his efforts, Billroth's name defines two important and well-known reconstructive surgical procedures—the Billroth I and Billroth II. Total or partial gastrectomies are usually performed to remove benign and cancerous tumors of the stomach and for treatment stomach or duodenal ulcers.

STOMACH ANATOMY AND FUNCTION

The stomach is composed of various tissues and cells that work together to initiate the digestion of food. The stomach is located in the left upper part of the abdomen between the esophagus and duodenum and can be divided into four parts: the cardia, the fundus, the body, and the antrum. The cardia is the portion of the stomach that is in direct continuity with the esophagus, the

Copyright 2004, Elsevier (USA). All rights reserved.

Swidsinski, A., and Lee, S. P. (2001). The role of bacteria in gallstone pathogenesis. *Front. Biosci.* **6**, E93–E103.
Swidsinski, A., Ludwig, W., Pahlig, H., and Priem, F. (1995). Molecular genetic evidence of bacterial colonization of cholesterol gallstones. *Gastroenterology* **108**, 860–864.
Van Erpecum, K. J., Wang, D. Q. H., Lammert, F., Paigen, B., Groen, A. K., and Carey, M. C. (2001). Phenotypic characterization of *Lith* genes that determine susceptibility to cholesterol cholelithiasis in inbred mice: Soluble pronucleating proteins in gallbladder and hepatic biles. *J. Hepatol.* **35**, 444–451.

Gastrectomy

ALEXANDER PEREZ AND MALCOLM K. ROBINSON
Harvard Medical School

antrectomy Type of partial gastrectomy in which the stomach's antrum is removed.
Billroth I and II Reconstructive surgical procedures used after partial gastrectomy in which the remaining part of the stomach is connected to the duodenum or jejunum, respectively.
endoscope Medical instrument used to examine the inside of the stomach and duodenum; consists of a thin, long, flexible tube that contains a light and camera, which can be easily passed through the mouth and into the stomach and duodenum. A small, flexible wirelike instrument can be inserted through the endoscope and out its end to sample or remove stomach masses such as polyps.
esophagojejunostomy Surgical creation of an artificial passage between the esophagus and jejunum; the passage is used to reconstruct the intestinal tract following a total gastrectomy.
gastrectomy Surgical procedure in which part (i.e., a "partial" gastrectomy) or all (i.e., a "total" gastrectomy) of the stomach is removed.
gastroduodenostomy Type of gastroenterostomy in which a portion of the stomach is connected to the duodenum.
gastroenterostomy Surgical creation of an artificial passage between the stomach and any part of the small intestine (e.g., the duodenum or jejunum).
gastrojejunostomy Type of gastroenterostomy in which a portion of the stomach is connected to the jejunum.
gastrostomy Artificial opening in the stomach wall, usually created by surgical means. Not to be confused with gastrectomy.
palliation Treatment designed to improve a patient's symptoms without curing the disease causing the symptoms.
sphincter Ringlike band of muscle fibers that can intermittently contract to constrict an opening, thereby controlling flow of substances out of a structure. For example, the pyloric sphincter controls flow of contents between the stomach and duodenum.
stomach polyp Raised mass of the inner lining of the stomach; protrudes into the lumen of the stomach.
tumor Abnormal collection of cells that form a mass or polyp; in the lay community, at times mistakenly assumed to be equivalent to "cancer," although tumors can be either malignant (i.e., cancerous) or benign (noncancerous). Malignant and benign tumors are distinguished by their appearance under the microscope and/or their propensity to spread throughout the body.
ulcer (stomach or duodenum) Local defect or erosion of the inner lining of the stomach or duodenum.

The word "gastrectomy" comes from the Greek *gaster*, meaning stomach, and the root word *ectomia*, meaning to surgically excise or remove. Christian Theodur Billroth, a well-known surgeon from Vienna, first described partial gastrectomy procedures in 1881. In recognition of his efforts, Billroth's name defines two important and well-known reconstructive surgical procedures—the Billroth I and Billroth II. Total or partial gastrectomies are usually performed to remove benign and cancerous tumors of the stomach and for treatment stomach or duodenal ulcers.

STOMACH ANATOMY AND FUNCTION

The stomach is composed of various tissues and cells that work together to initiate the digestion of food. The stomach is located in the left upper part of the abdomen between the esophagus and duodenum and can be divided into four parts: the cardia, the fundus, the body, and the antrum. The cardia is the portion of the stomach that is in direct continuity with the esophagus, the

Copyright 2004, Elsevier (USA). All rights reserved.

fundus is the uppermost portion of the stomach, the body is the middle and largest portion, and the antrum is the bottom portion, which is in continuity with the small intestine. The pylorus, a muscular sphincter located at the stomach's exit, regulates passage of partially digested food from the stomach into the first part of the small intestine or duodenum (see Fig. 1).

The cells of the stomach produce juices such as hydrochloric acid and enzymes (e.g., pepsin), which reduce food to a liquid form ready to enter the small intestine. The stomach rhythmically contracts to mix swallowed food with stomach juices, which aids in the digestive process over a few hours. Once reduced to liquid form, the stomach contracts and the pyloric sphincter relaxes, allowing the now liquefied contents of the stomach to flow into the small intestine. The small intestine then continues the digestion process and begins absorbing nutrients into the bloodstream for use by the body.

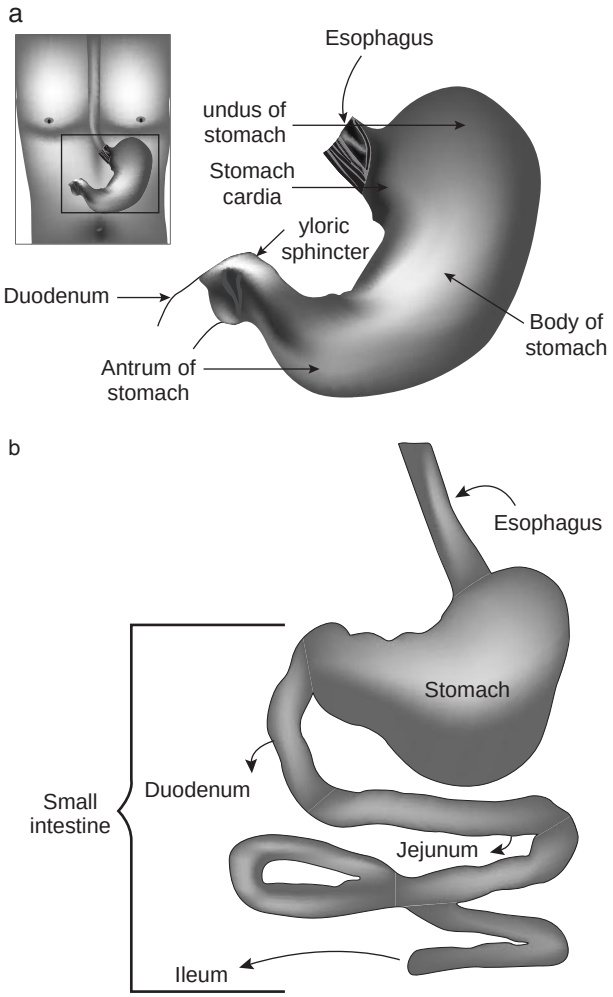


FIGURE 1 (a) Parts of the stomach. (b) Relationship between the stomach, esophagus, and small intestine.

BENIGN AND CANCEROUS STOMACH TUMORS

When treating diseases related to the stomach, the two main reasons to perform a gastrectomy are treatment of stomach tumors and ulcers. Partial gastrectomy may also be used as part of the treatment of other conditions, such as cancer of the pancreas and esophagus and treatment of obesity. Although these latter conditions are not the focus here, the principles of gastrectomy are similar regardless of the indication for performing such surgery.

The stomach can develop a variety of benign and malignant tumors. Benign (i.e., noncancerous) tumors are usually cohesive and have well-demarcated borders. They do not invade or infiltrate the surrounding normal tissues and do not spread to different areas of the body. Malignant (i.e., cancerous) tumors usually invade the surrounding normal tissues and have the potential to spread to distant sites in the body if they are allowed to continue to grow. The spread of malignant tumor cells to areas of the body that are distant from their site of origin is referred to as metastasis. Once metastasis has occurred, the chance of cure by surgery alone is highly unlikely. Hope for cure under such circumstances requires surgery to be combined with other cancer treatment modalities, such as chemotherapy or radiation therapy.

Stomach cancer, also known as gastric cancer, is a common cancer of the gastrointestinal tract. It occurs most frequently in men over age 50 years. Every year in the United States, 25,000 people develop gastric cancer. There are a variety of malignant stomach tumors, and the vast majority of stomach cancers require gastrectomy for potential cure. Although there have been many advances in the treatment of stomach cancer, the likelihood of survival depends on whether the cancer can be treated by gastrectomy while the tumor is still confined to the stomach. The proportion of all individuals diagnosed with stomach cancer who will survive for 5 years is approximately 5–15%. The survival rate for a particular individual, however, depends on the type of cancer, if it has invaded nearby structures, or if it is has metastasized to distant sites of the body, such as the liver. Hence, a person with a "favorable" stomach cancer may have a survival chance of greater than 90% whereas patients with "unfavorable" tumors may have no chance of survival.

Gastrectomy may also be used for palliation of patients who have advanced stomach cancer. Patients

whose stomach cancer blocks the passage of food or secretions, or causes significant bleeding and pain, may be treated by total or partial gastrectomy. Although there may be no hope of cure, the quality of these patients' lives may be significantly improved by performing a gastrectomy in such circumstances.

There are also a variety of benign tumors of the stomach, including benign stomach polyps. Polyps of small to medium size can be removed by an endoscope. However, polyps that are too large to be removed by the endoscope are usually treated by partial gastrectomy. In addition, patients whose polyps who have been removed by the endoscope and are found to be cancerous on microscopic analysis may require further treatment with gastrectomy.

STOMACH AND DUODENAL ULCERS

Another major reason to perform a gastrectomy is for treatment of stomach and duodenal ulcers. The ulcers may form as a result of acid and pepsin produced by the stomach eroding through the inner lining of the stomach and/or duodenum. Alternatively, nonsteroidal antiinflammatory drugs such as aspirin or ibuprofen may lead to ulcer formation. Finally, certain types of bacteria, such as *Helicobacter pylori*, may lead to ulceration of the stomach and duodenum. In the past, stomach and duodenal ulcers were routinely treated by partial, and in some cases total, gastrectomy. However, advances in medical treatments with a variety of drugs have made surgery unnecessary in most cases. These drugs include antacids and certain antibiotics.

Occasionally, stomach or duodenal ulcers do not heal with medical therapy. This not only causes persistent pain for the patient but also raises the possibility that cancer cells may be contained in the ulcer. In addition, complications from stomach and duodenal ulcers may develop due to failed drug therapy or progression of the ulcer disease before medical therapy is initiated. These complications include perforation, in which the ulcer erodes through all layers of the stomach or duodenum, leading to leakage of intestinal fluid into the abdomen and a life-threatening infection. Alternatively, there may be severe bleeding because the ulcer has eroded into a nearby blood vessel. Finally, swelling related to the ulcer may prevent passage of stomach contents into the small intestine. Total or partial gastrectomy may be necessary if a stomach or duodenal ulcer fails to heal with medical therapy, if there is any concern that an ulcer contains cancer, or if an ulcer complication occurs. The ulcer is removed as part of the gastrectomy in the case of a stomach ulcer. A gastrectomy alone does not remove a duodenal ulcer. However, gastrectomy decreases or eliminates the flow of acid and pepsin produced by the stomach into the duodenum, thereby allowing the duodenal ulcer to heal on its own.

GASTRECTOMY

There are two major steps involved in performing a gastrectomy: stomach resection and reconstruction of the intestinal tract. Resection involves determining the portion of the stomach to be removed or "resected." This determination is based on the location of the diseased portion of the stomach. When the entire stomach is removed, it is referred to as a "total" gastrectomy (Fig. 2). When part of the stomach is removed, it can be referred to as a "partial" gastrectomy. Often a partial gastrectomy is more precisely defined by the segment or amount of the stomach that is removed. Hence, an "antrectomy" is a partial gastrectomy in which the stomach antrum alone is removed. This is roughly 35% of the stomach (Fig. 3). A "distal" gastrectomy is a partial gastrectomy in which the distal 50% of the stomach (i.e., the part of the stomach closest to the duodenum) is removed. In a distal gastrectomy, all of the antrum and part of the body of the stomach are removed. Finally, a "subtotal" gastrectomy is a partial gastrectomy in which 80% of the distal stomach is removed. A subtotal gastrectomy is essentially just a more extensive form of a distal gastrectomy (Fig. 4).

Once a total or partial gastrectomy is performed, the remaining segments of the gastrointestinal tract must be reconnected to allow uninterrupted flow of food from the esophagus into the small intestine. This is referred to as reconstruction of the gastrointestinal tract. There are a wide variety of reconstruction procedures that can be done using sutures, surgical staples, or a combination of the two. Exactly, what reconstructive procedure is done and how it is performed largely depends on a surgeon's preference. Following an antrectomy, the surgeon may

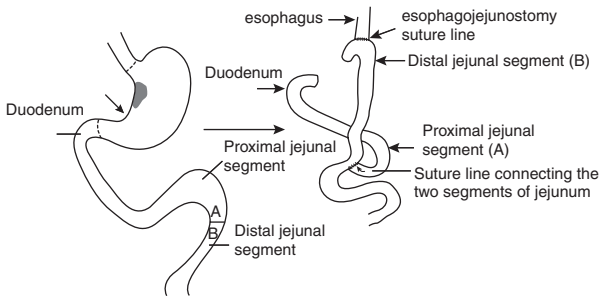


FIGURE 2 Total gastrectomy with esophagojejunostomy reconstruction.

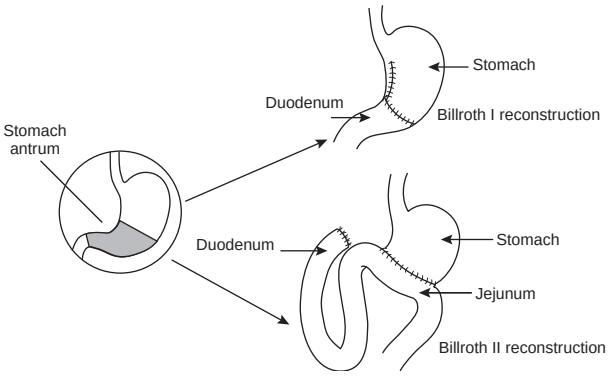


FIGURE 3 Antrectomy with Billroth I and II reconstructions.

choose to use a Billroth I reconstruction, in which a gastroduodenostomy is performed. Alternatively, the surgeon may use a Billroth II reconstruction, in which a type of gastrojejunostomy is performed (Fig. 3). For a distal or subtotal gastrectomy, another type of gastrojejunostomy may be used (Fig. 4).

A much more complex reconstruction is required following a total gastrectomy for the following reasons (see Fig. 2). The esophagus has very limited mobility and cannot be pulled down into the abdomen very far. In addition, the duodenum is fixed in position by its attachments to various other structures in the abdomen. Hence, the duodenum can not be pulled up to the esophagus. Thus, it is not possible to connect the esophagus to the duodenum after a total gastrectomy. In contrast, the jejunum is much more mobile than either the esophagus or the duodenum and is therefore used for reconstructive purposes following total gastrectomy. Although it may be possible to pull a segment of jejunum up to the esophagus without special surgical techniques, the connection between the esophagus and the jejunum may be under marked tension and may tend to tear or pull a part. Thus, the mobility of the jejunum is frequently optimized to allow safe connection to the esophagus in a two-step process. First, the jejunum is divided,

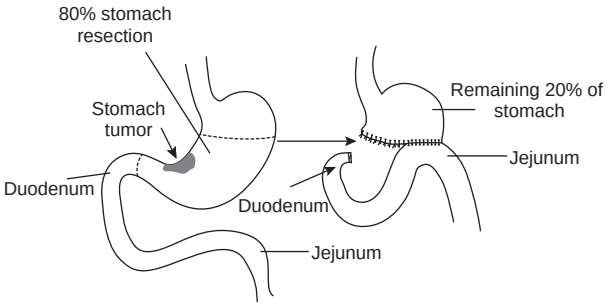


FIGURE 4 Subtotal gastrectomy with gastrojejunostomy reconstruction.

which allows the segment farthest from the duodenum (i.e., the distal segment) to easily be pulled up to the esophagus without tension. The reason this increases jejunal mobility is beyond the scope of this discussion, but has to do with certain anatomical properties of the jejunum and the blood vessels to which the jejunum is attached. This distal segment of jejunum is either sewn or stapled to the esophagus to form an esophagojejunostomy. The second step involves stapling or sewing the proximal segment of jejunum (i.e., the segment in continuity with the duodenum) to the side of the distal jejunum (see Fig. 2).

COMPLICATIONS FOLLOWING GASTRECTOMY

Unfortunately, removal of all or part of the stomach also decreases the stomach's reservoir and digestive capacity. Therefore, almost all patients have to alter their eating habits postgastrectomy. In addition, they may have nausea, vomiting, and diarrhea for several days or weeks after surgery. Changing to small and frequent meals with low sugar and high protein content helps many of these symptoms. Most people successfully adapt to the surgery and lead normal lives, assuming no other complications develop.

Unfortunately, 20% of patients have significant complications, with 5% of patients having lifelong complications and 1% with disabling complications. Early complications following surgery may include bleeding or injury to organs or structures in the area of the surgery. The connections made as part of the reconstruction procedure may leak intestinal fluids and lead to a mild, moderate, or life-threatening infection in the abdomen. Alternatively, the connections may swell or become constricted, resulting in partial or complete obstruction of the flow of food from the mouth into the distal small intestines. Finally, a superficial infection of the surgical skin incision may develop. Most of these complications are treatable without additional surgery. Occasionally, surgery is needed to correct long-term problems or fix problems that are identified immediately after surgery. In general, however, a gastrectomy results in high survival rates with low complications when performed for benign diseases or for early forms of cancer.

Further Reading

Ashley, S. W., Envoy, D., and Daly, J. M. (2001). Stomach. *In* "Principles of Surgery" (S. I. Schwartz, G. T. Shires, F. C. Spencer, J. M. Daly, J. E. Fischer, and A. C. Galloway, eds.), 7th Ed., pp. 1181–1215. McGraw-Hill, New York.

Pierie, J. P., and Ott, M. J. (2001). The stomach. *In* "Current Surgical Therapy" (J. L. Cameron, ed.), 7th Ed., pp. 105–112. Mosby Publ., Boston.

Sawyers, J. L. (2002). Gastric procedures. *In* "American College of Surgeons. ACS Surgery, Principles of Practice" (D. W. Wilmore, L. Y. Cheung, A. H. Harken, J. W. Holcroft, J. L. Meakins, and N. J. Soper, eds.), pp. 707–719. WebMD Professional Publ., New York.

Turner, D. J., and Mulholland, M. W. (1996). Gastric resection and reconstruction. *In* "Shackelford's Surgery of the Alimentary Tract" (G. D. Zuidema and C. J. Yeo, eds.), 5th Ed., Vol. 2, pp. 130–160. W. B. Saunders, Philadelphia.

Gastric Acid Secretion

CATHERINE S. CHEW
Medical College of Georgia, Augusta

afferent Describes the nerves that carry impulses from the outer body toward the brain or spinal cord.

autonomic nervous system Part of the nervous system in vertebrates that controls involuntary activity.

chyme Thick, fluid masses of partially digested food and gastric secretions that pass from the stomach to the small intestine.

efferent Describes the nerves that carry impulses away from the brain or spinal cord.

endocrine Related to cells that secrete hormones into the bloodstream.

enterochromaffin-like cell Synthesizes, stores, and secretes histamine; serves as a paracrine regulator in the gastric mucosa.

gastrin Gastrointestinal hormone that is produced by G cells in the gastric antral mucosa; stimulates gastric acid secretion.

neurotransmitter Chemical that carries messages between nerve cells.

paracrine Related to cells that secrete regulatory substances in the vicinity of target cells.

postganglionic Occurring after the ganglion.

preganglionic Occurring before the ganglion.

P-type ATPase Enzyme that utilizes ATP hydrolysis to transport cations or positively charged ions across cell membranes; "P" refers to the enzymatic requirement for a high-energy covalent β-aspartyl phosphate intermediate.

pyloropyloric Reflex that is initiated in the pylorus and acts in the pylorus.

sympathetic nervous system Part of the autonomic nervous system that is active during stress or danger.

Gastric glands and surface epithelial cells in the stomach produce 1–2 liters/day of gastric juice, the major components of which are water, hydrochloric acid (HCl), salts (Na^+, K^+, and Cl^-), pepsins, intrinsic factor, and mucus. The gastric mucosal glands in the fundus and body of the stomach contain parietal (oxyntic) cells that secrete the HCl, a process regulated by neural, endocrine, and paracrine mechanisms. There are approximately 1 billion parietal cells in the human gastric mucosa, and these metabolically active cells have the ability to generate more than 3×10^6 hydrogen ions/second. The result of this intense acid secretory activity is that the pH of gastric juice is quite acidic, dropping below pH 1 at high rates of secretion. HCl is necessary for the conversion of inactive pepsinogen to pepsin, which initiates the digestion of collagen; HCl also aids in the destruction of ingested bacteria, although some bacteria, such as *Helicobacter pylori*, which has been implicated in the development of peptic ulcer disease, manage to survive despite the hostile environment. The maximal rate of acid output is gender and age related. In human males, maximally stimulated acid secretion initially generates 10–30 mmol of HCl/hour. In females, the rate is 25–40% lower. In both males and females, maximal acid output declines with age.

INTRODUCTION

The recognition that gastric juice contains HCl and that the secretion of HCl is controlled by multiple factors has gradually evolved from observations made over the past 200 years or so. As early as 1802, the French physiologist, Pierre Jean Georges Cabanis, recognized the link between gastric function and emotion. It was not until 1833, however, that the physician William Beaumont

Copyright 2004, Elsevier (USA). All rights reserved.

Pierie, J. P., and Ott, M. J. (2001). The stomach. *In* "Current Surgical Therapy" (J. L. Cameron, ed.), 7th Ed., pp. 105–112. Mosby Publ., Boston.

Sawyers, J. L. (2002). Gastric procedures. *In* "American College of Surgeons. ACS Surgery, Principles of Practice" (D. W. Wilmore, L. Y. Cheung, A. H. Harken, J. W. Holcroft, J. L. Meakins, and N. J. Soper, eds.), pp. 707–719. WebMD Professional Publ., New York.

Turner, D. J., and Mulholland, M. W. (1996). Gastric resection and reconstruction. *In* "Shackelford's Surgery of the Alimentary Tract" (G. D. Zuidema and C. J. Yeo, eds.), 5th Ed., Vol. 2, pp. 130–160. W. B. Saunders, Philadelphia.

Gastric Acid Secretion

CATHERINE S. CHEW
Medical College of Georgia, Augusta

afferent Describes the nerves that carry impulses from the outer body toward the brain or spinal cord.

autonomic nervous system Part of the nervous system in vertebrates that controls involuntary activity.

chyme Thick, fluid masses of partially digested food and gastric secretions that pass from the stomach to the small intestine.

efferent Describes the nerves that carry impulses away from the brain or spinal cord.

endocrine Related to cells that secrete hormones into the bloodstream.

enterochromaffin-like cell Synthesizes, stores, and secretes histamine; serves as a paracrine regulator in the gastric mucosa.

gastrin Gastrointestinal hormone that is produced by G cells in the gastric antral mucosa; stimulates gastric acid secretion.

neurotransmitter Chemical that carries messages between nerve cells.

paracrine Related to cells that secrete regulatory substances in the vicinity of target cells.

postganglionic Occurring after the ganglion.

preganglionic Occurring before the ganglion.

P-type ATPase Enzyme that utilizes ATP hydrolysis to transport cations or positively charged ions across cell membranes; "P" refers to the enzymatic requirement for a high-energy covalent β-aspartyl phosphate intermediate.

pyloropyloric Reflex that is initiated in the pylorus and acts in the pylorus.

sympathetic nervous system Part of the autonomic nervous system that is active during stress or danger.

Gastric glands and surface epithelial cells in the stomach produce 1–2 liters/day of gastric juice, the major components of which are water, hydrochloric acid (HCl), salts (Na^+, K^+, and Cl^-), pepsins, intrinsic factor, and mucus. The gastric mucosal glands in the fundus and body of the stomach contain parietal (oxyntic) cells that secrete the HCl, a process regulated by neural, endocrine, and paracrine mechanisms. There are approximately 1 billion parietal cells in the human gastric mucosa, and these metabolically active cells have the ability to generate more than 3×10^6 hydrogen ions/second. The result of this intense acid secretory activity is that the pH of gastric juice is quite acidic, dropping below pH 1 at high rates of secretion. HCl is necessary for the conversion of inactive pepsinogen to pepsin, which initiates the digestion of collagen; HCl also aids in the destruction of ingested bacteria, although some bacteria, such as *Helicobacter pylori*, which has been implicated in the development of peptic ulcer disease, manage to survive despite the hostile environment. The maximal rate of acid output is gender and age related. In human males, maximally stimulated acid secretion initially generates 10–30 mmol of HCl/hour. In females, the rate is 25–40% lower. In both males and females, maximal acid output declines with age.

INTRODUCTION

The recognition that gastric juice contains HCl and that the secretion of HCl is controlled by multiple factors has gradually evolved from observations made over the past 200 years or so. As early as 1802, the French physiologist, Pierre Jean Georges Cabanis, recognized the link between gastric function and emotion. It was not until 1833, however, that the physician William Beaumont

Copyright 2004, Elsevier (USA). All rights reserved.

provided the first scientific evidence, based on his observations of a fistulous patient, Alexis St. Martin, that emotional states influence gastric acid secretion. Hydrochloric acid was identified as a component of gastric juice in the early 1800s by William Prout, but remained a controversial finding for the next 50 years because of the many claims that the gastric mucosa secreted phosphoric or lactic acid or other substances. Although direct evidence was lacking, it became generally accepted during this era that the parietal cell is the source of HCl secretion.

In the early twentieth century, major progress was made toward identifying the physiological regulators of the acid secretory response. The hormone gastrin was isolated by J. S. Edkins in 1905 and subsequently was shown to increase HCl secretion. Ivan Pavlov defined the "psychic phase" of gastric acid secretion by demonstrating that the sight, smell, and anticipation of eating were potent stimulants of acid secretion. Histamine was discovered in 1913 by Henry Dale and was shown, in 1916, to stimulate HCl secretion by Leon Popielski, a student of Pavlov's. This discovery, which was published in 1920, initiated a bitter controversy over the role of histamine in the regulation of acid secretion that continued for over 50 years. The first evidence linking brain pathways with gastric acid secretion was published in the 1930s by Cushing, who found that patients with intracranial lesions secreted high levels of gastric acid and developed peptic ulcers.

The sequencing of gastrin by Gregory and Tracy in 1964 added fuel to the scientific controversy over the respective roles of histamine and gastrin; the controversy continued long after the discovery of the gastric histamine-2 (H2) receptor by Sir James Black and colleagues in 1971. Two prominent opposing viewpoints were championed most strongly by Charles Code, who proposed histamine as the final common mediator, and Morton Grossman and colleagues, who believed that the gastric parietal cell possessed receptors for histamine, gastrin, and acetylcholine and that these secretagogues had potentiating interactions at the level of the parietal cell. In order to resolve these conflicting hypotheses, it became clear that cellular models were required. The first such model, the isolated gastric gland, was described in 1976 by Thomas Berglindh and Carl Johann Öbrink. The gland preparation continues to be a popular model for the study of acid secretory responses at the cellular level. An important characteristic of this cellular model is that normal connections between cells are maintained and cells within the glands retain their normal polarities. The presence of endocrine/paracrine cells in the glands can be an advantage or a disadvantage, depending on the research goals. Thus, this model is ideal for studies of paracrine regulation of parietal cells, but not for characterizing responses at the level of the parietal cell. To define "pure" parietal cell responses at both the physiological and biochemical levels requires preparations of highly enriched parietal cells. Shortly after the gland model was described, the first step toward addressing these shortcomings was made by Andrew Soll, who used centrifugal elutriation, a technique that separates cells in suspension based on size, to partially enrich parietal cells from mixed populations of gastric mucosal cells. These initial parietal cell isolates were enriched to ~50% purity. Since then, there has been a progression of improvements in the enrichment technique to the point where parietal cells can now be enriched to near homogeneity and maintained for several days in primary culture.

Another important milestone in defining acid secretory-related functions of gastric mucosal cells was the identification of histamine-containing enterochromaffin-like (ECL) cells by Rolf Håkanson and colleagues in 1986. Christian Prinz, George Sachs, and colleagues developed techniques to study these cells in isolation in the early 1990s and subsequent work has recently provided significant new information on the mechanisms involved in the regulation of gastric acid secretion.

GASTRIC MORPHOLOGY

General Organization of the Stomach and Regional Variations

Like the rest of the tubular gastrointestinal (GI) tract, the stomach is structurally organized into four concentric layers: the inner or luminal mucosa, the submucosa, the muscularis externa, and the outer serosa. The mucosa and submucosa form longitudinal folds, or rugae, that contain funnel-shaped invaginations at the mucosal surface. These invaginations, which are referred to as gastric pits, or foveolae, form the openings for ducts into which the gastric glands empty. Anatomically, the stomach is divided into four regions: the cardiac, which is a 2- to 3-cm-wide region at the esophageal junction; the fundus, which projects in a domelike fashion above the esophagus; the body, which forms the central and largest region of the stomach; and the pyloric antrum (*pyloros* being the Greek term for "gatekeeper"), which falls in the region between the angular notch and the duodenal junction (see Fig. 1 for an overview of the anatomy of the stomach).

Histologically, the gastric mucosa is divided into three, rather than four, separate regions: the cardiac, oxyntic, and antral or antropyloric. In this case, the

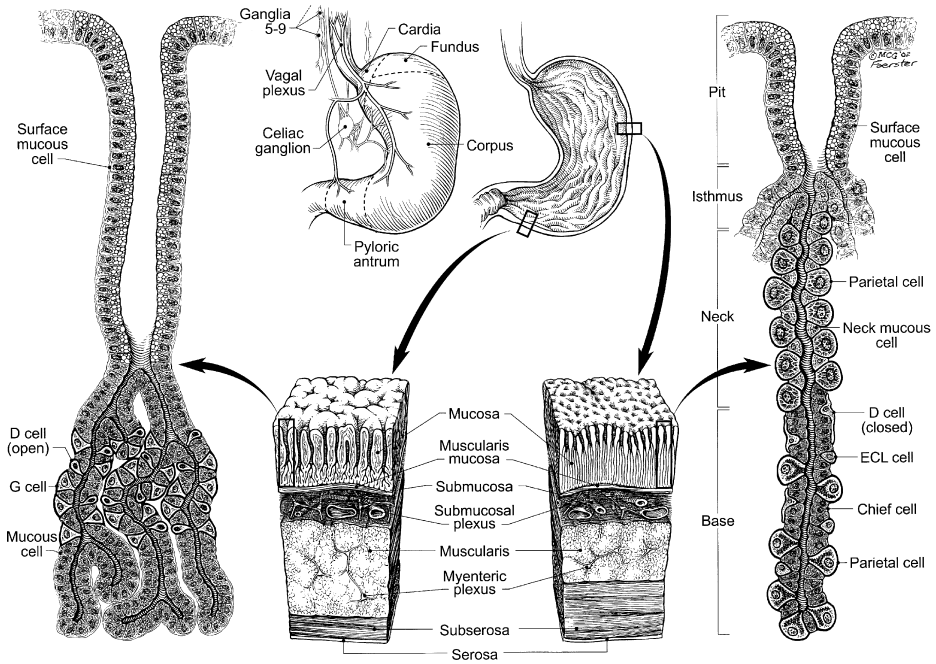


FIGURE 1 Anatomical overview of the stomach. Figure by John Forrester, copyright Medical College of Georgia.

regions are defined based on the structures of their glands. Although there is a common structural basis in gastric glands, the occurrence and distribution of different cell types within the glands vary from region to region. In humans, the cardiac region is a few centimeters wide and contains labyrinthine glands composed mainly of mucus-producing cells. The oxyntic mucosa, which is the acid-producing region of the stomach, extends through the fundus and body of the stomach. [It should be noted that this region is often referred to as the fundus rather than the oxyntic mucosa (see Table I, for example)]. Within the oxyntic mucosa, the glands are long and straight, and two or more glands open into a common pit. Columnar surface epithelial cells line the pit, extending down into the gland opening. The narrow region of the gland closest to the gastric pit is called the isthmus, or neck, region, and it contains mucous neck cells, stem cells, and a few immature parietal cells. The body of the gland begins immediately below the isthmus and extends to the base. Parietal cells are most numerous in the upper regions of the body, becoming less numerous near the base where pepsinogen-secreting chief cells predominate. ECL cells and other endocrine/paracrine cells, which are derived from the neural crest, are interspersed among secretory epithelial cells in oxyntic glands. In the antral mucosa, the gastric pits are deeper, and the glands have wide lumens with prominent branching at their bases. The predominant cell types in the antral mucosa are mucus-producing cells, gastrin-producing G cells, and serotonin-producing ECL cells. A few parietal cells are also present in this region.

Endocrine/Paracrine Cells Involved in the Regulation of HCl Secretion

ECL cells, G cells and somatostatin-producing D cells are known to play an important role in the regulation of HCl secretion. Other less well-characterized endocrine/paracrine cells are probably also involved in this process. In the gastric mucosa, endocrine/paracrine cells are almost exclusively localized within the glan-

TABLE I Distribution and Properties of Endocrine Cells in the Stomach of the Rat[a]

Cell type	Amine/amino acid	Regulator peptide	Location
ECL cell	Histamine	?	Fundus
D cell	?	Somatostatin	Fundus/antrum
G cell	GABA	Gastrin	Antrum
EC cell	Serotonin	?	Antrum
A-like cell	?	Ghrelin	Fundus/antrum
D1 cell	?	?	Fundus/antrum
P cell	?	?	Fundus/antrum

[a] Modified from Lindstrom *et al.* (2001), with permission. The many question marks illustrate that there is much to learn in gastric endocrine cell biology.

dular epithelium at sites below the surface epithelial cells and the pits. Generally, but not exclusively, endocrine-like cells in the oxyntic mucosa contain different secretory products as compared to those in the antral mucosa. There are also different anatomical relationships between endocrine-like cells and neighboring cells in the two regions. These anatomical differences may reflect differences in the modes of delivery of secretory products and/or differences in stimulatory mechanisms. The locations of known endocrine/paracrine cells and their secretory products are summarized in Table I.

The Histamine-Storing ECL Cell

There are two types of histamine-storing cells in the stomach, ECL cells and classical mast cells. ECL cells are involved in the regulation of gastric acid secretion whereas mast cells are involved in mediating the immune response. Mast cells originate in the bone marrow and migrate to the stomach. In contrast, ECL cells originate in the gastric mucosa. The distribution of both ECL and mast cells varies with the species. In large mammals, including dogs, cats, pigs, and humans, mast cells are numerous and are scattered throughout the mucosa. In contrast, in the rat, mast cells are relatively few in number and are localized not only within the mucosal surface but also within the submucosal and muscle layers. ECL cells predominate in the lower gland region in birds and mammals, but in fish, amphibia, and lizards, they are present primarily in the upper half of glands.

In rats, the species in which these cells have been most extensively studied, ECL cells are the predominant histamine-storing cell, containing $\sim$80% of the total gastric histamine. They are also particularly abundant, composing $\sim$30% of the total endocrine/paracrine cell population. Transmission electron microscopy (EM) studies indicate that secretory vesicles in ECL cells have electron-dense cores surrounded by large halos. At the light microscopic level, ECL cells can be identified with antibodies directed against the histamine-synthesizing enzyme, histidine decarboxylase, and against the vesicular monoamine transporter subtype 2 (VMAT-2). In the human stomach, a subpopulation of ECL cells has immunoreactivity for the α subunit of human chorionic gonadotropin (hCG-α) as well as other constituents that are found in neuroendocrine cells, such as chromogranin/pancreastatin and synaptophysin.

The Gastrin-Storing G Cell

The G cells are flask-shaped, with apical surface microvilli that project into the lumen of the antral glands. Secretory granules in G cells have variable structures and densities, which may reflect the accumulation of different gastrin precursors within these granules. In general, endocrine cells in the GI tract have been classified based on their silver staining characteristics as well as the presence of chromogranin A, an ubiquitously expressed endocrine cell protein that is thought to play a role in the packaging of secretory products. Antral G cells stain strongly with both methods. The distribution of G-cells in antral glands is species dependent. In rats, G cells predominate in the basal region, but in humans, dogs, and cats, they appear in stratified layers between the base of the pits and the mucous cells. In pigs, G cells are distributed along the length of the glands.

Immunohistochemical analyses suggest that G cells contain other peptides in addition to gastrin, possibly including peptide YY (PYY), thyrotropin-releasing hormone (TRH)-like peptides, vasoactive intestinal peptide (VIP), proenkephalin gene-derived peptides, xenopsin, adrenocorticotropic hormone (ACTH), and human CG-α (hCG-α). The presence of these peptides in G cells has not yet been confirmed biochemically nor have their roles in the regulation of acid secretion been defined. Such analyses are presently difficult because G cells can be enriched to only $\sim$35% with current isolation protocols.

Somatostatin-Storing D Cells in Oxyntic and Antral Glands

The main secretory product of D cells is somatostatin, an inhibitory regulator of acid secretion, and the predominant form of somatostatin that is secreted by gastric D cells is somatostatin-14. Compared to somatostatin-28, which is the form present in the general circulation, somatostatin-14 appears to be a more physiologically relevant inhibitor of acid secretion. In humans and most mammals, the granules in D cells are large (100–400 nm), round, and of medium density. The granules in rodent D cells are smaller. In the gastric antrum, D cells are the "open" type whereas those in the fundus are the "closed" type. Is has been assumed that the open morphology allows antral D cells to sense the presence of acid in the stomach lumen; however, this assumption has not been proved. Both antral and fundic D cells have long processes, which were originally thought to form direct contacts with parietal and G cells. More detailed transmission EM studies suggest, however, that thin extensions from other cell types block direct contacts. If this is the case, somatostatin may be delivered to parietal and G cells through the local circulation rather than by simple diffusion. Somatostatin released from intestinal D cells into the general

circulation can also serve as an endocrine inhibitor of acid secretion. Experimentally, gastric D cells can currently be enriched to ~90% purity based on their small size and low buoyant densities (1.040 g/ml) using a combination of density gradient and centrifugal elutriation techniques similar to those that were initially used to prepare highly enriched parietal cell populations.

CELLULAR DIFFERENTIATION WITHIN OXYNTIC GASTRIC GLANDS

Within the oxyntic mucosa, there is constant turnover and regeneration of the different cell types. Landmark studies by Karam and Leblond during 1993–1995 defined gastric glands as "zymogenic units" that contain multipotent, undifferentiated, granule-free stem cells from which several different progenitors are produced, including "prepit precursors," "preneck cell precursors," and "preparietal cells." These stem cells are present in the isthmus of the glands. Surface or pit cells arise from differentiation of the prepit precursor lineage and migrate upward from the isthmus during migration. Mucous neck cells are derived from the preneck cell lineage and are the precursors of chief or zymogenic cells, which are present in the base of the glands. Preparietal cells appear to be derived from granule-free stem cells as well as prepit and preneck precursors. Preparietal cells appear to migrate from the gland isthmus in both directions: upward toward the pit and downward toward the neck and base of the glands. In contrast to surface epithelial cells, which survive for only a few days, parietal cells have a life span of several months, depending on the species. Fully differentiated parietal cells are mitochondria rich and possess an extensive tubulovesicular or tubulocisternal membrane system as well as a complex, F-actin-rich, apically directed internalized membrane region or intracellular canaliculus. These cells also express several proteins that are either not expressed at all or are expressed at much lower levels in other gastric mucosal cell types. The most prominent example of the former is the H^+,K^+-ATPase, which is a P-type ATPase that is responsible for proton secretion by the parietal cell. In contrast to gastric epithelial cells, G and D cells appear to be derived from a common multihormonal precursor cell that is present within the isthmus of glands.

INNERVATION OF THE STOMACH

The stomach, like the rest of the gastrointestinal tract, is extensively innervated by the autonomic nervous system (ANS). There are two major sources of autonomic innervation, extrinsic and intrinsic. The extrinsic nerve supply feeds into a self-contained, highly complex intrinsic nervous system, which is also referred to as the enteric nervous system. In addition to regulating acid secretion, neuronal messengers modulate a number of other important gastric functions, including motility, local blood flow, lower esophageal and pyloric sphincter tone, and hormone/paracrine release. Numerous neurotransmitters are present within the gastric mucosa, including acetylcholine, serotonin, nitric oxide, γ-amino butyric acid, and pituitary adenylyl cyclase-activating peptide (PACAP). Neuropeptide Y (NPY) may also be present in extrinsic sympathetic fibers. Peptide-containing neuronal fibers have also been found to contain abundant vasoactive intestinal peptide and gastrin-releasing peptide (GRP), with lesser concentrations of substance P, calcitonin gene-related peptide (CGRP), somatostatin, NPY, and galanin. Mucosal fibers may contain multiple transmitters. There is also a high level of nitric oxide synthetase (NOS) in the gastric mucosa. A neuropeptide, orexin A, has been recently immunolocalized to neuronal structures that express VIP and/or NOS. However, although NOS-containing nerve fibers are abundant in the smooth muscle layer, few such fibers are found in the gastric mucosa. Thus, much of the NOS activity may originate from mucosal epithelial cells.

Extrinsic Innervation and CNS Control

There are two sources of external or extrinsic neuronal pathways: the parasympathetic, via the vagus nerve, and the sympathetic, via the celiac ganglion and splanchnic nerves. Of the two, the vagus is the most important regulator of gastric acid secretion. Vagal activation stimulates acid secretion. However, there are a few exceptions (see *In Vivo* Regulation of Gastric Acid Secretion: Cephalic Phase). The effects of sympathetic stimulation on acid secretion are generally inhibitory and indirect, resulting from effects on contractility and blood flow.

In general, sympathetic and parasympathetic fibers release either acetylcholine or norepinephrine. Fibers that release the neurotransmitter acetylcholine are cholinergic; those that release norepinephrine are adrenergic (from adrenalin). Preganglionic neurons in both the sympathetic and the parasympathetic nervous systems are exclusively cholinergic. Most postganglionic sympathetic neurons are adrenergic, but a few are cholinergic. In contrast to sympathetic fibers, most parasympathetic fibers, including the vagal fibers innervating the stomach, pass uninterrupted all the way

to the innervated organ and, thus, are preganglionic and cholinergic at the level of the stomach.

Vagal fibers that innervate the stomach originate in the medulla from neurons present in the dorsomotor nucleus of the vagus and, to a lesser extent, the nucleus ambiguus and the nucleus tractus solitarius. The posterior region of the stomach is innervated by the right thoracic vagal trunk, whereas the lesser curvature and ventral stomach are innervated by branches of the left vagus. Once vagal fibers enter the stomach, they synapse with ganglia of the enteric nervous system. Acetylcholine that is released by vagal fibers binds to nicotinic receptors present on the ganglionic neurons. Nicotinic receptors are stimulated by nicotine as well as by acetylcholine. These receptors are distinguished from muscarinic receptors by selective inhibitors such as atropine, which blocks muscarinic receptors but not nicotinic receptors. There are five known muscarinic subtypes, M_1–M_5. Recent studies of muscarinic receptor knockout mice suggest the muscarinic M_3 receptor subtype is involved in the regulation of basal and vagally stimulated acid secretion, but the M_1 receptor subtype is not required for vagally stimulated secretion, at least in this species.

The Enteric Nervous System

Similar to most regions of the gastrointestinal tract, the stomach contains an extensive array of interconnected autonomic ganglia and their associated nerve fibers. This intrinsic (enteric) nervous system is often referred to as the "second brain" based on its complexity and the presence of many of the same neurotransmitters that are found in brain. There are three distinct plexuses within the enteric nervous system: the submucosal (Meissner's), myenteric (Auerbach's), and serosal. The submucosal plexus is located within the submucosa just below the mucosal layer and the myenteric plexus is located between the outer longitudinal and inner circular muscle layers. Nerve fibers interconnect the dense network of ganglia within each plexus and also form interconnections between the different plexuses. In laboratory animals, myenteric ganglia are large and much more numerous than submucosal ganglia and are the major source of innervation for the gastric mucosa as well as the submucosa and smooth muscle. Myenteric neurons also appear to be most extensively innervated by vagal preganglionic fibers. In humans, the submucosal plexus in the antropyloric region appears to be more much more elaborate as compared to rats and guinea pigs. Serosal ganglia are not yet well characterized. In addition to acetylcholine, the intramural plexuses appear to contain at least 20 potential neurotransmitters. Many of the neurotransmitters present within these plexuses have also been localized within endocrine/paracrine cells in the gastric mucosa as well as in other regions of the gastrointestinal tract.

IN VIVO REGULATION OF GASTRIC ACID SECRETION

Between meals, the human stomach secretes acid at approximately 10% of the maximal rate. The volume of gastric juice is low, and the pH is generally below 2. After ingestion of a meal, gastric acid secretion increases dramatically. Classically, and rather artificially, three phases of meal-stimulated gastric acid secretion have been described based on selected experimental approaches. These include (1) the cephalic phase, which is activated by the sight, taste, smell, chewing, and swallowing of food; (2) the gastric phase, which is activated by the presence of digestion products in the gastric lumen and by the distension of the stomach; and (3) the intestinal phase, which is activated by the presence of chyme in the small intestine. These secretory phases, like other functions of the gastrointestinal tract, are, in reality, initiated more or less simultaneously following the ingestion of a meal. Secretion in all three phases is regulated by a complex process that involves the meal-induced activation of extrinsic and intrinsic neuronal pathways and release of hormones and paracrines.

In vivo, the major activators of HCl secretion are histamine, gastrin, and acetylcholine; however, many additional factors regulate the release of each of these secretagogues at several different levels. Known negative regulators include somatostatin, prostaglandins of the E2 series, epidermal growth factor (EGF), secretin, cholecystokinin (CCK), gastric inhibitory polypeptide (GIP), serotonin, neurotensin, interleukin-1, glucagon-like peptides, peptide YY, substance P, calcitonin gene-related peptide, bombesin, corticotropin releasing hormone (CRH), amylin, orexin A, and galanin. Other factors such as leptin, which is present in fundic glands, and ghrelin in the CNS, can stimulate HCl secretion. In most cases, the physiological role of these factors in regulating gastric acid secretion has not yet been established. Those factors that are best characterized are discussed in the following sections.

Cephalic Phase

Chemoreceptors and mechanoreceptors in the mouth, nasal cavities, and upper esophagus are activated by the sight, smell, and thought of food as well as by the chewing and swallowing reflexes. The resulting

afferent nerve impulses are relayed by vagal afferents mainly to the dorsal motor vagal nucleus in the central nervous system (CNS), which signals the stomach via vagal efferent pathways. Stress and hypoglycemia also activate the cephalic phase of secretion.

In humans, the cephalic phase accounts for less than half of the total acid secretory response to a meal. The magnitude of the response depends on how appetizing the food appears to the individual. Because vagal stimulation does not elicit the maximal possible rate of HCl secretion, it has been concluded that this response involves both stimulatory and inhibitory effects. A current view of the mechanisms associated with the dual effects of vagal activation is as follows. In the fundus and body of the stomach, vagal stimulation induces the release of acetylcholine by both preganglionic and postganglionic enteric neurons. The postganglionic release of acetylcholine stimulates parietal cells to secrete HCl. In the antrum, however, the preganglionic release of acetylcholine can induce the release of both stimulatory and inhibitory neurotransmitters by postganglionic neurons. As yet, these neurotransmitters have not been unequivocally identified. One candidate is gastrin-releasing peptide, which is structurally related to the amphibian neuropeptide, bombesin. GRP can release gastrin from G cells and somatostatin from D cells. Other transmitters may also be released by vagal stimulation. Thus, the maximal rate of acid secretion that can be achieved by vagal stimulation depends on the balance of neurotransmitters released in the vicinity of the G and D cells as well as parietal cells. Because, for example, vagal stimulation is a more potent activator of HCl secretion in dogs as compared to in humans, the inhibitory pathways may be more predominant in humans. The relative contribution of gastrin to the cephalic phase of secretion also appears to be species dependent.

The relationship between the CNS and the overall regulation of gastric acid secretion is quite complex. Thus far, most information has been derived from mapping studies in which various neurotransmitters are injected into different regions of the brain. Intracerebroventricular injection of thyrotropin-releasing hormone, a three-amino-acid peptide, activates the vagus nerve and mimics cephalic phase stimulation. Thus, TRH appears to play an important role in mediating vagal stimulation of acid secretion. This effect can be inhibited by the injection of peptide YY. Mapping studies in which various peptides are injected into regions of the brain known to send projections of vagal and sympathetic nerves involved in secretory inhibition indicate that the hypothalamus is an important site for central inhibition. Medullary and spinal sites also appear to be involved. Peptides that increase the activity of centrally acting inhibitory pathways in dogs and rats include bombesin, CGRP, ß-endorphin, neurotensin, prostaglandins, interleukin-1β, CRH, and orexin A. Neuropeptide Y has either stimulatory or inhibitory effects, depending on the site of injection.

Gastric Phase

As food enters the stomach, acid secretion is further stimulated by distension of the gastric wall as well as by the presence of food and certain beverages. Other factors, such as caffeine, alcohol, and luminal calcium also stimulate HCl secretion. Acid secretion in this phase accounts for at least 50% of the total and is primarily mediated by the hormone gastrin. Ingested nutrients, in particular peptides and amino acids, potently stimulate gastrin release, apparently by a direct effect on G cells, whereas distension induces the release of gastrin by both cholinergic and noncholinergic reflex mechanisms. Depending on the species, distension can account for 20–50% of the total acid secretory response. Distension reflexes may be local but can also involve the extrinsic nervous system. In the latter case, extramural or vago-vagal reflexes send signals via vagal afferents to the vagal nucleus, activating efferent vagal pathways from the CNS that then stimulate parietal cells in the oxyntic mucosa and G cells in the antrum. Intramural or local reflexes are mediated by one or more neurons within the enteric nervous system. These reflexes stimulate the parietal cell more effectively than they stimulate the G cell. All distension reflexes are blocked by the anticholinergic drug atropine, indicating that they are mediated at some point by the cholinergic neurotransmitter acetylcholine. Because vagotomy reduces, but does not completely abolish, gastrin release in response to pyloric distension, the gastrin response may be mediated by both vago-vagal and local, or pyloropyloric, reflexes.

After the ingestion of a meal, secreted acid is initially buffered by the presence of food and protein digestion products in the stomach. With continued HCl secretion, the pH drops and negative feedback mechanisms are called into play. Acid secretion is inhibited when the pH reaches 2 and is almost completely suppressed when the pH drops below 1. The pH-dependent inhibition of HCl secretion is correlated with the release of somatostatin. Distension of the stomach also releases somatostatin by activation of cholinergic and noncholinergic neuronal pathways. Antral somatostatin, which inhibits the release of gastrin from G cells, presumably serves as a counterbalance to the stimulatory effect of gastrin. Recent evidence suggests that atrial natriuretic peptide

(ANP) may also play a role in the regulation of gastric somatostatin secretion. ANP release can be induced by the neurotransmitter PACAP (which is present in the enteric nervous system in two different forms, PACAP-27 and -38) and suppressed by activation of intramural cholinergic neurons.

With gastrin, there is also a positive feedforward mechanism. In this case, a rise in luminal pH is correlated with increased gastrin secretion. This response occurs, for example, when acid secretion is inhibited by potent pharmacological agents such as omeprazole and related H^+,K^+-ATPase or proton pump inhibitors (PPIs), which can almost completely suppress parietal cell HCl secretion (see Receptor-Mediated Responses at the Cellular Level: Regulation of HCl Secretion at the Level of the Parietal Cell). A similar response occurs in atrophic gastritis, a condition in which parietal cell function is reduced. The prolonged elevation of serum gastrin is correlated with gastric mucosal hypertrophy and increased ECL numbers.

Intestinal Phase

The intestinal phase probably accounts for less than 5% of the total acid secretory response. It can be initiated by perfusion of the jejunum with amino acids and may involve the action of a putative factor called "enterooxyntin" as well as the hormone gastrin, which is released from G cells in the upper part of the small intestine. Once acidic chyme enters the duodenum, negative neural and hormonal feedback mechanisms are activated, leading to a reduction in HCl secretion. This serves to protect the small intestine from damage that could be caused by too much acid or excessive hyperosmolarity. The mechanisms responsible for negative feedback from the small intestine to the stomach are poorly characterized. The older literature refers to an inhibitory factor called "enterogastrone" that was thought to be released from the small intestine. Subsequently, three major inhibitors or enterogastrones were described in many textbooks: secretin, cholecystokinin, and gastric inhibitory peptide (GIP). These factors were thought to be released in response to the presence of fat, acid, and/or carbohydrates in the duodenum. Today, it is recognized that several inhibitory pathways, including neural, hormonal, and paracrine, are probably activated by the gastric emptying of acid, fat, and hyperosmolar solutions into the duodenum. The search for physiologically relevant inhibitory factors is ongoing.

Fat is the most potent inhibitor of acid secretion, suppressing both meal-stimulated gastric acid secretion and gastric emptying. Fat probably induces the release of several inhibitors of acid secretion, including CCK, which probably acts on CCK-1 or A-type receptors present on somatostatin-containing D cells in the gastric antrum. The hormone secretin, which is released from S cells in the upper small intestine in response to a pH $\leq$ 4, fatty acids, protein digestion products, and bile salts, potently inhibits both gastrin release and acid secretion. Secretin inhibition may be mediated by the release of both somatostatin and prostaglandins. Although the mechanism of inhibition is unclear, current evidence suggests that secretin is released from S cells following activation of a capsaicin-sensitive vagal afferent pathway. (Capsaicin is a neurotoxin that ablates C-type unmyelinated sensory nerve fibers. It is frequently used to study the afferent limbs of neural reflexes.) Because the neurotransmitter PACAP can also induce the release of secretin, somatostatin, and prostaglandins, it is possible that PACAP is released on activation of this pathway.

The systemic administration of peptide YY inhibits acid secretion, and circulating levels of PYY increase after a meal. However, recent evidence suggests that PYY is not a physiological mediator of this response. GIP and neurotensin are also released from the small intestine in response to fat. Because GIP is a very weak inhibitor in humans, it is not thought to be a major mediator of the inhibitory response. In contrast, neurotensin is a potent inhibitor and, thus, is a strong candidate enterogastrone. The colon also has the capacity to release acid inhibitory factors or colonogastrones. These factors have not yet been identified, although PYY is a candidate for this response.

RECEPTOR-MEDIATED RESPONSES AT THE CELLULAR LEVEL

Cellular and molecular studies with isolated cells have provided a wealth of information on the specific receptors and receptor subtypes present on the parietal cell as well as on G cells, D cells, and ECL cells. An overview of the current understanding of this area of research is presented in Fig. 2.

Regulation of HCl Secretion at the Level of the Parietal Cell

Parietal cells express histamine H2-type receptors, muscarinic M_3 receptors, and gastrin/CCK-2 or B-type receptors. Although gastrin plays a central role in the stimulation of gastric acid secretion, this hormone has minimal effects on the acid secretory response when administered to parietal cells in isolation. In contrast, histamine is a potent acid secretory agonist that induces

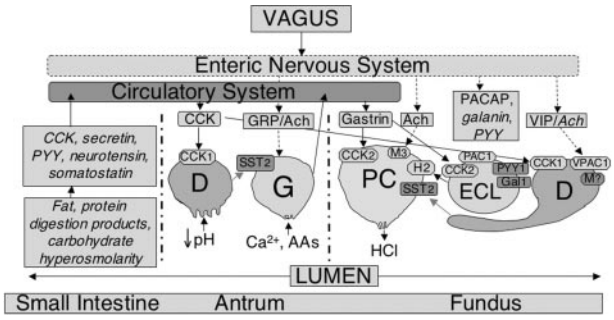


FIGURE 2 Overview of mechanisms of receptor-mediated secretory responses involving D, G, and parietal cells (PC) in the gastric mucosa and small intestine. CCK, Cholecystokinin; GRP, gastrin-releasing peptide; Ach, acetylcholine; PACAP, pituitary adenylate cyclase-activating peptide; PYY, peptide YY; VIP, vasoactive intestinal peptide; M_3, muscarinic type 3; H2, histamine-2.

a sustained acid secretory response and potentiates the secretory responses to gastrin and acetylcholine at the level of the parietal cell. When histamine binds to H2-type receptors on the parietal cell, the enzyme adenylyl cyclase is activated, leading to an increase in intracellular cyclic adenosine monophosphate (cAMP) concentrations. Histamine can also elevate intracellular calcium ($[Ca^{2+}]_i$) in parietal cells; however, the acid secretory response to this agonist appears to be mediated exclusively by the cAMP signaling pathway. The elevation of cAMP is correlated with dramatic morphological transformations in which the internalized, apically directed canalicular membrane expands and elongated microfilaments appear in the canalicular lumen. These morphological changes are accompanied by a substantial increase in oxygen consumption, which is required for the generation of protons. Proton secretion occurs following the activation of the proton "pump," or H^+,K^+-ATPase, which appears to depend on the translocation from an internal membrane complex, the tubulovesicular or tubulocisternal compartment, to the canalicular membrane. When the secretory stimulus is removed, oxygen consumption falls and the H^+,K^+-ATPase is returned to the tubulovesicular compartment by a poorly understood membrane retrieval process.

The magnitude of the acid secretory response to acetylcholine and analogues such as carbachol varies among species, but is generally transient when the effects of histamine are pharmacologically blocked by H2 receptor antagonists such as cimetidine (Tagamet) and ranitidine (Zantac). Compared to the H2 blockers, PPIs such as omeprazole are more potent and specific inhibitors of HCl secretion. This happens because (1) PPIs bind directly and irreversibly to the H^+,K^+-ATPase only in the highly acidic environment of the stomach and (2) H^+,K^+-ATPase activation is a common, terminal step in the acid secretory response.

In contrast to histamine, a rise in $[Ca^{2+}]_i$ is necessary to elicit an acid secretory response to acetylcholine. Acetylcholine and analogues such as carbachol elevate $[Ca^{2+}]_i$ by an inositol trisphosphate (IP_3)-dependent release of calcium from intracellular stores plus the influx of extracellular calcium through unidentified, nonexcitable calcium channels. IP_3 is produced along with diacylglycerol (DAG) on activation of the enzyme phospholipase C-β. DAG activates one or more protein kinase C isoforms. Because the general activation of protein kinase C (PKC) with phorbol ester inhibits the acid secretory response to histamine and acetylcholine, PKC may activate an inhibitory pathway. Such a dual mechanism may help to explain the transient nature of the cholinergic response. The potentiating effect of histamine on the cholinergic response can also be demonstrated indirectly in genetically engineered histidine decarboxylase-deficient mice in which carbachol stimulates only a weak and transient acid secretory response. (In mammals, histidine decarboxylase is an enzyme that is necessary for histamine synthesis.)

Gastrin also induces a rise in $[Ca^{2+}]_i$ but has little to no effect on the acid secretory response at the level of the parietal cell when the effects of endogenous histamine are blocked. Other endogenous inhibitors, such as prostaglandins of the E series, somatostatin, and EGF/transforming growth factor-α (TGF-α) partially inhibit histamine-stimulated acid secretion in isolated parietal cells and reduce the cAMP response to histamine. Thus, these inhibitors can regulate the acid secretory response by acting directly at the level of the parietal cell. They can also inhibit parietal cell HCl secretion indirectly by actions on other mucosal cell types (see later).

Regulation of Histamine Release from ECL Cells

Since it was discovered that ECL cells are distinct from classical mast cells, it has become clear that ECL cells play a central role in the regulation of gastric acid secretion. Thus far, experiments with isolated ECL cells have been performed mainly in rat and, to a lesser extent, rabbit models. Results to date indicate that ECL cells from rats secrete histamine in response to stimulation by gastrin (CCK-2 receptor subtype) and PACAP (PAC1 receptor subtype), but not by acetylcholine. PYY and the neurotransmitter galanin inhibit ECL cell histamine release by binding to PYY type 1 and galanin type 1 receptors, respectively. Somatostatin inhibits

histamine release by ECL cells by binding to a type 2 somatostatin (SST2) receptor; prostaglandin E2 inhibits by binding to an EP3-type receptor. Although both CGRP and γ-aminobutyric acid (GABA) stimulate histamine release in the intact stomach, they have no direct effect on ECL cells *in vitro*. Thus, these effectors must act indirectly to stimulate histamine release.

In contrast to its weak effect on the parietal cell, gastrin is a potent stimulant of ECL cell secretion. The differences in response patterns in these two cell types may be the result of gastrin-induced activation of mutually independent signaling pathways. For example, gastrin induces a significant rise in intracellular calcium concentrations and activates the Ras−mitogen-activated protein (MAP) kinase signaling pathway in ECL cells. In contrast, gastrin does not activate the Ras−MAP kinase pathway in isolated parietal cells and elevates $[Ca^{2+}]_i$ in only ∼one-third of these cells. Another important difference between parietal and ECL cells that may or may not be relevant to their differential responses to gastrin is that, unlike parietal cells, ECL cells possess excitable, voltage-dependent calcium channels. These differences in calcium channels may explain why the *in vivo* administration of classical calcium channel blockers partially inhibits acid secretion, although these inhibitors have no specific effect on isolated parietal cells.

The recent finding that PACAP induces histamine release from ECL cells conflicts with the *in vivo* observation that peripherally injected PACAP inhibits gastrin-stimulated acid secretion. This conflict appears to have been resolved by the findings that PACAP can also stimulate the release of inhibitors of gastric acid secretion, including somatostatin, PGE2, and secretin. Thus, PACAP probably acts on multiple cell types, with the local release of PACAP serving to sway secretion toward a stimulated or inhibited state. The observations that ECL cells are not stimulated *in vitro* by the muscarinic cholinergic agonist, acetylcholine, also conflict with findings that the *in vivo* administration of M_1 muscarinic receptor antagonists suppresses histamine release. The release of PACAP from nerve terminals impinging on ECL cells in response to cholinergic activation of the PACAP-containing neurons may explain these conflicting findings. However, it is more difficult to reconcile the absence of cholinergic receptors on ECL cells with earlier observations that the cholinergic agonist, carbachol, induces histamine release in isolated gastric gland preparations. As yet, it is not clear whether there are species differences in the cholinergic response, or whether the cholinergic release of histamine is mediated indirectly through the release of an unknown intermediate.

Synthesis and Release of Gastrin from G Cells and Gastrin Functions

The concept that gastrin regulates acid secretion at multiple levels is gaining acceptance. As discussed in the preceding section, gastrin acutely stimulates the release of histamine from ECL cells. However, from the early work of Leonard Johnson and others, it has been recognized for over 30 years that gastrin also regulates gastric mucosal growth. Disruption of gastrin gene expression in mice by targeted deletion suppresses the maturation of parietal cells and reduces their numbers. In contrast, overexpression of the gastrin gene is associated with an increased proliferation of the gastric epithelium that is accompanied by an increase in the number of parietal cells and an increase in secretion. This initial proliferative response is followed by a progressive decline in the number of parietal cells and acid secretion as the mice age. Because parietal cells do not undergo cell division, and the proliferating stem cells may or may not express gastrin receptors, the effects of gastrin on cell proliferation may be mediated either directly via a gastrin receptor or indirectly through the release of growth factors such as TGF-α, amphiregulin, and heparin-binding EGF, possibly from parietal cells. Gastrin may also release Reg-1-α from chief cells.

There are many different forms of gastrin. The major secretory product of G cells is amidated gastrin, which is one of several different peptides formed from progastrin by posttranslational modification. Although there are species differences in G cell processing of progastrin, the major initial cleavage product is G34-Gly, which is converted to G34 (34 amino acids). After the Gly-gastrins are produced, they are amidated by peptidyl-α-amidating monooxygenase. Cleavage of G34-Gly and G34 generates the respective G17 products (17 amino acids). In humans, G17 is the predominant product. Amidated gastrins appear to be to major stimulants of HCl secretion. The Gly-gastrins do not stimulate acid secretion directly; however, studies with gastrin knockout mice suggests that these forms of gastrin may serve to regulate the capacity of the parietal cell to respond to stimulation.

Studies with acutely isolated and cultured G cells have identified several potential regulatory pathways. Both GRP and the amphibian neurotransmitter bombesin (which has structural similarity to GRP) release gastrin. Acetylcholine also releases gastrin by binding to M_3-type muscarinic receptors. Luminal factors, including calcium and the aromatic amino acids phenylalanine, tryptophan, and tyrosine, may stimulate gastrin release by a direct action on luminal G cell receptors.

Until recently, the consensus was that the main stimulatory pathway for gastrin release was mediated by GRP that was released near G cells in response to vagal activation. However, recent studies with a GRP receptor antagonist have raised some disturbing questions. In essence, the GRP inhibitor was found to suppress acid secretion that was induced by test meals and sham feeding in human subjects, but was not found to inhibit the rise in plasma gastrin levels. In contrast, when GRP was infused, plasma gastrin levels did rise, and the GRP inhibitor blocked this response. These results suggest that although GRP is involved in mediating vagally stimulated HCl secretion, it is not the physiological mediator of vagally stimulated gastrin release. If these findings are validated, a new search will be necessary to identify the neurotransmitter involved in this response.

As in ECL cells, activation of SST2 receptors inhibits gastrin release from G cells. A low pH inhibits gastrin release whereas an elevated pH stimulates release. The pH effect appears to be indirect, possibly mediated by capsaicin-sensitive afferent neurons.

Regulation of Somatostatin Release from D Cells

Somatostatin released from D cells in both the oxyntic and the antral mucosa negatively regulates acid secretion. The inhibitory effect is tonic, and somatostatin can act in both a paracrine and endocrine fashion to suppresses the release of gastrin from G cells and histamine from ECL cells. Interestingly, uroguanylin, which is a peptide that appears to play a role in regulating water and electrolyte balance, has also been detected in gastric D cells. Thus, this peptide may also play a role in regulating acid secretion.

Although fundic D cells can currently be enriched to $\sim$90%, D cells in general are not yet as well characterized as ECL cells and parietal cells. The D cells in the fundus and antrum have different morphologies and differing responses to acidity and food, and, thus, may be differentially regulated. There may also be species differences associated with D cell regulation. To date, most data suggest that both fundic and antral D cells possess CCK-1 but not CCK-2 type receptors, the latter of which are present on parietal cells. In terms of the *in vivo* response, this appears to be logical because gastrin is a potent activator of the CCK-2, but not the CCK-1, receptor subtype, whereas the reverse is true for sulfated CCK. The presence of the CCK-1 receptor subtype on D cells provides an explanation for the mechanism through which CCK released from the intestine inhibits acid secretion. Thus, CCK stimulates somatostatin release from D cells via an action on CCK-1 receptors and gastrin stimulates acid secretion by acting on CCK-2 receptors present on both parietal and ECL cells.

Fundic D cells may also possess receptors for secretin, epinephrine, VIP, GRP and/or CGRP and the structurally similar peptide amylin, and acetylcholine. With the exception of acetylcholine, all of these agents stimulate the release of somatostatin in isolated D cells. Amylin may be released from D cells to act on the cells in an autocrine fashion. CGRP also stimulates somatostatin synthesis, which can indirectly regulate gastrin release. Acetylcholine appears to inhibit somatostatin release by acting through the heterotrimeric G protein, G_i.

EPILOGUE

From this brief discussion, it is apparent that the stomach is a highly complex organ that is controlled at multiple levels, ranging from the CNS, to the intestines, to local neuronal and cellular interactions. Over the past 20 years or so, work with isolated enriched populations of different gastric mucosal cell types has provided considerable information on the specific functions of these cells. The use of transgenic and knockout mice models has begun to open new avenues of exploration. The role that species differences will play in defining the mechanisms controlling gastric acid secretion in these new models is a question for the future.

See Also the Following Articles

Calcitonin Gene-Related Peptide (CGRP) • Cholecystokinin (CCK) • Enteric Nervous System • Enterochromaffin-like (ECL) Cells • Gastric H^+,K^+-ATPase • Gastrin • Gastrin-Releasing Peptide (GRP) • Histamine • Neurotensin • Pancreatic Polypeptide Family • Parietal Cells • Pituitary Adenylate Cyclase Activating Peptide (PACAP) • Proton Pump Inhibitors • Serotonin • Somatostatin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Bordi, C., D'Adda, T., Azzoni, C., and Ferraro, G. (2000). Classification of gastric endocrine cells at the light and electron microscopical levels. *Microsc. Res. Tech.* **48**, 258–271.

Chew, C. S. (1993). Peptidergic regulation of gastric acid secretion. *In* "Handbook of Experimental Pharmacology. Gastrointestinal Regulatory Peptides" (D. R. Brown, ed.), Vol. 106, pp. 199–252. Springer-Verlag, Berlin and Heidelberg.

Davenport, H. W. (ed.) (1992). "A History of Gastric Secretion and Digestion: Experimental Studies to 1975." Oxford University Press, New York.

Dockray, G. J., Varro, A., Dimaline, R., and Wang, T. (2001). The gastrins: Their production and biological activities. *Annu. Rev. Physiol.* **63**, 119–139.

Forte, J. G., and Soll, A. (1989). Cell biology of hydrochloric acid secretion. *In* "Handbook of Physiology. Section 6: The Gastrointestinal System" (J. G. Forte, ed.), pp. 207–228. Oxford University Press, New York.

Hersey, S. J., and Sachs, G. (1995). Gastric acid secretion. *Physiol. Rev.* **75**, 155–189.

Hildebrand, P., Lehmann, F. S., Ketterer, S., Christ, A. D., Stingelin, T., Beltinger, J., Gibbons, A. H., Coy, D. H., Calam, J., Larsen, F., and Beglinger, C. (2001). Regulation of gastric function by endogenous gastrin releasing peptide in humans: Studies with a specific gastrin releasing peptide receptor antagonist. *Gut* **49**, 23–28.

Hinkle, K. L., and Samuelson, L. C. (1999). Lessons from genetically engineered animal models. III. Lessons learned from gastrin gene deletion in mice. *Am. J. Physiol.* **277**, G500–G505.

Hirshowitz, B. I. (1989). Neural and hormonal control of gastric acid secretion. *In* "Handbook of Physiology. Section 6: The Gastrointestinal System" (J. G. Forte, ed.), pp. 127–157. Oxford University Press, New York.

Lindström, E., Chen, D., Norlen, P., Andersson, K., and Híkanson, R. (2001). Control of gastric acid secretion: The gastrin–ECL cell–parietal cell axis. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **128**, 505–514.

Lloyd, K., Kent, C., and Debas, H. (1994). Peripheral regulation of gastric acid secretion. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., Vol. 2, pp. 1185–1226. Raven Press, New York.

Prinz, C., Zanner, R., Gerhard, M., Mahr, S., Neumayer, N., Hohne-Zell, B., and Gratzl, M. (1999). The mechanism of histamine secretion from gastric enterochromaffin-like cells. *Am. J. Physiol.* **277**, C845–C855.

Walsh, J. H. (1994). Gastrointestinal hormones. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1–128. Raven Press, New York.

Gastric Cancer Surveillance

EMMY LUDWIG AND PETER H. R. GREEN
Columbia University College of Physicians and Surgeons

dysplasia An unequivocally neoplastic proliferation of epithelium.

early gastric cancer Cancer involving the mucosa or submucosa (± lymph node involvement).

metaplasia Conversion of tissue into a form that is not normally present in that area.

Gastric cancer is the second most common cause of cancer-related deaths worldwide. Most cases present at advanced stages. Surgical resection remains the only curative therapy. The goal in reducing mortality from gastric cancer is via diagnosis at a treatable stage. The >90% 5-year survival for early cancer and the dismal prognosis for advanced cancer (<10% 5-year survival) make surveillance a necessary mission. The aim is to identify dysplasia and early cancer. In Japan, a decline in the mortality from gastric cancer in those that have undergone screening has confirmed the value of screening in geographic areas of very high incidence.

RECOMMENDATIONS

Currently there is no role for screening of the general population for gastric cancer in the United States and other developed countries, apart from areas with an exceptional rate. There are, however, subpopulations at increased risk.

Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) carries an increased risk of upper gastrointestinal cancer, accounting for most of the mortality in patients who have undergone prophylactic total colectomy. These malignancies arise in adenomas of the duodenum and stomach. Most gastric polyps in FAP patients are of the fundic gland type, a subset of polyp without malignant potential in non-FAP patients. Adenomas occur mainly in the antrum. All polyps in FAP should be biopsied and larger polyps should be removed. All FAP patients should be surveyed endoscopically and those with a large number of polyps and proven dysplasia should be followed more often.

Helicobacter pylori

Many studies have found an association between *Helicobacter pylori* positivity, the development of

 Copyright 2004, Elsevier (USA). All rights reserved.

Forte, J. G., and Soll, A. (1989). Cell biology of hydrochloric acid secretion. *In* "Handbook of Physiology. Section 6: The Gastrointestinal System" (J. G. Forte, ed.), pp. 207–228. Oxford University Press, New York.

Hersey, S. J., and Sachs, G. (1995). Gastric acid secretion. *Physiol. Rev.* **75**, 155–189.

Hildebrand, P., Lehmann, F. S., Ketterer, S., Christ, A. D., Stingelin, T., Beltinger, J., Gibbons, A. H., Coy, D. H., Calam, J., Larsen, F., and Beglinger, C. (2001). Regulation of gastric function by endogenous gastrin releasing peptide in humans: Studies with a specific gastrin releasing peptide receptor antagonist. *Gut* **49**, 23–28.

Hinkle, K. L., and Samuelson, L. C. (1999). Lessons from genetically engineered animal models. III. Lessons learned from gastrin gene deletion in mice. *Am. J. Physiol.* **277**, G500–G505.

Hirshowitz, B. I. (1989). Neural and hormonal control of gastric acid secretion. *In* "Handbook of Physiology. Section 6: The Gastrointestinal System" (J. G. Forte, ed.), pp. 127–157. Oxford University Press, New York.

Lindström, E., Chen, D., Norlen, P., Andersson, K., and Híkanson, R. (2001). Control of gastric acid secretion: The gastrin–ECL cell–parietal cell axis. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **128**, 505–514.

Lloyd, K., Kent, C., and Debas, H. (1994). Peripheral regulation of gastric acid secretion. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 2nd Ed., Vol. 2, pp. 1185–1226. Raven Press, New York.

Prinz, C., Zanner, R., Gerhard, M., Mahr, S., Neumayer, N., Hohne-Zell, B., and Gratzl, M. (1999). The mechanism of histamine secretion from gastric enterochromaffin-like cells. *Am. J. Physiol.* **277**, C845–C855.

Walsh, J. H. (1994). Gastrointestinal hormones. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1–128. Raven Press, New York.

Gastric Cancer Surveillance

EMMY LUDWIG AND PETER H. R. GREEN
Columbia University College of Physicians and Surgeons

dysplasia An unequivocally neoplastic proliferation of epithelium.

early gastric cancer Cancer involving the mucosa or submucosa (± lymph node involvement).

metaplasia Conversion of tissue into a form that is not normally present in that area.

Gastric cancer is the second most common cause of cancer-related deaths worldwide. Most cases present at advanced stages. Surgical resection remains the only curative therapy. The goal in reducing mortality from gastric cancer is via diagnosis at a treatable stage. The >90% 5-year survival for early cancer and the dismal prognosis for advanced cancer (<10% 5-year survival) make surveillance a necessary mission. The aim is to identify dysplasia and early cancer. In Japan, a decline in the mortality from gastric cancer in those that have undergone screening has confirmed the value of screening in geographic areas of very high incidence.

RECOMMENDATIONS

Currently there is no role for screening of the general population for gastric cancer in the United States and other developed countries, apart from areas with an exceptional rate. There are, however, subpopulations at increased risk.

Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) carries an increased risk of upper gastrointestinal cancer, accounting for most of the mortality in patients who have undergone prophylactic total colectomy. These malignancies arise in adenomas of the duodenum and stomach. Most gastric polyps in FAP patients are of the fundic gland type, a subset of polyp without malignant potential in non-FAP patients. Adenomas occur mainly in the antrum. All polyps in FAP should be biopsied and larger polyps should be removed. All FAP patients should be surveyed endoscopically and those with a large number of polyps and proven dysplasia should be followed more often.

Helicobacter pylori

Many studies have found an association between *Helicobacter pylori* positivity, the development of

 Copyright 2004, Elsevier (USA). All rights reserved.

atrophic gastritis, and risk for the development of gastric cancer. Although *H. pylori* should be eradicated in patients with peptic ulcer disease, the extremely high worldwide infection rate (>50%) makes endoscopic surveillance for gastric cancer impractical.

Gastric Polyps in Non-FAP Patients

Adenomatous gastric polyps are unequivocally neoplastic and have an increased incidence of carcinoma. Polyps with the greatest malignant potential are larger, are sessile, or have a villous component. Adenomas should be removed and the patient followed with surveillance endoscopy every 1 to 3 years to detect recurrent polyps or interval development of a carcinoma. Nonpolypoid mucosa should be biopsied to detect dysplasia and the degree of risk for subsequent malignancy. Flat adenomas are difficult to detect but are more likely to be dysplastic.

Postgastrectomy ("Gastric Stump Carcinoma")

Gastric carcinoma is a late complication of partial gastrectomy for benign ulcer disease. The relative risk of gastric cancer mortality in these patients is small and in some studies is associated with other risk factors such as smoking. Screening of this population is not recommended.

Chronic Atrophic Gastritis with Intestinal Metaplasia

Gastric cancer usually arises in a background of chronic atrophic gastritis. Population-based studies indicate that the risk of gastric cancer is increased in those with chronic atrophic gastritis, intestinal metaplasia (especially type III), and pernicious anemia. The overall risk is, however, small compared to the general population. For this reason, periodic endoscopic surveillance is not recommended. Pathology laboratories do not routinely type intestinal metaplasia. As a result, those with intestinal metaplasia at greatest risk for progressing to carcinoma cannot be identified.

FUTURE DIRECTIONS

The most effective approach to decreasing mortality from gastric cancer is early detection. The identification of genetic markers in high-risk families will facilitate screening. A major reduction in gastric cancer mortality on a worldwide basis will, however, require changes in nutrition and *H. pylori* eradication.

See Also the Following Articles

Atrophic Gastritis • Cancer, Overview • Familial Adenomatous Polyposis (FAP) • Gastrectomy • Gastric Polyps • *Helicobacter Pylori* • Stomach, Adenomas and Carcinomas of the

Further Reading

American Society for Gastrointestinal Endoscopy. (1998). The role of endoscopy in the surveillance of premalignant conditions of the upper gastrointestinal tract: Guidelines for clinical application. *Gastrointest. Endosc.* **34**, 18S–20S.

Axon, A. T. R., Boyle, P., Riddell, R. H., Grandjouan, S., Hardcastle, J., and Yoshida, S. (1994). Summary of a working party on the surveillance of premalignant lesions. *Am. J. Gastroenteol.* **89**, S160–S168.

Green, P. H. R., O'Toole, K. M., Slonim, D., Wang, T., and Weq, A. (1988). Increasing incidence and excellent survival of patients with early gastric cancer: Experience in a United States medical center. *Am. J. Med.* **85**, 658–661.

Parsonnet, J., Friedman, G. D., Vandersteen, D. P., *et al.* (1991). *Helicobacter pylori* infection and the risk of gastric carcinoma. *N. Engl. J. Med.* **325**, 1127–1131.

Gastric Emptying

DEIRDRE O'DONOVAN, CHRISTINE FEINLE-BISSET, KAREN JONES, AND MICHAEL HOROWITZ
University of Adelaide and Royal Adelaide Hospital, South Australia

chyme Mass of partially digested food that moves from the stomach to the duodenum.
gastroparesis Stasis of food in the stomach in the absence of mechanical obstruction.

Gastric emptying is a complex process designed to deliver chyme into the small intestine at a rate that optimizes digestion and absorption. This process is controlled by complex neurohormonal mechanisms, the most important of which result from the interaction of nutrients with small intestinal receptors. Historically, disordered gastric emptying, especially delayed gastric emptying or gastroparesis, was considered a rare disorder. However, advances in radioisotopic gastric emptying techniques have established that it is a common clinical problem, especially among patients with diabetes mellitus.

INTRODUCTION

This article reviews the motor mechanisms involved in gastric emptying, patterns of gastric emptying for both liquids and solids, factors that regulate gastric emptying, methods that are available to quantify gastric emptying, and approaches to the investigation and treatment of disordered gastric emptying.

PATTERNS AND DETERMINANTS OF GASTRIC EMPTYING

Cannon, in 1911, was the first to describe in animal studies that transpyloric flow was pulsatile rather than continuous. Thus, most liquefied chyme is delivered into the duodenum in a series of small gushes rather than on a continuous basis. The mechanical determinants of individual flow episodes are poorly defined; however, recent advances in techniques to quantify gastric emptying have provided an increased understanding of emptying patterns on a second-by-second basis. With these advances in technology, it is now clear that transpyloric flow is bidirectional, so that both antegrade and retrograde flow occurs.

Meal Composition and Volume

Overall patterns of gastric emptying are dependent on several characteristics of the ingested material, including nutritive and physical properties, so that solids, nutrient liquids, and nonnutrient liquids empty from the stomach at different rates. A fundamental property of the stomach is its ability to discriminate between liquid and solid meal components. Although liquid emptying occurs with relative mechanical ease, solid emptying is dependent on an initial process of "trituration," or grinding, of solid particles in the antrum before emptying proceeds.

Liquids

Liquid emptying is critically dependent on volume ingested as well as on osmolarity and nutrient content (Fig. 1). Nonnutrient and low-nutrient liquids empty relatively rapidly from the stomach, with an overall monoexponential pattern; the volume of liquid that is emptied into the duodenum in a given time period is proportional to the volume left in the stomach. Consequently, the rate of emptying is influenced by intragastric volume; posture is also important. In contrast, high-nutrient liquids are retained in the distal stomach for longer periods and empty more slowly as a result of feedback from small intestinal nutrient receptors. Posture and intragastric volume contribute minimally in the latter case. Gastric emptying of a nutrient liquid consists of an initial rapid phase, followed by a slower phase during which 2–3 kcal/minute will be delivered into the duodenum, essentially irrespective of the source of those calories. In addition to caloric density, characteristics of the nutrient may be important in the rate of liquid emptying—e.g., fructose is a less potent inhibitor of gastric emptying compared to an isocaloric load of glucose.

Solids

The emptying of digestible solids from the stomach proceeds at a much slower pace compared to nutrient and non-nutrient liquid emptying. The overall emptying rate of digestible solids is characterized by a

 Copyright 2004, Elsevier (USA). All rights reserved.

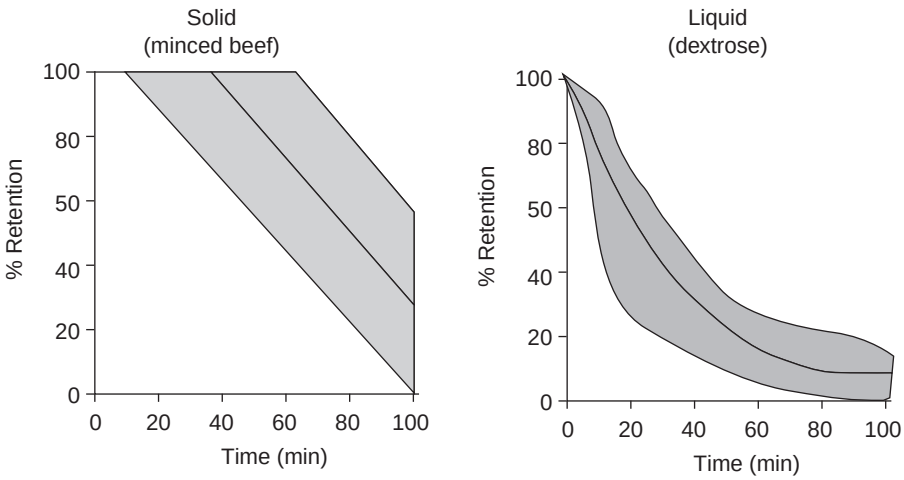


FIGURE 1 Gastric emptying for a meal consisting of both a solid (100 g of minced beef) and a liquid (150 ml of 10% dextrose), consumed in the sitting position, measured radioisotopically in a normal subject. The normal range (mean ± 2 SD) is shown in the shaded areas.

so-called lag phase during which little or no emptying occurs. The lag phase, which lasts from 20 to 60 minutes, is accounted for by an initial retention of the solid in the proximal stomach, followed by redistribution to the antrum, where, in most cases, solid particles undergo trituration before any emptying occurs. If the viscosity of the meal is increased sufficiently, the ability of the stomach to discriminate between large and small particles is impaired and much larger particles may be delivered into the duodenum. Following the lag phase, an emptying phase occurs that approximates a linear pattern with a relatively constant rate of emptying for at least the first 2–3 postprandial hours. The rate of emptying is to some extent load dependent; larger solid meals are delivered slightly more rapidly into the small intestine when compared with smaller meals. Larger indigestible solids have been thought not to empty until the return of migrating motor complex, during phase III, but a recent study has shown that they may empty earlier than this and during the postprandial period. The amount of liquid consumed with the solid can affect the rate of solid emptying. In a mixed meal of solid and liquids, up to 80% of the liquid phase empties before the solid, suggesting that the stomach is capable of discriminating between solids and liquids when presented simultaneously. Furthermore, the presence of liquid in the stomach, particularly if the liquid contains nutrient, prolongs the emptying of solids.

Fat

Foods high in fat are handled differently by the stomach and, therefore, are considered separately from liquids and solids. Fat represents a challenge to the stomach, because it is liquid at room temperature and coalesces to form large globules in the stomach. In the erect posture, fat is retained in the uppermost part of the stomach, floating on other gastric contents, and it is therefore not surprising that the gastric emptying of fat may be dependent on posture; due to the high nutrient density of fat, it can markedly delay the emptying of a low-nutrient liquid.

Small Intestinal Feedback

The major factor regulating gastric emptying of nutrients is feedback inhibition; this is triggered by receptors that are distributed throughout the small intestine and is mediated by both neural and hormonal mechanisms. The extent of small intestinal feedback is dependent on both the number and the site of the small intestinal receptors that are exposed, and is influenced by patterns of prior nutrient exposure; ingestion of a diet high in carbohydrate or fat accelerates subsequent emptying of glucose or fat. Fat and other nutrient-mediated slowing of gastric emptying is dependent on digestion taking place, e.g., hydrolysis of triglycerides to fatty acids, and therefore their effects on gastric emptying cannot proceed in the setting of impaired digestion, such as pancreatic exocrine insufficiency (Fig. 2). Cecoileal reflux of short-chain fatty acids and the presence of fat in the ileum, the so-called ileal brake, also contribute to the regulation of gastric emptying. Other factors that influence gastric emptying via a duodenal feedback loop include duodenal acidification and distension.

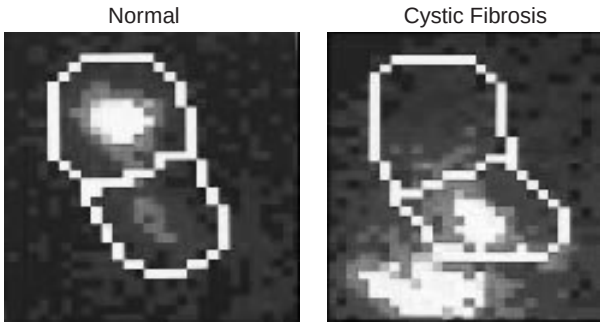


FIGURE 2 Scintigraphic images in a normal volunteer and a patient with cystic fibrosis, in the sitting position, showing the oil phase of a labeled liquid meal (60 ml of olive oil and 290 ml of low-nutrient beef consommé soup) 60 minutes postingestion. The total stomach region of interest is divided into proximal and distal regions. In the normal subject, most of the oil is in the proximal stomach, whereas in the cystic fibrosis patient, most of the oil has emptied from the stomach and most of the remainder is in the distal stomach.

Extrinsic Factors

A number of external factors influence the rate of gastric emptying in humans, including emotional state (anger, pain, and nausea may all slow gastric emptying). Chronic smoking is also associated with slower gastric emptying. Male gender and younger age are associated with slightly more rapid gastric emptying.

MEASUREMENT OF GASTRIC EMPTYING

During the past 25 years, the development and application of new techniques to quantify gastric emptying have greatly increased understanding of diagnosis and treatment of gastric emptying disorders in humans.

Scintigraphy

Scintigraphy remains the gold standard for the quantitative assessment of gastric emptying; it is relatively easy to perform and is noninvasive. The radiation dose approximates that received from a single abdominal radiograph. The technique involves the ingestion of a meal in which one (solid phase) or preferably two (solid and liquid phases) meal components have been labeled with a radiopharmaceutical, commonly gallium-67-labeled ethylenediaminetetra acetic acid (^{67}Ga-labeled EDTA) or indium-111-labeled diethylenetriaminepentaacetic acid (^{111}In-labeled DTPA) for the liquid phase and sulfur colloid labeled with technetium-99m (^{99m}Tc) for the solid phase (Fig. 3). An external gamma camera quantifies the disappearance of each isotope from the "region of interest," and from these data various parameters can be derived to describe gastric emptying. For both solids and liquids, the time taken for half of the meal to empty ($T_{1/2}$) is a frequently used parameter. In addition, intragastric meal distribution can also be assessed. Unfortunately, there is a lack of standardization of scintigraphic techniques, with substantial variation between different centers, particularly in relation to the volume and composition of the test meal, the posture of the subject during the gastric emptying measurement, the duration and mode of data acquisition, correction factors, and the calculation of gastric emptying rates. This renders comparisons between studies performed in different centers difficult, although there have been recent attempts to rectify this situation.

Carbon Breath Tests

Carbon breath tests (e.g., [^{13}C]octanoic acid) have recently been used to quantify solid and/or liquid gastric emptying. The [^{13}C]octanoic acid breath test involves ingestion of a test meal containing labeled octanoic acid, which, after emptying from the stomach, is absorbed in the small intestine and oxidized to CO_2, which is excreted by the lungs. Although these tests are

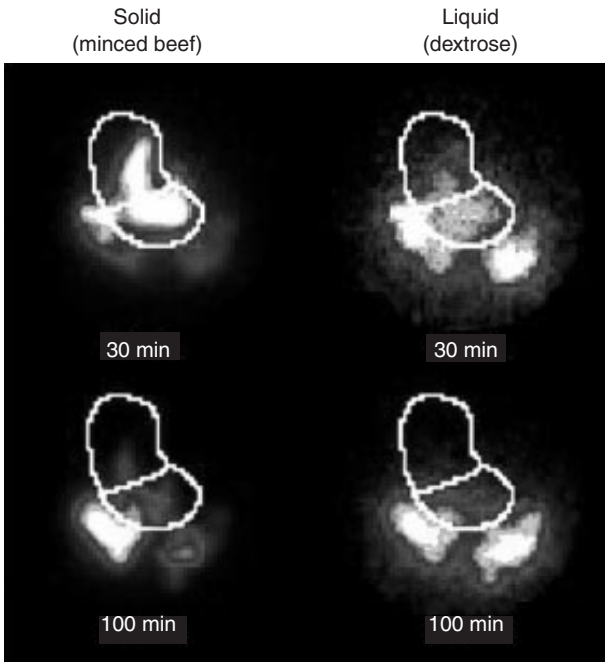


FIGURE 3 Scintigraphic images showing abdominal distribution of radioactivity at 30 and 100 minutes following a meal of 100 g of minced beef tagged with ^{99m}Tc-labeled sulfur colloid and 150 ml of 10% dextrose tagged with ^{67}Ga-labeled EDTA in a healthy volunteer. Total, proximal, and distal stomach regions of interest are illustrated.

cheaper and simpler than scintigraphy and, with the use of stable isotopes, avoid the use of irradiation, there is considerable debate as to the appropriate method of data analysis. Additional validation of these methods in patients with disordered gastric emptying, particularly those in whom gastric emptying is markedly delayed, is required before their use can be advocated. However, it seems likely that carbon breath tests will prove to be useful as a screening test for gastroparesis and in large epidemiological studies.

Ultrasound

High-resolution real-time ultrasound equipment has been used to measure gastric emptying and has the advantage of being noninvasive and does not involve radiation. Hence, it can potentially be repeated on many occasions. The development of Doppler techniques and three-dimensional ultrasound has also allowed more accurate assessment of transpyloric flow and intragastric volumes. The widespread application of ultrasound is limited by the high level of expertise required to perform and interpret the studies and the difficulty of performing the study in obese subjects or in the presence of excessive bowel gas. Ultrasound is, therefore, not recommended at present as a useful alternative to scintigraphy.

Radiological Measurement

Contrast studies with liquid barium sulfate have limited usefulness in the assessment of gastric emptying because of their nonphysiological nature, the high radiation exposure, and the inability to measure fractional stomach emptying. An abdominal X ray taken 6 hours after ingestion of radio-opaque markers has, however, been reported as a sensitive technique for assessing gastric emptying of nondigestible solids. This method probably assesses whether phase III of the interdigestive myoelectric complex is present in the stomach.

Electrical Impedance

Changes in electric impedance can be used to measure the volume of liquid meals remaining in the stomach. A large-volume liquid meal is administered and the impedance recorded at regular intervals. The values obtained over time represent the emptying of the meal. The technique is noninvasive and does not use radiation. Gastric emptying of solid meals, however, cannot be evaluated reliably. Acid secretion must be inhibited pharmacologically during measurements because the sensitivity of gastric contents changes when acid is secreted in the stomach. Impedance techniques are not suitable for measuring the rapid gastric emptying that may occur after gastric resections.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) allows detailed evaluation of gastric emptying (both total and regional) and intragastric distribution of a meal. However, its role in clinical practice is limited particularly as a result of its high cost and lack of accessibility.

Absorption Kinetics of Orally Administered Drugs

In humans, there is minimal gastric absorption of many orally administered drugs; the rate of absorption of a drug is therefore a measure of the rate of gastric emptying. Correct determination of the rate of gastric emptying by measurement of blood (or salivary) concentrations of intestinally absorbed solutes, such as paracetamol, is, however, fairly inaccurate and unsatisfactory for most circumstances in which there is a need to measure gastric emptying. However, there are recent developments with these techniques.

ETIOLOGY AND SEQUELAE OF DISORDERED GASTRIC EMPTYING

Many disorders are associated with abnormalities of gastric emptying and, potentially, with upper gastrointestinal symptoms, changes in oral drug absorption, and, in patients with diabetes, alterations in glycemic control. The causes of disordered gastric emptying are varied (Table I). Delayed gastric emptying, or gastroparesis, essentially describes stasis of food in the stomach in the absence of obstruction and represents the most common disorder of gastric emptying. Delayed gastric emptying can be divided into acute, and often reversible, and chronic (arbitrarily defined as lasting longer than 3 months) cases, and usually irreversible. It should be recognized that the magnitude of the delay in gastric emptying may be relatively modest and it has been suggested that a distinction should be made between "gastroparesis" and "delayed gastric emptying"; i.e., a diagnosis of gastroparesis should be reserved for those with grossly delayed gastric emptying. Patients with gastroparesis may present with symptoms such as nausea, vomiting, abdominal fullness, and early satiation, but patients also may be asymptomatic. Although historically it was assumed that symptoms were a result of delay in gastric emptying, it is now recognized that the relationship between severity of

TABLE I Causes of Gastroparesis and Rapid Gastric Emptying

Disorder	Cause
Transient delayed gastric emptying	Drugs (e.g., morphine, anticholinergics, nicotine, dopaminergics), postoperative ileus, viral gastroenteritis, electrolyte abnormalities (hyperglycemia, hypokalemia, hypomagnesemia), hypothyroidism, hyperthyroidism, hypopituitarism, Addison's disease, herpes zoster, critical illness, pregnancy
Chronic gastric stasis	Diabetes mellitus, idiopathic/functional dyspepsia, postsurgical (e.g., vagotomy), gastroesophageal reflux, atrophic gastritis, progressive systemic sclerosis, chronic idiopathic intestinal pseudo-obstruction, myotonia dystrophica, dermatomyositis/polymyositis, systemic lupus erythematosus, Duchenne's muscular dystrophy, amyloidosis, autonomic degeneration, spinal cord disease, tumor associated, anorexia nervosa and bulimia nervosa, central nervous system disease, brain stem lesions, Parkinson's disease, postirradiation, HIV infection, porphyria, liver disease
Rapid gastric emptying	Postsurgical, pancreatic insufficiency (fat), "early" type 2 diabetes, Zollinger–Ellison syndrome, duodenal ulcer disease

symptoms, at least when assessed as a total score, and the rate of gastric emptying is relatively weak. Certain symptoms, such as postprandial fullness and severe vomiting, may be potentially more sensitive. The causes of upper gastrointestinal symptoms in patients with gastroparesis are likely to be multifactorial; in patients with functional dyspepsia, symptoms may result from visceral hypersensitivity and impaired postprandial gastric relaxation. In patients with diabetes, symptoms may be modulated by the blood glucose concentration.

Acute Gastroparesis

Gastric emptying may be affected during generalized ileus due to surgery, severe infection, or metabolic derangement—e.g., in patients with diabetes mellitus, gastric emptying is slower during hyperglycemia and accelerated during hypoglycemia. In these situations, restoration of normal gastric function usually follows correction of the underlying problem. A number of drugs, such as levodopa and nicotine, may lead to delayed gastric emptying. In addition to delaying gastric emptying of other gastric contents, these drugs may affect their own absorption by delaying emptying of the remaining drug from the stomach. These effects are generally reversible on cessation of the drug.

Chronically Delayed Gastric Emptying

The most common causes of chronically delayed gastric emptying include diabetes mellitus, functional dyspepsia, gastroesophageal reflux disease, and postsurgical delayed gastric emptying, with prevalence rates between 30 and 50%. The main pathogenic factors involved in diabetes are autonomic neuropathy and acute changes in blood glucose concentrations. Even relatively minor changes in blood glucose concentrations within the physiological range affect gastric emptying; e.g., gastric emptying of a meal will be slower at a blood glucose concentration of 8 mmol/liter compared with 4 mmol/liter. There is evidence in patients with functional dyspepsia that the rate of emptying may also be associated with demographic factors, including gender and body weight. It is also well recognized that delayed gastric emptying may be a complication of malignancy in the absence of gastric or intestinal obstruction, particularly associated with carcinomas of the breast, lung, and pancreas. Gastric surgery, involving vagal interruption, gastric resection, and/or a drainage procedure (performed now much less frequently than before), may be associated with either delayed or more rapid gastric emptying; these cannot be discriminated on the basis of symptoms. Delayed gastric emptying occurs frequently in patients with anorexia nervosa, but often improves with recommencement of a significant nutritional intake and precedes significant weight gain.

Investigation of Disordered Gastric Emptying

Because of the poor correlation between symptoms and disordered gastric emptying, it is not always clear who should undergo clinical investigation and, if so, when. A general approach to investigation follows.

History and Physical Examination

Investigation of a patient with symptoms suggestive of disordered gastric emptying should start with a comprehensive history focusing on the course of symptoms and their relationship to the composition, volume, and timing of meals. Symptoms may follow a relapsing and remitting pattern. Anorexia, early satiation, weight loss, excessive abdominal bloating, nausea with food ingestion with or without vomiting, and vomiting of food

ingested more than 8 hours earlier may all be suggestive of delayed gastric emptying. Unfortunately, there is a poor correlation with symptoms (Fig. 4) and the presence of delayed gastric emptying, making assessment of the problem difficult. Physical examination is often unremarkable, but a succussion splash, suggestive of gastric dilatation, may be detected.

Further Investigations

Biochemical, hematologic, and urine analyses may reveal previously undetected metabolic disorders, including diabetes, hypothyroidism, hypokalemia, Addison's disease, and potential drug side effects. Following that assessment, conditions causing mechanical obstruction need to be excluded, usually by means of upper endoscopy. If negative, gastric emptying should ideally be measured, preferably by scintigraphy.

Approach to Treatment

Because the motor dysfunctions that contribute to delayed gastric emptying are poorly defined, targeted treatment to accelerate gastric emptying is often difficult. As a result, treatment is primarily aimed at symptom resolution or improvement. When symptoms are mild or absent, treatment may not be indicated, but many patients complain of persistent symptoms such as nausea, vomiting, and weight loss. It should be remembered that psychiatric abnormalities can contribute to symptom severity and should be addressed when appropriate.

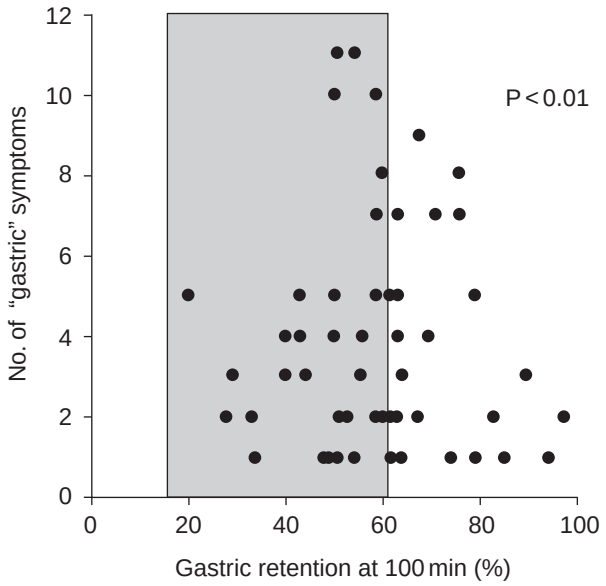


FIGURE 4 The relationship between symptoms (total score) referable to delayed gastric emptying and gastric emptying of a solid (minced beef) meal in patients with diabetes. The normal range for gastric emptying is shown in the shaded area.

Diet No formal evaluation of dietary modification has been performed. However, a stepwise approach is recommended, beginning with reducing the size and fat content of meals, aimed at decreasing the gastric load. Of vital importance is the maintenance of adequate nutrition. Supplemental calories can be supplied by liquid formula, and in severe cases, it may be necessary to start with liquid nutrition, either orally or intragastrically, because emptying of liquid is often less affected than that of solid. Parenteral nutrition should be avoided, if at all possible. However, dietary modification is successful alone in only a minority of patients and most will require drug treatment.

Prokinetic agents The use of prokinetic drugs currently forms the mainstay of treatment; these include cisapride, domperidone, metoclopramide, and erythromycin. In general, these drugs all provide dose-related improvements in gastric emptying, although their pharmacological mechanisms of action differ (Table II); the response to prokinetic therapy tends to be greater when gastric emptying is more delayed. With the exception of erythromycin, all of these drugs have been shown to improve symptoms and quality of life, but there is evidence that tolerance may develop to the gastrokinetic effects of erythromycin. However, erythromycin is the most potent drug when given intravenously (in doses < 3 mg/kg) and may be particularly useful in the early phase of treatment. It is worth noting that the response to erythromycin is attenuated during hyperglycemia; this effect probably applies to other prokinetic agents. The drug of choice for oral administration is probably cisapride, which acts by a 5-hydroxytryptamine isotype 4 (5-HT4) receptor-mediated effect. However, recent reports of adverse events, including serious cardiac arrhythmias (as a result of Q–T interval prolongation), have resulted in restrictions being placed on the use of cisapride in the United States and in many other countries. Alternative agents that act through a 5-HT4 effect, e.g., tegaserod, are now under development. At present, prolonged oral therapy can be achieved with the use of metoclopramide or domperidone (this latter drug is not available in the United States); the former has the advantage of being available as a subcutaneous preparation, which may be of use in those patients with intractable vomiting, but the possibility of potential central nervous system side effects with metoclopramide should be borne in mind. If symptoms are refractory to prokinetic therapy, placement of a feeding jejunostomy may be required to maintain nutrition.

Surgery In most cases, surgery is not recommended, and may be associated with clinical

TABLE II Prokinetic Drugs Used in the Treatment of Gastroparesis[a]

Drug	Mechanism of action	Route of administration	Oral dose
Cisapride	5-HT4 receptor agonist	po	10–20 mg tid
Domperidone	Dopamine D2 receptor antagonist	po	10–20 mg bid–qid
Metoclopramide	5-HT4 receptor agonist, D2 antagonist	po, sc, im, iv	10 mg tid
Erythromycin	Motilin receptor agonist	po or iv	250–500 mg tid

[a] Abbreviations: po, periorbital; sc, subcutaneous; im, intramuscular; iv, intravenous; tid, ter in die (three times/day); bid, bis in die (twice/day).

deterioration. However, with patients for whom medical treatment has failed and persisting weight loss is present, surgical treatment should be considered in specialist centers only, with a partial gastrectomy with Billroth I reconstruction being the procedure of choice. In patients with diabetes, there is evidence that pancreatic transplantation may improve both gastric emptying and symptoms.

Experimental approaches There has been renewed interest in gastric electrical stimulation as a therapy, whether using neural electrical stimulation at a high frequency, which probably stimulates vagal sensory nerves and may suppress the vomiting center, or gastric electrical pacing, in which the electrical stimulation of cholinergic motor neurones approximates the physiological frequency (~ 3 cycles/minute). Although the precise mode of action remains unclear, this method deserves further study. Preliminary results from ongoing trials suggest that the therapy provides significant relief of symptoms, including nausea and vomiting; however, its effects on gastric emptying are as yet unconfirmed.

See Also the Following Articles

Breath tests • Diabetic Gastroparesis • Duodenal Motility • Functional (Non-Ulcer) Dyspepsia • Gastric Motility • Ileal Break • Postprandial Motility • Ultrasonography

Further Reading

Berstad, A., Hausken, T., Gilja, O. H., Hveem, K., Nesje, L. B., and Odegaard, S. (1996). Ultrasonography of the human stomach. *Scand. J. Gastroenterol.* **220** (Suppl.), 75–82.

Champion, M. C. (1996). Treatment of gastroparesis. *In* "Evolving Concepts in Gastrointestinal Motility" (M. C. Champion and W. C. Orr, eds.)., pp. 108–147. Blackwell Science, London.

Chaudhuri, T. K., and Fink, S. (1991). Gastric emptying in human disease states. *Am. J. Gastroenterol.* **86**, 533–538.

Hasler, W. L. (2000). The brute force approach to electrical stimulation of gastric emptying: The future treatment for refractory gastroparesis. *Gastroenterology* **118**, 258–263.

Hebbard, G. S., Sun, W. M., Bochner, F., and Horowitz, M. (1995). Pharmacokinetic considerations in gastrointestinal motor disorders. *Clin. Pharmacokinet.* **28**, 41–66.

Hornbuckle, K., and Barnett, J. L. (2000). The diagnosis and workup of the patient with gastroparesis. *J. Clin. Gastroenterol.* **30**, 117–124.

Horowitz, M., and Dent, J. (1991). Disordered gastric emptying: Mechanical basis, assessment and treatment. *Balliere's Clin. Gastroenterol.* **5**, 371–407.

Horowitz, M., Collins, P. J., and Shearman, D. J. C. (1985). Disorders of gastric emptying in humans and the use of radionuclide techniques. *Arch. Intern. Med.* **145**, 1467–1472.

Horowitz, M., Dent, J., Fraser, R., Sun, W., and Hebbard, G. (1994). Role and integration of mechanisms controlling gastric emptying. *Dig. Dis. Sci* . **39**(Suppl.), 7S–13S.

Horowitz, M., Su, Y. C., Rayner, C. K., and Jones, K. L. (2001). Gastroparesis: Prevalence, clinical significance and treatment. *Can. J. Gastroenterol.* **15**, 805–813.

Kelly, K. A. (1980). Gastric emptying of liquids and solids: Roles of proximal and distal stomach. *Am. J. Physiol.* **239**, G71–G76.

Maughan, R. J., and Leiper, J. B. (1996). Methods for the assessment of gastric emptying in humans: An overview. *Diabet. Med.* **13**, S6–S10.

Moore, J. G. (1996). Gastroparesis: Pathogenesis and evaluation. *In* "Evolving Concepts in Gastrointestinal Motility" (M. C. Champion and W. C. Orr, eds.), pp. 87–107. Blackwell Science, London.

Plourde, V. (1999). Stress induced changes in the gastrointestinal motor system. *Can. J. Gastroenterol.* **13**(Suppl.), 26A–31A.

Quigley, E. M. (2000). Pharmacotherapy of gastroparesis. *Expert Opin. Pharmacother.* **1**, 881–887.

Smout, A., Horowitz, M., and Armstrong, D. (1994). Methods to study gastric emptying. *Dig. Dis. Sci.* **39**(Suppl.), 130S–132S.

Smout, A. J. P. M., Jebbink, H. J. A., Akkermans, L. M. A., and Bruijs, P. P. M. (1994). Role of electrogastography and gastric impedance measurements in the evaluation of gastric emptying and motility. *Dig. Dis. Sci.* **39**(Suppl.), 1105–1135.

Vantrappen, G. (1994). Methods to study gastric emptying. *Dig. Dis. Sci.* **39**, 91S–94S.

Gastric H^+,K^+-ATPase

JAI MOO SHIN*, OLGA VAGIN*, WOLFGANG SIMON†, AND GEORGE SACHS*

**University of California, Los Angeles and Veterans Affairs Greater Los Angeles Healthcare System and †Byk-Gulden Pharmaceutical, Germany*

gastric H^+,K^+-ATPase A P_2-type ion-motive ATPase enzyme that carries out an electroneutral exchange of cytoplasmic protons for extracytoplasmic potassium.

parietal cell The cell type in the gastric mucosa that contains H^+,K^+-ATPase enzyme and secretes HCl.

sarcoplasmic reticulum (SR) Ca^{2+}-ATPase A P_2-type ion-motive ATPase found in muscle sarcoplasmic reticulum that transports Ca^{2+} into the SR lumen.

Gastric H^+,K^+-ATPase is the enzyme responsible for the elaboration of HCl by the parietal cell of the gastric mucosa. It is a P_2-type ion-motive ATPase and allows the exchange of H^+ for K^+. This is an electroneutral exchange of cytoplasmic protons for extracytoplasmic potassium.

INTRODUCTION

Gastric H^+,K^+-ATPase consists of an α-subunit, which is composed of approximately 1034 amino acids, and a β-subunit, which is a glycoprotein composed of approximately 290 amino acids. The primary sequences of the α-subunits, as deduced from their cDNAs, have been reported. The pig gastric H^+,K^+-ATPase α-subunit sequence as deduced from its cDNA consists of 1034 amino acids and has a M_r of 114,285, the rat gastric H^+,K^+-ATPase consists of 1033 amino acids, and the rabbit and human gastric H^+,K^+-ATPases consist of 1035 amino acids. The degree of conservation among the α-subunits is extremely high (over 97% identity) among species. In addition, the gene sequence for the human H^+,K^+-ATPase α-subunit and the 5′ part of the rat H^+,K^+-ATPase α-subunit have been determined and indicate that the human gastric H^+,K^+ ATPase gene has 22 exons.

These H^+,K^+-ATPase α-subunits show a high level of homology ($\sim$60% identity) with the Na^+,K^+-ATPase catalytic α-subunit. The putative distal colon H^+,K^+-ATPase α-subunit has also been sequenced and shares 75% homology with both the H^+,K^+-ATPase and the Na^+,K^+-ATPase. The primary sequences of the H^+,K^+-ATPase β-subunits have been reported for rabbit, pig, rat, mouse, and human. The hydropathy analysis predicts that there is one membrane-spanning region, located at the region between amino acid sequence positions 38 and 63 near the N-terminus, and six or seven N-linked glycosylation sites. The function of the β-subunit is not clearly known, but this subunit appears to be required for the assembly and targeting of the catalytic subunit. The β-subunit plays a role in maintaining the structure of the α-subunit to enable effective binding of K^+ ions to the outside face of the α-subunit.

PHYSIOLOGY

The parietal cell generates acid secretion across the mucosal surface. These cells have histamine type 2, and muscarinic cholinergic receptors in the basolateral membranes. Histamine, is the major stimulant for acid secretion. Stimulation triggers dramatic morphological changes in the parietal cell from the resting state to the stimulating state. In the resting parietal cell, the gastric H^+,K^+-ATPase is present in smooth-surfaced cytoplasmic membrane tubules. On stimulation of acid secretion, the pump is found on the microvilli of the secretory canaliculus of the parietal cell. This morphological change results in a several-fold expansion of the canaliculus. In addition to this transition, there is activation of K^+ and Cl^- channels in the secretory membrane, which allows K^+ to access the extracytoplasmic face of the pump.

The gastric H^+,K^+-ATPase proton pump secretes gastric acid from the cytoplasmic region to the lumen, utilizing ATP. Activity of the gastric H^+,K^+-ATPase results in a primary secretion of 160 mM HCl into the secretory canaliculus. Since the pump is electroneutral, it is necessary to transfer the same amount of K^+ ions into the cytoplasmic region of the parietal cell during H^+ secretion. The K^+ is recycled from the cytoplasmic region to the lumen through a potassium channel in the apical membranes.

Gastric H^+,K^+-ATPase synthesis takes place in the endoplasmic reticulum of the parietal cell. Activation of the histamine-2 (H2) receptors by histamine results in a transient increase of mRNA of the H^+,K^+-ATPase

 Copyright 2004, Elsevier (USA). All rights reserved.

α-subunit, and H2 receptor antagonists suppress mRNA levels to slightly below basal levels. The α-subunit itself is not stable. The β-subunit is necessary to stabilize the α-subunit and to target it to the apical canalicular membrane. The half-life of the gastric H^+,K^+-ATPase is approximately 54 h in rats. However, H2 receptor antagonist treatment increases the half-life to 125 h. It seems that the resting state, when the gastric H^+,K^+-ATPase is located in tubular vesicles, enhances the enzyme stability dramatically compared to that of the stimulated state.

STRUCTURE AND ION TRANSPORT

The H^+,K^+-ATPase α-subunit has 10 transmembrane segments and the β-subunit has 1 transmembrane segment. The H^+,K^+-ATPase α-subunit has a strong association with the β-subunit. The lumenal loop between the seventh transmembrane segment (M7) and the eighth transmembrane segment (M8) of the α-subunit is associated with the β-subunit. The H^+,K^+-ATPase exists as an $(\alpha\beta)_2$ heterodimeric dimer. The cytoplasmic loop between the fourth and the fifth transmembrane segments is in close proximity to the two α-subunits. Structurally, based on the X-ray crystal structure of sarcoplasmic reticulum (SR) Ca^{2+}-ATPase, this enzyme can be divided into three domains, the cytoplasmic, membrane, and extracytoplasmic domains. The cytoplasmic domain contains four subdomains: a stalk domain, the N-domain, the P-domain, and the A-domain. The N-domain is a large cytoplasmic domain between M4 and M5 where ATP binds, the P-domain is a cytoplasmic domain near M4 where phosphorylation and dephosphorylation occur, the A-domain is a cytoplasmic domain between M2 and M3, which moves depending on conformational changes, and the stalk domain contains the extension of the transmembrane segments into the cytoplasmic region and forms a link for the passage of ions into the membranes. A proposed structure for the gastric α-subunit, based on the SR Ca^{2+}-ATPase structure, is shown in Fig. 1.

The H^+,K^+-ATPase catalyzes transport by means of conformational changes driven by cyclic phosphorylation and dephosphorylation of the catalytic subunit of the ATPase. The H^+ ion is transported from the cytoplasmic region to the lumen by a conformational change induced by phosphorylation and dephosphorylation. The H^+,K^+-ATPase initially binds hydronium ion (H_3O). On phosphorylation, the conformation changes from $E_1P \cdot H_3O^+$ to the $E_2P \cdot H_3O^+$ form. After H_3O^+ is released and K^+ is bound on the extracytoplasmic surface of the enzyme, the $E_2P \cdot K^+$ conformation is adopted. The $E_2P \cdot K^+$ conformation can be easily converted to the E_1K^+ conformation with dephosphorylation. The E_1K^+ conformation releases K^+ to the cytoplasmic side, allowing the new binding of H_3O^+.

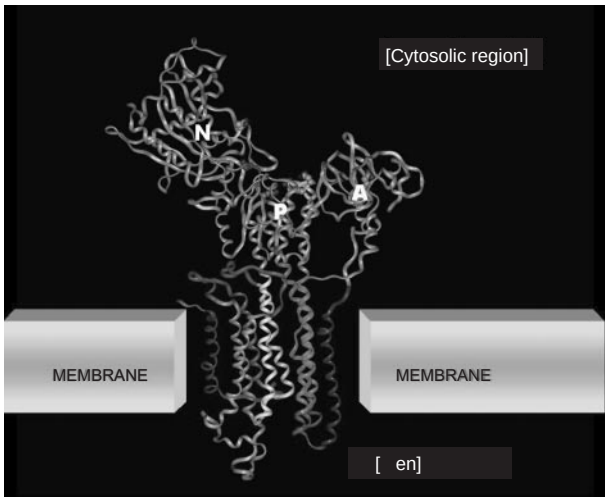


FIGURE 1 Proposed structure of the gastric H^+,K^+-ATPase α-subunit. The gastric H^+,K^+-ATPase α-subunit has 10 transmembrane segments (right represents 1st transmembrane segment, M1, and left shows 10th transmembrane segment, M10). A-, P-, and N-domains are proposed based on SR Ca^{2+}-ATPase crystal structure (E_1-Ca). ATP approaches the N-domain and phosphorylates Asp-386 in the P-domain and the N-domain tilts and docks on A-domain to form E_2-P.

The structural organization and changes in the cytoplasmic domains associated with the $E_1 \Leftrightarrow E_2$ conformational transitions involve several domain movements. The N-domain docks onto the P-domain with the A-domain displaced to one side in the E_1 or E_1P conformations, whereas the A-domain docks onto the P-domain and the N-domain is displaced from the P-domain in the $E_2(K)$ and E_2P conformations. Disruption of the interaction between the P-domain and the N-domain and formation of an A-domain to P-domain interaction during the $E_1 \Leftrightarrow E_1P \Leftrightarrow E_2P$ steps is associated with a change in ligation of Mg^{2+} ions from residues in the sequences TGDGVNDS (P-domain major) and near VAGDA (N-domain minor) in E_1 or E_1P to TGESE (A-domain major) and TGDGVNDS (P-domain minor) in E_2P. This change in ligation may be crucial in the change of reactivity of the phosphoenzymes, switching from ADP-sensitive in E_1P to water-sensitive in E_2P. Presumably, the cytoplasmic domain interactions accompanying E1 $\Leftrightarrow$ E_2 transitions are then relayed via the extended stalk helices of M4 and M5 to the transmembrane segments in which the transported cations are occluded. Changes in tilt, turn, or position of the transmembrane helices, driven by changes in the N-, P-, and A-domains, may allow the cations to dissociate at the opposite surface from that at which

they were occluded. The similarity of structural organization and conformational changes of the cytoplasmic domains of H^+,K^+-ATPase and Na^+,K^+-ATPase suggests that these features are common to other P-type pumps.

Many biochemical and mutagenesis studies provided evidence of a close association between M4 and M6 and between M5 and M7. This also indicates that M4, M5, and M6 are involved in the pathway transporting H_3O^+ outward and K^+ inward. When the Ca^{2+}-occluded E_1 conformation of the Ca^{2+}-ATPase is compared to the Ca^{2+}-free E_2 conformation, the dissociation of Ca^{2+} ions shows a dramatic rearrangement of 6 of 10 transmembrane segments (M1 to M6). M1 and M2 move upward to the cytoplasmic domain with a dramatic tilt of the region near the cytoplasmic domain (stalk domain). M3 and M4 shift downward to the lumen with a strong curve of M3. As the Ca^{2+} ion dissociates, helical bundles of M3, M4, M5, and M6 seem to be loose. The E_2 conformation shows more compact space for easy chelating of Ca^{2+} ion in the lumenal surface than the Ca^{2+}-occluded E1 conformation. It is likely that the gastric H^+,K^+-ATPase has a structural conformation very similar to that of Ca^{2+}-ATPase.

See Also the Following Articles

Gastric Acid Secretion • H2-Receptor Antagonists • Parietal Cells

Further Reading

Melle-Milovanovic, D., Lambrecht, N., Sachs, G., and Shin, J. M. (1998). Structural aspects of the gastric H^+,K^+ ATPase: The M5/M6 domain and alpha beta association. *J. Physiol. Scand.* **163**(Suppl. 643), 147–162.

Munson, K., Lambrecht, N., Shin, J. M., and Sachs, G. (2000). Analysis of the membrane domain of the gastric H^+,K^+-ATPase. *J. Exp. Biol.* **203**(Pt. 1), 161–170.

Sachs, G., Shin, J. M., Munson, K., Vagin, O., Lambrecht, N., Scott, D. R., Weeks, D. L., and Melchers, K. (2000). Review article: The control of the gastric acid and *Helicobacter pylori* eradication. *Aliment. Pharmacol. Ther.* **14**, 1383–1401.

Toyoshima, C., and Nomura, H. (2002). Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–612.

Gastric Infection (non-*H. pylori*)

ALLISON MOORE LIDDELL
Presbyterian Hospital of Dallas, Texas

gastritis Inflammatory condition of the stomach, acute or chronic, that is sometimes due to an infectious pathogen.
gastropathy Disorder of the stomach that includes gastritis and other noninflammatory conditions.

A variety of pathogens other than *Helicobacter pylori* cause gastric disease by direct infection of the gastric mucosa and subsequent acute and/or chronic inflammation. Most pathogens involve other organs and other parts of the gastrointestinal tract primarily, and only occasionally produce gastric pathology, even less often with signs or symptoms of gastric disease. A few, such as cytomegalovirus, *Histoplasma*, and the nematodes that cause anisakiasis, produce gastric manifestations frequently in their risk groups. Increasing numbers of immunocompromised patients, particularly those undergoing immunosuppressive therapy or infected with human immunodeficiency virus, have an increased incidence of gastric infections overall. Most of these clinical syndromes have been described rarely in apparently immunocompetent hosts. Multiple gastric malignancies have also been associated to varying degrees with infectious agents. These include Kaposi's sarcoma (human herpesvirus-6), lymphoma (Epstein–Barr virus, *H. pylori*), gastric adenocarcinoma (*H. pylori*, Epstein–Barr virus, *Mycoplasma*, *Streptococcus*, *Rhizopus*), and anisakiasis polyadenoma (nematodes). These tumors are not discussed further here, and the focus is instead on the most common infectious causes of gastritis and gastropathy.

BACTERIA

Phlegmonous gastritis is a rare infection involving the submucosal layer of the gastric wall. A subset of

Copyright 2004, Elsevier (USA). All rights reserved.

they were occluded. The similarity of structural organization and conformational changes of the cytoplasmic domains of H^+,K^+-ATPase and Na^+,K^+-ATPase suggests that these features are common to other P-type pumps.

Many biochemical and mutagenesis studies provided evidence of a close association between M4 and M6 and between M5 and M7. This also indicates that M4, M5, and M6 are involved in the pathway transporting H_3O^+ outward and K^+ inward. When the Ca^{2+}-occluded E_1 conformation of the Ca^{2+}-ATPase is compared to the Ca^{2+}-free E_2 conformation, the dissociation of Ca^{2+} ions shows a dramatic rearrangement of 6 of 10 transmembrane segments (M1 to M6). M1 and M2 move upward to the cytoplasmic domain with a dramatic tilt of the region near the cytoplasmic domain (stalk domain). M3 and M4 shift downward to the lumen with a strong curve of M3. As the Ca^{2+} ion dissociates, helical bundles of M3, M4, M5, and M6 seem to be loose. The E_2 conformation shows more compact space for easy chelating of Ca^{2+} ion in the lumenal surface than the Ca^{2+}-occluded E1 conformation. It is likely that the gastric H^+,K^+-ATPase has a structural conformation very similar to that of Ca^{2+}-ATPase.

See Also the Following Articles

Gastric Acid Secretion • H2-Receptor Antagonists • Parietal Cells

Further Reading

Melle-Milovanovic, D., Lambrecht, N., Sachs, G., and Shin, J. M. (1998). Structural aspects of the gastric H^+,K^+ ATPase: The M5/M6 domain and alpha beta association. *J. Physiol. Scand.* **163**(Suppl. 643), 147–162.

Munson, K., Lambrecht, N., Shin, J. M., and Sachs, G. (2000). Analysis of the membrane domain of the gastric H^+,K^+-ATPase. *J. Exp. Biol.* **203**(Pt. 1), 161–170.

Sachs, G., Shin, J. M., Munson, K., Vagin, O., Lambrecht, N., Scott, D. R., Weeks, D. L., and Melchers, K. (2000). Review article: The control of the gastric acid and *Helicobacter pylori* eradication. *Aliment. Pharmacol. Ther.* **14**, 1383–1401.

Toyoshima, C., and Nomura, H. (2002). Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–612.

Gastric Infection (non-*H. pylori*)

ALLISON MOORE LIDDELL
Presbyterian Hospital of Dallas, Texas

gastritis Inflammatory condition of the stomach, acute or chronic, that is sometimes due to an infectious pathogen.
gastropathy Disorder of the stomach that includes gastritis and other noninflammatory conditions.

A variety of pathogens other than *Helicobacter pylori* cause gastric disease by direct infection of the gastric mucosa and subsequent acute and/or chronic inflammation. Most pathogens involve other organs and other parts of the gastrointestinal tract primarily, and only occasionally produce gastric pathology, even less often with signs or symptoms of gastric disease. A few, such as cytomegalovirus, *Histoplasma*, and the nematodes that cause anisakiasis, produce gastric manifestations frequently in their risk groups. Increasing numbers of immunocompromised patients, particularly those undergoing immunosuppressive therapy or infected with human immunodeficiency virus, have an increased incidence of gastric infections overall. Most of these clinical syndromes have been described rarely in apparently immunocompetent hosts. Multiple gastric malignancies have also been associated to varying degrees with infectious agents. These include Kaposi's sarcoma (human herpesvirus-6), lymphoma (Epstein–Barr virus, *H. pylori*), gastric adenocarcinoma (*H. pylori*, Epstein–Barr virus, *Mycoplasma*, *Streptococcus*, *Rhizopus*), and anisakiasis polyadenoma (nematodes). These tumors are not discussed further here, and the focus is instead on the most common infectious causes of gastritis and gastropathy.

BACTERIA

Phlegmonous gastritis is a rare infection involving the submucosal layer of the gastric wall. A subset of

Copyright 2004, Elsevier (USA). All rights reserved.

phlegmonous gastritis cases develop emphysematous gastritis, and diagnosis is based on the clinical presentation of an acute abdomen with systemic toxicity, and on radiographs demonstrating gas bubbles within the stomach wall. Risk factors include recent abdominal surgery, gastroenteritis, ingestion of corrosive substances, and alcohol binge drinking. Patients present with acute abdominal pain, nausea, vomiting, melena, and/or hematemesis. Symptoms may be more subacute in immunocompromised patients. Endoscopy shows edematous, darkened gastric mucosa. In a case series of 28 patients, the organisms most commonly involved were *Escherichia coli* (six cases), *Streptococcus* species (six cases), *Enterobacter* species (five cases), and *Pseudomonas aeruginosa* (three cases). The mortality rate was 61%. Long-term morbidity due to gastric strictures was reported in 21% of cases. Other reports have implicated *Clostridium perfringens, Clostridium welchii*, other Enterobacteriaceae, and *Staphylococcus* spp. Computed tomography may be useful both in establishing the diagnosis and for following the resolution of emphysematous gastritis. Therapy includes broad-spectrum antibiotics directed against enteric pathogens and surgery with drainage of pus or resection of necrotic tissue.

Mycobacterium tuberculosis infection involves the stomach in less than 1% of cases, particularly in children, and produces nonspecific symptoms such as nausea, vomiting, abdominal pain, and dyspepsia. Half of patients have evidence of tuberculosis involving other organ systems. Gastric tuberculosis may arise from primary invasion after ingestion of the organism or from dissemination during reactivation disease. Appearance at endoscopy is variable, and may include gastritis, ulcerations (single or multiple), nodules, or larger masses. Endoscopic brush cytology may demonstrate the organism, and standard histopathology reveals granulomatous inflammation. Treatment involves prolonged therapy with multiple agents, and prognosis is favorable assuming proper adherence to the medical regimen and attention to nutrition.

Mycobacterium avium-intracellulare complex (MAC), a common opportunistic infection in advanced autoimmune deficiency syndrome (AIDS), presents with findings of mesenteric lymphadenopathy, bacteremia, and involvement of multiple organs. The site of inoculation is the gastric or respiratory mucosa, and the organism is then taken up by macrophages and reaches the bloodstream via lymphatic channels. Presenting symptoms include fever, weight loss, diarrhea, abdominal pain, and malabsorption. Gastrointestinal obstruction due to massive lymphadenopathy has been reported. Disease manifestations directly involving the gastric mucosa, however, are much less common than those of the small intestine. Clinical presentation may be nonspecific abdominal complaints or upper gastrointestinal (GI) bleeding. Endoscopic findings include gastric ulceration, and histopathology reveals foamy macrophages loaded with mycobacteria. Treatment, as with tuberculosis, involves prolonged therapy with multiple agents.

Actinomycosis is a chronic suppurative infection due to a number of gram-positive, non-spore-forming, anaerobic bacterial species. Abdominal actinomycosis most commonly involves the appendix, terminal ileum, and cecum. Gastric involvement is usually caused by *Actinomyces israelii*, a mouth commensal. In the setting of defects in gastric mucosal integrity, such as peptic ulcer disease or gastric surgery, perigastric actinomycotic masses may form. Intramural gastric actinomycosis has also been described rarely, and the portal of entry may not be apparent. The presentation is characterized by a subacute development of epigastric pain, upper GI bleeding, and occasionally fever, weight loss, and development of a fistulous tract through the skin or into the lumen of the stomach. Radiographic and endoscopic evaluation reveals a mass lesion with or without ulcer, and the mass may infiltrate surrounding structures. Unless a fistulous tract draining inflammatory material containing organisms is present, diagnosis relies on histopathology with characteristic sulfur granules, which are clumps of the branched gram-positive filamentous organisms. Occasionally culture is helpful. Treatment options include surgical resection and penicillin or other active antibiotics for at least 3 months.

Syphilis is a complex multisystem infection that has experienced a resurgence during the first years of the twenty-first century. Secondary syphilis is characterized by early and late manifestations, and the stomach may be involved in either phase. Overall, gastric syphilis occurs in $<1\%$ of syphilis cases. Symptoms are nonspecific and include abdominal pain, vomiting, and weight loss, in addition to symptoms related to nongastric manifestations such as generalized lymphadenopathy and rash. Complications reported include upper GI bleeding, perforation, and obstruction. Endoscopic findings may include diffuse enlargement of gastric rugae, mucosal nodularity, erosive gastritis, and gastric ulceration, most commonly involving the antrum. These findings may mimic gastric lymphoma, linitis plastica, and hypertrophic lymphocytic gastritis. The classic radiologic finding on contrast radiographs (found in $<25\%$ of patients) is "hourglass stomach," produced by multiple ulcerations and prepyloric narrowing. Histologic specimens reveal ulceration and gastritis with an

atypical polyclonal plasma cell infiltrate. Diagnosis is confirmed by specific treponemal serologic testing and demonstration of spirochetes on Warthin–Starry and immunofluorescent stains of gastric mucosal biopsy specimens specific for *Treponema pallidum*. Gastric syphilis responds completely to penicillin therapy, often promptly, and some clinicians recommend antisecretory therapy as well to promote healing.

Gastrointestinal anthrax may involve the gastric mucosa. Herbivores, the usual human source for anthrax, become infected in rural parts of the world from ingesting spores in the soil. Humans ingest or inhale spores found on hides or hair of such exposed animals. When swallowed, anthrax spores may cause lesions from the oral cavity to the cecum. This underreported manifestation of anthrax may be rapidly fatal. Nausea, vomiting, abdominal distension, and severe abdominal pain are presenting symptoms of gastric involvement. The ulcerative lesions may bleed; hemorrhaging in severe cases may be massive. Based on limited reports of GI anthrax, the disease spectrum ranges from the asymptomatic to the fatal, by shock or sepsis. Diagnosis is based on culture or histopathology. Treatment includes effective antibiotics and supportive measures.

FUNGI

Gastric involvement by *Candida* spp. is difficult to define in terms of pathogenicity. Colonization of preexisting lesions with *Candida* is common, and invasion by this route may result in disseminated infection, particularly in immunosuppressed hosts. However, these same hosts may also develop primary mucosal candidiasis of the stomach, with shallow ulcerations and plaques of exudate noted at endoscopy. Oropharyngeal and esophageal candidiasis cases are much more common. Treatment with fluconazole is highly effective, although resistance may develop in immunosuppressed patients on long-term fluconazole maintenance for other invasive candidal infections. These patients may require amphotericin. The finding of *Candida* in a gastric lesion, particularly in an immunocompetent host, may not warrant specific therapy.

Disseminated histoplasmosis often involves the gastrointestinal tract, and presentation with symptoms of such involvement occurs in about 20% of cases. However, presentation with predominantly GI symptoms is rare, and these symptoms are nonspecific. Gastric ulcers and infiltrative lesions have been described due to both *Histoplasma capsulatum* and *Histoplasma duboisii*. Histological analysis reveals granulomatous inflammation with intra- and extracellular yeast forms. Treatment involves systemic antifungal therapy with or without surgical resection.

Gastrointestinal *Pneumocystis carinii* (PCP) has been reported in AIDS patients maintained on aerosolized pentamidine for PCP prophylaxis. Organisms have been detected in ascites fluid, and biopsy specimens from erosive gastritis and duodenitis with multiple small scattered nodules have noted massive infiltration of foamy granular exudates of *P. carinii*. Patients may present with fever, anorexia, upper abdominal pain, or ascites plus symptoms attributable to nongastrointestinal organ involvement. Treatment is as for PCP pneumonia, with trimethoprim–sulfamethoxazole or second-line regimens if there is intolerance to sulfonamides.

Mucormycosis, caused by fungi of the order Mucorales in the class Zygomycetes, may involve the stomach and has been reported in diabetics (particularly with acidosis) as well as renal, liver, heart, and heart–lung transplant recipients, the severely malnourished, and patients with AIDS, uremia, and hematologic malignancies. Corticosteroid use in particular appears to be a major contributing factor. Direct ingestion of spores is the presumed portal of entry. Less than 10% of invasive mucormycosis cases present as gastrointestinal involvement, and of these, at least 60% involve the stomach. The disease is rapidly progressive, and mortality remains high, although early diagnosis improves outcome. Postmortem diagnosis is common in reported case series. Patients may present with upper abdominal pain, anorexia, nausea, vomiting, and gastrointestinal bleeding. Emphysematous gastritis and perforation have been reported, and may occur early posttransplantation. Large, deep, gastric ulcers with a gray–black eschar have been described, with histological appearance of extensive necrosis, vascular invasion, thrombosis, and nonseptate right-angular branching hyphae and cultures of tissue that grow mucormycetes. Colonization of ulcer surfaces without tissue invasion has a less clear significance and much better prognosis. Standard therapy involves an amphotericin formulation with or without surgery.

Gastric involvement in basidiobolomycosis, caused by a fungus in the Entomophthorales order of the class Zygomycetes, which usually produces cutaneous disease, was recently reported in a case series of seven patients from Arizona. The presenting complaint was abdominal pain, and less often fever. Diabetes, pica, and use of histamine-2 (H2) blockers have been associated with this slowly progressive infection, and therapy includes surgical resection and itraconazole. Amphotericin resistance has been documented. Diagnosis was

based on histopathology demonstrating the fungal morphology and surrounding eosinophilic inflammation, but all patients in this series demonstrated peripheral eosinophilia as well.

VIRUSES

Gastropathy is an uncommon presentation of cytomegalovirus (CMV) disease but can manifest as gastric ulcers, acute gastritis, chronic persistent gastritis, and protein-losing hypertrophic gastropathy. In patients with human immunodeficiency virus (HIV) specifically, gastritis ranks in incidence only behind retinitis, esophagitis, and colitis. One study of 497 HIV-infected patients who underwent esophagogastroduodenoscopy (EGD) for upper GI symptoms noted CMV infection in 8 of 16 patients with gastroduodenal ulcers who had evaluable histology specimens, and in 1 of 20 patients without ulceration. CMV gastropathy has also been well described in patients with hematologic malignancies, and in both bone marrow and solid organ transplant patients. Gastric involvement is a more common manifestation of CMV disease in transplant populations. Recently, an association between Menetrier's disease and CMV infection has been more fully described. Diagnosis of invasive CMV infection of the stomach requires demonstration of characteristic cytomegalic inclusions and/or positive immunostaining in tissue specimens of CMV virus antigens. Treatment of CMV gastric infection is similar to other presentations of invasive CMV infection, and includes valganciclovir, ganciclovir, or foscarnet for 21 days. Data regarding maintenance therapy specifically for gastric infection are lacking, although some researchers extrapolate from data for other types of CMV disease. (Indications for prophylaxis in specific groups of immunocompromised patients differ, and full discussion is beyond the scope of this discussion.)

Non-CMV herpesviruses documented to cause gastropathy include varicella zoster virus (VZV) and herpes simplex virus (HSV). Esophageal involvement is far more common with HSV, but gastric lesions have been reported in predominantly immunocompromised patients. Histologic findings include the typical inclusion bodies and ground glass nuclei, and differentiation between viruses requires culture, molecular, or serologic techniques. Radiologically, a cobblestone pattern is described due to the superficial interconnecting of linear ulcers alternating with bleblike edema. VZV gastritis may present with abdominal pain, nausea, and vomiting before the skin lesions of disseminated zoster become visible. Endoscopic and histologic appearance may be nonspecific, although diagnosis may be confirmed using molecular techniques. Treatment for HSV and VZV gastritis is intravenous acyclovir.

PARASITES

Strongyloides stercoralis is a soil-borne intestinal nematode with worldwide distribution. *Strongyloides* filariform larvae are ingested or penetrate human skin. After penetration, the invasive filariform larvae migrate to the lungs. Larvae in the lungs penetrate the bronchioles and are expectorated and swallowed. The larvae in the upper gastrointestinal tract then mature into adult females that produce eggs through parthenogenesis (the direct production of viable offspring from the female without fertilization by the male). Adults and thin-shelled eggs are confined to the mucosa and are not present in stool. Eggs that embryonate and hatch release rhabditiform larvae, which may be detected in stool. Rhaditiform larvae become infective filariform larvae either in stool or in the soil after excretion, completing the life cycle. Alternatively, if infection does not occur, *Strongyloides* rhabditiform larvae in the soil are able to mature into male and female adult worms. Adult worms exist in the soil as saprophytes, mate, and produce eggs. These eggs release rhabditiform larvae, which may differentiate into infective filariform larvae.

Strongyloides has been reported to involve almost every organ system and can disseminate to cause fatal illness in an immunosuppressed host. *Strongyloides* may involve the stomach and present with nausea, vomiting, anorexia, diarrhea, weight loss, and postprandial epigastric pain. Hematemesis and gastrointestinal obstruction have been reported as well. Diagnosis is made by stool examination, small bowel biopsy, or duodenal drainage. Examination of multiple stool samples increases sensitivity. Peripheral eosinophilia may be a diagnostic clue. Literature reports describe gastritis and numerous serpiginous, white, very superficial lesions that appeared to be exudates about 5 to 6 mm in size in the body and antrum of the stomach. Histologic examination confirms the presence of *Strongyloides* larvae and shows dense eosinophilic infiltrates, focal atrophy, and inflammation and destruction of glands, especially in the vicinity of the parasites. The mucosa is intact. Treatment is thiabendazole.

Recent evidence suggests that gastric involvement in AIDS-related cryptosporidiosis can be demonstrated in up to 90% of AIDS patients with cryptosporidium diarrhea, but isolated gastric symptoms are rare. Gastric ulcer, gastritis, and partial pyloric obstruction with thickened, inflamed antral folds have been described endoscopically, and gastric histology has noted cryptosporidial

macrogametes and trophozoites attached to the luminal surface of gastric epithelial cells. Treatment is difficult, but options include nitazoxanide or paromomycin with or without azithromycin.

Gastric anisakiasis is caused by ingestion of raw fish, particularly salmon and herring, containing third-stage nematodal larvae. The larvae invade the gastric mucosa, causing acute upper abdominal pain with or without bleeding. Symptoms usually occur within 7 days of ingestion. Endoscopy reveals a submucosal mass with a protruding larval form. Removal by forceps is commonly practiced, although patients may pass the larva spontaneously. Histologic examination reveals an intense eosinophilic infiltrate. Recurrent episodes are associated with more severe symptoms, and chronic gastritis may occur if the parasite is not recognized and removed. Cooking and freezing are adequate methods of killing the larvae.

See Also the Following Articles

Candidiasis • Cryptosporidium • Cytomegalovirus • Erosive and Hemorrhagic Gastritis (Gastropathy) • Gastritis • Mycobacterial Infection

Further Reading

Atten, M. J., Attar, B. M., Teopengo, E., *et al.* (1994). Gastric syphilis: A disease with multiple manifestations. *Am. J. Gastroenterol.* **89**, 2227–2229.

Lee, E., and Feldman, M. (2002). Gastritis and other gastropathies. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease, Pathophysiology/Diagnosis/Management" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., Chap. 42, pp. 810–828. W. B. Saunders, Philadelphia.

Long, B. W., Johnston, J. H., Wetzel, W., Flowers, R. H., and Haick, A. (1995). Gastric syphilis: Endoscopic and histological features mimicking lymphoma. *Am. J. Gastroenterol.* **90**(9), 1504–1507.

Singh, N. (1995). Invasive gastrointestinal zygomycosis in a liver transplant recipient: Case report and review of zygomycosis in solid-organ transplant recipients. *Clin. Infect. Dis.* **20**(3), 617–620.

Sirisanthana, T. (2002). Anthrax of the gastrointestinal tract. *Emerg. Infect. Dis.* **8**(7), 649–651.

Sugimachi, K., *et al.* (1985). Acute gastric anisakiasis. Analysis of 178 cases. *JAMA* **523**, 1012–1013.

Yeomans, N., and Lambert, J. (1995). Infections of the stomach and duodenum. *In* "Bockus Gastroenterology" (W. Haubrich, F. Schaffner, and J. Berk, eds.), 5th Ed., Chap. 47, pp. 805–815. W. B. Saunders, Philadelphia.

Gastric Motility

CHRISTINE FEINLE-BISSET, DEIRDRE O'DONOVAN, AND MICHAEL HOROWITZ
University of Adelaide and Royal Adelaide Hospital, South Australia

gastric accommodation Relaxation of the proximal stomach to "accommodate" food without a major increase in intragastric pressure.
migrating motor complex Cyclical motor pattern that occurs in the fasting gastrointestinal tract; consists of three phases (I–III) with distinct differences in activity.
motility Contractile activity of the wall of the gastrointestinal tract.

The fasting and postprandial motility patterns of the stomach, pylorus, and proximal small intestine are regulated by neural and humoral mechanisms, including those triggered by small intestinal nutrients. Various methods can be used to quantify gastric motility. The pathophysiology of disordered gastric motility is complex and poorly understood.

INTRODUCTION

In the early 1820s, Alexis St. Martin, a patient suffering from an abdominal gunshot wound, afforded his physician, William Beaumont, the opportunity to be the first to observe through an open gastric fistula that food, whether ingested or placed directly into the stomach through the fistula, induced gastric wall contractions that facilitated the breakdown of food into smaller particles. Two gastric regions with distinct motor activity were later identified by Walter Cannon, using roentgenography to investigate gastric motor function in his studies in dogs and cats.

The various compartments of the gastroduodenal region serve different functions. The proximal stomach has the capacity to relax, so that ingested food can be accommodated with minimal increase in intragastric pressure; distal stomach phasic contractions then mix and grind solid food and propel "liquefied" solids toward the pylorus. Tonic and phasic pyloric contractions regulate transpyloric flow, and the duodenum can modulate transit of chyme and, hence, the exposure of small intestinal receptors to nutrients. However, it is essential to recognize that the intragastric processing and subsequent delivery of nutrient into the small intestine at a rate that optimizes digestion and absorption are dependent on the integration of motility between these different regions.

PATTERNS OF FASTING AND POSTPRANDIAL GASTRIC MOTILITY

The motility of the stomach, as in other areas of the upper gastrointestinal tract, is organized into two basic patterns—fasting and fed—which have fundamental differences. In the fasted state, the proximal stomach is in a continuous state of partial contraction, maintained by a vagally mediated cholinergic input. In contrast, the contractile activity of the distal stomach is characterized by a pattern of cyclical activity, the so-called migrating motor complex (MMC) (Fig. 1), which consists of three distinct phases (I–III) and, in humans, has a cycle length of approximately 100 minutes. Approximately 80% of the cycles originate in the antrum,

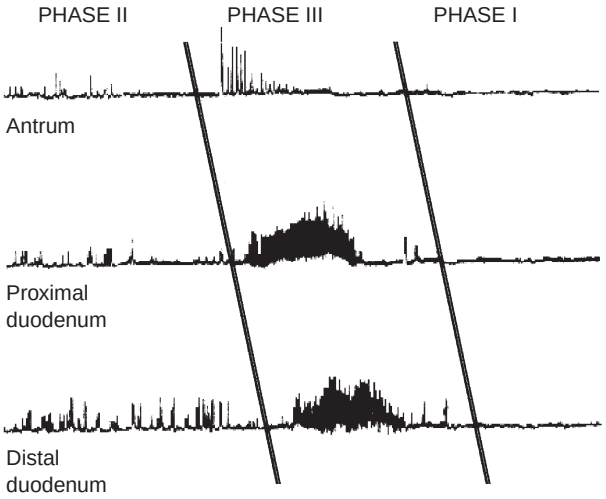


FIGURE 1 Motility patterns in the fasting gastrointestinal tract. The pattern of the fasting stomach and small intestine is characterized by cyclical activity, the so-called migrating motor complex, which consists of three distinct phases (I–III); in humans, the cycle length of the three phases is approximately 100 min. Reproduced with permission from Rao and Schulze-Delrieu (1993).

 Copyright 2004, Elsevier (USA). All rights reserved.

the remainder in the small intestine. Phase I is characterized by a period of quiescence in the antrum, lasting for ~40–60 minutes; phase II is a period of increasing, but irregular, phasic contractions and lasts for ~30–40 minutes; phase III consists of a short period of intense contractile activity for ~5–10 minutes. During phase III, the contractions are characterized by a high amplitude and occur at maximum frequency—that is, at the frequency of the electric pacemaker, which in the stomach is 3/minute. Phase III of the MMC is propulsive and, as it migrates from the stomach through the small intestine into the ileum, it "sweeps" undigested or indigestible food particles, dead cells, secretions, and bacteria, hence the terminology of the "gastrointestinal housekeeper."

Meal ingestion interrupts the cyclical motor pattern of the MMC and causes distinct changes in gastric motility, characterized by a relaxation of the proximal stomach, irregular, phasic contractile activity in the antrum, an increase in tonic and phasic pyloric pressures, and irregular contractile activity in the duodenum. The motor response to food is dependent on its physical state (solid or liquid) and nutrient content. Ingestion of water does not affect fasting motility, unless ingested in volumes >400 ml. Ingestion of a solid meal induces strong contractions in the antrum, whereas the same meal in homogenized form or nutrient-containing liquid meals induce less pronounced antral contractions. Hence, a solid meal is traditionally used for the clinical assessment of antral motor function.

The proximal gastric motor response to food ingestion may be divided temporarily into two phases. The first phase occurs during, and shortly after, the ingestion of a meal and consists of two relaxatory reflexes; "receptive relaxation" describes the relaxation of the fundus in response to swallowing, and this is followed by "gastric accommodation," whereby the proximal stomach relaxes so that a progressive increase in meal volume is not associated with a major increase in intragastric pressure. The second phase of postprandial proximal gastric motility consists of a continuous tonic contraction that gradually squeezes gastric content toward the distal stomach. However, because this tonic contractile activity is associated with little change in intragastric pressure, it cannot be recorded by standard manometric techniques. Changes in gastric tone can indirectly be quantified as a change in gastric volume using the electronic barostat (see later).

The functions of the distal stomach are accomplished by coordinated antral and pyloric contractile activity. In contrast to the proximal stomach, the motility of the antrum consists of phasic contractions that propagate toward the pylorus and serve to mix and grind food (Fig. 2). The pylorus exhibits both tonic and phasic pressures, with a narrow zone of contraction of approximately 2 mm. Stimuli that increase antral contractions include gastric distension; the physical consistency of gastric chyme (i.e., particulate vs. homogenized) also affects the depth of contractions. Each peristaltic wave originates in the midstomach, or corpus, as a shallow indentation and deepens as it moves more distally, but does not always entirely occlude the lumen, compressing liquids and solids within the distal antrum.

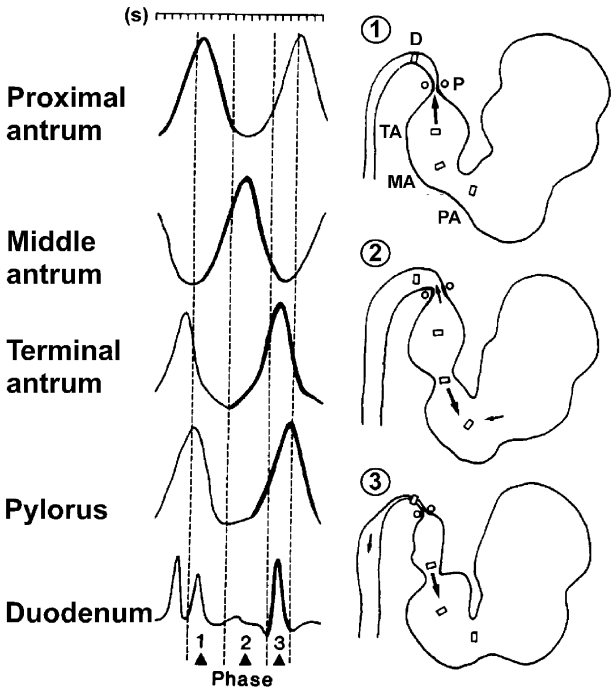


FIGURE 2 Relationship between antropyloroduodenal contractions (tracings, left) and movement of gastric content (diagrams, right) as recorded concurrently with fluoroscopy and strain gauge transducers (recording motility). The tracings show the progression of a peristaltic wave moving over the proximal (PA), the middle (MA), and the terminal (TA) antrum to the pylorus (P) and duodenum (D). (1) The wave moves over the proximal antrum. The terminal antrum and the pylorus relax. Gastric content is forced into the terminal antrum. (2) The wave travels over the middle antrum. Gastric content is evacuated through the pylorus and also retropelled into the proximal antrum. A subsequent wave moves over the gastric body, driving digesta into the proximal antrum. (3) The wave moves over the terminal antrum with increasing velocity. Contractions of the terminal antrum and pylorus enhance retropulsion and grinding. The subsequent wave advances the antrum. A propagative wave starts on the duodenal bulb. Reproduced from Ehrlein, H. J. and Schemann, M. (1990). Influence of food constituents. *In* "Gastropyloro-duodenal Coordination" (J. M. Van Nueten *et al.*, eds.). Wrightson Biomedical Publishing, with permission from H.-J. Ehrlein.

As the antral wave advances toward the pylorus, the pylorus is open, allowing some fluid and solid particles less than 1 mm in diameter to pass into the duodenum. Just before the antral wave reaches the pylorus, the latter closes, ceasing transpyloric flow. Pyloric closure is followed immediately by occlusion of the terminal antrum, generating pressures of more than 100 mmHg, and producing a forceful retropulsion of solid particles, which both grinds and mixes them with gastric juice. Relaxation of the terminal antrum is followed by pyloric opening and emptying of liquid and suspended particles, whereas the subsequent peristaltic wave is at the proximal antrum (Fig. 2).

Phasic and tonic pyloric contractions occur either in isolation or are coordinated with antral contractions and appear to control transpyloric flow by acting primarily as a brake, allowing flow to occur in a series of small gushes only when the pylorus is open. The mechanical determinants of individual flow episodes are still poorly defined. Flow could potentially reflect a local increase in antroduodenal pressure differences due to peristaltic antral contractions, or may be associated with a common-cavity pressure difference between the distal antrum and proximal duodenum during periods of relative antral motor quiescence, when either antegrade or retrograde flows could occur.

As already described, postprandial duodenal motility is characterized by a pattern of irregular contractile activity, consisting of isolated pressure waves that mix chyme with intestinal secretions, and some pressure wave sequences that move the chyme slowly more distally. In this way, absorption and the interaction between nutrients and small intestinal receptors are optimized.

NEURAL INNERVATION OF THE STOMACH

Digestive functions, including gastric motility, are controlled by a number of neural networks that are located in the wall of the gastrointestinal tract, prevertebral ganglia, spinal cord, and brain. These neural networks are fundamental to the integration of the functions of the proximal and distal stomach and the pylorus.

Gastric Electrical Pacemaker

The electrical stimulus for gastric (as well as small intestinal) smooth muscle contraction is provided by cells recently identified as a subtype of interstitial cells of Cajal. The latter generate so-called slow waves, or pacemaker potentials, and are located in the myenteric plexus (i.e., between the longitudinal and circular muscle layers) in the upper body. Slow waves, continuous, rhythmic changes in the membrane potential occurring in the human stomach at a maximum frequency of 3/minute, do not per se initiate contractions. The latter are generated by action potentials (as a result of rapid membrane depolarization), superimposed on the partial depolarization of the slow wave. Accordingly, phasic contractions in the antrum and pylorus are associated temporarily with the rhythm and propagation of the gastric pacemaker potential when stimuli enhance contractility above the basal level; the maximum contraction frequency is limited by the frequency of the slow wave, i.e., 3/minute in the stomach. Both neural and hormonal inputs can modulate the duration and the amplitude of the action potential, i.e., increase it above, or maintain it below, the threshold required for the triggering of a contraction.

Intrinsic Innervation of the Stomach

The enteric nervous system (ENS), also termed the "brain of the gut," because it is able to function without input from the central nervous system, is located within the walls of the stomach and small intestine. It consists of two systems of nerve plexus—the myenteric (or Auerbach) plexus, located between the circular and longitudinal muscles, and the submucosal (or Meissner) plexus, located beneath the mucosa. Within the ENS, there are three major types of neurons. Sensory neurons contain specialized receptors that detect changes in the environment (tension, contraction, chemical stimuli), which are then transformed into electrical signals, so-called action potentials, and transmitted along sensory nerve fibers to other parts of the nervous system. Interneurones form networks that process sensory information and control the behaviour of motor neurones, which are the final common pathway for the transmission of signals to the effector systems. Both excitatory (i.e., stimulating muscle contraction) and inhibitory (i.e., suppressing muscle contractions/stimulating muscle relaxation) exist. The major excitatory transmitters are acetylcholine and substance P, and the major inhibitory neurotransmitters are vasoactive intestinal peptide (VIP) and nitric oxide (NO). The enteric nervous system, thus, comprises local circuitries for the performance of integrative functions independent of extrinsic innervation.

Extrinsic Innervation of the Stomach

The stomach, like other parts of the gastrointestinal tract, is innervated by extrinsic nerves—the sympathetic and parasympathetic parts of the autonomic nervous system and sensory nerves that project to the

spinal cord (splanchnic and sacral afferents) and to the brain stem (vagal afferents). Gastric motility is controlled predominantly by the vagus nerve. Vagal afferents project to the nucleus tractus solitarius (nTS), where they form synapses with interneurons that project to the dorsal motor nucleus of the vagus (DMNV) and to higher brain centers. From the DMNV, efferent projections return to the stomach, modulating the activity of the muscle cells ("effector system") through activation of either inhibitory or excitatory motor neurons. This circuit is termed a "vago-vagal reflex." Vago-vagal reflexes are also modulated by input from other brain regions, including the forebrain and the area postrema (AP).

SMALL INTESTINAL REGULATION OF GASTRIC MOTILITY

Both nutrients and hormones, released in response to the presence of nutrients in the small intestine, modulate gastric motility.

Effects of Nutrients

The presence of dietary nutrients and osmotically active substances in the small intestine has pronounced effects on gastric and pyloric motility; the effect of nutrients is related to their energy content and the length / region of small intestine that is exposed to the nutrients. Direct infusion of nutrients into the small intestine is associated with relaxation of the proximal stomach, suppression of antral pressure waves, stimulation of tonic and phasic pyloric pressures, and slowing of gastric emptying (Fig. 3).

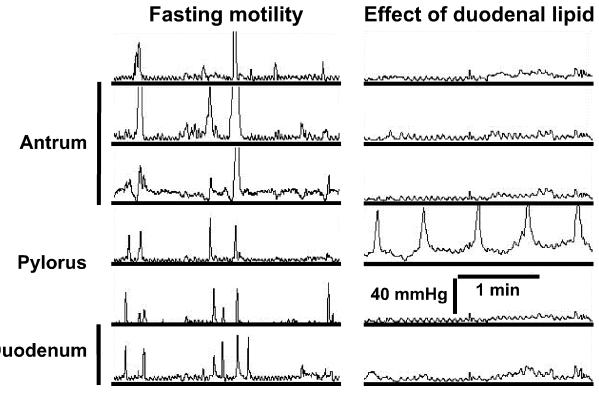


FIGURE 3 The effect on antropyloroduodenal motility of a duodenal lipid infusion. Duodenal lipid is associated with suppression of antral and duodenal pressure waves and stimulation of tonic and phasic pyloric pressures.

Humoral Mechanisms

Food ingestion stimulates the secretion of a large number of gastrointestinal hormones, which have modulatory effects on gastric motility. For example, fat and protein ingestion predominantly stimulates cholecystokinin (CCK) secretion and carbohydrate and fat ingestion stimulates glucagon-like peptide-1 (GLP-1) secretion. Exogenous administration of these peptides in apparently "physiological" concentrations induces relaxation of the proximal stomach, suppression of antral pressure waves, and stimulation of tonic and phasic pyloric pressures. Receptor antagonists for some of these peptides are available for use in humans. For example, the CCK-A receptor antagonist, loxiglumide, attenuates the effects of duodenal fat on the stomach and pylorus, suggesting that CCK mediates, at least in part, the effects of fat on gastric motility.

METHODS FOR THE ASSESSMENT OF GASTRIC MOTILITY

A number of techniques are available for the assessment of different aspects of gastric motility. According to their applications, they can be divided into techniques to assess gastric electrical activity (not discussed further here), gastric wall motion, proximal gastric relaxation, and intraluminal phasic pressures (Table I). In research studies, a number of techniques are frequently used concurrently. However, in contrast to the assessment of gastric emptying, a number of these techniques, including magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), and three-dimensional ultrasound, are currently too expensive and labor intensive to be used in the clinical setting.

Assessment of Gastric Wall Motion

Techniques to assess gastric wall motion include scintigraphy, ultrasound, and magnetic resonance imaging (Fig. 4). Scintigraphy, the "gold standard" technique to quantify gastric emptying, also allows an assessment of intragastric meal distribution and antral contractile activity. The clinical utility of this latter technique remains to be established. Ultrasonography is well suited to evaluate gastric wall motion and intragastric volume. The recent development of three-dimensional ultrasound techniques has advanced the capacity to evaluate relationships between antral area/volume and contractile activity, transpyloric flow, and duodenal contractile activity concurrently. MRI has recently been used to detect peristaltic occlusive and nonocclusive movements of the gastric wall. However, due to its limited accessibility and high cost, it is unlikely that MRI

TABLE I Techniques for Assessment of Gastrointestinal Motility

Technique	Parameters assessed
Electrogastrography	Myoelectrical activity
Scintigraphy	Gastric emptying, intragastric distribution, and antral contractions
Ultrasonography	Antral area (two dimensional), volume (three dimensional), antropyloric contractions, and pyloric flow
Magnetic resonance imaging	Gastric emptying, intragastric distribution, and antral contractions
Single-photon emission computed tomography	Gastric volume (relaxation)
Barostat	Proximal gastric volume (relaxation) and gastric pressure–volume relationship
Intraluminal manometry	Lumen-occlusive contractions

will be used clinically for this purpose in the foreseeable future.

Assessment of Gastric Relaxation

Until the advent of the barostat, no technique was available to quantify relaxation of the proximal stomach. The barostat consists of a pressure transducer linked by an electronic relay to an air injection system. An infinitely compliant bag, located in the proximal stomach, is then connected via a double-lumen tube to the barostat. Once a pressure is set in the system (frequently 2 mmHg above basal intragastric pressure), the barostat can indirectly measure gastric relaxation by monitoring changes in intrabag volume at that pressure. Thus, when the stomach relaxes, air is injected into the gastric bag to maintain the pressure, while air is withdrawn when the stomach contracts. The barostat has significant limitations, because the air-filled bag in the proximal stomach disturbs normal physiology—for example, the presence of the bag affects antral motility. More recently, other techniques, such as single-photon emission computed tomography imaging, three-dimensional ultrasound, and MRI have been used to quantify proximal gastric motility in research studies.

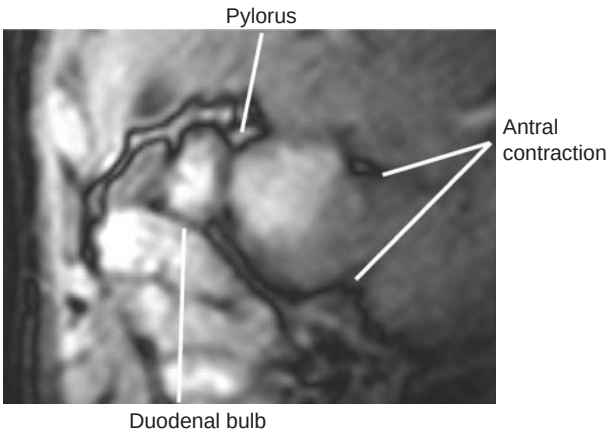


FIGURE 4 Magnetic resonance imaging scan of the antropyloroduodenal region. Wall motion in the antrum, reflecting an antral contraction, is clearly visible.

Measurement of Antropyloroduodenal Phasic Contractions

Manometry assesses intraluminal pressure changes due to contractile activity of the antropyloroduodenal region. In order to measure pressures over a segment of stomach or small intestine, a multilumen catheter with multiple, closely spaced side holes is used. Although both solid-state and water-perfused manometric catheters are potentially applicable, in most cases perfused systems are used; the system is connected to a pump with a water reservoir and perfused continuously. In order to be reliably detected, contractions must be lumen occlusive and occur at the site of a side hole. It is important to recognize that patterns of muscular contractions are not synonymous with patterns of lumen occlusion due to the variation in luminal diameters. For example, a recent study demonstrated that only 54% of contractile events, as recorded by MRI, were associated with a detectable pressure event, as assessed by manometry. It is impossible to position a recording side hole in the pylorus accurately, because it is a narrow and mobile structure, hence the so-called sleeve sensor or multiple closely spaced side holes are required to record pyloric pressures accurately. The sleeve sensor is a perfused channel in the manometric catheter that records the highest pressure point along its length. By spanning the length of the sphincter, the sleeve sensor is able to maintain contact to sense the maximum pressure reliably even if the catheter moves slightly.

PATHOPHYSIOLOGY OF GASTRIC MOTILITY

A number of acute and chronic conditions of diverse etiology are associated with disordered gastric emptying/gastroduodenal motility. The underlying pathology may potentially involve the gastric musculature, its extrinsic or intrinsic nervous innervation, or the interstitial cells of Cajal, although the number of histopathological studies has been limited. Increased small intestinal feedback could potentially also lead to slow gastric emptying.

Motor Dysfunctions Responsible for Slow Gastric Emptying

Abnormally slow gastric emptying, or gastroparesis, may theoretically result from defective mechanical breakdown of food, ineffective propulsion of intragastric content, and/or an abnormally high resistance to emptying. Postsurgical gastroparesis is associated with a reduction in gastric motility, particularly contractile activity of the antrum. The motor dysfunctions responsible for gastroparesis are poorly characterized; in part, this reflects the substantial technical difficulties associated with measurement of gastric motility—in most studies, only one or two motor components (most frequently the antrum) have been evaluated. It is, however, clear that, in many cases, the motor abnormalities are heterogeneous and that the organization of gastroduodenal contractile activity is frequently impaired. Proximal gastric motility is abnormal in many patients with impairment of gastric relaxation induced by a meal. Both fasting and postprandial antral hypomotility occur frequently, particularly a reduction in the number of antral pressure waves that are temporarily associated with duodenal waves. Increased pyloric motility does not appear to be a major factor contributing to gastroparesis. Abnormal proximal small intestinal motor function also occurs frequently, but has been poorly characterized. A knowledge of the motor dysfunctions responsible for disordered gastric emptying has potential implications for the efficacy of pharmacotherapy.

CONCLUSION

The motor activity of the stomach is characterized by two distinct motor patterns, fasting and fed, with markedly differing functions. Gastric motor patterns are governed by both extrinsic and intrinsic innervation and modulated by a number of factors, including nutrients and humoral factors. Abnormal gastric emptying is associated with a number of gastric motor dysfunctions.

See Also the Following Articles

Barostat • Basic Electrical Rhythm • Duodenal Motility • Enteric Nervous System • Gastric Emptying • Interstitial Cells of Cajal • Manometry • Migrating Motor Complex • Postprandial Motility • Pylorus • Small Intestinal Motility • Vagus Nerve

Further Reading

Azpiroz, F., and Malagelada, J. R. (1985). Intestinal control of gastric tone. *Am. J. Physiol.* **249**, G501–G509.

Azpiroz, F., and Malagelada, J. R. (1985). Physiological variations in canine gastric tone measured by an electronic barostat. *Am. J. Physiol.* **248**, G229–G237.

Berstad, A., Hausken, T., Gilja, O. H., Thune, N., Matre, K., and Odegaard, S. (1994). Volume measurements of gastric antrum by 3-D ultrasonography and flow measurements through the pylorus by duplex technique. *Dig. Dis. Sci.* **39**, 97S–100S.

Bühner, S., and Ehrlein, H. J. (1989). Characteristics of postprandial duodenal motor patterns in dogs. *Dig. Dis. Sci.* **34**, 1873–1881.

Camilleri, M., Brown, M. L., and Malagelada, J. R. (1986). Relationship between impaired gastric emptying and abnormal gastrointestinal motility. *Gastroenterology* **91**, 94–99.

Cannon, W. B. (1898). The movements of the stomach studied by means of the Roentgen rays. *Am. J. Physiol.* **1**, 359–382.

Cannon, W. B., and Lieb, C. W. (1911). The receptive relaxation of the stomach. *Am. J. Physiol.* **29**, 267–273.

Code, C. F., and Marlett, J. A. (1975). The interdigestive myoelectric complex of the stomach and small bowel of dogs. *J. Physiol.* **246**, 289–309.

Dent, J. (1976). A new technique for continuous sphincter pressure measurement. *Gastroenterology* **71**, 263–267.

De Wever, I., Eeckhout, C., Vantrappen, G., and Hellemans, J. (1978). Disruptive effect of test meals on interdigestive motor complex in dogs. *Am. J. Physiol.* **235**, E661–E665.

Faas, H., Hebbard, G. S., Feinle, C., Kunz, P., Brasseur, J. G., Indireshkumar, K., Dent, J., Boesiger, P., Thumshirn, M., Fried, M., and Schwizer, W. (2001). Pressure-geometry relationship in the antroduodenal region in humans. *Am. J. Physiol.* **281**, G1214–G1220.

Feinle, C., D'Amato, M., and Read, N. W. (1996). Cholecystokinin-A receptors modulate gastric sensory and motor responses to gastric distension and duodenal lipid. *Gastroenterology* **110**, 1379–1385.

Feinle, C., Grundy, D., and Read, N. W. (1997). Effects of duodenal nutrients on sensory and motor responses of the human stomach to distension. *Am. J. Physiol.* **273**, G721–G726.

Gilja, O. H., Hausken, T., Wilhelmsen, I., and Berstad, A. (1996). Impaired accommodation of proximal stomach to a meal in functional dyspepsia. *Dig. Dis. Sci.* **41**, 689–696.

Heddle, R., Dent, J., Read, N. W., Houghton, L. A., Toouli, J., Horowitz, M., Maddern, G. J., and Downton, J. (1988). Antropyloroduodenal motor responses to intraduodenal lipid infusion in healthy volunteers. *Am. J. Physiol.* **254**, G671–G679.

Huizinga, J. D., Berezin, I., Chorneyko, K., Thuneberg, L., Sircar, K., Hewlett, B. R., and Riddell, R. H. (1998). Interstitial cells of Cajal: Pacemaker cells? *Am J. Pathol.* **153**, 2008–2011.

Indireshkumar, K., Brasseur, J. G., Faas, H., Hebbard, G. S., Kunz, P., Dent, J., Feinle, C., Li, M., Boesiger, P., Fried, M., and Schwizer, W. (2000). Relative contributions of "pressure pump" and "peristaltic pump" to gastric emptying. *Am. J. Physiol.* **278**, G604–G616.

Jones, K., Edelbroek, M., Horowitz, M., Sun, W. M., Dent, J., Roelofs, J., Muecke, T., and Akkermans, L. (1995). Evaluation of antral motility in humans using manometry and scintigraphy. *Gut* **37**, 643–648.

Katschinski, M., Schirra, J., Begliner, C., Langbein, S., Wank, U., D'Amato, M., and Arnold, R. (1996). Intestinal phase of human antro-pyloro-duodenal motility: Cholinergic and CCK-mediated regulation. *Eur. J. Clin. Invest.* **26**, 574–583.

Kerlin, P. (1989). Postprandial antral hypomotility in patients with idiopathic nausea and vomiting. *Gut* **30**, 54–59.

Kuiken, S. D., Samsom, M., Camilleri, M., Mullan, B. P., Burton, D. D., Kost, L. J., Hardyman, T. J., Brinkmann, B. H., and O'Connor, M. K. (1999). Development of a test to measure gastric accommodation in humans. *Am. J. Physiol.* **277**, G1217–G1221.

Pröve, J., and Ehrlein, H. J. (1982). Motor function of gastric antrum and pylorus for evacuation of low and high viscosity meals in dogs. *Gut* **23**, 150–156.

Rao, S. S. C., and Schulze-Delrieu, K. (1993). The stomach, pylorus and duodenum. *In* "An Illustrated Guide to Gastrointestinal Motility, 2nd Ed. (D. Kumar and D. Wingate, eds.), pp. 373–392. Churchill Livingstone, Edinburgh.

Read, N. W., and Houghton, L. A. (1989). Physiology of gastric emptying and pathophysiology of gastroparesis. *Gastroenterol. Clin. North Am.* **18**, 359–373.

Wood, J. D., Alpers, D. H., and Andrews, P. L. (1999). Fundamentals of neurogastroenterology. *Gut* **45**, (Suppl. 2) II6–II16.

Gastric Outlet Obstruction

RICHARD H. TURNAGE, LESTER W. JOHNSON, AND QUYEN CHU

Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center, Shreveport

gastric outlet obstruction A mechanical obstruction of the stomach as it empties into the duodenum or through a surgically created stoma.

proximal gastric vagotomy Also known as highly selective vagotomy or parietal cell vagotomy; an operation in which the vagal branches of the nerves of Laterjet are divided as they enter the lesser curvature of the stomach. This ameliorates vagally mediated gastric acid secretion from the parietal cell mass while preserving the motor branches to the antrum and the pylorus.

Gastric outlet obstruction occurs due to obstruction at or near the pylorus or a surgically created gastroenterostomy. By far, the most common causes of gastric outlet obstruction are peptic ulcer disease and adenocarcinoma of the gastric antrum. This article discusses the causes of this condition, its clinical manifestations, and the diagnostic strategies and potential therapeutic modalities involved in treating these patients.

ETIOLOGY

Since the 1970s, gastric adenocarcinoma has supplanted peptic ulcer disease as the most common cause of gastric outlet obstruction. Unusual causes of gastric outlet obstruction include benign neoplasms of the pylorus and stomach, Crohn's disease of the stomach or duodenum, adult hypertrophy of the pylorus, heterotopic pancreatic tissue at the pylorus, and fibrous intra-abdominal adhesions. Peptic ulcer disease associated with gastric outlet obstruction is usually localized to the duodenum, but pyloric channel and gastric antral ulcers may also cause obstruction. In these instances, an active ulcer crater is associated with acute edema, inflammation, and pylorospasm. Rarely is the obstruction due solely to the presence of a fixed fibrous scar. This explains those occasions in which medical management of the ulcer results in resolution of obstructive symptoms. Pyloric channel ulcers are often associated with gastric retention; in one early study, nearly 50% of the patients with ulcers within the pyloric channel had symptoms of gastric outlet obstruction.

Adenocarcinoma, as well as other gastric malignancies such as lymphoma, may present with signs and symptoms of gastric outlet obstruction. Differentiation of a benign ulcer from a malignant gastric ulcer may be difficult on endoscopic (or radiographic) inspection and

 Copyright 2004, Elsevier (USA). All rights reserved.

Indireshkumar, K., Brasseur, J. G., Faas, H., Hebbard, G. S., Kunz, P., Dent, J., Feinle, C., Li, M., Boesiger, P., Fried, M., and Schwizer, W. (2000). Relative contributions of "pressure pump" and "peristaltic pump" to gastric emptying. *Am. J. Physiol.* **278**, G604–G616.

Jones, K., Edelbroek, M., Horowitz, M., Sun, W. M., Dent, J., Roelofs, J., Muecke, T., and Akkermans, L. (1995). Evaluation of antral motility in humans using manometry and scintigraphy. *Gut* **37**, 643–648.

Katschinski, M., Schirra, J., Begliner, C., Langbein, S., Wank, U., D'Amato, M., and Arnold, R. (1996). Intestinal phase of human antro-pyloro-duodenal motility: Cholinergic and CCK-mediated regulation. *Eur. J. Clin. Invest.* **26**, 574–583.

Kerlin, P. (1989). Postprandial antral hypomotility in patients with idiopathic nausea and vomiting. *Gut* **30**, 54–59.

Kuiken, S. D., Samsom, M., Camilleri, M., Mullan, B. P., Burton, D. D., Kost, L. J., Hardyman, T. J., Brinkmann, B. H., and O'Connor, M. K. (1999). Development of a test to measure gastric accommodation in humans. *Am. J. Physiol.* **277**, G1217–G1221.

Pröve, J., and Ehrlein, H. J. (1982). Motor function of gastric antrum and pylorus for evacuation of low and high viscosity meals in dogs. *Gut* **23**, 150–156.

Rao, S. S. C., and Schulze-Delrieu, K. (1993). The stomach, pylorus and duodenum. *In* "An Illustrated Guide to Gastrointestinal Motility, 2nd Ed. (D. Kumar and D. Wingate, eds.), pp. 373–392. Churchill Livingstone, Edinburgh.

Read, N. W., and Houghton, L. A. (1989). Physiology of gastric emptying and pathophysiology of gastroparesis. *Gastroenterol. Clin. North Am.* **18**, 359–373.

Wood, J. D., Alpers, D. H., and Andrews, P. L. (1999). Fundamentals of neurogastroenterology. *Gut* **45**, (Suppl. 2) II6–II16.

Gastric Outlet Obstruction

RICHARD H. TURNAGE, LESTER W. JOHNSON, AND QUYEN CHU

Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center, Shreveport

gastric outlet obstruction A mechanical obstruction of the stomach as it empties into the duodenum or through a surgically created stoma.

proximal gastric vagotomy Also known as highly selective vagotomy or parietal cell vagotomy; an operation in which the vagal branches of the nerves of Laterjet are divided as they enter the lesser curvature of the stomach. This ameliorates vagally mediated gastric acid secretion from the parietal cell mass while preserving the motor branches to the antrum and the pylorus.

Gastric outlet obstruction occurs due to obstruction at or near the pylorus or a surgically created gastroenterostomy. By far, the most common causes of gastric outlet obstruction are peptic ulcer disease and adenocarcinoma of the gastric antrum. This article discusses the causes of this condition, its clinical manifestations, and the diagnostic strategies and potential therapeutic modalities involved in treating these patients.

ETIOLOGY

Since the 1970s, gastric adenocarcinoma has supplanted peptic ulcer disease as the most common cause of gastric outlet obstruction. Unusual causes of gastric outlet obstruction include benign neoplasms of the pylorus and stomach, Crohn's disease of the stomach or duodenum, adult hypertrophy of the pylorus, heterotopic pancreatic tissue at the pylorus, and fibrous intra-abdominal adhesions. Peptic ulcer disease associated with gastric outlet obstruction is usually localized to the duodenum, but pyloric channel and gastric antral ulcers may also cause obstruction. In these instances, an active ulcer crater is associated with acute edema, inflammation, and pylorospasm. Rarely is the obstruction due solely to the presence of a fixed fibrous scar. This explains those occasions in which medical management of the ulcer results in resolution of obstructive symptoms. Pyloric channel ulcers are often associated with gastric retention; in one early study, nearly 50% of the patients with ulcers within the pyloric channel had symptoms of gastric outlet obstruction.

Adenocarcinoma, as well as other gastric malignancies such as lymphoma, may present with signs and symptoms of gastric outlet obstruction. Differentiation of a benign ulcer from a malignant gastric ulcer may be difficult on endoscopic (or radiographic) inspection and

 Copyright 2004, Elsevier (USA). All rights reserved.

is dependent on histologic examination of biopsy specimens of the margins of the ulcer.

CLINICAL PRESENTATION

The principal symptoms of gastric outlet obstruction are nausea and vomiting, which often develop over a period of weeks to months. The vomiting is often copious in nature and consists of poorly digested food. Many patients with gastric outlet obstruction from ulcer disease have a long history of ulcer-related pain; in several early studies, patients had symptoms of ulcer disease for nearly 10 years. Recent onset of upper abdominal pain or obstruction in the absence of abdominal pain should raise the possibility of a malignant etiology. Weight loss and malnutrition are common findings in patients presenting with either benign or malignant obstruction. Physical examination may identify a gastric succussion splash, upper abdominal distension (from a markedly dilated stomach), and perhaps visible gastric peristalsis. Physical evidence of poor nutrition and intravascular volume depletion may also be apparent. Intubation of the stomach with a nasogastric tube will yield at least 500 ml, and often significantly greater quantities, of fluid and undigested particulate matter.

Patients with gastric outlet obstruction will often be depleted of intravascular volume and have electrolyte disturbances associated with the loss of large volumes of gastric juice. Gastric juice in patients with gastric outlet obstruction contains approximately 100 mEq per liter of chloride, 45 mEq per liter of sodium, and 10 mEq per liter of potassium. The electrolyte disturbance associated with gastric outlet obstruction consists initially of a decrease in plasma chloride and an increase in bicarbonate. Over time, a sodium deficit develops (due to sodium losses in the vomitus and its enhanced excretion into the urine) and on depletion of the sodium reserves, hydrogen and potassium ions are excreted with the bicarbonate as sodium is preserved. This results in the paradoxical excretion of acidic urine in the setting of a metabolic alkalosis. The classic electrolyte disorder of patients with long-standing gastric outlet obstruction is a hypochloremic, hypokalemic metabolic alkalosis.

DIAGNOSTIC STRATEGY

Aspiration of gastric contents through a nasogastric tube will reveal substantial quantities of retained fluid. A gastric aspirate volume of more than 300 ml 4 h after a meal or greater than 200 ml after an overnight fast suggests impaired gastric emptying. In the past, the diagnosis was suspected based on a saline load test in which 750 ml of saline was placed into the stomach through a nasogastric tube; aspiration of more than 300 ml of fluid 30 min after instillation suggests delayed gastric emptying.

Mechanical obstruction of the gastric outlet may be confirmed with endoscopy and/or barium contrast radiographs. Endoscopy will allow visualization of the site of obstruction and often provide evidence of its etiology. This procedure may be helpful in differentiating a mechanical outlet obstruction from a functional obstruction (e.g., postvagotomy or postresection gastric stasis or diabetic gastroparesis). The importance of biopsy cannot be overestimated in differentiating benign from malignant gastric ulcers. Prior to this procedure, retained gastric contents should be evacuated with a large-bore gastric tube (e.g., an Ewald tube) to allow visualization of the gastric mucosa.

A plain abdominal radiograph may suggest gastric stasis by demonstrating a large gastric shadow with retained food particles that appear as patchy translucencies. An upright roentgenogram will show a high gastric air—fluid level. A barium contrast study of the stomach also may confirm the presence of outlet obstruction and provide evidence of its etiology. For example, the normal stomach will largely empty in approximately 2 h with a small residual amount of barium present after 4 to 6 h. In instances of gastric outlet obstruction, more than 50% of the barium will be retained within the stomach 4 to 6 h after the barium is swallowed. A barium contrast study may elucidate postoperative anatomy in patients who previously have undergone gastric operations for ulcer disease. Barium studies are occasionally helpful in differentiating a benign from a malignant ulcer or in demonstrating ulcer disease in the presence of postoperative gastric deformities.

TREATMENT

Many patients with gastric outlet obstruction require operative management. In general, this is preceded by a period of meticulous medical preparation, the aims of which are decompression of the dilated stomach, healing of the ulcer, correction of fluid and electrolyte abnormalities, and improvement in the general condition of the patient. A nasogastric tube is placed at the time of admission to decompress the stomach; a large-bore gastric tube may be necessary to remove large particulate matter prior to diagnostic studies. Intravenous fluid resuscitation with normal saline and potassium chloride will correct the intravascular volume deficit and replete the total body sodium, chloride, and potassium pools. Patients should receive appropriate medical treatment for peptic ulcer disease with either a proton pump

inhibitor or a histamine-2 receptor antagonist, given intravenously until the nasogastric tube is removed. Those patients with evidence of *Helicobacter pylori* infection should receive appropriate antibiotic therapy. Most clinicians will begin parenteral nutrition given the incidence of malnutrition in these patients and the often prolonged period of time prior to the patient again being able to tolerate oral feedings. Within a few days of this treatment, most patients will demonstrate significant improvement in their gastric emptying with a reduction in the volume of aspirate from their nasogastric tube. This will allow diagnostic studies and planning of definitive treatment.

Patients suffering gastric outlet obstruction from a distal gastric cancer should undergo operative treatment. The only chance for cure comes from total resection with or without adjuvant chemoradiation therapy (depending on the stage of the disease). Patients with more advanced disease may be palliated by the endoscopic placement of a stent across the obstruction or by the performance of a gastrojejunostomy at a site well proximal to the malignancy.

Some patients with gastric outlet obstruction from benign ulcer disease will respond to medical management of the ulcer alone. The long-term efficacy of medical management combined with endoscopic dilation of the stricture is unclear. Two recent reports suggest that nearly 70% of patients with gastric outlet obstruction due to peptic ulcer disease could be successfully managed medically along with endoscopic dilation. Kozarek reported that 16 of 23 patients undergoing balloon dilation of gastric outlet obstruction were asymptomatic up to 2.5 years later. However, medical management alone has a very high failure rate on long-term follow-up. Jaffin, for example, reported that 92% of patients with gastric outlet obstruction who were treated medically required operative management for recurrent or persistent symptoms within 3 years.

If surgery is required due to failure of medical/dilation therapy, the best operation for gastric outlet obstruction from benign ulcer disease is debated. Several papers, including one prospective randomized trial, demonstrated that proximal gastric vagotomy and gastrojejunostomy was superior to vagotomy and antrectomy or proximal gastric vagotomy and gastroduodenostomy. Others have reached similar conclusions, based on retrospective comparisons, that proximal gastric vagotomy and pyloroplasty is superior to vagotomy and antrectomy or truncal vagotomy and pyloroplasty. Donahue *et al*. reported that 93% of their patients undergoing proximal gastric vagotomy with drainage did well with no recurrent ulcers, no significant gastric atony, and minimal postoperative symptoms for as long as 4.6 years following surgery. Of note, there is significant experience to suggest that proximal gastric vagotomy with intraoperative dilation of the stricture alone is associated with an extremely high incidence of recurrent obstruction.

See Also the Following Articles

Duodenal Obstruction • Duodenal Ulcer • Gastric Surgery • Gastric Ulcer • Gastroenterostomy • Pyloric Stenosis • Pyloroplasty • Pylorus • Stomach, Adenomas and Carcinomas of the

Further Reading

Csendes, A., Maluenda, F., Braghetto, I., Schutte, H., Burdiles, P., and Diaz, J. C. (1993). Prospective randomized study comparing three surgical techniques for the treatment of gastric outlet obstruction secondary to duodenal ulcer. *Am. J. Surg.* **166**, 45–49.

Donohue, P. E., Yoshida, J., Richter, H. M., Liu, K., Bombeck, C. T., and Nyhus, L. M. (1988). Proximal gastric vagotomy with drainage for obstructing duodenal ulcer. *Surgery* **104**, 757–764.

Gleysteen, J. J., and Droege, E. A. (1988). Expedient surgical treatment of chronic ulcer stenosis: A case for proximal gastric vagotomy. *J. Clin. Gastroenterol.* **10**, 619–622.

Jaffin, B. W., and Kaye, M. D. (1985). The prognosis of gastric outlet obstruction. *Ann. Surg.* **201**, 176–179.

Kozarek, R. A., Botoman, V. A., and Patterson, D. J. (1990). Long-term follow-up in patients who have undergone balloon dilation for gastric outlet obstruction. *Gastrointest. Endosc.* **36**, 558–561.

Gastric Polyps

STEFFEN MUEHLDORFER*, CHRISTIAN ELL†, AND MANFRED STOLTE‡

University of Erlangen, Nuremberg, †Horst-Schmidt-Kliniken, Wiesbaden, and ‡Institute of Pathology, Bayreuth, Germany

gastric polyp Benign or malignant lesion in the stomach that is elevated above the surrounding gastric mucosa.

polyposis Condition in which numerous polyps are present, usually due to a genetic defect.

Recommendations for diagnosis and treatment of gastric polyps remain controversial; no consensus exists regarding their malignant potential or the reliability of endoscopic forceps biopsy. According to the classification of gastric tumors and polyps established by the World Health Organization, the frequency of malignant transformation depends on histologic type. The cancer risk has been reported to range from 0 to 8.6% (average 2.1%) for hyperplastic polyps, to be approximately 5% for tubular adenomas, and to range from 28.5 to 40% for villous adenomas and pyloric gland adenomas. Therefore, a precise histologic classification is essential to therapeutic action. Especially important is the reliability of the histologic rating of forceps biopsy sampling with regard to the total polyp.

TABLE I Incidence of Various Gastric Polyps[a]

Type	Number	Incidence
Neoplasias		
Tubular adenoma	600	(16.7%)
Tubulopapillary adenoma	77	(2.2%)
Papillary adenoma	7	(0.2%)
Pyloric gland adenoma	13	(0.4%)
Adenocarcinoma	423	(11.8%)
Carcinoid tumor	95	(2.6%)
Total	1215	(33.9%)
Tumorlike lesions		
Hyperplastic polyp	2036	(56.7%)
Inflammatory fibroid polyp	192	(5.4%)
Heterotopia	111	(3.1%)
Peutz–Jeghers polyp	21	(0.6%)
Juvenile polyp	11	(0.3%)
Cronkhite–Canada polyp	2	(0.1%)
Total	2373	(66.1%)

[a] $n = 3588$; includes polyps seen in patients at the Institute of Pathology, Bayreuth, in the period 1969–1992, excluding fundic gland polyps, mesenchymal tumors, and polypous lymphomas.

EPIDEMIOLOGY

Gastrointestinal polyps are by definition lesions elevated above the surrounding mucosa. In the stomach, they are found in 1–5% of all gastroscopic examinations. The reported incidence varies because fundic gland polyps, which comprise of about 50% of all polyps, have been only inconsistently included in the surveys. More than one polyp can be found in around 25% of the patients. One of the largest histopathologic series evaluated almost 3600 gastric polyps. The absolute numbers and relative percentages of the various types of gastric polyps, excluding fundic gland polyps, are shown in Table I. Concerning the location of gastric polyps, it has been shown that pancreatic and Brunner's gland heterotopia as well as inflammatory fibroid polyps arise mainly in the antrum, whereas carcinoid tumors are predominantly found in the corpus. Hyperplastic polyps and adenomas distribute rather equally between antrum and corpus.

The age distribution of patients with gastric polyps differs according to the histologic type. Peutz–Jeghers polyps, juvenile polyps, and pancreatic heterotopias are usually diagnosed in patients younger than 50 years, and fundic gland polyps occur most commonly in the age range between 40 and 69 years; all other histologic types are found mainly in the age group 60 years and older.

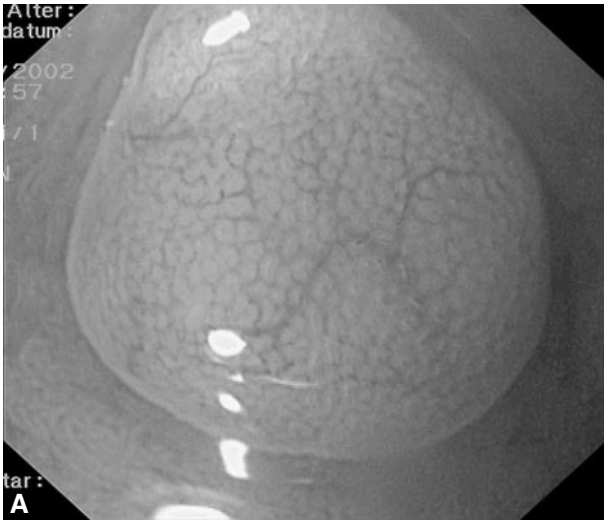
PATHOLOGY

Polyps in the stomach are a heterogeneous group of tumors that comprise nonneoplastic polyps (such as fundic gland polyps), neoplastic polyps (such as adenomas or adenocarcinomas), "reactive" polypoid lesions (such as foveolar hyperplasia), and polypoid intramural masses. In contrast to the situation in the colon, most gastric polyps are nonneoplastic.

 Copyright 2004, Elsevier (USA). All rights reserved.

Fundic gland polyps are the most frequent types of polyps in the stomach (Fig. 1). They can develop after a shift of the proliferation zone into the gland area. Fundic gland polyps are multiple in most cases, are potentially reversible, and carry no neoplastic risk. The additional development of adenomatous changes and dysplasia in fundic gland polyps has been found in 1% of sporadic cases and in 25–44% of patients with familial adenomatous polyposis. Although fundic gland polyps are nonneoplastic, it can be shown that they signal an increased risk in a patient of harboring a colorectal adenoma or carcinoma.

Histologically, hyperplastic polyps are characterized by an elongation, twisting, branching, and cystic dilatation of the foveolae (Fig. 2). The surface is often eroded and the epithelium may show regeneration that



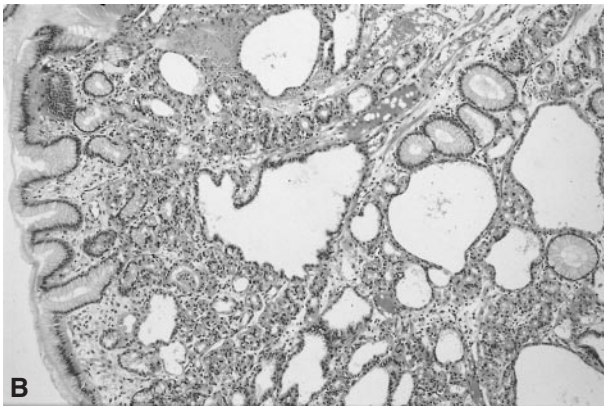
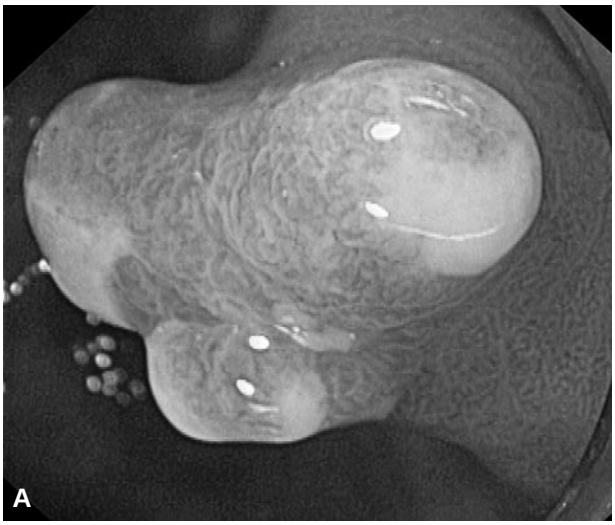


FIGURE 1 Fundic gland polyp. The diagnostic histological finding is the occurrence of cysts within the fundic glands. In this case, most of them are lined by columnar epithelium. (A) Magnification endoscopy (×115). (B) Hematoxylin and eosin stain, original magnification ×400.



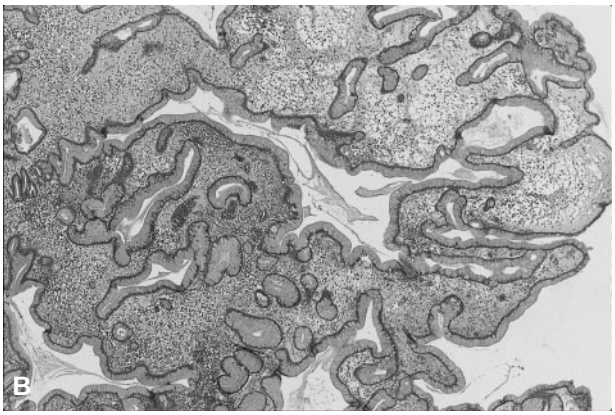


FIGURE 2 Hyperplastic polyp. In the center of the polyp, bordering the luminal area, the epithelial cells show stratification of the nuclei. The glands are cribriform in appearance, suggesting the development of a highly differentiated adenocarcinoma. (A) Hyperplastic polyp with erosions. (B) Hematoxylin and eosin stain, original magnification ×400.

can be misinterpreted as adenomatous or dysplastic changes. They are usually solitary, sessile lesions less than 1.5 cm in size. Focal foveolar hyperplasia is considered a potential precursor of hyperplastic polyps by the World health Organization (WHO). However, this is still a matter of debate. Although some pathologists claim to recognize differences on the basis of architectural and cytological criteria in forceps biopsy material, this does not hold true in common practice. Although the hyperplastic polyp is nonneoplastic, dysplastic changes and/or gastric adenocarcinoma may develop within the lesion. The risk of developing an adenocarcinoma within a hyperplastic polyp ranges from 0 to 8% (mean 2.1%). It has been recently demonstrated that hyperplastic polyps may disappear after *Helicobacter*

pylori eradication. *Helicobacter pylori* eradication therapy may, therefore, be a therapeutic option for hyperplastic polyps occurring in association with *H. pylori* gastritis. Hyperplastic polyps develop in atrophic gastric mucosa in 40–75% of cases and may thus reflect the presence of atrophic gastritis. The risk of patients with multiple hyperplastic polyps of developing a gastric carcinoma appears to be increased and may be as high as 3.6%.

Histologically, adenomas are characterized by columnar epithelium that is pseudostratified and shows elongated atypical nuclei and increased mitotic activity (Fig. 3). Adenomas represent approximately 10% of gastric polyps. Adenomas appear in tubular, tubulovillous, and villous forms or as pyloric gland adenomas.

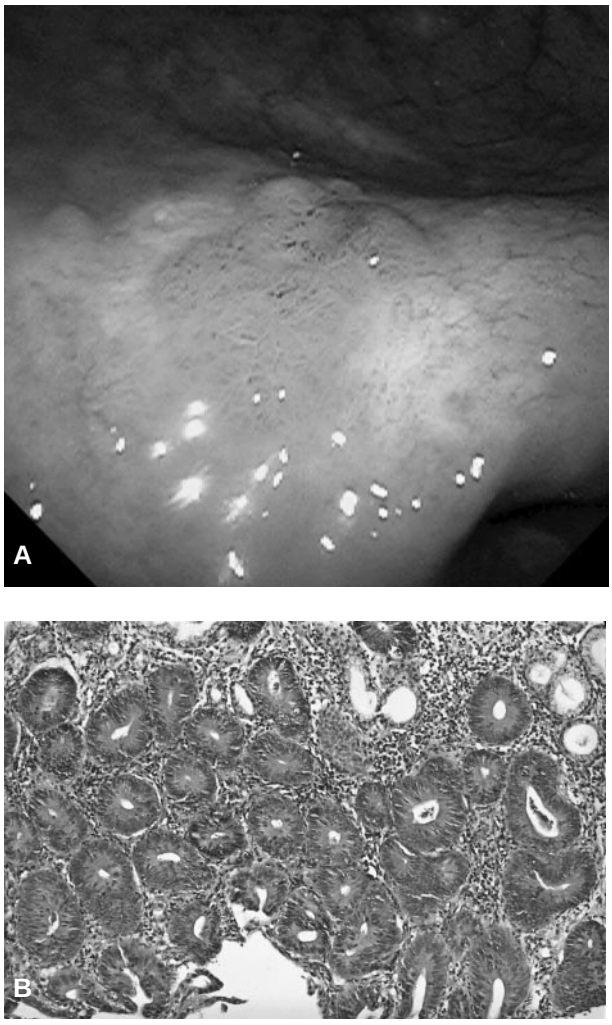


FIGURE 3 A pyloric gland adenoma composed of small tubules lined by gastric-type epithelium. (A) Adenoma in the cardia. (B) Adenoma; hematoxylin and eosin stain, original magnification ×400.

Adenomas are precancerous lesions, comparable to colonic adenomas. Therefore, high-grade dysplasia and frank carcinoma may develop in adenomas. The risk of developing adenocarcinoma in an adenoma correlates with its size and structure. Polyps more than 2 cm in diameter may harbor a focal adenocarcinoma in 40% of cases. Gastric adenomas are also a marker for an increased risk to develop a gastric carcinoma in the remainder of the stomach. According to the literature, coincident gastric carcinomas occur at a rate of 8–59%.

Mesenchymal polypoid tumors are characterized by a spindle cell or epitheloid appearance. Most of them are classified as gastrointestinal stroma tumors (GISTs). Generally, tumors with a diameter less than 5 cm, fewer than two mitoses/10 high-power microscope fields, and a lack of necrosis are held to have no or only a low malignant potential. The remainder are considered high-risk GISTs. Leiomyomas can be observed in 0.9% of gastric resection specimens. Rarely, leiomyosarcomas may develop in lesions greater than 2–3 cm in diameter. When 10 mitotic figures are found per 50 high-power microscope fields, the lesion is considered to be a leiomyosarcoma.

CLINICAL MANIFESTATIONS–DIAGNOSIS

In most cases, polyps are an incidental finding during routine endoscopy, because they only rarely produce symptoms, such as gastrointestinal bleeding with or without anemia or delayed gastric emptying. Bleeding is the consequence of an eroded surface of the polyp. Hyperplastic polyps, inflammatory fibroid polyps, adenocarcinoma, and leiomyomas are predisposed to bleeding. Obstruction of the gastric outlet caused by gastric polyps is a rare event. The symptoms are usually intermittent in case the polyp is pedunculated.

On endoscopy, some polyps can be differentiated by macroscopic characteristics. Fundic gland polyps present as multiple 2- to 3-mm sessile lesions coating the body and fundus; these can be plucked like grapes with the biopsy forceps. They may be difficult to differentiate from carcinoid tumors, which display similar characteristics but have a firm consistency. The presumptive endoscopic diagnosis of gastric tubular adenoma is relatively certain, because this lesion is mainly found in the form of a flat, only slightly elevated, polyp. Hyperplastic polyps are usually softer and more "shiny" compared to other polyps, and their surface is almost always eroded. Although the endoscopic appearance of some polyps may already be diagnostic, the final diagnosis must be based on a histological examination.

MANAGEMENT—TREATMENT

With the introduction of endoscopic polypectomy in the stomach in the early 1970s, an alternative to surgical ablation became available. Because forceps biopsies had been considered unreliable, the procedure enjoyed widespread use. However, complications soon led to a reconsideration of clearing the stomach of all polyps and to the use of forceps biopsy for decision-making.

Because most gastric polyps are not pedunculated, endoscopic mucosal resection techniques must be applied. The most widespread technique is the "strip-off biopsy," which consists of injection of saline in the submucosal layer to form a bleb, with the polyp on top, which is then cut by snare strangulation. Another important technique is endoscopic mucosal resection (EMR) using a transparent plastic cap (EMRC procedure). After a sufficient volume of saline has been injected, the mucosa, including the target lesion, is sucked into a cap attached to the tip of the endoscope, strangulated by a snare wire, and resected by electrocauterization. Endoscopic mucosal resection can generally be performed safely in the stomach because it has a relatively thick muscle layer. But at the lesser curvature in the upper and middle thirds, special attention should be paid to avoid muscle involvement, because stretching of the mucosa is primarily limited. Low-concentration epinephrine saline solution is effective in the control of bleeding during EMR in most cases. However, in case of a spurting bleeding in the stomach, placing a hemostatic clip is the most reliable and safe therapeutic modality to control the bleeding.

In many cases, a meticulous examination of forceps biopsy samples by an experienced pathologist leads to a correct histologic diagnosis without a complete polypectomy. However, foci of carcinoma are present in a small percentage of hyperplastic polyps and may be missed with biopsy sampling. In a recent survey, four such polyps (3.1%) were found in a total of 131 hyperplastic polyps, ranging from 10 to 50 mm (median 15 mm). In only one of these, the malignant focus was detected by forceps biopsy. Since the first description in 1978 of a hyperplastic polyp containing a carcinoma, several case reports have been published. Although at first it was thought that only hyperplastic polyps larger than 20 mm could contain malignant foci, this finding has more recently been observed in polyps of 5 mm.

A recommendation based on these data to remove all stomach polyps completely with the snare classified as hyperplastic on biopsy must be weighed against the possible risks of polypectomy, which seem to be higher than risks of polypectomy in the colon. In a recent prospective study, bleeding occurred in 7.2% of the procedures, but the bleeding could usually be managed endoscopically. This underscores the importance of proficiency in endoscopic hemostatic techniques by the endoscopic team as a prerequisite for performing polypectomy in the stomach. In view of the risk—benefit profile, complete endoscopic removal of even smaller polyps rated as hyperplastic on biopsy must be considered.

An alternative therapeutic option in patients with multiple small hyperplastic polyps or risk factors for polypectomy seems to be the eradication of *H. pylori*. In one study, 35 patients who had hyperplastic polyps at least 3 mm in diameter and evidence of infection with *H. pylori* were randomized to no treatment or to antibiotic therapy aimed at eradicating *H. pylori*. After 12 to 15 months, complete regression of polyps was observed more often in patients in the treated group in whom *H. pylori* was cured.

Regarding focal foveolar hyperplasia, this recommendation should be adapted to the endoscopic appearance of the polyp. For polyps with characteristic macroscopic features, i.e., size < 5 mm, multiple occurrence, and location in the antrum, no further action seems to be justified. For all other polyps rated as focal foveolar hyperplasia on biopsy, endoscopic polypectomy should be considered also, because, in the clinical, setting no unequivocal differentiation from other histological polyp types ,based on forceps biopsy specimens, seems to be possible at present. Of course, the patient's general condition (especially age and concomitant illnesses that might be exacerbated by possible complications) should also be a factor.

No controversy exists for the indication to remove adenomatous polyps, because the adenoma—carcinoma sequence holds true in gastric as well as in colorectal polyps. In one series, the concomitant presence of adenomatous and carcinomatous portions in a single polyp was observed in two specimens. Here—in contrast to carcinomatous foci in hyperplastic polyps—an underestimation on biopsy is not an issue, because all adenomas must be removed completely. Because the malignant potential of mesenchymal tumors is difficult to assess and most of them are not suitable for endoscopic resection, local surgical excision of tumors larger than 3—5 cm, with a margin of approximately 2 cm, is recommended.

Surveillance endoscopy 1 year after removing adenomatous gastric polyps is reasonable to assess recurrence at a prior excision site, new or previously missed polyps, and/or supervening early carcinoma in gastric mucosa apart from the site of coincident polyps. If

this examination is negative, surveillance endoscopy should be repeated no more frequently than 3- to 5-year intervals. No surveillance endoscopy is necessary after removal of nonadenomatous gastric polyps.

See Also the Following Articles

Gastric Cancer Surveillance • *Helicobacter pylori* • Stomach, Adenomas and Carcinomas of the

Further Reading

American Society for Gastrointestinal Endoscopy Guidelines. (1998) The role of endoscopy in the surveillance of premalignant conditions of the upper gastrointestinal tract. *Gastrointest. Endosc.* **48**, 663–668.

Inoue, H., Fukami, N., Yoshida, T., and Kudo, S. (2002). Endoscopic mucosal resection for esophageal and gastric cancers. *J. Gastroenterol. Hepatol.* **17**, 382–388.

Muehldorfer, S. M., Stolte, M., Martus, P., Hahn, E. G., and Ell, C. (2002). Diagnostic accuracy of forceps biopsy versus polypectomy for gastric polyps: A prospective multicentre study. *Gut* **50**, 465–470.

Oberhuber, G., and Stolte, M. (2000). Gastric polyps: An update of their pathology and biological significance. *Virchows Arch.* **437**, 581–590.

Ohkusa, T., Takashimizu, I., Fujiki, K., Suzuki, S., Shimoi, K., Horiuchi, T., *et al.* (1998). Disappearance of hyperplastic polyps in the stomach after eradication of *Helicobacter pylori*. A randomized, clinical trial. *Ann. Intern. Med.* **129**, 712–715.

Stolte, M. (1995). Clinical consequences of the endoscopic diagnosis of gastric polyps. *Endoscopy* **27**, 32–37.

Stolte, M., Sticht, T., Eidt, S., Ebert, D., and Finkenzeller, G. (1994). Frequency, location, and age and sex distribution of various types of gastric polyp. *Endoscopy* **26**, 659–665.

Gastric Reservoirs

JAN TACK

University Hospital Gasthuisberg and University of Leuven, Belgium

dyspepsia Symptoms originating in the upper gastrointestinal tract, including upper abdominal pain/discomfort, early satiety, postprandial abdominal bloating/distension, and nausea with or without vomiting.

functional dyspepsia Dyspeptic symptoms with no definable organic cause after evaluation usually consisting of history and physical examination, blood tests, and upper endoscopy.

nitric oxide A gas produced by cells of the brain, blood vessels, and immune system.

vasoactive intestinal polypeptide A polypeptide from the small intestine; it induces systemic vasodilation, hypotension, increased cardiac output, respiratory stimulation, and hyperglycemia.

During fasting, muscle fibers of the proximal stomach maintain a vagally mediated tonic contractile activity, which generates gastric fundus tone. During and after ingestion of a meal, relaxation of the proximal stomach occurs, which provides the meal with a reservoir and enables a gastric volume increase without a rise in pressure. This also allows the stomach to retain food and to allow passage to the duodenum at a rate that matches the duodenal absorptive capacity. Tone in the gastric antrum and its ability to function as a reservoir have not been established; the antrum seems to act predominantly as a muscular pump that grinds the food and promotes evacuation.

CONTROL OF THE ACCOMMODATION REFLEX

Studies in animals have demonstrated that the gastric accommodation reflex, which provides the meal with a reservoir and enables a gastric volume increase without a rise in pressure (Fig. 1), is mediated via a vagovagal reflex pathway that activates nonadrenergic noncholinergic neurons in the gastric wall. Several lines of evidence suggest a role for nitric oxide and vasoactive intestinal polypeptide as inhibitory neurotransmitters mediating gastric relaxation.

Copyright 2004, Elsevier (USA). All rights reserved.

this examination is negative, surveillance endoscopy should be repeated no more frequently than 3- to 5-year intervals. No surveillance endoscopy is necessary after removal of nonadenomatous gastric polyps.

See Also the Following Articles

Gastric Cancer Surveillance • *Helicobacter pylori* • Stomach, Adenomas and Carcinomas of the

Further Reading

American Society for Gastrointestinal Endoscopy Guidelines. (1998) The role of endoscopy in the surveillance of premalignant conditions of the upper gastrointestinal tract. *Gastrointest. Endosc.* **48**, 663–668.

Inoue, H., Fukami, N., Yoshida, T., and Kudo, S. (2002). Endoscopic mucosal resection for esophageal and gastric cancers. *J. Gastroenterol. Hepatol.* **17**, 382–388.

Muehldorfer, S. M., Stolte, M., Martus, P., Hahn, E. G., and Ell, C. (2002). Diagnostic accuracy of forceps biopsy versus polypectomy for gastric polyps: A prospective multicentre study. *Gut* **50**, 465–470.

Oberhuber, G., and Stolte, M. (2000). Gastric polyps: An update of their pathology and biological significance. *Virchows Arch.* **437**, 581–590.

Ohkusa, T., Takashimizu, I., Fujiki, K., Suzuki, S., Shimoi, K., Horiuchi, T., *et al.* (1998). Disappearance of hyperplastic polyps in the stomach after eradication of *Helicobacter pylori*. A randomized, clinical trial. *Ann. Intern. Med.* **129**, 712–715.

Stolte, M. (1995). Clinical consequences of the endoscopic diagnosis of gastric polyps. *Endoscopy* **27**, 32–37.

Stolte, M., Sticht, T., Eidt, S., Ebert, D., and Finkenzeller, G. (1994). Frequency, location, and age and sex distribution of various types of gastric polyp. *Endoscopy* **26**, 659–665.

Gastric Reservoirs

JAN TACK

University Hospital Gasthuisberg and University of Leuven, Belgium

dyspepsia Symptoms originating in the upper gastrointestinal tract, including upper abdominal pain/discomfort, early satiety, postprandial abdominal bloating/distension, and nausea with or without vomiting.

functional dyspepsia Dyspeptic symptoms with no definable organic cause after evaluation usually consisting of history and physical examination, blood tests, and upper endoscopy.

nitric oxide A gas produced by cells of the brain, blood vessels, and immune system.

vasoactive intestinal polypeptide A polypeptide from the small intestine; it induces systemic vasodilation, hypotension, increased cardiac output, respiratory stimulation, and hyperglycemia.

During fasting, muscle fibers of the proximal stomach maintain a vagally mediated tonic contractile activity, which generates gastric fundus tone. During and after ingestion of a meal, relaxation of the proximal stomach occurs, which provides the meal with a reservoir and enables a gastric volume increase without a rise in pressure. This also allows the stomach to retain food and to allow passage to the duodenum at a rate that matches the duodenal absorptive capacity. Tone in the gastric antrum and its ability to function as a reservoir have not been established; the antrum seems to act predominantly as a muscular pump that grinds the food and promotes evacuation.

CONTROL OF THE ACCOMMODATION REFLEX

Studies in animals have demonstrated that the gastric accommodation reflex, which provides the meal with a reservoir and enables a gastric volume increase without a rise in pressure (Fig. 1), is mediated via a vagovagal reflex pathway that activates nonadrenergic noncholinergic neurons in the gastric wall. Several lines of evidence suggest a role for nitric oxide and vasoactive intestinal polypeptide as inhibitory neurotransmitters mediating gastric relaxation.

Copyright 2004, Elsevier (USA). All rights reserved.

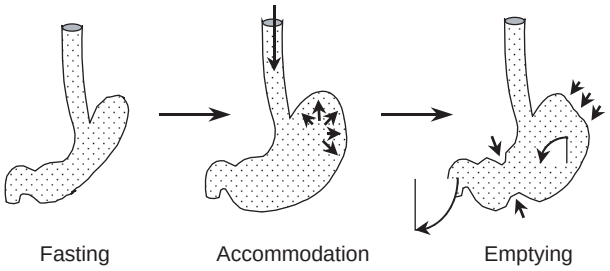


FIGURE 1 Schematic illustration of the role of accommodation to a meal. The accommodation reflex consists of relaxation of the proximal stomach, providing the meal with a reservoir and enabling a volume increase without a rise in pressure. Subsequently, the meal is emptied from the stomach at a rate that matches the absorbtive capacity of the duodenum.

In the mouse and in the guinea pig, involvement of 5-hydroxytryptamine (5-HT) receptors on intrinsic neurons in vagally mediated gastric relaxation has been demonstrated. More recently, it was demonstrated that 5-HT-induced relaxation of the guinea pig stomach is mediated via the release of nitric oxide through activation of a 5-HT_1-like receptor.

In humans, it is unknown whether the accommodation reflex also involves serotonergic activation of nitrergic neurons. The lack of suitable selective 5-HT ligands for *in vivo* use in humans hampers the investigation of this issue. However, the role of 5-HT in the control of gastric motility can be investigated in an alternative way. As in the central nervous system, a serotonin reuptake system is present in the enteric nervous system (ENS). The serotonin reuptake system in the ENS is also inhibited by selective serotonin reuptake inhibitors. Therefore, 5-HT reuptake inhibitors may increase the availability of synaptically released 5-HT, not only in the central nervous system, but also in the enteric nervous system. Physiological processes involving the release of 5-HT in the ENS are enhanced shortly after pretreatment with a selective serotonin reuptake inhibitor. This property has been used to investigate the role of 5-HT in the control of gastric motility in humans.

Twelve healthy volunteers underwent a gastric barostat study on two occasions, after pretreatment with placebo or the selective serotonin reuptake inhibitor paroxetine at 20 mg/day. Paroxetine significantly enhanced the amplitude of the meal-induced fundus relaxation. This observation suggests involvement of 5-HT in the gastric accommodation reflex in humans also. The 5-HT receptor involved and its localization cannot be determined from the present study. Selective serotonin reuptake inhibitors may act on neurons that are located centrally as well as peripherally. It was observed, however, that peripherally acting 5-HT_{1P} agonists or 5-HT_4 agonists are also able to enhance gastric accommodation to a meal, supporting involvement of one or both of these peripheral 5-HT receptors.

N^G-monomethyl-L-arginine (L-NMMA) is an inhibitor of nitric oxide synthase, suitable for use in humans. Recent studies using this agent have been able to demonstrate the involvement of nitric oxide in the control of interdigestive motility and in mediating transient lower esophageal sphincter relaxations in humans. These properties were used to test the hypothesis that nitric oxide is involved in the control of postprandial gastric tone in humans. Pretreatment with L-NMMA decreased the gastric accommodation to a meal in a dose-dependent manner, thereby establishing that the gastric accommodation reflex in humans involves activation of nitrergic neurons.

Taken together, these studies suggest that release of 5-HT, probably at the level of the enteric nervous system, with subsequent activation of nitrergic motor neurons, is involved in the control of the accommodation reflex in humans (Fig. 2). Animal studies have shown that intrinsic neurons in the stomach can be activated through a 5-HT_1-like receptor.

In the guinea pig, a 5-HT_{1P} receptor was shown to be present on nitrergic neurons in the myenteric plexus of the stomach, where it mediates a prolonged depolarization in response to the application of 5-HT. Recently, it was demonstrated that sumatriptan, a 5-HT_1 receptor agonist that is used in the treatment of migraine in humans, is an agonist at 5-HT_{1P} receptors on nitrergic myenteric neurons in the stomach. It was therefore investigated whether sumatriptan is able to induce relaxation of the proximal stomach in human and whether this relaxation is mediated through the activation of nitrergic neurons.

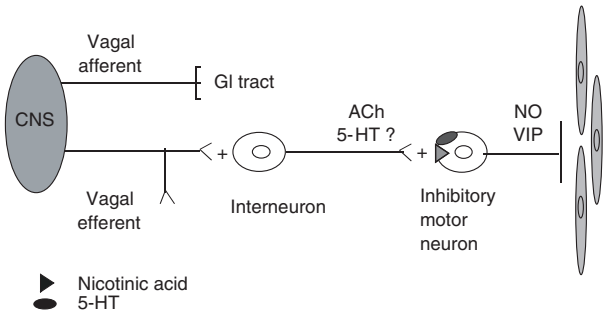


FIGURE 2 Neural control pathway of the accommodation reflex, based on animal studies. The accommodation reflex is mediated via a vagovagal reflex pathway that activates inhibitory motor neurons in the gastric wall to release nitric oxide and vasoactive intestinal polypeptide. Involvement of a 5-hydroxytryptamine receptor on intrinsic inhibitory neurons has been demonstrated.

Using a gastric barostat, it was demonstrated that the administration of sumatriptan induces relaxation of the gastric fundus in human. Pretreatment with L-NMMA decreased the gastric relaxatory response to sumatriptan in a dose-dependent manner, thereby establishing that the drug relaxes the proximal stomach in humans through a nitrergic pathway.

PATHOPHYSIOLOGICAL ROLE OF IMPAIRED ACCOMMODATION

Scintigraphic and ultrasonographic studies have demonstrated an abnormal intragastric distribution of food in patients with functional dyspepsia, with preferential accumulation in the distal stomach. This finding suggests defective postprandial accommodation of the proximal stomach. Consistently, studies using a gastric barostat have shown reduced proximal gastric relaxation in response to a meal in patients with functional dyspepsia. In view of the role of gastric accommodation in providing a reservoir during meal intake, it was hypothesized that the absence of normal accommodation might lead to early satiety.

Using a gastric barostat in 40 consecutive dyspeptic patients, it was shown that impaired gastric accommodation was present in 40% of these patients and this impairment was independently associated with early satiety. The relationship between impaired gastric accommodation and early satiety is also apparent from the correlation between the amplitude of the meal-induced relaxation and the number of calories ingested at maximum satiety in patients with early satiety.

These data suggest that impaired accommodation is the mechanism underlying the symptom of early satiety. According to this hypothesis, causing impaired accommodation in healthy subjects should induce symptoms of early satiety. Previously, it was reported that gastric accommodation is inhibited by pretreatment with L-NMMA. Administration of the motilin agonist erythromycin was shown to cause a contraction of the proximal stomach in the postprandial phase, thereby reducing the accommodation to a meal. These properties were used to investigate the hypothesis that inhibition of accommodation or administration of a fundus-contracting drug might enhance meal-induced satiety in humans.

Eight healthy subjects underwent a standardized satiety drinking test twice, after pretreatment with placebo or erythromycin at the start of the meal in a double-blind randomized cross-over fashion. Erythromycin significantly decreased the amount of food ingested and significantly increased the average satiety scores for the same number of kilocalories ingested (see Fig. 3). In 11 healthy subjects, the influence of L-NMMA on meal-induced satiety was tested using the satiety drinking test. Compared to placebo, L-NMMA significantly decreased the amount of food ingested to reach maximum satiety and significantly increased the average satiety scores for the same number of kilocalories ingested. These observations are in support of the hypothesis that impaired accommodation is a mechanism underlying the symptom of early satiety.

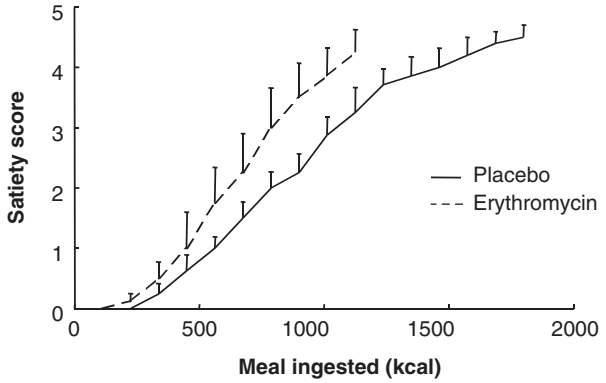


FIGURE 3 Influence of erythromycin on meal-induced satiety in eight healthy controls. Satiety scores for the same number of calories were significantly higher after pretreatment with erythromycin. Data are shown up to the highest number of ingested calories reached by all subjects.

The mechanism underlying impaired postprandial relaxation of the proximal stomach is unknown. Several possible pathways may be involved. Theoretically, impaired relaxation can result from a disorder at the level of the sensory apparatus, the vagal reflex pathway, or the intrinsic inhibitory innervation. Recently, it was demonstrated that patients with presumed postinfectious dyspepsia (acute-onset dyspepsia accompanied by infectious symptoms) have a high prevalence of impaired accommodation. It was also observed that these patients had an impaired relaxatory response to sumatriptan, whereas the stomach relaxed adequately in response to administration of the nitric oxide donor amylnitrite. This finding suggests a defect at the level of gastric intrinsic nitrergic neurons.

CONCLUSIONS

Gastric accommodation provides the meal with a reservoir and enables a gastric volume increase without a rise in pressure. Gastric accommodation in humans is controlled by a vagal reflex pathway that involves the release of serotonin and the activation of a nitrergic motor neuron. This is mimicked by administration of

the 5-HT_{1P} receptor agonist sumatriptan, which causes a relaxation of the proximal stomach in humans through a nitrergic pathway.

Almost half of the patients with functional dyspepsia have an impaired accommodation reflex and this is associated with early satiety and weight loss. Drug-induced inhibition of the accommodation reflex is able to induce early satiety in healthy subjects, thereby supporting the hypothesis that impaired accommodation is the mechanism underlying the symptom of early satiety. In a subset of patients, impaired accommodation seems to occur as a prolonged sequela of an acute gastrointestinal infection. In these patients, the defect seems to occur at the level of gastric intrinsic nitrergic neurons.

See Also the Following Articles

Functional (Non-Ulcer) Dyspepsia • Nitric Oxide • Small Intestinal Motility • Vasoactive Intestinal Peptide (VIP)

Further Reading

Boeckxstaens, G. E., Pelckmans, P. A., Bogers, J. J., Bult, H., De Man, J. G., Oosterbosch, L., Herman, A. G., and Van Maercke, Y. M. (1991). Release of nitric oxide upon stimulation of nonadrenergic noncholinergic nerves in the rat gastric fundus. *J. Pharmacol. Exp. Ther.* **256**, 441–447.

Desai, K. M., Sessa, W. C., and Vane, J. R. (1991). Involvement of nitric oxide in the reflex relaxation of the stomach to accommodate food or fluid. *Nature* **35**, 477–479.

Gilja, O. H., Hausken, T., Wilhelmsen, I., and Berstad, A. (1996). Impaired accommodation of proximal stomach to a meal in functional dyspepsia. *Digest. Dis. Sci.* **41**, 689–696.

Grundy, D., Gharib-Naseri, M. K., and Hutson, D. (1993). Role of nitric oxide and vasoactive intestinal polypeptide in vagally mediated relaxation of the gastric corpus in the anaesthetized ferret. *J. Auton. Nerv. Syst.* **43**, 241–246.

Hirsch, D. P., Holloway, R. H., Tytgat, G. N. J., and Boeckxstaens, G. E. E. (1998). Involvement of nitric oxide in human transient lower esophageal sphincter relaxations and esophageal primary peristalsis. *Gastroenterology* **115**, 1374–1380.

Li, C. G., and Rand, M. J. (1990). Nitric oxide and vasoactive intestinal polypeptide mediate non-adrenergic, noncholinergic inhibitory transmission to smooth muscle of the rat gastric fundus. *Eur. J. Pharmacol.* **191**, 303–309.

Meulemans, A. L., Eelen, J. G. M., and Schuurkes, J. A. J. (1995). NO mediates gastric relaxation after brief vagal stimulation in anesthetized dogs. *Am. J. Physiol.* **269**, G255–G261.

Meulemans, A. L., Helsen, L. F., and Schuurkes, J. A. J. (1993). The role of nitric oxide (NO) in 5-HT induced relaxations of the guinea-pig stomach. *Naunyn-Schmiedeberg Arch. Pharmacol.* **384**, 424–430.

Michel, K., Sann, H., Schaaf, C., and Schemann, M. (1997). Subpopulations of gastric myenteric neurons are differentially activated via distinct serotonin receptors: Projection, neurochemical coding, and functional implications. *J. Neurosci.* **17**, 8009–8017.

Russo, A., Fraser, R., Adachi, K., Horowitz, M., and Boeckxstaens, G. (1999). Evidence that nitric oxide mechanisms regulate small intestinal motility in humans. *Gut* **44**, 72–76.

Salet, G. A. M., Samsom, M., Roelofs, J. M. M., van Berge Henegouwen, G. P., Smout, A. J. P. M., and Akkermans, L. M. A. (1998). Responses to gastric distention in functional dyspepsia. *Gut* **42**, 823–829.

Scott, A. M., Kellow, J. E., Shuter, B., Cowan, H., Corbett, A.-M., Riley, J. W., Lunzer, M. R., Eckstein, R. P., Holschl, R., Lam, S.-K, and Jones, M. P. (1993). Intragastric distribution and gastric emptying of solids and liquids in functional dyspepsia. *Digest. Dis. Sci.* **38**, 2247–2254.

Tack, J., Broeckaert, D., Coulie, B., and Janssens, J. (1998). Influence of cisapride on gastric tone and on the perception of gastric distension. *Aliment. Pharmacol. Ther.* **12**, 761–766.

Tack, J., Coulie, B., Wilmer, A., Andrioli, A., and Janssens, J. (2000). Effect of sumatriptan on gastric fundus tone and on the perception of gastric distension in man. *Gut* **46**, 468–473.

Tack, J., Piessevaux, H., Coulie, B., Caenepeel, P., and Janssens, J. (1998). Role of impaired gastric accommodation to a meal in functional dyspepsia. *Gastroenterology* **115**, 1346–1352.

Tack, J. F., Janssens, J., Vantrappen, G., and Wood, J. D. (1992). Actions of 5-hydroxytryptamine on myenteric neurones in the gastric antrum of the guinea-pig. *Am. J. Physiol.* **263**, G838–G846.

Troncon, L. E. A., Bennett, R. J. M., Ahluwalia, N. K., and Thompson, D. G. (1994). Abnormal distribution of food during gastric emptying in functional dyspepsia patients. *Gut* **35**, 327–332.

Gastric Stapling

MALCOLM K. ROBINSON
Harvard Medical School

abdomen Portion of the body that lies between the chest and the pelvis and contains several organs, including the stomach, intestines, liver, spleen, pancreas, kidneys, and bladder.

bariatric Field of medicine or surgery that encompasses the study and prevention of obesity, as well as the evaluation and treatment of obese individuals. It is derived from the Greek "baros," meaning weight, and "iatrike," meaning medicine or surgery.

body mass index Formulation that describes relative weight for height; used to define weight categories (e.g., underweight, normal, overweight, or obese).

bypass Procedure in which one structure is artificially connected to another so as to circumvent a normally intervening segment of tissue. For example, in a gastric bypass procedure, the stomach is connected to the second portion of small intestine such that food circumvents or "bypasses" the main part of the stomach and the first portion of the small intestine.

equilibrium State of balance in which opposing forces or processes exactly counteract each other.

gastric Pertaining to, affecting, or originating in the stomach. Derived from the Latin "gastricus," meaning stomach.

gastroplasty Surgical procedure in which the stomach is reshaped, reconfigured, or reconstructed with sutures or surgical staples.

morbid obesity Severe form of obesity that is frequently associated with a variety of conditions such as diabetes, heart disease, and high blood pressure. Any person with a body mass index equal to or greater than 40 kg/m^2 is considered morbidly obese.

stomach Part of the gastrointestinal tract that lies between the esophagus and first part of the small intestine; serves as a reservoir for swallowed food and is where food digestion is initiated. Although the terms "stomach" and "abdomen" are frequently used interchangeably, they are distinct entities and should not be confused.

Gastric stapling is the name used for a variety of surgical procedures that produce weight loss in obese people. As part of these procedures, surgical staples are used to partition the stomach so that the capacity of the stomach is drastically reduced. The amount of food consumed can be limited to 2–3 ounces per meal, resulting in significant weight loss over 12 to 15 months.

INTRODUCTION

Gastric stapling is most frequently used to refer to a specific procedure called "vertical banded gastroplasty" (VBG). Gastric stapling has recently been described in a more general way to include all types of weight loss surgery procedures, even though some of the procedures are quite different from a VBG.

Gastric stapling is used in reference to the medical field of "bariatric surgery," which encompasses the evaluation, surgical treatment, and management of morbidly obese individuals who have had or who are contemplating surgical treatment of their obesity. Bariatric surgeons perform a variety of bariatric surgical procedures, which are specifically designed to promote weight loss. It is estimated that 40,000 of these procedures were performed in the year 2000 in the United States alone. It is anticipated that this number will continue to increase rapidly given the dramatic increase in obese adults in the United States.

OBESITY AND ITS TREATMENT

The National Institutes of Health recommends using the body mass index (BMI) to determine weight category as the first step in determining what, if any, intervention is necessary to promote weight loss. The BMI generally correlates with total body fat content, although this index may be less useful in highly trained athletes, who have an unusually large proportion of muscle. The BMI is calculated by taking weight measured in kilograms and dividing by the square of height measured in meters (i.e., BMI = kilograms/meters^2).[1] A BMI of less than 18.5 kg/m^2 is considered underweight; 18.5–25 kg/m^2 is normal weight, 25.0–29.9 kg/m^2 is overweight, and greater than 30 kg/m^2 is considered obese. Extreme or morbid obesity is defined as a BMI greater than or equal to 40 kg/m^2. For the individual of

[1]Pounds can be converted to kilograms by dividing by 2.2 (e.g., 220 lb equals 100 kg). Inches can be converted to meters by multiplying by 0.0254 (e.g., 60 inches equals 1.524 m). Thus, a 220-lb person who is 5 feet tall has a BMI of 43.1 kg/m^2.

average height, a BMI of 30 kg/m^2 is approximately 30–35 pounds over normal weight and a BMI of 40 kg/m^2 is 100 pounds over normal weight. It is estimated that 97 million adults in the United States, or approximately 55%, are overweight or obese.

Weight gain or loss depends on the number of calories (i.e., food) consumed by a person relative to the number of calories that their body uses to fuel its many functions. When caloric consumption exceeds caloric utilization on a long-term basis, weight gain occurs. Conversely, when caloric utilization exceeds caloric consumption, weight loss occurs. When caloric consumption equals caloric utilization, a state of equilibrium is achieved and weight is neither gained nor lost.

The treatment of an overweight or obese person can include diet, behavior modification, exercise, prescription drugs, and/or surgery. All of these modalities, with the exception of exercise, result in decreased calorie intake. Exercise increases calorie utilization. Treatment recommendations depend not only on how overweight a person is but also on whether they have weight-related health problems. These conditions—such as high blood pressure, heart disease, and diabetes—significantly increase the risk of death and the development of other diseases.

People with a BMI of 35–39.9 kg/m^2 with weight-related health problems and those with a BMI of 40 kg/m^2 even without weight-related health problems are considered at extremely high risk for developing or exacerbating a variety of diseases. Not surprisingly, these individuals are also at extremely high risk of dying earlier compared to people of normal weight. Because of these risks, bariatric surgical procedures may be appropriate for those who are in the highest weight categories and fail to lose weight with other methods. Although surgery is the most aggressive and invasive form of obesity therapy, such therapy has a high success rate and low complications in the appropriately selected individual. In addition, surgery in these groups of people results in the most durable and greatest average weight loss compared to any other weight loss strategy. However, it is imperative to assess surgical risk, as well as patient motivation and psychological suitability, on a case by case basis before recommending surgery, because serious and life-threatening complications may occur with this intervention.

GASTRIC STAPLING AND GASTRIC BYPASS

Although several techniques are used to reconstruct the digestive system in a way that promotes weight loss, all bariatric procedures work by limiting food intake, limiting food absorption, or both, thus reducing calorie intake. As a person loses weight, the number of calories absorbed by their body will eventually equal the number of calories burned up by their body, at which point weight loss will stop. For patients undergoing bariatric surgery, this equilibrium generally occurs when they lose 50–80% of their excess body weight.

Gastric stapling is one major prototype of bariatric surgical procedures. The medical procedure for gastric stapling, vertical banded gastroplasty, involves a procedure in which the front of the stomach is stapled to the back of the stomach, so as to create a small stomach pouch, which is partitioned from the remainder of the stomach. The pouch is oriented vertically relative to the side of the stomach and only holds 2–3 ounces of food. The passageway or outlet leading out of this pouch is encircled with a plastic band to prevent the outlet from dilating over time. This is done to improve the durability of weight loss (see Fig. 1). Food absorption from the intestines is not altered because the rest of the intestinal tract is not manipulated in this procedure. Thus, a VBG promotes weight loss purely by limiting food intake.

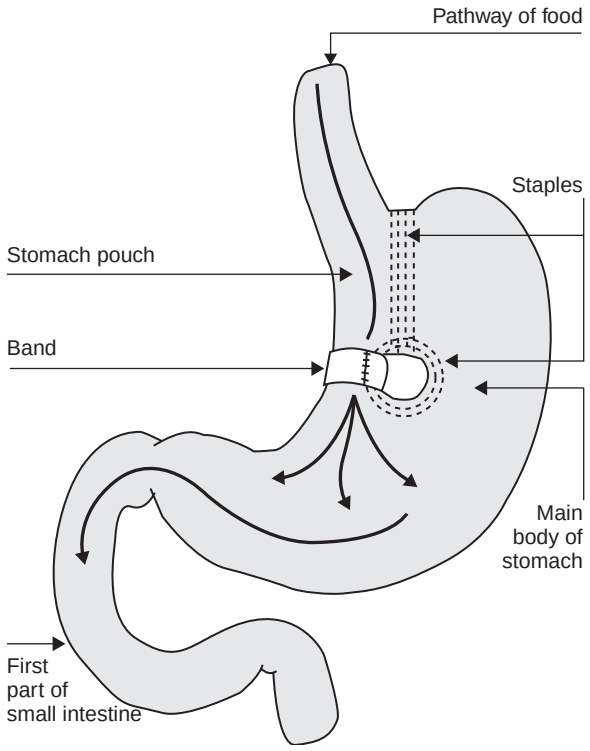


FIGURE 1 Gastric stapling, or vertical banded gastroplasty.

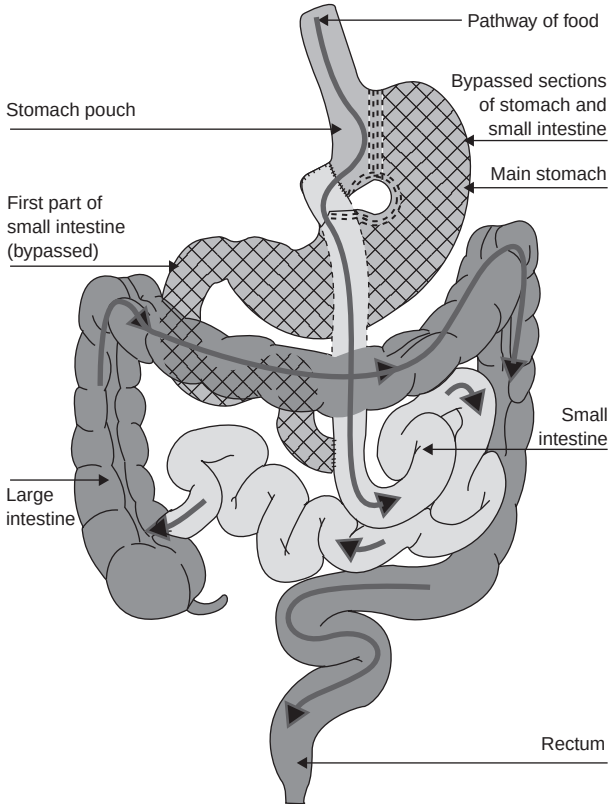


FIGURE 2 Gastric bypass.

The gastric bypass is another major prototype of bariatric surgical procedures. In this procedure, a small stomach pouch is constructed. A section of small intestine is attached directly to this pouch such that food goes from the mouth into the pouch and directly into the small intestine, "bypassing" the majority of the stomach and a significant portion of the beginning of the small intestines (see Fig. 2). Food intake is limited due to the small stomach pouch. In addition, any food that is consumed is not completely absorbed because it bypasses a significant portion of the stomach and intestines, which ordinarily would aid in complete food digestion and absorption. There are many ways to perform gastric bypass procedures, with variations in how the stomach pouch is constructed and the amount of intestine that is bypassed. However, all of the techniques share the principle of a small stomach pouch attached to a section of small intestine, which limits both food intake and food absorption. Currently, the gastric bypass is the most frequently performed type of bariatric surgical procedure in the United States.

See Also the Following Articles

Gastric Surgery • Obesity, Treatment of

Further Reading

Buchwald, H. (2002). Overview of bariatric surgery. *J. Am. Coll. Surgeons* **194**, 367–375.

Deitel, M. (1998). "Surgery for the Morbidly Obese Patient." Lea and Febiger, Philadelphia.

Pi-Sunyer, F. X., *et al.* (1998). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—The evidence report. *Obesity Res.* **6**(Suppl. 2), 51S–210S.

Gastric Surgery

QUYEN D. CHU, GAZI ZIBARI, AND RICHARD H. TURNAGE
Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center, Shreveport

gastric bypass Operative procedure performed for the treatment of morbid obesity; the stomach is divided proximally into a small gastric pouch, which is drained into a Roux-en-Y jejunal limb.

morbid obesity Defined by having a body mass index greater than 40 kg/m^2; this condition is associated with a number of metabolic and cardiopulmonary complications, including diabetes mellitus, hyperlipidemia, hypertension, sleep apnea, cardiomyopathy, coronary artery disease, and joint disease.

proximal gastric vagotomy Also called highly selective vagotomy and parietal cell vagotomy; specific division of the vagal nerve fibers as they innervate the lesser curvature of the stomach from the esophageal hiatus to the gastric incisura; commonly performed to prevent vagally mediated gastric acid secretion without interrupting the antral and pyloric motor function.

truncal vagotomy Division and resection of part of the vagus nerves as they exit the esophageal hiatus; commonly performed to prevent vagally mediated gastric acid secretion.

Gastric surgery is indicated for benign and malignant diseases of the stomach as well as to assist in the management of morbid obesity. The indications for operation, the types of commonly used operative procedures, and the results of surgical therapy for the management of patients with peptic ulcer disease, morbid obesity, and gastric malignancy are all important considerations.

OPERATIONS FOR PEPTIC ULCER DISEASE

Indications for Operation

The principal indications for operation in patients with peptic ulcer disease are the treatment of medically intractable symptoms and the management of acute life-threatening ulcer-related complications. In the 1960s and 1970s, nearly 50% of the operations for duodenal ulcer disease were performed to relieve chronic symptoms, principally intractable ulcer pain. Now, nearly all operative procedures for peptic ulcer are performed to treat an acute life-threatening ulcer complication, i.e., bleeding, perforation, or gastric outlet obstruction. Many factors are undoubtedly important in producing this change. Of particular note, a better understanding of the pathophysiology of peptic ulcer disease has led to refinements in medical management, such as the development of potent inhibitors of gastric acid secretion and effective antibiotic regimens for *Helicobacter pylori*. Recently, these refinements in medical management have become incorporated into the perioperative strategies for managing acute complications of ulcer disease, such as the eradication of *H. pylori* in patients undergoing omental patching of a perforated duodenal ulcer.

Operative Procedures

The most appropriate operation for patients with peptic ulcer disease depends on the location of the ulcer (gastric versus duodenal), the presence of ulcer-related complications, and the condition of the patient at operation. Patients undergoing operation for the elective management of gastric ulcer generally require partial gastrectomy (to include the ulcer) with either a Billroth I gastroduodenostomy (less commonly) or a Billroth II gastrojejunostomy (more commonly). Patients with duodenal ulcer disease requiring operation for intractable ulcer symptoms may be treated with truncal vagotomy and antrectomy, truncal vagotomy and drainage (either pyloroplasty or gastrojejunostomy), or proximal gastric vagotomy. The results of these procedures are summarized in Table I.

Truncal vagotomy denervates the parietal cell mass, thereby diminishing vagally mediated acid secretion, whereas antrectomy removes the antral G cells, the source of the potent gastric acid secretogogue, gastrin. Combination of truncal vagotomy and antrectomy reduces gastric acid secretion by 80–90% and is associated with a very low incidence of recurrent ulcer formation (1–2%). However, this more extensive procedure is also associated with higher rates of perioperative and postoperative morbidity (15–30%)

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Results of Elective Operations for the Treatment of Duodenal Ulcer Disease[a]

Procedure	Ulcer recurrence rate (%)	Perioperative morbidity rate (%)	Perioperative mortality rate (%)	Visick grades I+II (%)[b]
Truncal vagotomy and antrectomy	2	10–30	1–2	60–80
Truncal vagotomy and pyloroplasty	10–15	10	0.5–1	75
Proximal gastric vagotomy	10–20	5–10	0.2–0.7	80–95

[a] Morbidity and mortality rates are much higher for emergency operations performed for ulcer complications.
[b] Visick grades I and II indicate patients who are asymptomatic and require no medication (grade I) or asymptomatic and require over-the-counter medications occasionally (grade II).

and mortality (1–2%) compared to lesser procedures, particularly in medically compromised patients.

The proximal gastric vagotomy (also called parietal cell vagotomy or highly selective vagotomy) selectively denervates the parietal cell mass while preserving innervation of the antral pump and pylorus as well as the hepatic and celiac branches of the vagus. This procedure preserves normal (or near-normal) gastric emptying and hence no gastric emptying procedure is necessary. The ulcer recurrence rate associated with this procedure is higher than that associated with resectional therapies; in Jordan and Thornby's prospective study, the risk of recurrence after this procedure was 14% in 102 patients followed for up to 20 years. Prospective trials have demonstrated that the ulcer recurrence rates associated with a truncal vagotomy and drainage procedure vary from 7 to 28% with followup from 4 to 12 years. The perioperative risks and long-term complications are certainly greater than that associated with a proximal gastric vagotomy and in some series approximate that of vagotomy and antrectomy.

Truncal vagotomy and pyloroplasty are particularly well suited for the management of patients who are bleeding from erosion of a duodenal ulcer into the gastroduodenal artery. In this setting, the first portion of the duodenum and the pylorus are incised longitudinally and the bleeding vessel is oversewn through the posterior wall of the duodenum with a "U-stitch." Truncal vagotomy is performed to reduce the risk of ulcer recurrence and the duodenal and pyloric incision is closed transversely (Heineke–Mikulicz pyloroplasty). The longer operative times required for a truncal vagotomy and antrectomy or a proximal gastric vagotomy dissuade most surgeons from performing these procedures in patients suffering massive hemorrhage from a duodenal ulcer. Patients presenting with perforation of a duodenal ulcer require, at the very least, operative patching of the ulcer (Graham patch) with omentum or falciform ligament. Failure to treat underlying *H. pylori* infection is associated with an ulcer recurrence rate of 38% within 1 year, compared with 5% when treated with appropriate antibiotics. Patients with a long history of complicated ulcer disease may also benefit from a proximal gastric vagotomy, provided that the patient is able to tolerate a longer operative procedure and there is relatively limited soiling of the peritoneal cavity. The operative options for managing patients with gastric outlet obstruction from ulcer disease include truncal vagotomy and antrectomy (or partial gastrectomy) with a Billroth II gastrojejunostomy or truncal vagotomy and gastrojejunostomy.

Perioperative mortality is dramatically increased by the presence of acute ulcer-related complications. At the Massachusetts General Hospital, the death rate after elective ulcer procedures has been reported to be 1 in 200; the mortality rates associated with emergent operations for ulcer bleeding and perforation have been 1 in 5 or greater.

Complications of Operative Management

The complications of operations for ulcer disease include ulcer recurrence, mechanical obstructions, functional disorders of gastric emptying, and metabolic sequelae (see Table II).

TABLE II Long-Term Complications of Operations for Peptic Ulcer Disease

Ulcer recurrence	**Disorders of gastrointestinal motility**
Incomplete vagotomy	Gastric stasis with poor gastric emptying
Retained, excluded antrum	Early satiety
Mechanical obstructions	Dumping
Afferent limb obstruction	Chronic diarrhea
Efferent limb obstruction	
Gastric outlet obstruction	
Other	**Metabolic sequelae**
Weight loss/anorexia	Iron deficiency
Bile reflux gastritis	Calcium malabsorption with osteopenia
	Cholelithiasis

OPERATIONS FOR MORBID OBESITY

Indications for Operation

Individuals with morbid obesity are those with a body mass index (BMI) greater than 40 kg/m^2 or who weigh more than 100 pounds over their ideal body weight (IBW). These persons have a reduced life expectancy due to the severe metabolic and cardiovascular disorders associated with this degree of obesity. A National Institutes of Health (NIH) Consensus Development Conference held in March 1991 recommended that patients whose BMI exceeded 40 kg/m^2 are candidates for weight reduction procedures if they strongly desired substantial weight loss because of its impact on the quality of their lives. Furthermore, patients whose BMI was 35–40 kg/m^2 were also considered candidates for operative management if they had significant obesity-related comorbid medical conditions, such as severe sleep apnea, obesity-related cardiomyopathy, severe joint disease, and diabetes mellitus. This conference noted that these patients should first be treated with nonsurgical therapies, including dietary regimens, exercise routines, behavior modification, and psychologic support modalities associated with extremely high failure rates in this particular patient group.

Operative Procedures

The two principal gastric operations performed to facilitate weight loss in morbidly obese individuals, and the only two operations endorsed by the NIH Consensus Conference, are the vertical banded gastroplasty and the gastric bypass. The vertical banded gastroplasty (VBG) creates a 20-ml vertical gastric pouch with the outlet restricted to 32-French (around 1 cm) by a 1.5-× 5.0-cm strip of polypropylene mesh sutured to itself. The gastric bypass (GB) is characterized by a 15- to 20-milliliter gastric pouch, a vertical staple line dividing the stomach from the pouch, and reconstruction of gastrointestinal continuity with a Roux-en-Y gastrojejunostomy. These procedures have been found to improve weight-associated comorbid medical conditions, including sleep apnea, obesity-related hypoventilation, diabetes mellitus, hypertension, and hyperlipidemia. It is not yet known whether these beneficial effects are maintained long enough to prevent renal disease, stroke, myocardial infarction, and congestive heart failure. Multiple trials have demonstrated superior early and late weight loss in patients undergoing GB when compared with VBG. In one collected review, the average loss of excess weight was 53 and 69% at 1 year and 46 and 60% at 5 years for VBG and GB, respectively.

Complications of Operative Management

The immediate perioperative mortality rate is relatively low for both VBG and GB ($\leq$ 1%). Early postoperative morbidity has been reported to occur in as many as 10% of patients, with most complications due to wound infections and wound dehiscence, staple line dehiscence, stomal stenosis, marginal ulceration, various pulmonary complications (including pulmonary embolism and pneumonia), and deep venous thrombosis. Later complications include pouch and distal esophageal dilation, persistent vomiting (with or without stomal obstruction), cholecystitis due to preexisting gallstones or stones that develop due to rapid weight loss, incisional hernias, and, most importantly, failure to lose weight or maintain weight loss.

OPERATIONS FOR GASTRIC MALIGNANCY

The most common primary gastric malignancy is adenocarcinoma, representing 90% of gastric malignancies. At present, cure is possible only with resection. Chemotherapy and radiation therapy have been shown to extend survival when given in an adjuvant setting but have no such effect as primary therapy. The principal objectives of operations for gastric malignancies are to cure the patient of their disease or to ameliorate complications of advanced disease, specifically hemorrhage, obstruction, perforation, and/or intractable pain. The principal determinant of survival following resection for gastric cancer is the stage of the disease. Wanebo and colleagues reviewed the American College of Surgeons (ACS) Patient Care Evaluations and reported a 5-year survival of 60% for patients who had Stage I tumors, approximately 35% for patients with Stage II disease, 16–18% for Stage III disease, and less than 10% for patients with Stage IV disease.

Preoperative Evaluation and Indications for Operation

Preoperative assessment should attempt to detect local complications of the malignancy and evidence of advanced disease. The patient's performance status may be assessed using the Eastern Cooperative Oncology Group (ECOG) performance score. Physical

TABLE III Lymph Node Stations According to the Japanese Gastric Cancer Association

Dissection	Levels of lymph node stations or echelons removed
D1	Resection of stations 1–6: Nodes of the right (1) and left cardial (2), lesser curvature (3), gastroepiploic (4), suprapyloric (5), and infrapyloric (6) regions
D2	Resection of stations 1–11: D1+nodes along the left gastric artery (7), common hepatic artery (8), celiac axis (9), splenic hilum (10), and splenic artery (11)
D3	Resection of stations 1–14: D2+nodes of the hepatoduodenal ligament (12), the retropancreatic region (13), and the root of the mesentery (14)
D4	Resection of Stations 1–16: D3+nodes in the transverse mesocolon (15) and paraaortic regions (16)

examination may provide evidence of advanced disease with supraclavicular lymphadenopathy (Virchow's node), a mass at the umbilicus (Sister Mary Joseph's node), and/or pelvic peritoneal metastases palpable on rectal examination (Blumer's shelf). Radiologic examination should include a chest X ray, abdominal and pelvic computed tomography (CT) scans, and, in selected cases, an endoscopic ultrasound exam of the tumor. Preoperative endoscopic ultrasonography will accurately determine the T stage of the tumor in 70–80% of cases. Staging laparoscopy has been shown to be more sensitive in detecting hepatic, nodal, and peritoneal metastases as compared to ultrasound and CT in one series, but has been considered to be of limited clinical utility in other studies.

Operative Procedures

Adequate surgical resection of gastric cancer requires a wide gastric resection to achieve margins free of tumor and an en bloc resection of regional lymph nodes. The extent of resection is determined by the location and extent of the tumor and involvement of adjacent organs. The two most common operations for the curative resection of gastric cancer are subtotal gastrectomy for tumors located in the distal third of the stomach and total gastrectomy for tumors of the proximal and middle third as well as for gastric cancers of the diffuse type (*linitus plastica*). Tumors of the cardia and gastroesophageal junction are generally treated as distal esophageal tumors, with an esophagogastrectomy and gastric tube reconstruction with the gastric remnant based on the gastroepiploic artery. Gross resection margins of 6 cm are necessary to achieve reproducibly negative histologic margins and to minimize local recurrence. Frozen section examinations of the resection margins are obtained to ensure that the margins are free of microscopic disease.

The involvement of adjacent organs by the tumor is not a contraindication to resection, but such involvement clearly affects the likelihood of cure. A curative en bloc resection of (T4) lesions may be achieved if there is no more than one organ involved, there is less than or equal to N2 lymph node involvement, there is no diffuse carcinoma, and there are no distant metastases. In cases meeting these criteria, 5-year survival rates of 25–30% have been reported. Unless directly involved with tumor, there is no survival advantage afforded by splenectomy and/or distal pancreatectomy in patients undergoing curative gastric resections; there is substantial evidence to suggest that these procedures significantly enhance the risks of perioperative morbidity and mortality.

The extent of lymph node dissection performed in patients with gastric adenocarcinoma is debated. The Japanese Classification for Gastric Carcinoma (JCGC) has defined the extent of lymphadenectomy according to the levels of lymph node stations or echelons removed (Table III). Results from several European randomized trials have failed to demonstrate a survival benefit with more extensive lymph node dissections and hence most centers in the United States and Europe perform a D1 or lesser lymph node dissection (D0) as part of their gastric resection. In one of these studies, patients undergoing D2 dissection had significantly higher rates of complications, perioperative mortality, and longer hospital stays than did those undergoing lesser lymphadenectomies (D1). Proponents for a D2 lymphadenectomy argue that 30% of the patients undergoing a curative D2 dissection are found to have disease outside of area encompassed by a D1 resection. Furthermore, they suggest that the higher rate of complications associated with more extensive lymphadenectomies may be decreased as the surgeons gain experience with these techniques.

A recent multiinstitutional randomized trial performed by the Southwest Oncology Group (SWOG) suggests that patients receiving adjuvant chemoradiation therapy may have a modest survival advantage over patients not receiving adjuvant therapy. Patients

randomized to undergo resection with adjuvant chemoradiation therapy (5-fluorouracil and leukovorin plus 4500 cGY) had a median survival of 36 months compared with 27 months in patients undergoing surgery alone.

See Also the Following Articles

Duodenal Ulcer • Gastrectomy • Gastric Cancer Surveillance • Gastric Outlet Obstruction • Gastric Stapling • Gastric Ulcer • Gastroenterostomy • Gastrostomy • Marginal Ulcer • Obesity, Treatment of • Pyloroplasty • Stomach, Adenomas and Carcinomas of the

Further Reading

Bonenkamp, J. J., Songun, I., Hermans, J., *et al.* (1995). Randomized comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet* **345**, 745–748.

Cuschieri, A., Fayers, P., Fielding, J., *et al.* (1996). Postoperative morbidity and mortality after D1 and D2 resections for gastric cancer: Preliminary results of the MRC randomized controlled surgical trial. *Lancet* **347**, 995–999.

Jordan, P. H., and Thornby, J. (1994). Twenty years after parietal cell vagotomy or selective vagotomy antrectomy for treatment of duodenal ulcer: A final report. *Ann. Surg.* **220**, 283–296.

MacDonald, J. S., Smalley, S. R., Benedetti, J., Hundahl, S. C., Estes, N. C., *et al.* (2001). Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N. Engl. J. Med.* **345**, 725–730.

Marujama, K., Okabayashi, K., and Kinoshita, T. (1987). Progress in gastric cancer surgery in Japan and its limits of radicality. *World J. Surg.* **11**, 418–425.

Meyer, J. H. (1989). Chronic morbidity after ulcer surgery. *In* "Sleisinger & Fordtran. Gastrointestinal Disease. Pathophysiology, Diagnosis, Management," 4th Ed., pp. 962–987. W. B. Saunders, Philadelphia.

Ng, E. K. W., Lam, Y. M., Sung, J. J. Y., Yung, M. Y., To, K. F., Chan, A. C. W., Lee, D. W. H., Law, B. K. B., Lau, J. Y. W., Ling, T. K. W., Lau, W. Y., and Chung, S. C. S. (2000). Eradication of *Helicobacter pylori* prevents recurrence of ulcer after simple closure of duodenal ulcer perforation. *Ann. Surg.* **231**, 153–158.

Roder, J. D., Bonenkamp, J. J., Craven, J., *et al.* (1995). Lymphadenectomy for gastric cancer. I. Clinical trials: Update. *World J. Surg.* **19**, 546–553.

Roukos, D. H., Lorenz, M., and Encke, A. (1998). Evidence of survival benefit of extended (D2) lymphadenectomy in western patients with gastric cancer based on a new concept: A prospective long-term follow-up study. *Surgery* **123**, 573–578.

Stabile, B. E. (2000). Redefining the role of surgery for perforated duodenal ulcer in the *Helicobacter pylori* era. *Ann. Surg.* 231, 159–160.

Turnage, R. H., Sarosi, G., Cryer, B., Spechler, S., Peterson, W., and Feldman, M. (2003). The evaluation and management of patients with recurrent ulcer disease following acid-reducing operations: A systematic review. *J. Gastrointest. Surg.* (in press).

Wanebo, H. J., Kennedy, B. J., Winchester, D. P., Fremgen, A., and Stewart, A. K. (1996). Gastric carcinoma: Does lymph node dissection alter survival? *J. Am. Coll. Surg.* **183**, 616–624.

Welch, C. E., Rodkey, G. V., and von Ryll Gryska, P. (1986). A thousand operations for ulcer disease. *Ann. Surg.* 204, 455–467.

Gastric Ulcer

ARIC JOSUN HUI,* RUPERT WING-LOONG LEONG,† AND FRANCIS KA-LEUNG CHAN*
*Chinese University of Hong Kong, Prince of Wales Hospital and †Fremantle Hospital, Perth

biopsy urease test Provides indirect evidence, in gastric mucosal biopsies, of the presence of *Helicobacter pylori*, which produces the urease that metabolizes urea to ammonia and changes the pH indicator color.

cyclooxygenase Enzyme that exists in two isoforms, cyclooxygenase-1 and cyclooxygenase-2. Cyclooxygenase-1 is constitutive and regulates normal function in many organs. Cyclooxygenase-2 is inducible and is expressed by inflammatory cells to mediate inflammatory responses. A variant of cyclooxygenase-1 has been found in the central nervous system.

erosion Histologically defined as a defect that does not penetrate the muscularis mucosae; endoscopically, superficial mucosal breaks with a diameter less than 3–5 mm and with no perceptible depth.

Helicobacter pylori Spiral-shaped gram-negative bacilli with four to seven flagella at one pole; produce large amounts of urease.

nonsteroidal antiinflammatory drugs Compounds that nonselectively inhibit the cyclooxygenase enzyme and are potent antiinflammatory and antipyretic agents.

ulcer Defined histologically as a defect that penetrates the muscularis mucosa; the exact endoscopic definition remains controversial, with disagreement on the lesion size (usually 3–5 mm), but most agree perception of lesion depth is necessary.

Peptic ulcer formation in either the stomach or duodenum is due to an imbalance between erosive factors such as hydrochloric acid and pepsin and the protective mechanisms of the mucosa. Unlike duodenal ulcers, in which the importance of acid secretion is indisputable, gastric (stomach) ulcers can develop despite only minimal amounts of acid. Indeed, past studies have shown that the basal and maximal acid outputs in patients with gastric ulcers are no different than those in normal controls. The gastric mucosa has evolved to tolerate the high acidity of the stomach lumen via an intricate equilibrium of protective mechanisms. The gastric protective mechanisms (preepithelial, epithelial, and subepithelial factors) act in concert.

PHYSIOLOGY

Preepithelial protective gastric factors include the mucus–bicarbonate coating produced by the surface epithelial cells. This layer of mucus, which varies in thickness from 50 to 400 μm in humans, is composed of 95% water and 5% phospholipids and is effective in maintaining a near normal pH at the epithelial cell surface. Epithelial cells also secrete bicarbonate into the mucus layer, which creates a pH gradient from the lumen to the epithelial cell. Mucosal bicarbonate secretion is stimulated by luminal acid and prostaglandins and is dependent on an adequate vascular supply. Prostaglandins promote secretion of mucin from the epithelial cells and have an inhibitory role in the regulation of acid secretion.

Epithelial factors are the second line of defense. The surface epithelial cells are bound by tight junctional complexes to maintain a barrier to the luminal contents. Even if the mucus–bicarbonate layer is breached, the epithelial cells can still maintain a normal intracellular pH via Na^+/H^+ and Cl^-/HCO_3^- exchangers on the cell membrane. These transport mechanisms are also dependent on adequate HCO_3^- from the blood. When minor mucosal damage occurs, the epithelial cells can rapidly repair the damage by a process known as mucosal restitution. A mucoid cap formed over the breach provides an alkaline microenvironment that enables epithelial cells to migrate and repair the breach. This process does not require cell replication and is dependent on an adequate vascular supply. The repair of larger defects necessitates cell proliferation, which is dependent on endogenous prostaglandin and growth factors.

Subepithelial factors also contribute to mucosal defense against the luminal acid, pepsin. For each molecule of H^+ secreted by parietal cells, one molecule of HCO_3^- is transported across the basolateral membrane, creating (at least transiently) a "reservoir" of HCO_3^- in the submucosa. This reserve of HCO_3^- is transported by the rich submucosal vascular supply to the surface epithelial cells and is critical in providing an adequate quantity of HCO_3^- to maintain epithelial intracellular pH.

CAUSATIVE FACTORS IN GASTRIC ULCER FORMATION

Nonsteroidal Antiinflammatory Drugs

On a regular basis worldwide, more than 30 million people consume nonsteroidal antiinflammatory drugs

 Copyright 2004, Elsevier (USA). All rights reserved.

(NSAIDs), including low-dose aspirin. Low-dose aspirin is widely used because of its efficacy in reducing cardiovascular events. This has lead to an epidemic of ulcer complications. In the United States, NSAID use accounts for over 25% of all reported adverse drug reactions, with an estimated 16,500 arthritic patients dying every year from gastrointestinal toxicity. The mechanisms of NSAID-induced mucosal injury are incompletely understood but include both topical and systemic components.

NSAID users with negative baseline endoscopies who then have followup endoscopies 2–3 months after starting NSAIDs have a gastric ulcer incidence of 6–12%. In cross-sectional studies on chronic NSAID users, the prevalence of gastric ulcers range from 9 to 31%. In case-control studies in which NSAID users are compared with the matched control population, the relative risk of gastric ulceration is 4.0.

Topical Effects

NSAIDs can accumulate to very high levels within the surface epithelial cells by a phenomenon called "ion-trapping." Aspirin and many NSAIDs are weak organic acids that remain in the nonionized form in the strong acidic environment of the gastric lumen and can freely diffuse across the cell membrane. Once across the membrane, the high intracellular pH causes the H^+ to dissociate, trapping the negatively charged organic compound in the cell. As the nonionized form of the NSAID/aspirin remains in equilibrium across the cell membrane, more of the drug diffuses into the cell, leading to a much higher total drug concentration than outside the cell. The exact mechanism of NSAID-induced cellular toxicity is uncertain but may involve uncoupling of the oxidative phosphorylation necessary for energy metabolism, inhibition of ion transport, or alteration of membrane permeability.

NSAIDs also decrease mucus–bicarbonate secretion, compromise the hydrophobic properties of the mucus coat, and inhibit mucosal prostaglandin secretion. Some NSAIDs enter the enterohepatic circulation, leading to repeated exposure of the gastrointestinal mucosa to high intestinal NSAID concentrations. However, there is good evidence to suggest that these topical effects are not the major cause of NSAID-mediated injury. Previous attempts to minimize topical injury, such as the use of prodrugs, enteric-coated preparations, or parenteral administration, have failed to reduce the risk of ulcer.

Systemic Effects

The nonselective inhibition of cyclooxygenase (COX) by NSAIDs leading to depletion of endogenous prostaglandins is the most important cause of ulcer formation. Prostaglandins regulate mucosal blood flow, epithelial cell proliferation, epithelial restitution, mucosal immunocyte function, mucus and bicarbonate secretion, and basal acid secretion. Inhibition of prostaglandin synthesis probably weakens the gastric mucosal defense to resist luminal irritants. Animal studies involving immunization against prostaglandins (PGE2, PGI2, and PGF-2α), leading to their depletion, have produced gastric ulcers that macroscopically and histologically resemble NSAID-induced ulcers. Large-scale, placebo-controlled trials have shown that coadministration of the prostaglandin E1 analogue (misoprostol) with NSAIDs significantly reduces the incidence of gastric ulcer and its complications.

NSAIDs probably also damage the gastric mucosa via other mechanisms. Acute animal experiments have shown that neutrophil adherence to the endothelium of gastric microcirculation is critical in initiating NSAID injury. Neutrophil adherence damages the mucosa by liberating oxygen free radicals, releasing proteases, and obstructing capillary blood flow.

Role of Luminal Acid

Gastric acid plays a secondary role in the pathogenesis of NSAID-induced ulcer. Acid exacerbates NSAID injury by disrupting the basement membrane, producing deeper injury, interfering with platelet aggregation, and impairing ulcer healing. Concomitant use of high-dose histamine-2 (H2) antagonists or proton pump inhibitors has been shown to reduce the incidence of NSAID ulcers.

Helicobacter pylori

Helicobacter pylori accounts for about 70% of gastric ulcers. *Helicobacter pylori* infection induces mucosal injury not by direct invasion, but instead by the bacterial release of enzymes that destroy the surface mucin and phospholipids. *Helicobacter pylori* infection causes chronic inflammation through the release of cytokines and chemokines. Further injury is mediated by the production of ammonia, release of virulence factors, and induction of a hyperacidic state in some patients.

The patterns of gastritis associated with gastric and duodenal ulcers may differ. Duodenal ulcers are associated with an antrum-predominant gastritis whereas gastric ulcers are associated with a diffuse gastritis. Diffuse gastritis is associated with low acid output, gastric atrophy, and adenocarcinoma. Gastric acid appears to be an important factor in the distribution of *H. pylori* colonization. Despite being one of the most efficient producers of urease, which generates ammonia to

neutralize acid, *H. pylori* often colonizes the antrum because of the high acidity in the corpus. It has been demonstrated that acid suppression with proton pump inhibitors results in a proximal shift of *H. pylori* in the stomach, with increased colonization of the corpus. This leads to increased mucosal inflammation and further reduction in acid secretion. It is therefore postulated that the type of gastritis acquired may be dependent on various environmental factors, such as age of exposure, concurrent infection, and malnutrition, all of which may affect the parietal cell mass and acid-secreting capability.

Stress

Patients suffering from severe stress, such as critically ill patients in intensive care units, may develop stress ulcers. Stress-induced gastric mucosal injury has been reproduced in animal studies. Presence of coagulopathy, hypotension, and respiratory failure are key risk factors to development of ulcer bleeding. The mechanism of injury is unrelated to *H. pylori* and no inflammatory response is noted on biopsy of such ulcers. Although certain conditions may lead to acid hypersecretion, such as head injury or burns, it is believed that mucosal ischemia and reperfusion injury, with impairment of mucosal defense, are the main causative factors. Current evidence indicates that prophylaxis with proton pump inhibitors is effective in preventing bleeding from stress ulcers.

Other Causative Factors

Smoking

Cigarette smokers have a twofold increase in risk of developing ulcers compared to nonsmokers and there is a dose-dependent relationship between the amount of smoking and ulcer incidence. Risk of ulcer complications, such as perforation, has also been shown to be related to smoking in a dose-dependent relationship. Epidemiological studies have shown a strong association between *H. pylori* infection and smoking. Smoking may have an additive effect on the ulcer risk in chronic NSAID users. The exact mechanism of injury is unknown, but smoking reduces prostaglandin synthesis and increases generation of free radicals.

Genetic Predisposition

Patients who possess the blood group O antigen, the Lewis phenotype (a−b−), and the ABH nonsecretor trait have increased risk of peptic ulcers. However, it is believed that genetics plays only a minor role in the pathogenesis of peptic ulcers.

Other Factors

Alcohol intake, psychological factors, and diet have all been previously implicated as possible risk factors for ulcerogenesis. However, there is no compelling evidence to support a major role for these factors in ulcer formation.

EPIDEMIOLOGY

Gastric ulcers were the main form of ulcer disease until the early 1900s, when there was a marked increase in the incidence of duodenal ulcers. The incidence of gastric ulcers has remained relatively stable; a modest decrease in gastric ulcer mortality occurred from 1976 to 1986, but the mortality rate from gastric ulcer decreased from 3.5 per 100,000 in 1962 to 1 per 100,000 in 1979.

Epidemiological surveys conducted in the United States have demonstrated a significant increase in the hospitalizations of elderly patients for complicated ulcers that present with bleeding or perforation. This has been associated with an increase in mortality from ulcer disease in patients older than 75 years (despite the overall mortality being relatively static). This rise has been attributed to the widespread use of NSAIDs in the geriatric population. The aging of the population in developed countries and with the increased usage of NSAIDs in older patients have led to a marked increase in hospitalization rates for peptic ulcers in patients older than 65 years. The availability of H2 receptor antagonists, prostaglandin analogues, and proton pump inhibitors has not decreased the rate of ulcer complications.

Risk Factors for NSAID-Associated Ulcer Complications

The risk of developing ulcer complication in chronic NSAID users is about 2% annually. Epidemiological studies have identified a number of risk factors for NSAID-associated ulcer complications:

Past History of Ulcer Disease or Gastrointestinal Bleeding

A history of ulcer complications is the most important risk factor for recurrent ulcer complications associated with NSAID use. Ulcers tend to recur at the site of previous ulceration. This may be due to NSAID-related inhibition of ulcer healing, leading to more scarring and less reconstitution of normal gastric mucosa.

Elderly Population

Studies on age-related changes in gastric mucosal defenses have shown that prostaglandin synthesis,

mucosal blood flow, and bicarbonate secretion all decrease with age.

Dosage of NSAIDs

The risk of ulcer complications increases with higher doses of NSAIDs or use of multiple NSAIDs.

Concomitant Use of Anticoagulants or Corticosteroids

Although the use of corticosteroids alone does not increase ulcer risk, concomitant use of corticosteroids and NSAIDs substantially increases the risk of ulcer complications compared with the use of NSAIDs alone. The mechanism for this phenomenon is not currently understood. Use of anticoagulants in NSAID users increases the incidence of gastrointestinal bleeding, presumably by the NSAID antihemostatic properties.

DIAGNOSIS OF GASTRIC ULCERS

Clinical Features

Gastric ulcer disease is characterized by chronic relapsing and remitting epigastric pain that is dull in nature. However, at least 10% of ulcers are clinically silent, particularly in elderly patients and with the use of NSAIDs/aspirin. Many patients describe a nonspecific type of upper abdominal pain with no definite relation to food. The differential diagnosis of epigastric pain includes gastroesophageal reflux, gastric or pancreatic neoplasms, mesenteric ischemia (abdominal angina), or pancreatic or biliary disease. Pain may radiate from other areas (cardiac ischemia or aortic aneurysms/dissection). Ulceration of the stomach may be secondary to other systemic illness. Other presentations of gastric ulcer are the development of ulcer complications, such as hemorrhage, perforation, or gastric outlet obstruction. Examination of the patient may reveal nonspecific tenderness, but often the abdomen is normal to palpation.

Diagnosis

Endoscopy is currently the investigation of choice for the diagnosis of gastric ulcers. Although previous studies have shown that double-contrast studies performed by experienced radiologists have an equal accuracy in diagnosing gastric ulcers, use of endoscopy allows for biopsies or cytological brushings of any lesions detected. This is important because ulcerating carcinomas cannot be reliably distinguished from benign ulcers radiographically or endoscopically. Biopsies do not always exclude malignancy; 4% of ulcers initially labeled as benign eventually turn out to be malignant on repeated examinations. Hence, followup endoscopy is recommended to confirm ulcer healing for patients with gastric ulcers.

Complications of Gastric Ulcers

Hemorrhage

Ulcer bleeding is the most common complication of gastric ulcer and is more common in the elderly and in those on anti-platelet agents or anticoagulants. Mild cases may present only with iron deficiency or anemia, and in more severe cases present with coffee ground vomiting, melena, hematemesis, fresh rectal bleeding and hemodynamic compromise.

Perforation

Ulcer perforation is increasing in frequency in the elderly population, probably due to the increase in NSAID use. Chest radiographs may show free gas under the diaphragm but it should be stressed that absence of free gas does not exclude perforation. The ulcer can penetrate into adjacent organs or may be spontaneously sealed off, which may not give rise to signs of peritoneal irritation.

Gastric Outlet Obstruction

Ulcers near the gastroduodenal junction may lead to gastric outlet obstruction, either by inducing local inflammation and swelling or by mechanical obstruction from scarring. Gastric outlet obstruction may present with weight loss, poor appetite, abdominal pain, and vomiting. In severe obstruction, patients may have projectile vomiting and may not tolerate oral intake.

MANAGEMENT

Acute Management of Ulcers

Acid Suppressive Agents

Despite the fact that most gastric ulcer patients have a normal or reduced level of acid secretion, acid suppressive therapy is the mainstay of treatment. Gastric ulcer healing is related to the degree and duration of acid suppression. Antacids in large doses improve the healing of ulcers but have a high incidence of side effects. H2 receptor antagonists and proton pump inhibitors are currently the drugs of choice for the promotion of gastric ulcer healing. Gastric ulcers tend to heal more slowly than duodenal ulcers and smoking is known to retard the healing of ulcers. The ulcer healing rate with the use of H2 receptor antagonists is around 88% in 8 weeks. Proton pump inhibitors heal ulcers faster, and their ulcer healing rate is over 90%. Both

H2 receptor antagonists and proton pump inhibitors have excellent safety profiles.

Drugs Promoting Mucosal Defense

Sucralfate is a complex aluminum salt that forms an insoluble coating over the gastric and duodenal mucosa. The main mechanism of action of sucralfate is likely to be a barrier function against harsh luminal factors such as acid, pepsinogen, and bile salts. Sucralfate also has tropic effects on the gastric mucosa and stimulates secretion of prostaglandin and mucin. Studies have shown that sucralfate is as effective as H2 receptor antagonists, with an average ulcer healing rate of 88% in 8 weeks. Because sucralfate is only minimally absorbed, it has an excellent safety profile; the most common side-effect is constipation.

Misoprostol is a PGE1 analogue that has been approved in the United States. The main action of misoprostol is a local topical effect on the gastric mucosa; it acts by reducing acid secretion and stimulating the various mucosal defense mechanisms. Prostaglandin analogues also stimulate mucosal blood flow, which is important in ulcer healing. Misoprostol is very effective in the prophylaxis of gastroduodenal ulcers in patients taking NSAIDs. However, diarrhea is a common side effect, occurring in up to 30% of patients, and its effects on uterine pressure contraindicate its use in pregnancy.

Surgery

The development of effective and safe acid suppressive agents has relegated surgery to a secondary role in the management of benign gastric ulcer. Surgery is currently primarily used in gastric ulcer complications.

Management of Ulcer Complications

Hemorrhage

The first-line management of bleeding gastric ulcer is therapeutic endoscopy. Visible blood vessels are often visualized at the ulcer base during endoscopy. Treatment modalities include epinephrine injection (which creates a local tamponade effect and vasoconstriction), thermocoagulation with heater probe to coagulate the blood vessel, and application of endoclips to clamp the blood vessel. Intravenous acid suppressive agents alone cannot replace the role of endoscopy in hemostasis. However, use of intravenous proton pump inhibitors as an adjuvant therapy can reduce the risk of rebleeding after endoscopy in high-risk patients. In refractory bleeding that cannot be controlled with endoscopic therapy or after iatrogenic perforation of the ulcer during hemostasis, surgery is required. The ulcer is most often plicated, but in severe bleeding, a partial gastrectomy may be necessary.

Perforation

In spontaneous or iatrogenic perforation of peptic ulcers, surgery is required to lavage the peritoneal cavity and to repair the perforation. There have been reports of endoscopic closing of perforations with endoclips at specialist endoscopic centers, but the safety and efficacy of such procedures are still not established.

Gastric Outlet Obstruction

Gastric outlet obstruction can often be treated conservatively with acid suppressive therapy to allow ulcer healing and gradual reduction of inflammation and swelling around the pylorus. In refractory cases, endoscopic dilatation or surgical pyloroplasty may be necessary.

Prevention of Gastric Ulcer Recurrence in *Helicobacter pylori*-Infected Patients

Eradication of *H. pylori* has been shown to reduce recurrence of ulcers and is more cost-effective than providing long-term maintenance therapy. Effective and safe drug regimes for *H. pylori* eradication are available.

Prevention of Gastric Ulcer Recurrence in NSAID Users

Antiulcer Therapy

NSAID-induced gastric ulcers heal rapidly with conventional ulcer therapy once the NSAID is discontinued. By removing the causative factor, risk for ulcer recurrence is also low, provided there are no other causative factors, such as *H. pylori* infection. However, in patients who require long-term use of NSAIDs or aspirin, maintenance antiulcer therapy is necessary. A meta-analysis of 33 randomized controlled trials assessing the role of misoprostol, H2 receptor antagonist, and proton pump inhibitors in ulcer prophylaxis in chronic NSAID users showed that high-dose misoprostol (800 μg daily) is the only prophylactic agent to reduce ulcer complications. Standard doses of H2 receptor antagonists do not reduce the risk of gastric ulcers. Double doses of H2 receptor antagonist or proton pump inhibitors are effective in reducing gastric ulcers. High-dose misoprostol is likely the most effective prophylactic agent because it corrects NSAID-induced prostaglandin depletion and restores the normal gastric physiology. In patients with good drug compliance, it has been shown that 93% of patients on high-dose misoprostol do not develop ulcers at

12 weeks, in comparison to 82% of patients on proton pump inhibitors. However, the high incidence of side effects and drug noncompliance of misoprostol makes its efficacy comparable to proton pump inhibitors in the clinical setting.

COX-2 Inhibitors

Cyclooxygenase, the key enzyme in the production of prostaglandin, exists in three isoforms. COX-1 is constitutively expressed and is found in many organ systems, including the gastric mucosa, and is important in the maintenance of normal gastric function, COX-2 is proinflammatory and is up-regulated only during inflammation. COX-3 is a variant of COX-1 that has recently been identified in the brain. NSAIDs nonselectively inhibit all isoforms and hence have both beneficial antiinflammatory effects and harmful side effects (including gastrointestinal toxicity). Different NSAIDs have been observed to confer different risks of ulcer formation and ulcer complication. NSAIDs that are relatively selective for COX-2 produce a lower relative risk of ulcer complications.

Highly selective COX-2 inhibitors have been rapidly developed and marketed as effective antiinflammatory agents with reduced gastrointestinal side effects. Two large-scale studies (the CLASS study and the VIGOR study) have demonstrated that use of COX-2-selective inhibitors such as celecoxib and rofecoxib reduce the risk of ulcer complication compared to nonselective NSAIDs. These studies were largely based on average-risk patients and there are insufficient data on the gastric safety of COX-2 inhibitors in high-risk cases (i.e., patients with a history of peptic ulcer disease). The gastrointestinal protective effect of COX-2-selective inhibitors is offset by the use of low-dose aspirin and hence maintenance acid suppressive therapy may still be necessary for patients who require these agents for cardiovascular prophylaxis. Finally, in the VIGOR study, patients receiving rofecoxib had a significant increase in the incidence of acute myocardial infarction compared with patients receiving naproxen. It remains unclear whether this finding is due to the lack of antiplatelet effect or to unopposed inhibition of COX-2.

The use of selective COX-2 inhibitors is an exciting development in the management of chronic NSAID users who are at risk of ulcer disease. However, the safety profile of these drugs and the determination of which population of patients will actually benefit from COX-2 inhibitors are still under investigation.

Nitric Oxide-Releasing NSAIDs

There is currently much interest in the effectiveness of nitric oxide (NO) to reduce NSAID-induced gastric mucosal damage. Studies comparing the incidence of complications in NO plus NSAIDs versus conventional NSAIDs are currently underway.

See Also the Following Articles

Duodenal Ulcer • Epithelium, Repair of • Gastric Acid Secretion • Gastric Outlet Obstruction • H2-Receptor Antagonists • *Helicobacter Pylori* • Hemorrhage • NSAID-Induced Injury • Perforation • Proton Pump Inhibitors • Smoking, Implications of

Further Reading

Chan, F. K., and Leung, W. K. (2002). Peptic ulcer disease. *Lancet* **360**(9337), 933–941.

Cotton, P. B., and Williams, C. (1996). "Practical Gastrointestinal Endoscopy," 4th Ed. Blackwell Publ., Oxford.

Domschke, W., and Konturek, S. (eds.) (1993). "The Stomach: Physiology, Pathophysiology and Treatment." Springer-Verlag, Berlin.

Hawkey, C. J. (2000). Nonsteroidal anti-inflammatory drug gastropathy. *Gastroenterology* **119**(2), 521–535.

Yamada, T. (ed.) (1999). "Textbook of Gastroenterology," 3rd Ed. Lippincott Williams & Wilkins, Philadelphia.

Gastric Volvulus

WILLIAM HARFORD
University of Texas Southwestern Medical Center, Dallas

Borchardt's triad The combination of pain, unproductive vomiting, and the inability to pass a nasogastric tube in a case of acute gastric volvulus.
gastric volvulus Condition in which the stomach twists upon itself.
mesenteroaxial gastric volvulus Condition in which the stomach folds upon itself along its short axis.
organoaxial gastric volvulus Condition in which the stomach twists along its long axis.

Gastric volvulus occurs when the stomach twists upon itself. Such an event may be transient and produce few symptoms or it may lead to chronic or acute obstruction. If the vascular supply is compromised, acute gastric ischemia may occur.

ETIOLOGY AND PATHOGENESIS

In approximately two-thirds of cases, gastric volvulus is associated with a large diaphragmatic hernia. Sliding hiatal hernias do not lead to gastric volvulus.

Gastric volvulus may be organoaxial or mesenteroaxial (Fig. 1). Organoaxial volvulus is the most common type. The stomach twists along its long axis, which usually passes through the gastroesophageal and gastropyloric junctions. The antrum rotates anteriorly and superiorly and the fundus rotates posteriorly and inferiorly, twisting the greater curvature at some point along its length (Fig. 1, parts 3A and 3B). Less commonly, the long axis passes through the body of the stomach itself. In these cases, the greater curvatures of both the antrum and the fundus rotate anteriorly and superiorly (Fig. 1, parts 2A and 2B). Organoaxial volvulus is usually acute and vascular compromise with gastric ischemia may occur.

In mesenteroaxial gastric volvulus, the stomach folds on its short axis running across from the lesser curvature to the greater curvature. The antrum twists anteriorly and superiorly (Fig. 1, parts 1A and 1B). In rare cases, the antrum and pylorus rotate posteriorly. Mesenteroaxial volvulus is more likely than organoaxial volvulus to be chronic, with symptoms of incomplete and intermittent obstruction.

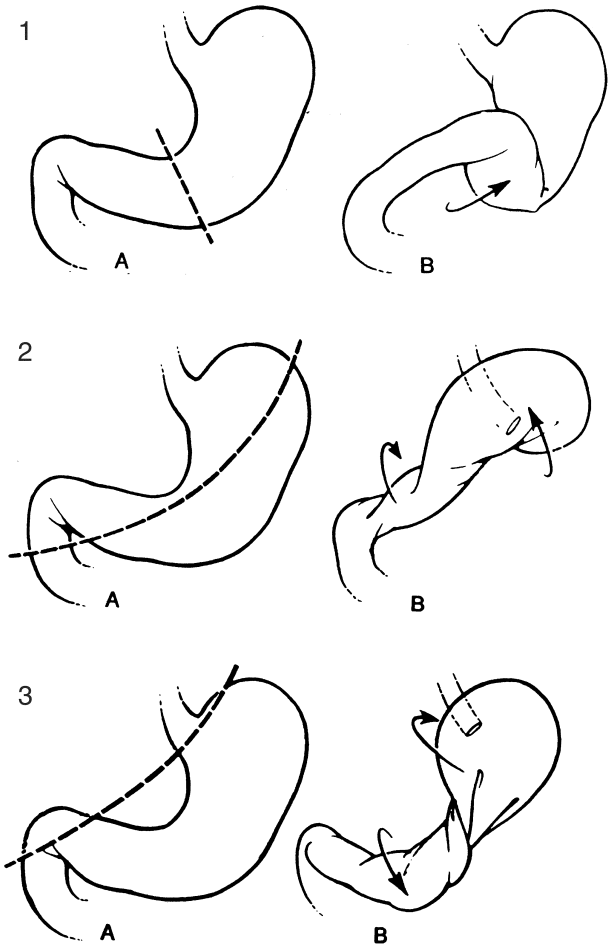


FIGURE 1 Gastric volvulus. (1A) Axis for potential mesenteroaxial volvulus bisecting the lesser and greater curvatures. (1B) Mesenteroaxial volvulus resulting from anterior rotation of the antrum along this axis. (2A) Axis for potential organoaxial volvulus passing through the body of the stomach. (2B) Organoaxial volvulus resulting from anterior superior rotation of the antrum along this axis. (3A) Axis for potential organoaxial volvulus passing through the gastroesophageal junction and the pylorus. (3B) Organoaxial volvulus resulting from anterior superior rotation of the antrum and posterior inferior rotation of the fundus along this axis. Redrawn from Carter, R., Brewer, L. A., and Hinshaw, D. B. (1980). Acute gastric volvulus. *Am. J. Surgery* 140, 101–102, with permission.

 Copyright 2004, Elsevier (USA). All rights reserved.

INCIDENCE AND PREVALENCE

Since many cases of gastric volvulus present with chronic, intermittent, and nonspecific symptoms, it is likely that many go undiagnosed and thus the incidence and prevalence are unknown. Approximately 15 to 20% of reported cases occur in young children with a congenital diaphragmatic defect. Among adults, men and women are equally affected and the peak incidence is in the fifth decade.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The presentation of acute gastric volvulus includes sudden severe pain in the upper abdomen or lower chest and persistent but unproductive vomiting. Because the volvulus causes gastric obstruction, it is often impossible to pass a nasogastric tube into the stomach. This combination of pain, unproductive vomiting, and inability to pass a nasogastric tube is termed Borchardt's triad. The differential diagnosis includes myocardial infarction, biliary obstruction, and acute pancreatitis. Physical examination may reveal evidence that the stomach is in the left chest. A large gas-filled structure may be seen in the chest on X ray. A barium upper gastrointestinal X ray or computerized tomography with oral contrast will confirm the diagnosis. If upper endoscopy is performed, it may show twisting of the gastric folds, but endoscopy is not prudent if gastric ischemia is suspected.

In contrast to acute gastric volvulus, chronic gastric volvulus is often associated with mild and nonspecific symptoms. These may include dysphagia, epigastric discomfort or fullness, bloating, and heartburn, particularly after meals. Symptoms may be present for months to years and it is likely that many cases are never diagnosed. Since the volvulus may be intermittent and transient, the diagnosis should be suspected if an upper gastrointestinal X ray shows a large diaphragmatic hernia, even if the stomach is not twisted at the time of the X ray.

TREATMENT AND PROGNOSIS

Acute gastric volvulus is an emergency and should generally be treated surgically once the diagnosis is confirmed. Surgery for gastric volvulus may be performed by open or laparoscopic techniques. Elective surgical repair is commonly performed with laparoscopic techniques. Once gastric torsion is reduced, the stomach must be fixed by gastropexy or tube gastrostomy and any associated diaphragmatic hernia must be repaired. Acute gastric volvulus has been associated with a high mortality in the past, but the prognosis has markedly improved in recent years.

See Also the Following Articles

Diaphragmatic Hernia • Hernias • Volvulus

Further Reading

Harford, W. V., and Jeyarajah, R. (2002). Abdominal hernias and their complications, including gastric volvulus. *In* "Gastrointestinal Disease, Pathophysiology, Diagnosis, and Management" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 7th Ed., Chap. 20, 369–385. W. B. Saunders Co., Philadelphia, PA.

Gastrin

CHRIS J. DICKINSON
University of Michigan Medical Center, Ann Arbor

cholecystokinin Peptide hormone produced by endocrine cells of the upper small intestine; secreted on ingestion of a meal; the major hormone responsible for pancreatic enzyme secretion and gallbladder contraction.
gastric development How the stomach grows and develops in a fetus.
gastrin Peptide hormone produced in the stomach; secreted into the bloodstream, enhances gastric acid secretion.
gastrointestinal cancers Tumors that arise from the gut tissues.
Helicobacter pylori Bacterium that lives in the stomach of humans; can cause peptic ulcers in some infected individuals.
H^+,K^+-ATPase Protein that pumps protons across the cell membrane in a 1 : 1 exchange for potassium ions.
parietal cell Acid-secreting cell of the stomach.
peptic ulcer disease Disorder caused by ulcers in the stomach or upper intestine (usually the duodenum); can arise from the action of gastric juices that contain acid and digestive enzymes such as pepsin.
signal transduction mechanisms Chemicals and other mechanisms by which a peptide binding to its receptor on the cell membrane results in some physiologic effect within a cell.
somatostatin Peptide hormone produced in the stomach; inhibits gastric acid secretion and gastrin secretion.

After the discovery of secretin by Bayliss and Starling in 1902, they surmised that a similar hormone might exert control over gastric acid secretion. Edkins then reported the identification of a substance, "gastrin," found in antral but not fundic mucosal extracts that stimulated gastric acid secretion. In 1962, Gregory and Tracy, purified an amidated heptadecapeptide now referred to as gastrin. Shortly thereafter, a gastrin peptide was synthesized, allowing investigators to examine the physiological actions of gastrin. The subsequent development of a gastrin radioimmunoassay helped researchers examine its secretion in physiologic and pathophysiologic states. It quickly became apparent that the regulation of gastric acid secretion was quite complex and that to elucidate those mechanisms, it would be necessary to examine all of the regulatory cells in isolation to determine the precise biological actions of gastrin and other gut peptides. Isolation of acid-secreting parietal cells, gastrin-producing G cells, and somatostatin-producing D cells from the stomach in primary culture became an important step in understanding gastrin secretion and biological actions. To explore the mechanisms of gastrin biosynthesis, investigators initially tried to identify larger molecular forms of gastrin that were presumed to be precursors of the mature peptide. Isolation of a complementary DNA encoding gastrin allowed investigators to hypothesize the presence of various precursors and then develop antisera to confirm the existence of the precursors in tissues and their relationship to the mature peptide. Finally, characterization and isolation of the receptor(s) for gastrin on various cells have allowed researchers to elucidate the mechanisms of gastrin's biological actions on target cells.

GASTRIN DURING DEVELOPMENT

Gastrin is transiently expressed in the neonatal pancreatic islets. However, the physiologic role of gastrin in the developing islet is unclear and gastrin-deficient mice do not have a known defect in islet function. Gastrin appears later in the stomach of the developing rat, as pancreatic gastrin diminishes. Gastrin-producing G cells can be detected in the human fetus as early as 12 weeks of gestation and increase to reach a maximum between 26 and 36 weeks of gestation.

Neonatal human plasma gastrin levels are approximately three times higher than those found in adults. Unfortunately, the physiological significance of neonatal hypergastrinemia is not known. It has been suggested that these elevated levels may play a role in the growth and development of the gastrointestinal (GI) tract. To explore the role of gastrin in gut development, two groups have developed gastrin-deficient mice. These mice are viable and develop normally with normal weight gain, suggesting that gastrin is not a vital hormone. However, both basal and stimulated gastric acid secretions are severely impaired in homozygous animals. Furthermore, there is a reduction in gastric mucosal thickness. Although cholecystokinin (CCK)

 Copyright 2004, Elsevier (USA). All rights reserved.

can bind to the gastrin receptor, CCK does not substitute for gastrin in knockout animals because gastrin-deficient and gastrin receptor-deficient mice share a similar gastric phenotype.

Antral gastrin content increases in the developing rat until the time of weaning. Early weaning or administration of corticosteroids or thyroid hormone increases antral gastrin content and promotes the maturation of the gastric mucosal gastrin receptors. It is unknown whether a similar phenomenon also occurs in the human stomach with a change in diet, but human plasma gastrin levels do remain elevated for several months after birth, when infants are primarily milk fed. The cause of neonatal hypergastrinemia is multifactorial and the rat seems to provide a suitable model for study. A milk-based diet that is high in known gastrin secretagogues such as amino acids and calcium may account in part for the hypergastrinemia. In both rats and humans, there is diminished gastric acid secretion due to a diminished number of functional acid-secreting parietal cells. In mature humans, this inhibits somatostatin secretion, which increases gastrin secretion. Consistent with this is the finding that the somatostatin content of the rat gastric antrum is low at birth and increases gradually until weaning, at around day 21. Moreover, infusion of somatostatin does not inhibit gastrin release in newborn rat pups until day 18 of life, suggesting that somatostatin receptor or postreceptor mechanisms are functionally mature until weaning. The combined effect of low somatostatin content and diminished response to exogenous somatostatin prior to weaning suggests that somatostatin plays a major role in neonatal hypergastrinemia.

GASTRIN RECEPTOR AND SIGNALING

Gastrin requires its carboxyl-terminal amide moiety for full biological activity. Thus, the isolation of a cDNA encoding the receptor that recognizes this motif represented a major advance in the understanding of gastrin physiology. This discovery allowed investigators to more fully elucidate the signal transduction mechanisms, structure/activity relationships, tissue distribution, and biological significance of gastrin. The gastrin/CCK-B, or CCK-2, receptor is a seven-transmembrane "heptahelical" receptor that is similar to the CCK-A, or CCK-1, receptor. The amino acid sequences of gastrin and CCK are quite similar and they share four identical amino acids at the amidated, C terminus of the mature peptide. CCK can bind to both the CCK-1 and CCK-2 receptors whereas gastrin binds to only the CCK-2 receptor at physiologic concentrations. Under normal physiologic circumstances, the serum concentration of CCK is far less than that of gastrin and thus CCK exerts almost no physiologic effects via CCK-2 receptors.

On ligand binding, the gastrin/CCK-B receptor binds to a G protein, which leads to the activation of phospholipase C (PLC); PLC in turn catalyzes the formation of inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 effects the release of intracellular calcium, and DAG activates protein kinase C (PKC). Unlike CCK binding to the CCK-A receptor, gastrin binding to its receptor does not result in the accumulation of cyclic adenosine monophosphate (cAMP). This binding initiates a complex signaling cascade, as shown in Fig. 1. Several studies have shown that gastrin can stimulate the proliferation of gut cells as measured by an increase in either cell number, [^{3}H]thymidine incorporation, ornithine decarboxylase (ODC) activity, or expression of two genes known to regulate cell proliferation (c-*fos* and c-*jun*). Clinical examples of this effect can be seen with the overproduction of gastrin in tumors in the Zollinger–Ellison syndrome, which produces significant gastric hyperplasia. Many of these tropic effects can be blocked by gastrin/CCK-B

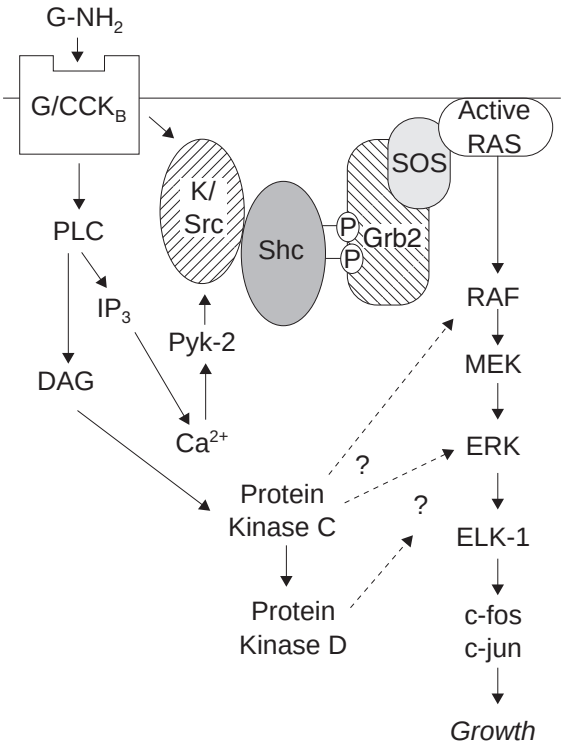


FIGURE 1 Model of gastrin signal transduction mechanisms.

receptor antagonists, suggesting that they are mediated by this receptor.

Although much work remains, progress has been made at characterizing the intracellular events that occur between the initial events at the cell membrane and the resultant actions in the nucleus (Fig. 1). It appears that in addition to activating the serine/threonine kinase (PKC) at the cell membrane, gastrin also activates tyrosine kinases (TKs; e.g., Src), which phosphorylate an adapter protein (Shc). Phosphorylated Shc then associates with two other proteins (Grb2 and a Ras activator, Sos). This complex can then stimulate the Ras-dependent activation of the mitogen-activated protein (MAP) kinase as well as the upstream activator of MAP kinase, a MAP and ERK kinase (MEK). ERK then phosphorylates a nuclear transcription factor (Elk-1) that enhances c-fos expression, after which c-fos combines with c-jun to induce transcription of other genes that regulate growth. Although gastrin receptor antagonists completely reverse the effects of c-fos expression, PKC inhibitors result in only a 40–50% reduction in gastrin-induced c-fos expression. This suggests that the tropic effects of gastrin are mediated by PKC-dependent and -independent pathways. The IP_3-mediated release of intracellular calcium induces tyrosine phosphorylation of the focal adhesion kinase Pyk-2 and Src. PKC can also activate protein kinase D, although the mechanisms by which PKD stimulates growth remain unknown. Other investigators have identified gastrin-mediated activation of the GTP-binding protein, Rho, which can stimulate growth via PKC- and MEK-independent pathways, presumably via changes in the cytoskeleton. It is anticipated that these and other mechanisms will soon be elucidated and will aid in understanding gastrin's potentially important growth-promoting effects.

GASTRIN PROCESSING AND PHYSIOLOGICAL RELEVANCE OF OTHER PROGASTRIN-PROCESSING PRODUCTS

Gastrin, as is the case with other regulatory neuropeptides, is initially synthesized as a large, biologically inactive precursor, progastrin (Fig. 2). Endoproteolytic processing of progastrin within G cells results in the formation of a glycine-extended form of gastrin (G-Gly) that is found in high concentrations in the developing rat gastric antrum. G-Gly serves as a substrate for the amidation enzyme, peptidylglycine α-amidating monooxygenase (PAM), that catalyzes the formation of fully processed mature, amidated

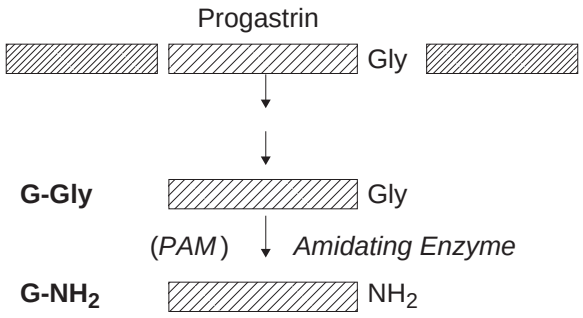


FIGURE 2 Progastrin posttranslational processing. Progastrin is sequentially processed to a glycine-extended form of gastrin (G-Gly). G-Gly then serves as a substrate for the amidation enzyme peptidylglycine α-amidating monooxygenase (PAM), which catalyzes the formation of amidated gastrin (G-NH_2). The C-terminal amide moiety is necessary for G-NH_2 binding to the gastrin/CCK-B receptor.

gastrin. Gastrin (G-NH_2) requires its carboxyl-terminal amide moiety for full biological activity mediated by gastrin/CCK-B receptors. Indeed, removal of the carboxyl-terminal amide in gastrin completely abolishes its acid stimulatory effects mediated by standard gastrin/CCK-B receptors, and the immediate precursor of amidated gastrins, G-Gly, is at least four orders of magnitude less potent than G-NH_2 in stimulating acid secretion from gastric parietal cells. Nevertheless, interest in the physiologic effects of G-Gly has been fueled by the observations that G-Gly is stored in brain and gut tissues, secreted with G-NH_2 from antral G-cells into the circulation, and achieves concentrations in plasma roughly equivalent to those of G-NH_2. Moreover, G-Gly is seen in greater concentrations than is G-NH_2 during development and in some malignant tissues that express gastrin, such as Zollinger–Ellison tumors and colon cancers. Finally, biosynthetic studies suggest that G-Gly may be a distinct end product of progastrin processing. Thus, the evidence points to a role for G-Gly as a growth factor but not as a direct acid secretagogue.

To further characterize the potential tropic effects of G-Gly, a study examined whether G-Gly might function as a growth factor in a fashion that could be distinguished from its relatively weak effects on the standard gastrin/CCK-B receptor. The abilities of G-NH_2 and G-Gly to stimulate DNA synthesis were compared. Both G-NH_2 and G-Gly stimulated [^{3}H]thymidine incorporation in a dose-dependent fashion. As expected, the stimulation induced by G-NH_2 was completely reversed by selective gastrin/CCK-B receptor antagonists. Further studies characterized a distinct receptor and signaling cascade for G-Gly that resulted in

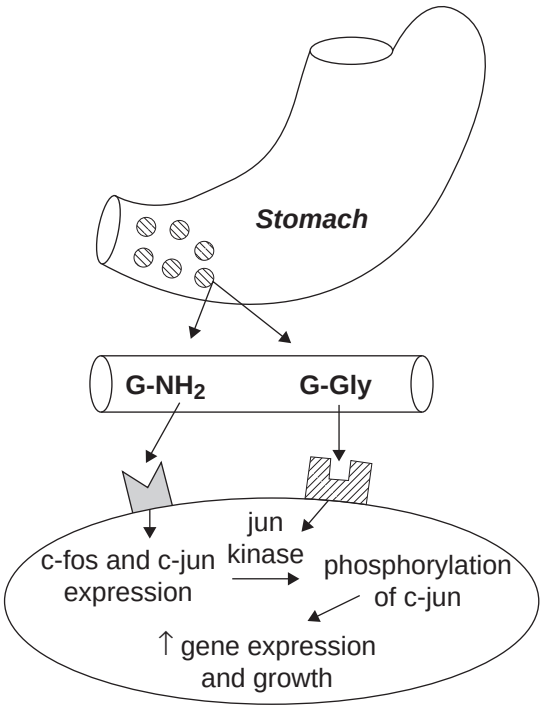


FIGURE 3 Mechanisms of G-NH_2 and G-Gly tropic effects. Amidated gastrin (G-NH_2) and its glycine-extended precursor (G-Gly) are released from G cells in the gastric antrum into the circulation. G-NH_2 acts through gastrin/CCK-B receptors to enhance the expression of c-fos and c-jun. G-Gly, via another distinct receptor and jun kinase, phosphorylates and thus bioactivates c-jun. c-fos and phosphorylated c-jun then induce the expression of other genes necessary for the proliferation of gut cells.

phosphorylation and activation of jun kinase but no changes in c-fos expression (Fig. 3).

The inability of investigators to isolate a cDNA encoding the putative G-Gly receptor has hampered further study in this area. Other data help to confirm the role that G-Gly may play in gastrointestinal physiology. G-Gly can stimulate growth in an autocrine fashion. Although acute administration of G-Gly has no effect on gastric acid secretion, chronic administration of G-Gly markedly enhances stimulated but not basal acid secretion from isolated parietal cells and *in vivo* via an increase in the expression of H^+,K^{+t}-ATPase within gastric parietal cells. G-Gly may play a role in the growth of the colon. Indeed, gastrin-deficient mice have a diminished colonic proliferative index and transgenic mice overexpressing G-Gly have enhanced colonic proliferation. Taken as a whole, these observations suggest that growth-related receptors for G-Gly work in concert with G-NH_2 to enhance the functional development of the gut.

GASTRIN, PEPTIC ULCER DISEASE, AND CANCER

Overview

The studies just described greatly expanded knowledge of the mechanisms of gastrin biosynthesis, release, and receptor/ligand interactions. However, the central clinical issue that initially drove these studies was elucidation of the mechanisms of peptic ulcer formation. For many years, peptic ulcer disease was a major cause of morbidity and mortality, and early studies seemed to confirm the notion that gastric acid was requisite to the formation of duodenal ulcers. Thus, it was hoped that a more thorough understanding of the regulation of gastric acid secretion would elucidate the pathogenesis of peptic ulcer disease. Early investigators identified gastrin as the primary postprandial regulator of gastric acid secretion.

The G cell is a classic gut endocrine cell, organized with microvilli on a luminal surface that allow the G cell to detect the presence of food within the stomach. Gastrin released from G cells into the circulation has three major effects mediated by gastrin/CCK-B receptors. First, gastrin can stimulate the release of acid directly from the parietal cell. Second, gastrin can also stimulate acid secretion by enhancing the release of histamine from enterochromaffin-like (ECL) cells, which then bind to histamine-2 (H2) receptors on the parietal cell (see Fig. 4). Third, gastrin also stimulates the release of somatostatin from gastric D cells. Somatostatin is a potent inhibitor of gastric acid secretion, acting directly on the parietal cell or blocking the continued release of

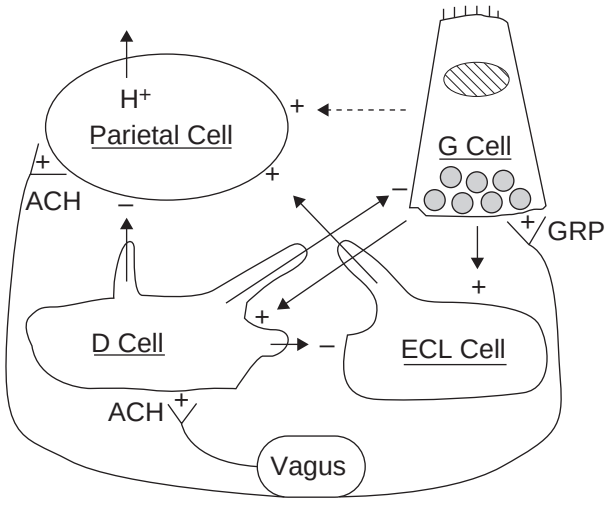


FIGURE 4 Interaction in the lumen between the G cell and other gastric cells regulating acid secretion. ACh, Acetylcholine; GRP, gastrin-releasing peptide; ECL, enterochromaffin-like cell.

gastrin from G cells and histamine release from ECL cells. Detailed reviews on the regulation of gastrin release have been published; a summary is provided herein.

Neural Control of Gastrin Secretion

During the cephalic and oropharyngeal phases of a meal, gastrin release is directly stimulated by acetylcholine (ACh) released from vagal fibers via muscarinic receptors. ACh released from other vagal fibers also inhibits the release of somatostatin from antral D cells, which indirectly enhances gastrin secretion. In addition to ACh, gastrin-releasing peptide (GRP) is also secreted from postganglionic vagal nerve fibers and directly stimulates gastrin release via GRP receptors on G cells.

During the gastric phase of a meal, antral distension evokes low levels of gastrin release; this is inhibited by low doses of atropine but is enhanced by atropine at higher doses. Because the gastrin release in response to antral distension is depressed by the β-adrenergic antagonists, an adrenergic sympathetic reflex may be involved. Antral distension in humans also activates inhibitory pathways of gastrin release because it inhibits GRP-stimulated release by an unknown mechanism.

Hormonal Control of Gastrin Secretion

A variety of peptide hormones also regulate gastrin release (detailed discussions of all of the mechanisms can be found elsewhere). Somatostatin is released from bulbous swellings at the end of cytoplasmic extensions of antral D cells onto G cells in many species. This intimate relationship of the D cell with the G cell provides a tonic inhibition to gastrin release and makes somatostatin the primary hormonal regulator of gastrin release. It also provides the mechanism for many of the indirect effects of other hormones on gastrin release.

Chemical Control of Gastrin Secretion

In accord with the presence of G cell microvilli extending into the lumen of the stomach, the G cell does respond to the presence of food within the stomach. In the perfused rat stomach or in antral mucosal segments, peptone stimulates gastrin and inhibits somatostatin release. Neurotoxins, such as tetrodotoxin, GRP antagonists, and atropine, abolish both stimulation of gastrin release and inhibition of somatostatin release, indicating that peptone-induced gastrin release is mediated, in part, by both cholinergic and GRP neurons. However, other groups have shown that amino acids and amines also stimulate gastrin release from isolated canine G cells. Other chemical stimuli, such as an increase in gastric acid production, completely suppress gastrin release. However, this is not a direct effect on the G cell, but is rather mediated by an increase in somatostatin secretion.

Helicobacter pylori, Ulcers, and Gastrin

Studies on the regulation of gastric acid secretion led to the discovery of H2 receptor antagonists that blocked the actions of histamine on the parietal cell. Although these and other gastric-acid-suppressive agents could heal the majority of duodenal ulcers with 6–8 weeks of therapy, the ulcer would generally reappear with cessation of therapy. Thus, the search for the cause of peptic ulcers continued. The discovery of a spiral bacterium (*Helicobacter pylori*) in the inflamed stomachs of patients with duodenal ulcers has led to a revolution in our understanding of ulcer pathogenesis and therapy. Because so many people are afflicted with peptic complaints, this discovery received a great deal of attention from scientists and the lay press. Although there was some initial skepticism regarding the relevance of this discovery, it is now widely accepted that the majority of patients with duodenal ulcers and 60–70% of those gastric ulcers are infected with *H. pylori*. Moreover, failure to eradicate the infection often results in recurrence of the ulcers, even with optimal antisecretory therapy. In contrast, eradication of *H. pylori* results in complete healing and a very low recurrence rate in the absence of reinfection.

Initially, these observations were difficult to mesh with our long-standing view that ulcers were secondary to a disorder in the regulation of gastric acid secretion. Not surprisingly, the view today is that both infection and alterations in gastric acid secretion appear to play a role in peptic ulcer pathogenesis. One observation that leads to this view is that although *H. pylori* is almost always present in patients with duodenal ulcers, the bacterium is more likely to be found in an inflamed gastric antrum than in the duodenum. Although initial infection with *H. pylori* is associated with gastritis and hyposecretion of acid, chronic infection leads to a moderate hypersecretion of gastric acid. It is believed that this chronic hypersecretion of acid results in ulcer formation in some patients. Patients infected with *H. pylori* have higher serum gastrins and decreased levels of antral somatostatin compared to noninfected controls, suggesting that *H. pylori* has some effect on the cells regulating gastrin release and acid secretion. Thus, recent investigations have focused on the mechanisms by which chronic *H. pylori* infection alters the regulation of gastric acid secretion.

When compared with *H. pylori*-negative subjects, patients with duodenal ulcers and *H. pylori* infection have several abnormalities in the regulation of gastric acid secretion. These include a two- to fourfold increase in basal acid secretion, an increased maximal response to exogenous gastrin, a marked increase in acid secretion to GRP, and an increased ratio of basal acid output to maximal gastrin-stimulated acid output. Interestingly, although *H. pylori*-infected patients have increased acid secretion and moderate hypergastrinemia, the level of hypergastrinemia cannot fully account for the increase in acid secretion. *Helicobacter pylori*-positive duodenal ulcer patients also have impaired acid inhibitory mechanisms. Because somatostatin-producing D cells mediate many of the gastric acid inhibitory pathways, the reduced amounts of antral somatostatin seen with chronic infection can contribute to the reduced acid. All of these abnormalities resolve within 1 year after eradication of infection except for the increased acid response to exogenous gastrin. It is doubtful that this response is due to an up-regulated acid secretory mechanism due to prolonged hypergastrinemia, because the *H. pylori*-induced hypergastrinemia resolves completely within 1 month after eradication. Furthermore, in patients with hypergastrinemia secondary to a gastrinoma, the enhancement of parietal cell mass fully resolves within 6 months after tumor resection. Thus, it is felt that patients who develop duodenal ulcers with *H. pylori* infection may have factors that increase their gastrin-stimulated acid secretory response.

To determine how *H. pylori* produces these effects in infected gastric mucosa, investigators have focused on the effects of bacterial products or inflammatory mediators on acid secretion. Because *H. pylori* infection is associated with an underlying gastritis, investigators have examined the effects of inflammatory mediators on the regulation of acid secretion. Interleukin-1β, interleukin-8, and tumor necrosis factor α augment gastrin release from G cells. These and other cytokines are elevated in infected gastric mucosa. Other recent investigations have suggested that inflammation, in general, rather than that specific to *H. pylori* infection, can elevate gastrin production and gastric acid secretion.

Cancer

The role of gastrin in gut tumors has been controversial for some time. Thus, it is useful to review the clinical conditions associated with hypergastrinemia. Hypergastrinemia produced by G cell neoplasms leads to the Zollinger–Ellison syndrome. In afflicted patients, serum gastrin concentrations exceed physiologic levels by severalfold. These slow-growing tumors are often found in the pancreas or duodenum and may not prove fatal for 20 or more years. Patients do exhibit markedly elevated gastric acid secretion and hypertrophy of the histamine-producing ECL cells, but rarely a frank ECL or carcinoid tumor, compatible with a model that gastrin alone is not a carcinogen.

Proton pump inhibitors are now widely prescribed because of their ability to block gastric acid secretion by virtue of their ability to bind to H^+,K^+-ATPase. Although treated patients do develop a mild hypergastrinemia while on medication, there seems to be little evidence of an increased cancer risk, even with long-term administration. In an analogous situation, patients with prolonged gastric inflammation (usually from *H. pylori* infection) often develop hypergastrinemia. A higher percentage of infected patients will develop gut tumors compared to noninfected controls. However, only a small minority ($< 1\%$) of infected patients develop gastric cancer, again suggesting that other factors must be involved.

Transgenic mice overexpressing fully processed, amidated gastrin have an early increased in parietal cell mass and ECL cell numbers similar to that found in human models of hypergastrinemia. Unlike the human condition, transgenic mice develop a late gastritis, and many succumb to gastric cancer. The progression to cancer can be significantly accelerated by concomitant infection with a mouse form of *H. pylori*, i.e., *Helicobacter felis*. Conversely, transgenic mice overexpressing G-Gly, but not amidated gastrin, have no alteration in gastric function, but rather develop an increased colonic proliferative index.

The role of gastrin in the development of colorectal cancer has been controversial for many years. Several studies have demonstrated a growth-stimulatory effect of gastrin or pentagastrin on colon cancers *in vivo* and *in vitro*, whereas others have failed to confirm this effect. Other human studies have been performed with conflicting results, but a recent epidemiological study suggests that prolonged hypergastrinemia associated with *H. pylori* infection may play a role in the development of some human colonic tumors. Because progastrin is often poorly processed in nonendocrine cells such as those found in the colon, it explains why "gastrin" antisera fail to detect significant quantities of amidated peptide. Consistent with this notion is that colon cancers contain large amounts of progastrin and other progastrin processing intermediates such as G-Gly, but little fully processed peptide, suggesting a role for G-Gly in colon cancer. In published phase II studies, investigators have developed a method of inducing antigastrin antibodies in patients with colon and pancreatic cancer. The antibody recognizes both amidated gastrin and G-Gly. Preliminary data suggest prolonged survival

of treated patients, furthering the claim that gastrin plays a role in gut cancers. However, many more data are needed to confirm this finding and to elucidate the mechanisms and circumstances under which gastrin exhibits these actions.

See Also the Following Articles

Cholecystokinin (CCK) • Duodenal Ulcer • Enterochromaffin-like (ECL) Cells • Gastric H^+,K^+-ATPase • Gastrinoma • Gastrin-Releasing Peptide (GRP) • *Helicobacter pylori* • Histamine • Pancreatic Enzyme Secretion (Physiology) • Parietal Cells • Somatostatin

Further Reading

Boel, E., Vuust, J., Norris, F., Norris, K., Wind, A., *et al.* (1983). Molecular cloning of human gastrin cDNA: Evidence for evolution of gastrin by gene duplication. *Proc. Natl. Acad. Sci. U.S.A* **80**, 2866–2869.

Dockray, G. J., Varro, A., Dimaline, R., and Wang, T. C. (2001). The gastrins: Their production and biological activities. *Annu. Rev. Physiol.* **63**, 119–139.

Edkins, J. S., and Tweedy, M. (1909). The natural channels of absorption evoking the chemical mechanism of gastric secretion. *J. Physiol. Lond.* **38**, 263–267.

Feldman, M., Walsh, J. H., Wong, H. C., Richardson, C. T. (1978). Role of gastrin heptadecapeptide in the acid secretory response to amino acids in man. *J. Clin. Invest.* **61**, 308–313.

Friis-Hansen, L. (2001). Lessons learned form the gastrin and gastrin receptor knockout mice. *Scand. J. Clin. Lab. Invest.* **234**, 41–46.

Gregory, R. A., and Tracy, H. J. (1961). The preparation and properties of gastrin. *J. Physiol. Lond.* **156**, 523–543.

Kopin, A. S., Lee, Y. M., McBride, E. W., Miller, L. J., Lu, M., *et al.* (1992). Expression cloning and characterization of the canine parietal cell gastrin receptor. *Proc. Natl. Acad. Sci. U.S.A.* **89**.

Marshall, B. J., and Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**(8390), 1311–1315.

McColl, K. E., Gillen, D., and El-Omar, E. (2000). The role of gastrin in ulcer pathogenesis. *Best Pract. Res. Clin. Gastroenterol.* **14**, 13–26.

McGuigan, J. E. (1968). Immunochemical studies with synthetic human gastrin. *Gastroenterology* **54**, 1005–1011.

Rozengurt, E., and Walsh, J. H. (2001). Gastrin, CCK, signaling, and cancer. *Annu. Rev. Physiol.* **63**, 49–76.

Sachs, G., Shin, J. M., Munson, K., Vagin, O., Lambrecht, N., Scott, D. R., and Weeks, D. L. (2000). Review article: The control of gastric acid and *Helicobacter pylori* eradication. **14**, 1383–1401.

Seva, C., Dickinson, C. J., and Yamada, T. (1994). Growth promoting effects of glycine-extended progastrin. *Science* 265, 410–412.

Soll, A. H. (1977). Studies on the actions and interactions of secretagogues on isolated mammalian parietal cells as reflected in changes in oxygen consumption and aminopyrine uptake. *Gastroenterology* **73**(4, Pt. 2), 899.

Walsh, J. H., Yallow, R. S., and Berson, S. A. (1971). The effect of atropine on plasma gastrin response to feeding. *Gastroenterology* **60**, 16–21.

Gastrin-Releasing Peptide

ROBERT T. JENSEN
National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health

bombesin-related peptides Peptides related to the 14-amino-acid peptide, bombesin, which was originally isolated from the skin of the frog, *Bombina bombina*. Many of these peptides (ranatensin, phyllolitorin, litorin) were isolated from the skin of other frogs.

gastrin-releasing peptide A 27-amino-acid mammalian peptide that is closely related to bombesin and was isolated from porcine stomach.

neuromedin B A 10-amino-acid mammalian peptide isolated from porcine spinal cord that resembles the frog peptides, ranatensin and litorin.

Gastrin-releasing peptide (GRP) is a 27-amino-acid peptide that was isolated in 1978 from porcine nonantral gastric tissue. It is named GRP for its ability to stimulate gastrin release, which was used to monitor its activity during purification. GRP and a 10-amino-acid peptide, neuromedin B (NMB), isolated in 1983 from porcine spinal cord, are now known to be the mammalian members of the bombesin family of peptides previously described in invertebrates. Beginning in 1970 when the 14-amino-acid peptide bombesin was isolated from the skin of two European frogs, *Bombina bombina* and *Bombina variegata*, a number of related peptides were discovered and were named after the frog from which they were isolated. GRP closely resembles bombesin, differing in only 1 amino acid in its last 10 carboxyl-terminal residues, which explains their similar biological activities. NMB more closely resembles the 11-amino-acid peptide ranatensin, which was originally isolated from the skin of the North American frog, *Rana pipiens*, in 1970 and the nonapeptide, litorin, isolated from an Australian frog. NMB differs from the last 10 carboxyl-terminal amino acids of GRP (GRP-10) by 3 amino acid substitutions. A third subfamily of bombesin-related peptides exists: the phyllolitorins, which have no mammalian equivalents at present. Phyllolitorins are nonapeptides isolated from a South American frog and differ from GRP and NMB by having a serine instead of a histidine, 3 amino acids from the carboxyl terminus. These small structural differences between GRP and NMB are important because both a GRP-preferring receptor and an NMB-preferring receptor (NMB-R) have been identified in mammals. This article will concentrate on GRP and discuss NMB/NMB-R only briefly.

GASTRIN-RELATED PEPTIDE STRUCTURE AND GENE SEQUENCE

Gastrin-releasing peptide (GRP), like all bombesin (Bn)-related peptides, is amidated at its carboxyl-terminus (Fig. 1), which is the biologically active portion of the peptide. The carboxyl-terminal decapeptide (GRP-10) in most assays has biologic activity equal to that of the native GRP. Structure–function studies of GRP demonstrate that the Trp-21, His-25, and the carboxyl-terminal amidation are particularly important for high-affinity receptor interaction. The sequence for the precursor of human GRP has been cloned. The mRNA is approximately 850 bases in length with a 300-base 3′-untranslated region, a 100-base 5′-untranslated region, a 148-amino-acid precursor containing a signal peptide, a single copy of GRP-27, and a carboxyl-amino extension peptide. The gene is located on chromosome 18 and comprises three exons and two introns, which, because of alternate splicing at the second intron, produces three mRNAs that encode an identical GRP, but with variable carboxyl-terminal extension peptides present. The importance of the alternate splicing is unknown. GRP is produced from the 148-amino-acid precursor by proteolytic cleavage at both the amino- and carboxyl-termini and removal of the carboxyl-terminal Met-Gly extension by an amidation enzyme.

GASTRIN-RELATED PEPTIDE RECEPTORS, PHARMACOLOGY, AND THEIR CELL TRANSDUCTION MECHANISMS

A GRP-preferring receptor (GRP-R; also called the BB2 receptor) has been characterized in human and a number of other species (Table I). The human GRP-R (hGRP-R) is a 384-amino-acid protein with a predicted molecular weight of 43 kDa (Table I). It is a member of a G-protein-coupled receptor superfamily with hydropathy plots predicting the characteristic 7-transmembrane

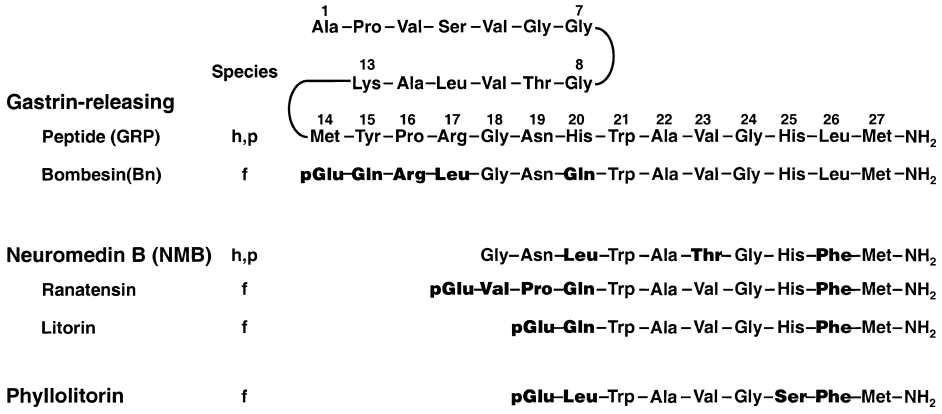


FIGURE 1 Amino acid sequences of gastrin-releasing peptide, neuromedin B, and phyllolitorin family of bombesin-related peptides. Sequence in boldface indicates differences from GRP sequence. Numbers at the top denote the position in GRP starting from the amino-terminus. f, frog; h, human; p, porcine.

domains. The GRP-R shows 55% amino acid identity with the neuromedin B-preferring receptor (NMB-R). The native hGRP-R has two potential glycosylation sites and in cells is 35% glycosylated with a molecular weight of 60 kDa. Studies of the murine GRP-R (mGRP-R), which has four potential glycosylation sites, demonstrate that all four sites are glycosylated and the glycosylation consists of four tri-antenary and/or tetra-antenary complex oligosaccharide chains containing no sialic acid. GRP-R glycosylation is necessary for receptor expression, ligand affinity, and G-protein coupling.

The hGRP-R has a higher affinity for GRP than for NMB, whereas the converse is true for the NMB-R (Table I). Recent studies demonstrate that the GRP-R's selectivity for GRP is primarily determined by amino acid differences between the third extracellular domains of GRP-R and NMB-R. Of the 20 amino acid differences between the GRP-R and NMB-R in the third extracellular domain, 2 differences are particularly important: the substitution in NMB-R of isoleucine for phenylalanine at position 185 of the GRP-R and, to a lesser extent, the substitution of isoleucine for alanine at position 198 of the GRP-R. Mutagenesis studies show the presence of an aromatic ring on the amino acid at GRP-R position 185 (mGRP-R) and, to a lesser extent, the size of the backbone substitution on the amino acid at position 198 are likely the important factors contributing to the different GRP affinities for these two receptors. It has been proposed that cation$-\pi$ or $\pi-\pi$ ligand receptor interactions may be important for mediating the marked effect of the aromatic amino acid at position 185 of GRP-R.

Five different classes of receptor antagonists have been described, two of which have a high degree of selectivity for the GRP-R over the NMB-R. These five classes include substituted substance P analogues, [D-Phe-12]Bn analogues, 13- to 14-amino-acid pseudopeptide analogues of GRP or bombesin (the CONH of Leu-13 is replaced by CH_2NH), des-Met-14 analogues of bombesin (or des-Met-27 GRP), and peptoids such as PD168368, which are nonpeptide analogues based on a single amino acid template. Some of the pseudopeptide analogues and des-Met-14 Bn analogues (Table I) have a greater than 1000-fold selectivity for the GRP-R over the NMB-R. These antagonists have proved useful for exploring the role of GRP-R in both physiological processes and pathological processes. For example, in human volunteers using the GRP-R antagonist, BIM26226

TABLE I Comparison of Characteristics of the Two Mammalian Bombesin (Bn) Receptors (GRP-R and NMB-R)

	hGRP-R (BB_2)	hNMB-R (BB_1)
Natural ligand	GRP	NMB
Gene location	Xp22	6q21
Amino acids	384	390
Pharmacology (affinity)		
Agonists		
GRP	0.3 n*M*	450 n*M*
NMB	290 n*M*	7 n*M*
Bn	0.5 n*M*	44 n*M*
Antagonists		
[D-Phe^6]Bn(6-14)methyl ester	3 n*M*	>10,000 n*M*
PD 168368	630 n*M*	15 n*M*

([D-F5-Phe-6, D-Ala-11]Bn(6–13)methyl ester), GRP was shown to be a physiological regulator of gastric acid secretion; however, gastrin release was not under the control of GRP in human. The role of these antagonists in the possible treatment of various disorders, especially cancer growth, will be discussed in a later section.

GRP primarily couples to $G_{q/11}$ and its activation results in the stimulation of phospholipase C with the generation of inositol phosphates, mobilization of cellular calcium, and release of diacylglycerol with activation of protein kinase C (PKC). GRP-R activation also results in the activation of phospholipase D, protein kinase D, mitogen-activated protein kinases, Src kinases, and stimulation of tyrosine phosphorylation of a number of intracellular proteins including p125 focal adhesion kinase, p130Cas, and paxillin. Studies demonstrate that bombesin-induced tyrosine phosphorylation of these proteins occurs through a PKC- and Ca^{2+}-independent pathway and this pathway depends on the integrity of the actin cytoskeleton and participation of the small GTP-binding protein, Rho. GRP-R activation results in increased expression of c-fos, c-jun, and jun-B with increased activator protein-1 expression and DNA synthesis. In addition to G-proteins, studies support the conclusion that a number of low-molecular-weight GTP-binding proteins play an important role in GRP-R-induced stimulation of intracellular pathways including Rho, p21ras, and Rac.

DISTRIBUTION OF GASTRIN-RELATED PEPTIDE AND ITS RECEPTOR

GRP immunoreactivity (GRP-IR) exists widely in the alimentary tract in nerve fibers and cell bodies. In the stomach, GRP-IR is found in cell bodies and nerve fibers of both the oxyntic and antral mucosa and the circular muscle layer. Similarly, in both the small and large intestine, GRP-IR is found in nerve fibers and cell bodies of the myenteric plexus. In the colon, GRP-IR is found in myenteric neurons projecting to both the circular muscle and caudad. Both nerve fibers and cell bodies in the pancreas contain GRP-IR and the fibers project to both the acini and ducts, with a spare projection to islets. Sympathetic prevertebral ganglia have a dense innervation of fibers containing GRP-IR.

GRP-IR is also widespread in the central nervous system (CNS). It is present in a subset of spinal sensory ganglia, in layers 1 and 11 of the posterior horn, and in nerve terminals surrounded by large motor neurons in the anterior horn. Cell bodies in the brain containing GRP-IR are found in the pons (dorsolateral tegmental nucleus), hypothalamus (paraventricular nucleus), nucleus of the solitary tract, dorsal parabrachial nucleus, superior olivary nucleus, mesencephalic central gray matter, and anterior aspect of the interpeduncular nucleus. Nerve fibers in the brain containing GRP-IR in high density are found in the thalamic periventricular nucleus, hypothalamic nuclei, amygdala, dorsolateral tegmental nucleus, trigeminal complex, dorsal vagal nucleus, and nucleus of the tractus solitarius.

GRP-R receptors and (or) GRP-R mRNA have been shown to be widely distributed in the CNS, the gastrointestinal (GI) tract, and the urogenital tract. In the CNS, high densities of GRP-R mRNA are found in isocortex (layer 11), dentate gyrus, N-lateral olfactory tract, magnocellular preoptic nucleus, numerous hypothalamic nuclei (suprachiasmatic, supraoptic, paraventricular, medial preoptic, lateral mammillary), and the nucleus ambiguus. In the urogenital system, GRP-R has been identified in the bladder, seminal vesicles, uterus, and oviduct. In the GI tract, the highest density of GRP-Rs occurs over gastric antral endocrine cells. GRP-Rs also occurred on circular smooth muscle cells of the gastric fundus, gastric antrum, ileum, and the longitudinal smooth muscle of the gastric fundus and gastric antrum as well as on neural elements in the myenteric plexus in the stomach, antrum, and small intestine. In the colon, GRP-Rs are found in submucous plexi and the circular smooth muscle.

GASTRIN-RELATED PEPTIDE RELEASE AND METABOLISM

GRP is a neuropeptide that is primarily localized in neural structures; therefore, its plasma levels may not be an accurate reflection of its local release in tissues. Splanchnic nerve stimulation in the calf, but not the pig, results in the release of GRP-IR into plasma and interstitial lymph. GRP-IR is also released by vagal stimulation of the isolated perfused pig stomach and pancreas. Electrical stimulation of gallbladder nerves in the pig results in the release of GRP-IR and contraction of the gallbladder, which is inhibited by tetrodotoxin.

GRP-IR is rapidly removed from the plasma with a half-life of 1.4 min in the pig. The clearance of fragments of GRP, such as GRP-10, is faster than that of GRP-27. Neutral endopeptidase [NEP, EC3–4-24–11 (also called endopeptidase-24–11, enkephalinase)] is the major peptidase responsible for GRP degradation. GRP-10 is degraded by hydrolysis of the His8Leu9 peptide bond, which results in its inactivation. NEP is

expressed on the cell surface of many cells that are regulated by GRP including CNS neurons, pancreatic acinar cells, gastrointestinal smooth muscle cells, and gastric mucosal cells.

PHYSIOLOGICAL/PHARMACOLOGICAL ROLES OF GASTRIN-RELATED PEPTIDE

A brief list of the effects of GRP is given in Table II. Many of the reported effects are limited by two factors. First, which effects are physiological and which are pharmacological has, in general, not been addressed because only recently have potent specific antagonists been described or the use of immunoneutralizing antibodies applied. Second, many of the older studies were performed with bombesin, which has only limited selectivity for GRP-R over NMB-R (see Table I) and therefore, in some cases, activation of NMB-Rs could be contributing to the described responses. Two recent studies in human volunteers using a highly selective GRP-R antagonist demonstrate that GRP is a physiological regulator involved in meal-stimulated acid secretion and gallbladder contraction as well as gastric emptying and decreasing small bowel transit time. However, GRP was not a physiological regulator of gastrin release in humans as suggested by extensive studies in animals. GRP functions as a neurotransmitter and neuromodulator.

GRP is a potent contractant of almost all gastrointestinal and urogenital smooth muscle (Table II) and is also a potent stimulant of the release of many hormones and neurotransmitters such as gastric inhibitory peptide, cholecystokinin (CCK), pancreatic polypeptide, glucagon, insulin, prolactin, and growth hormone (Table II). GRP has a potent stimulatory effect on pancreatic secretion in all species examined. GRP can stimulate CCK release; however, GRP's pancreatic stimulatory effect in rats is not blocked by a CCK-A receptor antagonist and in almost all species examined including humans, the pancreatic acini possess GRP receptors, leading to the conclusion that GRP's pancreatic stimulatory action is due to a direct interaction with pancreatic acinar cells.

GRP has numerous effects in the CNS and the most well-studied is its ability to stimulate a decrease in body temperature. It also is involved in the regulation of circadian rhythm and in the regulation of feeding. Extensive studies in both animals and humans provide strong support for GRP as a satiety factor. Furthermore, GRP-R knockout mice not only no longer respond to the satiety effects of GRP, these mice eat significantly more at each meal than normal littermates and show a significant elevation in body weight.

TABLE II Physiological and Pharmacological Effects of GRP

- CNS effects
  - Regulation of circadian rhythm
  - Thermoregulation (↓ body temperature)
  - Food intake (↓)
  - Behavioral effects (↑ grooming)
  - Sleep (↓); memory (↑)
  - CNS–gut effects (gastric antisecretory effect, stimulates gastrin and mucus release, inhibits gastric emptying and motility)
- Gastrointestinal tract
  - Gastric
    - ↑ Acid secretion
    - ↑ Gastrin somatostatin, GIP release
    - Alter gastric antral (↑), corpus (↑, ↓) motility
    - ↑ Gastric emptying
  - Intestinal
    - ↑ CCK, motilin, neurotensin, VIP, enteroglucagon release
    - ↓ Intestinal transit time
    - ↑ Contraction isolated muscle
  - Colon
    - ↑ Contraction of isolated muscle
    - ↑ Na^+, K^+ secretion in distal colonic epithelium
  - Pancreatic
    - ↑ Enzyme secretion
    - ↑ PP, insulin, glucagon release
    - ↑ Pancreatic growth
  - Gallbladder
    - ↑ Contraction of GB/isolated muscle
  - Liver
    - ↑ Fluid and bicarbonate secretion from cholangiocytes
- Urogenital tract
  - ↑ Contraction of urinary bladder
  - ↑ Contraction isolated muscle of uterus, urinary bladder
- Endocrine
  - Adrenal
    - ↑ Corticosterone release
    - ↑ ACTH release
  - Pituitary
    - ↑ Prolactin, growth hormone, LH release
    - ↓ TSH release
  - Other endocrine
    - ↑ Calcitonin, renin release
- Respiratory
  - ↑ Growth of bronchial epithelial cells
  - Involved in lung development (↑ branching morphogenesis, cell proliferation, differentiation)
- CVS effects
  - Blood pressure (↑)

Note. ACTH, adrenocorticotropic hormone; CCK, cholecystokinin; CNS, central nervous system; CUS, cardiovascular system; GIP, gastric inhibitory peptide; LH, luteinizing hormone; PP, pancreatic polypeptide; TSH, thyroid-stimulating hormone; VIP, vasoactive intestinal peptide.

One of the most important effects of GRP-R activation is the stimulation of growth of normal cells and in pathologic conditions. The latter will be discussed fully in the next section. GRP has been shown to have a growth stimulatory effect on normal pancreatic acinar cells (increase in pancreatic weight, DNA content), stomach mucosa, colonic mucosa, bronchial endothelial cells, growth plate chondrocytes, keratinocytes, and melanocytes. Results from detailed studies of Swiss 3T3 fibroblasts and other tissues show that the growth stimulatory effect is due to a direct activation of the GRP-R by GRP and not an indirect effect due to the release of another stimulatory agent.

GASTRIN-RELATED PEPTIDE IN PATHOLOGIC CONDITIONS

The most important pathologic conditions in which GRP may be involved are listed in Table III. There is considerable experimental data supporting its possible role in cancer growth and lung diseases, whereas a single study suggests that GRP release may play a decisive role in the maintenance of intestinal microvascular integrity after ischemia and with reperfusion. No studies have established that alterations of GRP are clearly involved in human feeding, obesity, or body weight disorders, but animal studies as well as studies on humans have firmly established GRP as an important satiety factor. It has been proposed that GRP may play a role in various ingestive disruptions associated with anorexia (anorexia nervosa, acquired immunodeficiency syndrome, cancer anorexia), bulimia, obesity, and depression. This proposal has received some support recently from a study of cerebrospinal fluid (CSF) GRP concentrations in normal females and patients recovered from anorexia nervosa or bulimia nervosa. CSF GRP-IR levels were significantly lower in recovered bulimia patients than either normal controls or recovered anorexia nervosa patients. It was proposed that this persistent GRP abnormality in recovered bulimia nervosa patients could be contributing to their episodic hyperphagia.

TABLE III Pathological Conditions in Which GRP May Be Involved

Cancer autocrine growth factor
Small cell and non-small-cell lung cancer
Prostate cancer
Neuroblastomas
Squamous cell carcinoma of head and neck
Renal cell carcinomas
Breast cancer
Gastric adenocarcinomas
Pancreatic adenocarcinomas
Astrocyte–glial brain tumors
Colon cancer—? Growth factor or morphogen
Feeding/body weight abnormalities
Lung diseases
Hyperoxic lung injury
Bronchopulmonary dysplasia
Tobacco-related injury
Intestinal microcirculation after ischemia–reperfusion

In 1985, GRP-related peptides were shown to be an autocrine growth factor in human small cell lung cancer. Not only did small cell lung cancer secrete GRP-related peptides, the tumors possess GRP-R and its activation stimulated the growth of the tumor cells, which could be blocked by a neutralizing antibody to bombesin or GRP's carboxyl-terminus, the biologically active portion. Subsequent studies demonstrated that a large number of different tumors possess GRP-R or GRP/bombesin and these could be a growth stimulant for prostate cancer, neuroblastomas, astrocytic–glial brain tumors, squamous cell carcinoma of the head and neck, some renal cell carcinomas, breast cancers, gastric adenocarcinomas, and some pancreatic adenocarcinomas (Table III). In one recent study of 161 human tumors, 100% of the prostate cancers ($n = 12$), 100% of gastrinomas ($n = 5$), 72% of breast cancers ($n = 57$), 38% of renal cell cancers ($n = 16$), 33% of small cell lung cancers ($n = 9$), and 0% of intestinal carcinoid tumors ($n = 24$) expressed GRP-R that could be detected by autoradiographic studies. *In vitro* GRP-R antagonists or GRP neutralizing antibodies inhibit the growth of many of these tumors and have raised the possibility that such treatment might be effective in patients with these malignancies; however, few studies have been carried out to date. In one study, 13 patients with small cell lung cancer were treated with a monoclonal antibody (2All) that neutralizes GRP's action over 1 h, three times a week for 4 weeks. One of 12 (8%) evaluable patients had a complete resolution of radiographically detectable tumor lasting 4 months and 4/12 (33%) had stable disease. Whether this approach, which requires large amounts of GRP antibody or the use of GRP receptor antagonists, will be a useful therapeutic approach for these tumors is at present unclear.

The exact role of GRP-R in tumor growth has recently been raised from studies primarily of human colonic adenocarcinomas. Colon cancers ectopically express GRP-Rs and GRP has been shown in some studies to stimulate their growth. In one study, 84% of human colon cancers aberrantly expressed GRP or GRP-R with 62% expressing both the ligand and receptor, whereas expression was not observed in adjacent normal mucosa. However, GRP/GRP-R co-expression was seen equally frequently in stage A or D cancers and was rarely

detected in metastases. Furthermore, no difference in survival occurred in patients expressing or not expressing cancers with GRP/GRP-R. In fact, GRP/GRP-R was found exclusively in well-differentiated regions of the colon cancers. From these observations, it has been proposed that GRP is unlikely to be a clinically significant growth factor in human colon cancer but instead functions as a differentiation factor or morphogen. From these studies and a review of GRP/GRP-R expression in a variety of tumors, it has been proposed that GRP and GRP-R are likely onco-fetal antigens whose reexpression in these tumors reflects a recapitulation of their normal role in regulatory organogenesis.

Because many tumors overexpress GRP-R, the possibility of imaging tumors using radiolabeled GRP analogues or using cytotoxic GRP conjugates to target an antitumor treatment is receiving considerable attention. Cytotoxic GRP analogues such as AN-215, which consists of a GRP analogue (20−27) linked to 2 pyrrolino-doxorubicin-14−0-hemiglutarate, have been shown *in vitro* to inhibit growth of pancreatic cancers, lung cancers, prostate cancers, and gastric cancers. In a study of pancreatic cancers in golden hamsters, the cytotoxic GRP analogue AN-215 had significant antitumor activity and lower toxicity than the unconjugated cytotoxic radical. Recent studies have also reported the development of 99m techneticum (Tc)-labeled GRP analogues that can be used to image GRP-R-overexpressing tumors. In two phase 1 studies, 99m Tc-GRP analogues are reported to image small cell lung cancer, prostate cancer, and breast cancer in patients. In addition, biologically active GRP analogues have recently been reported with an attached chelator group that can bind various radionuclides with high affinity. This strategy to develop analogues that can image tumors overexpressing receptors as well as deliver receptor-targeted radionuclide therapy has been extensively investigated with analogues of somatostatin for the treatment of human neuroendocrine tumors. Using a similar approach, diethylenetriaminepentaacetic acid GRP conjugates that efficiently couple ^{111}In have been reported. ^{111}In is a γ and auger electron emitter and can be used for imaging tumors as well as for cytotoxicity. Also, dodecanetretraacetic acid GRP analogues that couple more efficiently β-emitters such as ^{90}Y and may be more effective for GRP-targeted radiotherapy have been synthesized. At present, the usefulness of these GRP analogues for imaging tumors or for delivering receptor−target cytotoxicity has not been established.

GRP is the major endogenous peptide implicated in promoting growth and maturation during fetal lung development in humans, rats, and mice. Not only does administration of GRP-related peptides in fetal mice increase cell proliferation and differentiation, administration of GRP-R neutralizing antibodies inhibits lung differentiation. During lung development, GRP is found in pulmonary neuroendocrine cells (PNECs). Numerous studies have demonstrated that lung injury in adult animals, whether it is caused by smoke inhalation, asbestos, or other causes, leads to hyperplasia of PNEC with resultant increases in the levels of GRP-related peptides. Using a baboon model of bronchopulmonary dysplasia (BPD), GRP-related peptides were shown to mediate the lung injury. In this postnatal model, GRP-like peptides acted as proinflammatory mediators of BPD and the use of GRP neutralizing antibodies markedly decreased the severity of the chronic lung disease. It has been proposed that one of the important differences between smokers who develop lung disease and those who do not develop lung disease may be the response of their PNECs in releasing GRP-like peptides and other growth factors when exposed to cigarette smoke.

See Also the Following Articles

Cholecystokinin (CCK) • Colorectal Adenocarcinoma • Enteroglucagon • Gastric Acid Secretion • Gastrin • Glucose-Dependent Insulinotropic Polypeptide (GIP) • Growth Hormone • Motilin • Pancreatic Enzyme Secretion (Physiology) • Pancreatic Polypeptide Family • Somatostatin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Aguayo, S. M. (1993). Pulmonary neuroendocrine cells in tobacco-related lung disorders. *Anat. Rec.* **236**, 122−127.

Battey, J., and Wada, E. (1991). Two distinct receptors for mammalian bombesin-like peptides. *Trends Neurosci.* **14**, 524−528.

Benya, R. V., Kusui, T., Pradhan, T. K., Battey, J. F., and Jensen, R. T. (1995). Expression and characterization of cloned human bombesin receptors. *Mol. Pharmacol.* **47**, 10−20.

Breeman, W. A., de Jong, M., Erion, J. L., Bugaj, J. E., Srinivasan, A., Bernard, B. F., Kwekkeboom, D. J., Visser, T. J., and Krenning, E. P. (2002). Preclinical comparison of (111)In-labeled DTPA- or DOTA-bombesin analogs for receptor-targeted scintigraphy and radionuclide therapy. *J. Nucl. Med.* **43**, 1650−1656.

Bunnett, N. (1994). Gastrin-releasing peptide. *In* "Gut Peptides" (J. H. Walsh and G. J. Dockray, eds.), pp. 423−445. Raven Press, New York.

Cullen, A., Emanuel, R. L., Torday, J. S., Asokananthan, N., Sikorski, K. A., and Sunday, M. E. (2000). Bombesin-like peptide and receptors in lung injury models: Diverse gene expression, similar function. *Peptides* **21**, 1627−1638.

Hildebrand, P., Lehmann, F. S., Ketterer, S., Christ, A. D., Stingelin, T., Beltinger, J., Gibbons, A. H., Coy, D. H., Calam, J., Larsen, F., and Beglinger, C. (2001). Regulation of gastric function by endogenous gastrin releasing peptide in humans: Studies with a

specific gastrin releasing peptide receptor antagonist. *Gut* **49**, 23–28.
Jensen, J. A., Carroll, R. E., and Benya, R. V. (2001). The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. *Peptides* **22**, 689–699.
Jensen, R. T., and Coy, D. H. (1991). Progress in the development of potent bombesin receptor antagonists. *Trends Pharmacol. Sci.* **12**, 13–19.
Kelley, M. J., Linnoila, R. I., Avis, I. L., Georgiadis, M. S., Cuttitta, F., Mulshine, J. L., and Johnson, B. E. (1997). Antitumor activity of a monoclonal antibody directed against gastrin-releasing peptide in patients with small cell lung cancer. *Chest* **112**, 256–261.
Kroog, G. S., Jensen, R. T., and Battey, J. F. (1995). Mammalian bombesin receptors. *Med. Res. Rev.* **15**, 389–417.
Merali, Z., McIntosh, J., and Anisman, H. (1999). Role of bombesin-related peptides in the control of food intake. *Neuropeptides* **33**, 376–386.
Preston, S. R., Miller, G. V., and Primrose, J. N. (1996). Bombesin-like peptides and cancer. *Crit. Rev. Oncol. Hematol.* **23**, 225–238.
Rozengurt, E. (1998). Signal transduction pathways in the mitogenic response to G protein-coupled neuropeptide receptor agonists. *J. Cell. Physiol.* **177**, 507–517.
Schally, A. V., and Nagy, A. (1999). Cancer chemotherapy based on targeting of cytotoxic peptide conjugates to their receptors on tumors. *Eur. J. Endocrinol.* **141**, 1–14.
Tokita, K., Hocart, S. J., Coy, D. H., and Jensen, R. T. (2002). Molecular basis of the selectivity of the gastrin-releasing peptide receptor for gastrin-releasing peptide. *Mol. Pharmacol.* **61**, 1435–1443.

Gastritis

MARK FELDMAN
University of Texas Southwestern Medical School at Dallas and Presbyterian Hospital of Dallas

achlorhydria The inability of the stomach to produce any hydrochloric acid.
gastritis A disorder of the stomach characterized by acute and/or chronic inflammation of the mucosal lining.

The term gastritis refers to a group of diseases that have in common the symptom of inflammation of the stomach. In most cases, the inflammation exclusively or predominately involves the gastric mucosal lining. By definition, a gastric biopsy is required to diagnose gastritis.

CLASSIFICATION AND ETIOLOGY

There is only a modest correlation between gastritis and symptoms such as pain and nausea. In fact, many patients with gastritis have no symptoms referable to the stomach. A classification of gastritis is presented in Table I.

By far, the most common category worldwide is infectious gastritis. *Helicobacter pylori* infection accounts for the majority of cases of gastritis. Cytomegalovirus (CMV) gastritis is most typically seen in immunocompromised patients, such as patients with the acquired immunodeficiency syndrome, patients with organ transplants receiving immunosuppressive drugs, and patients with hematologic malignancies, many of whom receive chemotherapy agents.

ASSOCIATIONS WITH OTHER DISEASES

Certain types of chronic gastritis are associated with an increased risk of cancer. For example, *H. pylori* gastritis is a risk factor for both gastric adenocarcinoma and B-cell gastric mucosa-associated lymphoid tissue lymphoma. Lymphocytic gastritis, which can be associated with *H. pylori* infection or with celiac sprue, may be a precursor of gastric lymphoma. Autoimmune atrophic gastritis is associated with an increased risk of gastric adenocarcinoma, possibly as a consequence of gastric achlorhydria and bacterial overgrowth of the stomach. Patients with autoimmune atrophic gastritis also frequently have reduced intrinsic factor secretion and may develop cobalamin (vitamin B12) deficiency, with hematological sequelae (pernicious anemia) and neurological consequences (optic neuropathy, dementia, subacute combined degeneration of the spinal cord).

Copyright 2004, Elsevier (USA). All rights reserved.

Gastrinoma

PAUL N. MATON
Digestive Disease Specialists Incorporated, Oklahoma City

enterochromaffin-like cells Gastric fundic mucosa cells that release histamine in response to gastrin and other stimuli.

multiple endocrine neoplasia type 1 Autosomal dominant condition that is characterized by the development of hyperparathyroidism and pancreatic and pituitary endocrine tumors.

neuroendocrine tumors Growths that may occur in many tissues throughout the body; produce a variety of peptides and other agents, some of which are biologically active and produce distinctive hormonal syndromes.

octreotide Synthetic peptide that is related to the naturally occurring peptide somatostatin; has therapeutic uses and can be radiolabeled for use as a diagnostic agent.

secretin Naturally occurring peptide that is found in the duodenum; useful as a diagnostic test in gastrinoma.

Zollinger–Ellison syndrome Clinical syndrome defined by two physicians in 1995; described symptoms of two patients with severe ulcer disease, massive gastric acid hypersecretion, and islet cell tumors. This term has become synonymous with the clinical features of all gastrinomas.

Gastrinomas are tumors that produce and secrete a peptide, gastrin. Gastrinomas are a type of neuroendocrine tumor and usually arise near the pancreas. They may be benign or malignant; tumor growth and spread my cause direct effects, but the clinical effects are primarily due to the release of gastrin, which stimulates the production of large amounts of acid by the stomach. The definition of gastrinomas is restricted here to tumors that release gastrin into the circulation. Other neuroendocrine tumors may contain immunoreactive gastrin but release no biologically active peptide.

TUMOR BIOLOGY

Secretion of Gastrin

Gastrin is released into the circulation in normal individuals. Physiologically, gastrin is a mixture of two peptides, one of 17 amino acid residues and the other of 34 amino acid residues, with similar actions. Gastrin is released from the gastrin-producing cells (G cells) of the gastric antral and duodenal mucosa by various physiological stimuli, mainly the ingestion of food. In normal individuals, gastrin is carried in the circulation to the enterochromaffin-like (ECL) cells in the gastric body. Stimulated by gastrin, ECL cells release histamine, which acts on the parietal cells (the acid-producing cells) to stimulate acid secretion. The acid released into the gastric lumen subsequently comes into contact with the G cells in the gastric glands of the antrum and switches off gastrin release. There is thus a negative feed back control on gastrin release such that gastrin release is phasic and limited.

Gastrinomas are not subject to physiological control and thus patients with gastrinomas suffer the effects of continuously high circulating concentrations of gastrin. Gastrins secreted by gastrinomas are more heterogeneous compared to the physiological forms. Nevertheless, those that contain the C-terminal tetrapeptide sequence can interact with the gastrin receptor and are fully biologically active. These forms are also measured in the generally used gastrin radioimmunoassays. The main short-term effect of gastrin is to stimulate gastric acid secretion, and patients with gastrinomas have markedly increased acid output. Normal basal acid outputs are up to 4 mEq/hour in women and up to 7 mEq/hour in men. Patients with gastrinomas typically secrete more than 15 mEq/hour and often much more. This increased acid secretion in these patients initiates many of the clinical manifestations that result in the Zollinger–Ellison syndrome. This syndrome was described in 1955 as a triad of severe peptic ulcer disease, massive acid hypersecretion, and an islet cell tumor of the pancreas. Although it is now known that none of these is necessarily part of the syndrome of gastrinomas, the term has been retained.

Gastric acid hypersecretion can result in ulcer disease in the absence of gastric infection with *Helicobacter pylori* or ingestion of nonsteroidal antiinflammatory drugs. The ulcer disease may be extremely severe. Initial descriptions included multiple ulcers in unusual sites such as the distal duodenum, jejunum, and esophagus. Furthermore, bleeding and perforation are frequent. Indeed, in the years between the first description of

 Copyright 2004, Elsevier (USA). All rights reserved.

the syndrome and the advent of histamine-2 (H2) receptor antagonists, the complications of ulcer disease were responsible for 60–80% of the deaths due to gastrinomas. Probably due in part to the availability of anti-secretory drugs, ulcer disease even at presentation is typically less severe now, and patients are likely to have a single typical bulbar ulcer.

Increased gastric acid secretion has other effects within the gastrointestinal tract. The increase in acid secretion, accompanied by a corresponding increase in fluid secretion, may result in a secretory diarrhea. Indeed, some 25% of patients with gastrinomas present with diarrhea and never have any ulcer disease. The large amounts of acid may overcome the buffering capacity of the upper intestine and may inactivate pancreatic enzymes, resulting in malabsorption. Likewise, bile acids (particularly the glycine conjugates) may be precipitated, contributing to the steatorrhea. Gastric acid may also directly damage the intestinal mucosa and interfere with absorption. Patients have been described with vitamin B_{12} malabsorption, suggesting that even the terminal ileum may be damaged in some cases.

Gastrin has other long-term effects that are responsible for some of the clinical manifestations of Zollinger–Ellison syndrome. Gastrin is trophic to the gastric mucosa. There is an increase in the number of parietal cells, resulting in an increase in the maximal acid output, up to three or more times the upper limit of normal (20–50 mEq/hour). Gastrin also has marked trophic effects on the gastric ECL cells, causing ECL cell hypertrophy. This can progress to the development of ECL cell carcinoid tumors in some cases.

Although gastrin does not cause increase in numbers of cells in other gastrointestinal or extraintestinal tissues, it can cause increased cell turnover as measured by increased DNA synthesis. The mucosa of the colon is influenced in this way. This has led to the investigation of hypergastrinemia as an increased the risk for colon polyps and colon cancer. The studies suggest that there is no large effect, but a modest effect cannot be excluded.

Cell of Origin

Although gastrinomas were originally classed as islet cell tumors, normal G cells occur in the human pancreas only in embryonic life. A majority of the G cells occur in the gastric antrum and duodenum. In the antrum, they are found in the deeper parts of the gastric glands. In the duodenum, they occur in the crypts and Brunner's glands. If G cells are in sites where acid cannot reach the membrane of the cell, then those cells are not subject to physiological control. Thus, a nest of normal G cells protected from acid by an embryological accident or surgery could in theory give rise to the Zollinger–Ellison syndrome. However, studies suggest that at least in sporadic Zollinger–Ellison syndrome there is always a distinct gastrin-producing tumor.

For many years it was thought that most gastrinomas arose within the pancreas. However, it has become clear that in fact only a minority are pancreatic. This knowledge has come about in part through increasingly detailed prospective imaging studies, but mainly through the extremely detailed exploratory surgery performed by a small number of surgeons. Such studies have shown that patients may be cured of their disease by resection of extrapancreatic gastrinomas.

It had been known for many years that Zollinger–Ellison syndrome (and thus gastrinomas) occurred in two circumstances. In the majority of patients, Zollinger–Ellison syndrome occurs as an isolated disease. These patients have no family history and no other endocrine diseases. However, about 20–30% of patients with Zollinger–Ellison syndrome come from families in which a variety of endocrine diseases occur in each generation. These families have multiple endocrine neoplasia type 1 (MEN 1), an autosomal dominant disease in which >90% of affected individuals develop hyperparathyroidism, 66% develop Zollinger–Ellison syndrome, and smaller percentages develop other pancreatic endocrine tumors and pituitary tumors. Studies around 1990 demonstrated differences in site of origin in gastrinomas in patients with MEN 1 compared to gastrinomas in sporadic Zollinger–Ellison syndrome, and it was shown that these differences have implications for management.

In patients with sporadic gastrinomas, the tumors are nearly always solitary. About 40% occur in the mucosa of the duodenum, another 40% arise in the pancreas, 10–20% arise in peripancreatic lymph nodes, and 5% arise in other sites. Primary gastrinomas have been described in the liver, the bile duct, the jejunum, the kidney, the ovary, the heart, and the lung. Despite the fact that most G cells are in the stomach, gastric gastrinomas do not seem to occur.

In patients with Zollinger–Ellison syndrome and MEN 1, about 50% of gastrinomas are found in the mucosa of the duodenum. However, in contrast to sporadic duodenal gastrinomas, those in patients with MEN 1 are nearly always multiple and are distributed throughout the duodenum. In about 20% of cases the primary tumor is in peripancreatic lymph nodes, and in only 10% are primary tumors found in the pancreas. In 20% of patients with MEN 1 and Zollinger–Ellison syndrome, no gastrinoma can be found.

DNA Abnormalities in Gastrinomas

In recent years, much has been learned about abnormalities in DNA and the possible genesis of gastrinomas. Many of these insights have arisen from studies of patients with MEN 1. The genetic defect in MEN 1 is the loss of function of a single tumor suppressor gene on chromosome 11 at the 11q13 locus. In affected individuals, all cells are heterogeneous for this gene, with one functional copy and one defective copy. Different defects occur in different families.

In patients with MEN 1, analysis of the DNA of gastrinomas has demonstrated that in some, but not all, tumors, there is a loss of heterozygosity at the 11q13 gene locus, indicating that the sole functioning copy of the tumor suppressor gene has been rendered nonfunctional. In MEN 1 gastrinomas, mutations are found throughout the nine exons of the gene. Furthermore, not only do mutations differ in different individuals, but they also differ in different gastrinomas from the same individual.

Similar but different findings have been noted in sporadic gastrinomas. Deletions of the 11q13 locus are found in some but not all sporadic gastrinomas. Mutations include truncating, missense, and in-frame types. Unlike MEN 1 gastrinomas, most mutations in sporadic tumors occur between amino acid residues 66 and 163. In pancreatic and nodal gastrinomas, the mutations are commonly in exon 2, but mutations at this site are rare in duodenal tumors. In patients with sporadic tumors at more than one site, all the tumors have the same mutations, indicating that unlike MEN 1, the tumors in any one individual are monoclonal.

However, factors other than defects in the MEN 1 gene are important in gastrinoma development. In both sporadic and MEN 1 gastrinomas, only about 40% have abnormalities of the MEN 1 gene. Other defects have been found on chromosome 1, including aneuploidy and microsatellite instability, and in a study of various types of pancreatic endocrine tumors, defects were found on chromosomes 3, 11, 16, and 22. As yet no clinical correlations with allelic loss have been identified in gastrinomas or other pancreatic endocrine tumors.

Pathology

Gastrinomas may behave in a benign fashion or may be malignant and metastasize. In large series, approximately 30–50% metastasize. In some cases, the patient has had Zollinger–Ellison syndrome for many years and at surgery a single small tumor is found and resection results in cure. In other patients, there is a short history of <2 years and at presentation patients have disease metastatic to the liver. Initial metastases are to the lymph nodes, with subsequent spread to the liver and bone.

Gastrinomas that occur as part of MEN 1 appear to be malignant less often, compared to sporadic gastrinomas. However, malignant gastrinomas do occur in patients with MEN 1. In a recent study, 57 patients were followed for a mean period of 8 years. Hepatic metastases developed in 24%, with aggressive tumor growth in 13%. In general, liver metastases in patients with MEN 1 are associated with a long history of Zollinger–Ellison syndrome.

Histology

Histology of gastrinomas is typical for neuroendocrine tumors. Uniform cuboidal cells can be arranged in a solid nest of cells or in focal glandular, trabecular, or gyriform patterns. Histological features do not predict tumor behavior. Likewise, ultrastructural findings have not proved helpful. Gastrinomas usually, but not always, stain for gastrin. They generally also stain for neuron-specific enolase and for chromogranin A. In 20–50%, staining may also be positive for pancreatic polypeptide, glucagon, somatostatin, or ACTH.

Secretion of Other Products

At least 50% of sporadic gastrinomas contain other peptides as detected by immunoreactive staining. However, release of other peptides is less common and second endocrine syndromes are rare. The most frequently identified second peptides are chromogranins. The α and β chains of FSH are also released by some gastrinomas, as are motilin and pancreatic polypeptide. None of these peptides produces a second endocrine syndrome. About 4% of all sporadic gastrinomas and 17% of those with hepatic metastases secrete adrenocorticotropic hormone (ACTH), leading to the development of ectopic Cushing's syndrome. This is associated with a poor prognosis. Other peptides, including insulin and vasoactive intestinal peptide (VIP), are secreted much more rarely.

Gastrinomas may be only one type of endocrine tumor among many in a patient with MEN 1. It may thus be impossible in a patient with MEN 1 to be certain if the gastrinoma is secreting multiple peptides. These patients may have pancreatic tumors that secrete insulin, glucagon, pancreatic polypeptide, growth hormone-releasing hormone, or somatostatin. Furthermore, patients with MEN 1 may have circulating parathyroid hormone, or growth hormone, prolactin, follicle-stimulating hormone/luteinizing hormone (FSH/LH), or ACTH released from a pituitary tumor. Indeed,

Cushing's syndrome in a patient with Zollinger–Ellison syndrome and MEN 1 is more likely to be due to pituitary Cushing's than to ectopic ACTH production.

DIAGNOSIS OF ZOLLINGER–ELLISON SYNDROME

There are several steps in the diagnosis of a patient with Zollinger–Ellison syndrome. The most important step is to recognize that the patient may have the syndrome. Second, the diagnosis must be confirmed. Third, it is important to determine whether the patient has a sporadic gastrinoma or whether they have a gastrinoma as part of MEN 1. Finally, the site and extent of the gastrinoma need to be defined.

Clinical Features Suggesting Gastric Acid Hypersecretion and Zollinger–Ellison Syndrome

The features that suggest gastric acid hypersecretion are shown in Table I. Hypersecretion of acid gives rise to all the symptoms and clinical features in a patient with a gastrinoma unless the tumor is advanced. A duodenal ulcer is the commonest finding. Gastric ulcers do not occur in patients with gastrinoma. About 75% of patients present with abdominal pain with or without other symptoms. About 75% have diarrhea, 45% have heartburn, 17% experience weight loss, and only 5–10% have scarring of the esophagus or duodenum. About 25% have had gastrointestinal bleeding.

All members of families with MEN 1 are potentially at risk for gastrinoma. In most families, it is now possible to screen for the MEN 1 gene to determine whether an individual is affected. Most affected individuals will develop hyperparathyroidism first, but some develop Zollinger–Ellison syndrome or pituitary problems as the first manifestation. About 66% of patients with MEN 1 will develop gastrinoma at some stage.

Once the diagnosis of Zollinger–Ellison syndrome has been seriously entertained, the patient should have a fasting serum gastrin drawn and then be given 60–80 mg of a proton pump inhibitor per day while the rest of the work up is being performed.

TABLE I Clinical Features That Suggest Gastric Acid Hypersecretion

Recurrent or multiple duodenal ulcers
Helicobacter pylori negative?
No intake of nonsteroidal antinflammatory drugs?
Ulcers in unusual sites
Complicated ulcer disease with bleeding, perforation, or obstruction
Any patient who needs ulcer surgery
Large-volume nasogastric tube output after gastric surgery
Recurrent ulcers after surgery
Secretory diarrhea
Malabsorption
Family history of kidney stones, ulcers, and/or endocrinopathies
Ulcers and kidney stones
Thickened gastric mucosa seen on computer tomography scans or at endoscopy

Confirming the Diagnosis of Zollinger–Ellison Syndrome

To meet the criteria for Zollinger–Ellison syndrome, the patient should have hypergastrinemia and acid hypersecretion. A fasting measurement of serum gastrin is normally <100 pg/ml in the most commonly used immunoassay. The patient should be off all antisecretory medication, but care needs to be taken in doing this because a patient with Zollinger–Ellison syndrome can develop bleeding or a perforation in less than 48 hours off medication. Causes of hypergastrinemia are shown in Table II. If the gastrin concentration is >1000 pg/ml and gastric pH measured at endoscopy or by nasogastric aspiration is <2, then the diagnosis of Zollinger–Ellison syndrome is certain (if the patient does not have one of the very rare surgical causes of this combination—see later). In many patients with gastrinoma, however, the gastrin concentration is <1000 pg/ml, in which case gastric acid secretion must be measured. A nasogastric tube is placed in the dependent part of the stomach and the position is checked by recovery of >90% of 20 ml of water. Then gastric juice is aspirated in four 15-minute aliquots. Acid secretion is calculated from the titratable acidity (measured by back titration or calculated from standard tables) multiplied by the volume. Secretion of >15 mEq/hour (or >5 mEq/hour in a patient who has had a partial gastrectomy or a vagotomy) indicates gastric acid hypersecretion. Causes of gastric acid hypersecretion are shown in Table III. Mean acid outputs in a recent large series of patients with gastrinomas were 41 mEq/hour in patients with intact stomachs and 28 mEq/hour in patients with prior gastric surgery.

The only conditions that produce hypergastrinemia and hypersecretion of acid other than Zollinger–Ellison syndrome due to a gastrinoma are the retained gastric antrum syndrome, chronic partial gastric outlet obstruction, and, transiently, after massive small bowel resection. These conditions are generally easy to differentiate from gastrinoma.

The secretin test has been used in the past to help make the diagnosis of Zollinger–Ellison syndrome. The

TABLE II Causes of Hypergastrinemia

With hypochlorhydria	With normochlorhydria	With hyperchlorhydria
Gastric atrophy	Chronic renal failure	Zollinger–Ellison syndrome
Autoimmune (pernicious anemia)	*Helicobacter pylori* gastritis	Sporadic
Nonautoimmune (atrophic gastritis)	Vagotomy	As part of MEN 1
Helicobacter pylori-induced gastritis	Pheochromocytoma	Retained antrum syndrome
Chronic renal failure		Chronic gastric outlet obstruction
Proton pump therapy		Massive resection of small intestine (transient)
		Helicobacter pylori gastritis
		G cell hyperfunction

test involved administering a bolus of 2 IU/kg body weight of secretin in one arm and drawing blood from the other arm for serum gastrin assay at 15, 1, 2, 5, 10, and 20 minutes after injection. An increase in gastrin concentration of >200 pg/ml indicates Zollinger–Ellison syndrome. However, there is a 15% false negative rate, and false positives can also occur.

Diagnosis of the Type of Gastrinoma

Only about 25% of patients with gastrinoma will have MEN 1, and often this will be obvious. Most will have a family history or will have hyperparathyroidism or both (Table IV). However, sometimes Zollinger–Ellison syndrome may be the first manifestation, and occasionally there may be no family history. If there is any question as to whether the patient has MEN 1, plasma calcium and parathyroid hormone levels should be determined and imaging of the pituitary fossa should be done with magnetic resonance imaging (MRI). Plasma concentrations of prolactin, growth hormone, and LH/FSH should be measured.

Determination of Gastrinoma Localization and Spread

Many studies have examined the usefulness of a variety of imaging modalities for identifying the primary tumor and metastases. These include ultrasound, computer tomography (CT) scan, MRI scan, angiography, and portal venous sampling. Endoscopic ultrasound has not been subjected to a formal blinded trial. However, it is now clear that the single most useful imaging study is octreotide scanning. This is more sensitive and more specific than all the other modalities for detection of the primary tumor. Sensitivity is around 66% (less with smaller tumors and greater with larger tumors) and selectivity is 85%, with 12% false positives. Extraabdominal false positives can occur with thyroid disease, breast disease, and granulomatous disease. Within the abdomen, accessory spleens, previous operation sites, and renal parapelvic cysts may cause false positives. However, the only procedure more sensitive and specific than octreotide scanning is a very detailed laparotomy by an experienced surgeon (using Kocherization of the duodenum, intraoperative ultrasound, duodenal transillumination, and a duodenotomy).

For detection of metastases, octreotide scanning is also the best modality. Scanning detects more than 90% of hepatic metastases, whereas CT detects about 40% and MRI detects about 70%. Furthermore, octreotide is more specific than other modalities. In addition, octreotide scanning is the best modality for detection of bone metastases.

TABLE III Causes of Gastric Acid Hypersecretion

Idiopathic
Zollinger–Ellison syndrome
Sporadic
As part of MEN 1
Systemic mastocytosis
Retained antrum syndrome
Gastric outlet obstruction
Massive small bowel resection (transient)
Gastric carcinoid
Basophilic leukemia
Non-gastrin-producing ulcerogenic tumors

MANAGEMENT OF GASTRINOMA

In a patient with a gastrinoma, there are two management considerations. The first and most important is to control the acid hypersecretion. Once that has been achieved, treatment of the tumor should be considered.

Management of Gastric Acid Hypersecretion

Proton pump inhibitors have made the management of acid hypersecretion quite simple. Treatment with 60 mg/day of omeprazole or lansoprazole has been

TABLE IV Clinical Features That Suggest MEN 1

Family history of ulcer, kidney stones, and/or endocrinopathies
Past history of ulcers kidney stones and/or endocrinopathies
Hyperparathyroidism (>90%)
Islet cell tumors
Gastrinoma (66%)
Insulinoma
Nonfunctioning
Others
Pituitary tumors
Nonfunctioning
Acromegaly
Cushing's disease
Prolactinoma
Tumors producing follicle-stimulating hormone/ luteinizing hormone
Adrenal hyperplasia (usually nonfunctioning)
Multiple lipomas

shown to be a safe initial approach in virtually every patient with a gastrinoma. Fewer data are available for the other proton pump inhibitors, but 40–160 mg of pantoprazole or 60–120 mg of rabeprazole is sufficient, and 80 mg of esomeprazole will certainly be effective. However, the adequacy of acid control should be checked by the measurement of gastric acid output in the last hour before the next dose of drug. Output should be <10 mEq/hour in patients who have not had previous gastric surgery, <5 mEq/hour in patients who have had a vagotomy, and <1 mEq/hour in patients who have had a partial gastrectomy. In the latter patients, adequacy of therapy should also be confirmed by endoscopy. In patients with normal anatomy, drug dose can probably be slowly reduced over weeks or months if acid output is monitored carefully.

For patients who are unable to take oral medication, intravenous pantoprazole (80 mg every 12 hours, or occasionally every 8 hours) will control acid hypersecretion safely in all patients. This is therefore the best option in the perioperative period or if patients are vomiting.

The role of surgery to control acid hypersecretion is now mainly of historic interest. A total gastrectomy is rarely indicated and a partial gastrectomy is dangerous in these patients and should never be performed. A highly selective vagotomy has been shown to reduce drug requirements in patients for whom a surgical cure is not possible. A parathyroidectomy is very useful for patients with a gastrinoma and hyperparathyroidism. If such patients are rendered normocalcemic, acid output and serum gastrin will fall, sometimes to within normal limits.

Management of Tumor Growth and Spread

Localized Sporadic Gastrinomas

Patients with sporadic gastrinomas that are localized are potential candidates for surgical cure. Sporadic gastrinomas are usually single and experienced surgeons can now expect to find gastrinomas even when imaging studies are negative. In a recent series, tumors were found in the last 83 patients that were operated on. Patients with sporadic gastrinoma that has metastasized to local lymph nodes are also candidates for surgery. Life table analysis of such patients shows that such patients have a life expectancy of >80% over 10 years, which is not different from those who are cured by surgery.

Localized Gastrinoma in Patients with MEN 1

Patients with MEN 1 and localized gastrinomas are not generally candidates for surgery. These patients have multiple tumors and are not curable. Furthermore, they may have other types of neuroendocrine tumors that may produce peptides or be nonfunctioning. Thus, it may not be clear at surgery whether the tumors removed are indeed gastrinomas. At one major center, patients with MEN 1 were operated on if there were no hepatic metastases and one or more tumors were >2.5 cm in diameter. The rationale was that in other neuroendocrine tumors the propensity to metastasize is related to the size of the primary. At surgery, most of these patients had multiple tumors and 70% had tumors in lymph nodes. A small percentage of those operated on were found unexpectedly to have a tumor in the liver. None of the patients operated on was cured, and the 15-year survival in those patients with liver tumor was 52%. However, the 90% survival rate in the others was similar survival of patients with no tumor identified or with small tumors. It was concluded that surgery was useful in patients with MEN 1 and large localized tumors.

Metastatic Gastrinoma

Patients with metastatic gastrinoma have a reduced life expectancy. In one large survey, the 15-year survival rate in patients with localized tumor was 93%, the survival rate in those who developed hepatic metastases after diagnosis was 68%, and in those who presented with hepatic metastases the 10-year survival was 26%. Patients with metastases beyond the liver had a shorter survival rate. Overall, 10-year survival in those with liver metastases alone was 82% and in those with liver and bony metastases 20%. Some of these patients have undergone resection of localized hepatic metastases. Others have received chemotherapy, interferon α, or octreotide or had embolization of the liver or a combination of these. None of these approaches has proved

efficacious. Clearly, it is important to identify patients with gastrinomas before they metastasize to the liver. The recent report suggesting that the widespread use of proton pump inhibitors has led to delay in the diagnosis of Zollinger–Ellison syndrome is cause for concern.

MANAGEMENT OF OTHER ISSUES IN PATIENTS WITH GASTRINOMAS AND MEN 1

Patients with MEN 1 may have complex problems concerning pituitary tumors that may be functioning or non-functioning. They may also have a variety of problems related to the different types of pancreatic islet cell tumors that may occur. Although these issues are outside the scope of this article, two problems are of concern with respect to the gastrinoma.

As mentioned above, hyperparathyroidism exaggerates the hypergastrinemia and hyperacidity, and correction of the hyperparathyroidism lowers both acid and gastrin. These patients require either a subtotal parathyroidectomy or total parathyroidectomy with an implant placed in the forearm. Otherwise, recurrent hyperparathyroidism is almost universal.

Gastric ECL cell carcinoid tumors occur in up to 30% of patients with gastrinomas and MEN 1, but are much rarer in patients with sporadic gastrinomas. Gastrin is trophic to the ECL cells, and ECL cell carcinoids occur in other hypergastrinemic states, including pernicious anemia. However, in patients with MEN 1, the ECL cell carcinoids demonstrate loss of heterozygosity of the tumor suppressor gene at the 11q locus and thus share this abnormality with other neuroendocrine tumors in patients with MEN 1. Gastric ECL cell carcinoid tumors may be multiple but are usually benign and generally require no therapy. Antral carcinoids also occur in patients with Zollinger–Ellison syndrome and MEN 1, but these have the characteristics of sporadic carcinoids.

See Also the Following Articles

Duodenal Ulcer • Enterochromaffin-like (ECL) Cells • Gastric Acid Secretion • Gastrin • Histamine • Hyperparathyroidism • Multiple Endocrine Neoplasia (MEN) • Parietal Cells • Proton Pump Inhibitor

Further Reading

Gibril, F., Reynolds, J. C., Doppman, J. L., Chen, C. C., Venzon, D. J., Termanini, B., Weber, H. C., Stewart, C. A., and Jensen, R. T. (1996). Somatostatin receptor scintigraphy: Its sensitivity compared with that of other imaging methods in detecting primary and metastatic gastrinomas. A prospective study. *Ann. Intern Med.* **125**, 26–34.

Jensen, R. T., and Gardner, J. D. (1992). Zollinger–Ellison syndrome; clinical presentation, pathology, diagnosis and treatment. *In* "Peptic Ulcer Disease and Other Acid Related Disorders" (D. Zakim and A. J Dannenberg, eds.), pp. 117–212. Academic Research Associates, New York.

Lew, E. A., Pisegna, J. R., Starr, J. A., Soffer, E. F., Forsmark, C., Modlin, I. M., Walsh, J. H., Beg, M., Bochenek, W., and Metz, D. C. (2000). Intravenous pantoprazole rapidly controls gastric acid hypersecretion in patients with Zollinger–Ellison syndrome. *Gastroenterology* **118**, 696–704.

Maton, P. N. (1986). Cushing's syndrome in patients with the Zollinger–Ellison syndrome. *N. Engl. J. Med.* **15**, 1–5.

Maton, P. N. (1999). Acid hypersecretory states. *In* "Clinical Practice of Gastroenterology" (L. J. Brandt, ed.), pp. 315–323. Current Medicine Inc., Philadelphia.

Norton, J. A., Fraker, D. L., Alexander, H. R., Venzon, D. J, Doppman, J. L., Serrano, J., Goebel, S. U., Peghini, P. L., Roy, P. K., Gibril, F., and Jensen, R. T. (1999). Surgery to cure Zollinger–Ellison syndrome. *N. Engl. J. Med.* **341**, 635–644.

Walsh, J. H. (1994). Gastrin. *In* "Gut Peptides; Biochemistry and Physiology" (J. H. Walsh and G. J. Dockray, eds.), pp. 75–121. Raven Press, New York.

Yu, F., Lubensky, I. A., Mahlamaki, E. H., Zheng, Y. L., Herr, A. M., and Ferrin, L. J. (2000). Survey of genetic alterations in gastrinomas. *Cancer Res.* **60**, 5536–5542.

Yu, F., Venzon, D. J., Serrano, J., Goebel, S. U., Doppman, J. L., Gibril, F., and Jensen, R. T. (1999). Prospective study of the clinical course, prognostic factors, causes of death, and survival in patients with long-standing Zollinger–Ellison syndrome. *J. Clin. Oncol.* **17**, 615–630.

specific gastrin releasing peptide receptor antagonist. *Gut* **49**, 23–28.
Jensen, J. A., Carroll, R. E., and Benya, R. V. (2001). The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. *Peptides* **22**, 689–699.
Jensen, R. T., and Coy, D. H. (1991). Progress in the development of potent bombesin receptor antagonists. *Trends Pharmacol. Sci.* **12**, 13–19.
Kelley, M. J., Linnoila, R. I., Avis, I. L., Georgiadis, M. S., Cuttitta, F., Mulshine, J. L., and Johnson, B. E. (1997). Antitumor activity of a monoclonal antibody directed against gastrin-releasing peptide in patients with small cell lung cancer. *Chest* **112**, 256–261.
Kroog, G. S., Jensen, R. T., and Battey, J. F. (1995). Mammalian bombesin receptors. *Med. Res. Rev.* **15**, 389–417.
Merali, Z., McIntosh, J., and Anisman, H. (1999). Role of bombesin-related peptides in the control of food intake. *Neuropeptides* **33**, 376–386.
Preston, S. R., Miller, G. V., and Primrose, J. N. (1996). Bombesin-like peptides and cancer. *Crit. Rev. Oncol. Hematol.* **23**, 225–238.
Rozengurt, E. (1998). Signal transduction pathways in the mitogenic response to G protein-coupled neuropeptide receptor agonists. *J. Cell. Physiol.* **177**, 507–517.
Schally, A. V., and Nagy, A. (1999). Cancer chemotherapy based on targeting of cytotoxic peptide conjugates to their receptors on tumors. *Eur. J. Endocrinol.* **141**, 1–14.
Tokita, K., Hocart, S. J., Coy, D. H., and Jensen, R. T. (2002). Molecular basis of the selectivity of the gastrin-releasing peptide receptor for gastrin-releasing peptide. *Mol. Pharmacol.* **61**, 1435–1443.

Gastritis

MARK FELDMAN
University of Texas Southwestern Medical School at Dallas and Presbyterian Hospital of Dallas

achlorhydria The inability of the stomach to produce any hydrochloric acid.
gastritis A disorder of the stomach characterized by acute and/or chronic inflammation of the mucosal lining.

The term gastritis refers to a group of diseases that have in common the symptom of inflammation of the stomach. In most cases, the inflammation exclusively or predominately involves the gastric mucosal lining. By definition, a gastric biopsy is required to diagnose gastritis.

CLASSIFICATION AND ETIOLOGY

There is only a modest correlation between gastritis and symptoms such as pain and nausea. In fact, many patients with gastritis have no symptoms referable to the stomach. A classification of gastritis is presented in Table I.

By far, the most common category worldwide is infectious gastritis. *Helicobacter pylori* infection accounts for the majority of cases of gastritis. Cytomegalovirus (CMV) gastritis is most typically seen in immunocompromised patients, such as patients with the acquired immunodeficiency syndrome, patients with organ transplants receiving immunosuppressive drugs, and patients with hematologic malignancies, many of whom receive chemotherapy agents.

ASSOCIATIONS WITH OTHER DISEASES

Certain types of chronic gastritis are associated with an increased risk of cancer. For example, *H. pylori* gastritis is a risk factor for both gastric adenocarcinoma and B-cell gastric mucosa-associated lymphoid tissue lymphoma. Lymphocytic gastritis, which can be associated with *H. pylori* infection or with celiac sprue, may be a precursor of gastric lymphoma. Autoimmune atrophic gastritis is associated with an increased risk of gastric adenocarcinoma, possibly as a consequence of gastric achlorhydria and bacterial overgrowth of the stomach. Patients with autoimmune atrophic gastritis also frequently have reduced intrinsic factor secretion and may develop cobalamin (vitamin B12) deficiency, with hematological sequelae (pernicious anemia) and neurological consequences (optic neuropathy, dementia, subacute combined degeneration of the spinal cord).

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Gastritis

Infectious gastritis
Bacterial
Helicobacter pylori
Mycobacterium tuberculosis or *M. avium/ M. intracellulare* complex
Treponema pallidum (syphilis)
Others (rare)
Viral
Cytomegalovirus
Herpes simplex virus (rare)
Varicella/zoster virus (rare)
Others (rare)
Fungi
Candida albicans and other *Candida* species
Others (rare)
Parasites (rare)
Granulomatous gastritis (noninfectious)
Crohn's disease
Sarcoidosis
Others (rare)
Idiopathic (rare)
Distinctive forms of gastritis (based on microscopic appearance)
Eosinophilic
Collagenous (rare)
Lymphocytic
Graft-versus-host disease
Gastritis cystica profunda (rare)
Autoimmune gastritis
Atrophic
Others

Note. Common forms of gastritis are shown in boldface type.

EOSINOPHILIC GASTRITIS

Eosinophilic gastritis is a relatively uncommon condition of uncertain etiology characterized by peripheral eosinophilia, eosinophilic infiltration of the gastrointestinal tract, and upper gastrointestinal symptomatology. Eosinophilic gastritis is classified according to the layer of gastrointestinal tract involved (i.e., mucosal disease, muscle layer disease, and subserosal disease). Patients with mucosal involvement present symptoms similar to those of other patients with gastritis. Patients with muscle layer disease generally have pyloric obstruction and related symptoms such as nausea, vomiting, and early satiety. Patients with subserosal eosinophilic infiltration develop eosinophilic ascites. Individuals with eosinophilic gastritis often have food allergies and other allergic manifestations. Other parts of the gastrointestinal tract can be involved, such as the small intestine or colon. Eosinophilic gastritis typically responds to therapy with oral glucocorticoids.

THERAPY

Treatment of other forms of gastritis depends upon the underlying etiology (e.g., broad-spectrum antibiotics plus a proton pump inhibitor for *H. pylori* gastritis, gancyclovir for CMV gastritis, penicillin for syphilitic gastritis, and corticosteroids and/or other anti-inflammatory drugs for Crohn's disease or sarcoidosis).

See Also the Following Articles

Atrophic Gastritis • Candidiasis • Cobalamin Deficiency • Crohn's Disease • Cytomegalovirus • Eosinophilic Gastroenteritis • *Helicobacter pylori* • Mycobacterial Infection • Vitamin B12: Absorption, Metabolism, and Deficiency

Further Reading

Appelman, H. (1994). Gastritis: Terminology, etiology, and clinicopathological correlations: Another biased view. *Hum. Pathol.* **25**, 1006.

Correa, P. (1988). Chronic gastritis: A clinico-pathological classification. *Am. J. Gastroenterol.* **83**, 504.

Lee, E., and Feldman, M. (2002). Gastritis and other gastropathies. *In* "Sleisenger and Fordtran: Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 810–827. W. B. Saunders Co., Philadelphia.

Misiewicz, J. (1991). The Sydney system: A new classification of gastritis. *J. Gastroenterol. Hepatol.* **6**, 207.

Gastritis and *Helicobacter pylori*, Pediatric

SAMRA S. BLANCHARD, STEVEN J. CZINN, AND GISELA CHELIMSKY
Rainbow Babies, Children's Hospital, and Case Western Reserve University, Cleveland, Ohio

gastritis Inflammation of the mucous membrane of the stomach.

Helicobacter pylori A bacterium that can cause a transmissible infection of the gastric mucosal surface.

ulcer A break in the lining of the stomach, duodenum, or esophagus, where hydrochloric acid and pepsin are present is known as a peptic ulcer. Ulcers formed in the stomach are known as gastric or stomach ulcers.

Peptic ulcer disease affects 5 million Americans yearly and is a significant cause of morbidity in the United States. The economic cost of this disease now exceeds $4 billion annually. The recent literature suggests that *Helicobacter pylori* plays a major role in the pathogenesis of this disease and a significant correlation between the presence of *H. pylori* and histologic gastritis, duodenal ulcers, and gastric cancer has now been established. Since its first isolation 22 years ago, *H. pylori* infection continues to generate considerable interest in the medical and scientific community. Today, *H. pylori* is considered one of the most common bacterial infections of mankind, infecting nearly one-half of the world's population. Despite this worldwide distribution, the pathogenesis of *H. pylori*-associated gastroduodenal disease remains poorly understood. What is clear is that *H. pylori* infection rarely resolves spontaneously and that the chronic gastritis that accompanies infection plays an etiologic role in the development of peptic ulcer disease and gastric cancer. Evidence confirming that *H. pylori* infection plays a role in the pathogenesis of peptic ulcers comes from clinical studies demonstrating a significant reduction in ulcer relapse rates if the affected patient is cured of the *H. pylori* infection. Similarly, a number of epidemiological studies have demonstrated that chronic *H. pylori* infection can ultimately predispose an individual to a significantly increased risk of developing gastric cancer. These studies were so compelling that the World Health Organization has classified *H. pylori* as a type I human carcinogen.

INTRODUCTION

Effective antimicrobial therapies to treat *Helicobacter pylori* have been developed and, with proper compliance, are over 85% successful. However, individual and population-based problems with pharmaceutical treatment support a role for vaccination in the control of *H. pylori* infection. Current therapies require the patient to ingest multiple agents several times a day for at least 1 week. Treatment can be accompanied by nausea, diarrhea, abdominal pain, and pseudomembranous colitis. These adverse effects contribute to patient noncompliance and reduce the efficacy of the therapy. Additionally, these therapies are prohibitively expensive in developing nations where *H. pylori* infection is endemic, with rates of infection as high as 80–90%. Widespread treatment of *H. pylori* could also result in the development of antibiotic-resistant strains, which has already been observed in patients treated with triple therapy who failed to be cured of infection. Recent research has focused on preventing or curing the infection as well as identifying the mechanisms by which chronic *H. pylori* infection promotes the development of gastroduodenal disease; these findings will be summarized below.

Prior to the discovery of *H. pylori* by Marshall and Warren, a number of investigators identified "spirochetal" bacteria in the gastric mucosa of animal species and similar microorganisms were identified in human gastric tissue. Despite occasional reports documenting the presence of spiral microorganisms in association with the gastric mucosa, skepticism by the medical community was fueled by studies that were unable to confirm the presence of spiral organisms in human gastric tissue. In 1982, Marshall and Warren demonstrated an association between these microorganisms and gastroduodenal disease by culturing spiral *Campylobacter*-like organisms (later renamed *H. pylori*) from human gastric tissue. *H. pylori* is now routinely isolated from both adults and children with gastritis and peptic ulcer disease.

Support for a pathogenic role for *H. pylori* in humans initially came from independent studies performed by Marshall and Morris, who satisfied Koch's postulates by creating histologically confirmed gastritis following the ingestion of a suspension of viable organisms.

MICROBIOLOGY

Morphologically, *H. pylori* resemble the *Campylobacter* species and were initially called *Campylobacter*-like

 Copyright 2004, Elsevier (USA). All rights reserved.

organisms. The organism underwent a number of name changes including *Campylobacter pyloridis* and finally *Campylobacter pylori*. However, morphologic, biochemical, and genetic differences between these spiral organisms and other *Campylobacter* species were noted, indicating that these organisms were unique. Unlike most *Campylobacter* organisms, *H. pylori* has a smooth cell wall with multiple unipolar sheathed flagella. In addition to these morphological differences, *H. pylori* contains a number of enzymes, including a high concentration of urease, that are not present in other *Campylobacter* species. Accordingly, *C. pylori* was given a new genus and species name, *H. pylori*. *Helicobacter* organisms do not stain well with common histologic stains, such as hematoxylin and eosin, but can be demonstrated using silver containing stains, such as Warthin-Starry or Steiner stain.

CLINICAL FEATURES AND EPIDEMIOLOGY OF *H. PYLORI* DISEASE

H. pylori are most abundant in the mucus coat overlying the gastric epithelium and are rarely observed intracellularly. Thus, *H. pylori* infection is always associated with histologic gastritis, which is characterized by the presence of acute and chronic inflammation localized within the gastric antrum. Since the isolation and culture of *H. pylori* 22 years ago, the terms diffuse antral gastritis and multifocal atrophic gastritis have been introduced in an effort to explain the fundamental role that this organism plays in the development of gastroduodenal disease. According to the Sydney system for grading gastritis, individuals at risk for developing peptic ulcers (particularly duodenal ulcers) will develop diffuse antral gastritis, which is characterized by minimal inflammation within the fundus and little or no evidence of intestinal metaplasia within the antrum. In this situation, *H. pylori* can often be found in the duodenum, colonizing islands of gastric metaplasia that are present within the duodenum. Epidemiological studies have suggested that an individual with both *H. pylori* gastritis and gastric metaplasia of the duodenum has a 50-fold increased risk of developing a duodenal ulcer. The majority of patients diagnosed with peptic ulcer disease have evidence of *H. pylori* infection. Finally, combinations of acid-suppressive drugs and antimicrobial agents not only eradicate *H. pylori* and heal duodenal ulcers, but markedly decrease the recurrence rate of duodenal ulcers. Twelve-month recurrence rates of $<10\%$ have been reported in patients successfully treated for *H. pylori* infection.

A number of epidemiological studies have also demonstrated that chronic long-lasting *H. pylori* infection, particularly when acquired in childhood, can significantly increase the risk of developing gastric cancer and/or mucosa-associated lymphoid tissue lymphoma. *H. pylori*-infected individuals at risk for the development of gastric adenocarcinoma usually develop multifocal atrophic gastritis, which is characterized by pangastritis and the presence of significant intestinal metaplasia. These studies were so compelling that the World Health Organization has recently classified *H. pylori* as a human carcinogen.

Most epidemiological studies of *H. pylori* infection have been performed in adults, who likely had been infected for decades before diagnosis. Studies from all continents have documented that the acquisition of *H. pylori* infection occurs primarily during childhood. It is now well established that the overwhelming majority of individuals in developing countries are infected with *H. pylori* during childhood. Data to support the early acquisition of *H. pylori* infection come from retrospective studies that estimate the incidence of new *H. pylori* infections to be 0.37% per patient year. At present, it is known that *H. pylori* infection clusters within families and that the acquisition of this infection is strongly linked to conditions associated with lower socioeconomic status during childhood, such as residential crowding. Observations such as these support the hypothesis of person-to-person transmission of *H. pylori* either via the oral–oral route or via the fecal–oral route. The oral–oral route of transmission is supported by isolated reports confirming the transmission of *H. pylori* from the use of contaminated endoscopes and pH probes. Furthermore, there have been reports confirming the presence of viable *H. pylori* organisms recovered from the oropharynx and dental plaque. Arguments for a fecal–oral route of transmission are based on studies confirming that viable *H. pylori* can be cultured from human feces. Water and pets have also been suggested to be environmental sources of *H. pylori* infection. Additional studies to determine the route of transmission of *H. pylori* are under way so that adequate preventive strategies for this extremely common infection of mankind can be developed.

ROLE OF *H. PYLORI* INFECTION IN CHILDREN WITH ULCERS AND GASTRITIS

Although all children infected with *H. pylori* appear to develop chronic active gastritis, the majority of these children are asymptomatic and never have clinically evident disease. Only a minority of children develop peptic ulceration or gastric cancer, the more severe

manifestations of *Helicobacter* infection. The only known risk factors for the acquisition of *H. pylori* in childhood are minimal education and low socioeconomic status. The association of *H. pylori* infection and recurrent abdominal pain in children is controversial. With the exception of one study by Chong and Lou in 1995, there does not appear to be an association between *H. pylori* infection and an increased prevalence of recurrent abdominal pain. Therefore, routine evaluation for *H. pylori* in patients without symptoms of peptic ulcer disease is not warranted. There is no clear evidence that treatment of *H. pylori* will alleviate symptoms in children who have *H. pylori* gastritis and recurrent abdominal pain.

BACTERIAL PATHOGENESIS

The entire *H. pylori* genome has now been sequenced. Analysis of the genome has lent support to previously described biochemical and physiological observations that led to the designation of the new *Helicobacter* genus. The genome sequence has also suggested the presence of several characteristics common to other mucosal pathogens, including systems for motility, iron scavenging, complex adhesin systems for host–pathogen interactions, and regulatory networks. Additionally, the genetic machinery is present not only for DNA restriction and modification but most probably to provide mechanisms of antigenic variation.

Several virulence factors of *H. pylori* that either play a direct role in the ability of the bacteria to colonize the gastric mucosa or may contribute to the pathogenesis of disease have now been identified. The enzyme urease, which gastric *Helicobacter* species produce in large quantities, is located on the bacterial surface and was one of the first colonization/virulence factors described. It has been suggested that urease, which hydrolyzes urea into CO_2 and NH_4OH, may allow *H. pylori* to survive in the highly acidic environment of the gastric lumen by surrounding itself with a "cloud" of ammonia. Studies in gnotobiotic pigs using a chemically induced urease-negative mutant of *H. pylori* demonstrated that urease is essential for the colonization of the gastric mucosa by *H. pylori*. The ammonia produced by the hydrolysis of urea may also have direct toxic effects on the gastric epithelium. Additionally, some strains of *H. pylori* produce a potent cytotoxin, VacA, which causes vacuolization of tissue culture cells and most likely causes epithelial damage *in vivo*. It is also likely that motility is important in allowing the organism to penetrate the gastric mucus and localize to the gastric epithelium. Recent studies in the pig model using isogenic flagellin knockout strains of *H. pylori* indicated that motility is involved in the virulence and pathogenesis of *H. pylori*. Although *H. pylori* is not generally considered to be an invasive pathogen, it does bind to gastric epithelium both *in vitro* and *in vivo*. There is growing evidence that one or more of a relatively large number of putative adhesins may be important in the pathogenesis of *H. pylori* infections.

In addition to the colonization factors listed above, it has been suggested that phenotypic or genotypic differences among bacterial isolates may be important in the progression of disease. Studies investigating differences among bacterial strains were prompted by observations indicating that although approximately 50% of the world's population is infected with *H. pylori*, less than 10% of the infected population appears to exhibit any disease symptoms. A 120 kDa protein originally identified by gastric immunoglobulin A (IgA) antibodies in infected subjects was found to be coexpressed in bacterial strains expressing large amounts of VacA and was subsequently named CagA (for cytotoxin-associated gene A). There is accumulating evidence that individuals infected with strains of *H. pylori* that express VacA and/or CagA are more likely to develop peptic ulcers or gastric cancer than are individuals who are infected with VacA- or CagA-negative strains. Studies examining the relationship between the presence of CagA and the expression of interleukin-8 (IL-8) by epithelial cells suggest that the CagA protein is a marker for a series of genes within an *H. pylori* "pathogenicity island." Other putative virulence genes have also been identified. *IceA*, an *H. pylori* gene induced by contact with the epithelium, is selectively up-regulated following contact with the gastric epithelium. Two distinct allelic families have been identified (*IceA*1, *IceA*2), with *IceA*1 being the predominant allele in *H. pylori* strains recovered from patients with peptic ulcer disease; it is also associated with increased mucosal concentrations of IL-8. Although these bacterial factors probably play important roles in the diseases associated with *H. pylori* infection, by themselves they cannot explain the observed variations in clinical outcome. Fifty percent or more of all *H. pylori* strains fall into the "type I" designation yet only a small minority of individuals infected with type I strains develop serious disease. Some researchers believe that the host response to infection also plays an important role in determining *H. pylori* disease outcome.

IMMUNE RESPONSES TO *H. PYLORI* INFECTION ARE NOT PROTECTIVE

The inflammatory infiltrate associated with *H. pylori* infection primarily affects the antral region of the

stomach and is accompanied by epithelial degenerative changes. Ultimately, the inflammation associated with *H. pylori* infection is characterized by the infiltration of polymorphonuclear cells and mononuclear cells into the gastric lamina propria and hyperproliferation of the gastric epithelium. The mononuclear cells consist primarily of IgA plasma cells, B cells, and $CD4^+$ T cells. Interestingly, organized lymphoid aggregates not generally seen in the normal stomach are also present in the submucosa and may in part explain the nodularity noted in the gastric antrum of children infected with *H. pylori*.

Since *H. pylori* is a noninvasive bacterium that resides in the gastric mucus, its exposure to immune effector mechanisms is minimal. However, it is able to induce an inflammatory response that cannot eliminate it. A prominent feature of this frustrated inflammatory response is the enhanced presence of IL-8, a neutrophil-activating chemokine, in the gastric mucosa of infected individuals. The recruitment of neutrophils to the lamina propria initiates a cascade of pro-inflammatory events that eventually results in chronic active gastritis. As mentioned above, *H. pylori* strains carrying a pathogenicity island that encodes factors that increase IL-8 production tend to be more inflammatory than those strains that lack the pathogenicity island.

In addition to initiating pro-inflammatory events in the gastric mucosa, *H. pylori* itself is somehow "seen" by the host immune system, which results in *H. pylori*-specific antibody and T-cell responses. It is well documented that infection with *H. pylori* is accompanied by both gastric and serum antibody responses and, as discussed below, the presence of *H. pylori*-specific serum antibodies is often used as an indicator of infection. Local, gastric antibodies have been observed to bind to *H. pylori* in immunohistochemical staining analysis yet the response does not clear the infection.

In 1990, it was demonstrated that the T cells from *H. pylori*-infected patients specifically proliferate in response to *H. pylori* antigens. However, subsequent studies have yielded conflicting results regarding the role of cell-mediated immune responses in *Helicobacter* infections. Although some investigators have detected a *Helicobacter*-specific cellular immune response in infected patients, other investigators have found that lymphocytes from both *H. pylori* seropositive and *H. pylori* seronegative individuals will proliferate *in vitro* in response to *Helicobacter* antigens. In the latter case, *H. pylori* seropositive patients sometimes actually show a decrease in responsiveness compared to *H. pylori* seronegative individuals. This had led to the suggestion that *H. pylori* may down-regulate the cellular immune response toward itself. Until recently, the T cells used in these proliferation studies have employed peripheral blood mononuclear cells. However, several investigators have now examined the local gastric T-cell responses from *H. pylori*-infected patients. The T cells isolated from gastric biopsies are predominantly interferon-γ (IFN-γ)-secreting cells. Since IFN-γ is a pro-inflammatory T helper 1 (T_H1)-type cytokine, it has been suggested that T-cell responses may actually play a role in the gastritis associated with *Helicobacter* infections.

The authors' laboratory has used murine models of *Helicobacter* infection to demonstrate that the host response plays a role in *Helicobacter*-associated pathogenesis. One murine model that has been used extensively to study host responses is the *H. felis* model described by Lee and colleagues. Despite the fact that *H. felis* apparently lacks the *cag* pathogenicity island, it does yield a vigorous inflammatory response in several strains of mice. A series of inbred strains of mice (BALB/c, C3H, C57BL/6, and MHC congenic partners on the BALB/c and C57BL/6 backgrounds) infected with a single strain of *H. felis* were studied. Following infection, these strains of mice exhibited differences in both the extent and magnitude of infection and the nature of the inflammation. The BALB/c background had minimal inflammatory responses after *Helicobacter* infection, but mice with a C57BL/6 background had more severe inflammation. In addition, C57BL/6 mice, like humans infected with *H. pylori*, also exhibited an inflammatory response that increased in severity over time. Additional studies examining the host inflammatory response to *Helicobacter* infection using congenic partners on the BALB/c and C57BL/6 backgrounds suggest that non-major histocompatibility complex (MHC) background genes play the major role in determining disease character. These results have been confirmed by others who have shown similar results with *H. pylori*. These results suggest that the nature of the host immune or inflammatory response is also important in determining disease outcome after *H. pylori* infection of humans.

Participation of both innate and adaptive immune mechanisms in the *H. felis*-associated gastritis model was shown by experiments performed in severe combined immunodeficiency (SCID) mice. The magnitude of the inflammatory response in SCID mice and that of wild-type immunocompetent control mice on a BALB/c background (a low inflammatory mouse strain) were comparable but low. However, when SCID mice on a C57BL/6 background were infected with *H. felis*, they demonstrated a much lower inflammatory response than the wild-type immunocompetent C57BL/6 mice (a high inflammatory mouse strain). In fact, the inflammation in the C57BL/6 SCID mice was equivalent to that in the BALB/c wild-type mice and the BALB/c SCID mice. These results demonstrate that the immune

response contributes to the observed differences in disease outcome among various strains of mice infected with *H. felis* and strongly suggest that the immune response contributes to the different clinical disease outcomes seen in humans chronically infected with *H. pylori*.

Either antibodies or cell-mediated immune responses could play a role in the disease process. A role for antibodies in exacerbating *Helicobacter*-associated gastritis was suggested by the discovery of anti-*Helicobacter* antibodies in *H. pylori*-positive patients and mice that were cross-reactive for antigens in the gastric mucosa. Recently, the autoantibody theory of *Helicobacter* pathogenesis has fallen out of favor. Studies using μMt mice infected with *H. felis* that demonstrated no reduction in gastric inflammation also tends to rule out a role for antibodies in the pathogenesis of this infectious disease.

As mentioned above, the T cells present in the inflammation associated with an *H. pylori* infection are primarily IFN-γ producers. These cells are generally believed to be pro-inflammatory in nature. The authors' experiments using the *H. felis* mouse model are consistent with these observations. Adoptive transfer of *Helicobacter*-specific T_H1 cells into mice results in increased inflammatory responses in the gastric mucosa following challenge with *H. felis*. The extent of the T-cell-mediated inflammation seems to be dictated by the genetics of the host. When Balb/c mice (T_H2 biased) and C57BL/6 mice (T_H1 biased) are infected with *Helicobacter* simultaneously, only the C57BL/6 mice will develop significant inflammation of the gastric mucosa. This difference is not due to MHC usage; by treating infected Balb/c mice with IL-12, a T_H1 regulatory cytokine, inflammation levels rise to those observed in C57BL/6 mice.

The authors' laboratory has also used the *H. felis* mouse model to further explore the role of T cells in *Helicobacter* immunity. It has been observed that, as in humans, splenic lymphocytes from both infected and naive mice exhibited significant proliferation in response to heat-killed antigens and that the T cells from these mice are predominantly IFN-γ secretors. When C57BL/6 mice (T_H1 bias) were infected with *H. felis*, the resulting *Helicobacter*-specific T cells were IFN-γ producers. The levels of IFN-γ produced correlated with the level of inflammation in individual mice. The magnitude of the inflammation was reduced if mice were treated with anti-IFN-γ antibodies. Additionally, when *H. felis*-specific T_H1-cell lines were adoptively transferred into recipient mice that were then infected with *H. felis*, the recipient animals demonstrated enhanced gastric inflammation. Thus, the mouse model supports the data from human studies indicating that an inflammatory T_H1 cellular immune response plays a role in the gastric inflammation observed in *Helicobacter*-infected individuals.

DIAGNOSTIC METHODS

Diagnosis of *H. pylori* infection can be accomplished by several different methods. Although *H. pylori* infection appears to be a chronic lifelong infection, there is a robust host systemic and mucosal humoral immune response to this organism. Therefore, serologic tests have now become the mainstay for noninvasive screening diagnosis of this infection. Commercially available serologic tests can now be performed in a central laboratory, on whole blood using a rapid office-based test, and there are even suggestions that the serologic diagnosis can be made by screening saliva. All of these tests appear to be capable of initially diagnosing an *H. pylori* infection but have limited ability to confirm a cure. Histologic demonstration of *H. pylori* initially was dependent on silver staining (Warthin-Starry) techniques. Using this time-consuming, expensive, and technically difficult technique, spiral microorganisms can be seen clearly in the mucus coat overlying the gastric epithelium. Although the Gram stain is a simple and rapid technique, it is not recommended for *H. pylori* since these gram-negative rods stain poorly and are difficult to visualize. A number of other stains have been evaluated, including Giemsa, phase contrast, acridine orange, Brown-Hopps, and Genta stain, in an effort to develop a cost-effective and reproducible stain to identify *H. pylori* in human tissue.

Since *H. pylori* produces a large amount of urease, the immediate placement of biopsy material into urea-containing broth or agar yields an immediate (within 2 h) color change from yellow to pink, allowing rapid identification. Additional diagnostic techniques have been developed in an attempt to eliminate the need for endoscopy and biopsy. The stable isotopic ^{13}C nonradioactive urea breath test has been used diagnostically since urease is not present in mammalian gastric cells. The recovery of labeled $^{13}CO_2$ in exhaled air following the ingestion of stably labeled urea indicates *H. pylori* infection. The breath test is a useful, noninvasive, nonradioactive means of diagnosing *H. pylori* gastroduodenal disease.

As noted above, several enzyme-linked immunosorbent assay-based commercial kits that measure anti-*H. pylori* serum IgG antibody titers in adults. When used in children, the sensitivity and specificity of these tests is quite variable and seems to depend a great deal on the test used. Children under the age of

12 years also seem to have more false-negative serologies. Therefore, a negative serology does not rule out infection in young children. In children, the [^{28}C]urea breath test may offer some important advantages over the serologic tests. It is noninvasive, easily performed, and easily repeated in order to evaluate response to treatment. This test measures the actual bacterial colonization in gastric mucosa. However, this test is currently not validated or approved by the Food and Drug Administration (FDA) for use in children. Recently, guidelines for the diagnosis and treatment of *H. pylori* infection in children were endorsed by the Academy of Pediatrics and published. Based on the current available data, the use of serology or a urea breath test for the diagnosis of *H. pylori* infection in children is not recommended. These guidelines recommend endoscopy with biopsy for the evaluation of children with symptoms consistent with peptic ulcer disease. The purpose of the endoscopy is to identify organic etiologies for the dyspepsia, not simply to screen for *H. pylori*.

TREATMENT

Despite the fact that the National Institutes of Health consensus conference and numerous authorities all agree that patients with *H. pylori* infection and peptic ulcers should receive therapy targeted at eradicating the *H. pylori* infection, no therapeutic combinations were officially approved by the FDA until recently. However, the FDA has now officially approved a number of therapies for the treatment of *H. pylori*-associated peptic ulcer disease in adults. Specifically, a combination of clarithromycin 500 mg tid and omeprazole 40 mg qid for 2 weeks followed by 2 additional weeks of omeprazole therapy or a second combination of clarithromycin 500 mg tid and ranitidine/bismuth citrate 400 mg bid for 4 weeks has received approval. Finally, a PeptoBismol/metronidazole/tetracycline triple therapy has also been approved by the FDA for the eradication of *H. pylori* infection. Despite these approvals, numerous research groups continue to search for therapies giving better eradication rates. The addition of metronidazole to the omeprazole/claritromycin dual therapy appears to be superior and may allow physicians to shorten the length of therapy.

Compliance is a very important factor in achieving eradication of *H. pylori* in children. Several regimens were published, including 1-week therapy with colloidal bismuth subcitrate, clarithromycin, and metronidazole (eradication in 21 of 22 subjects) or 1-week treatment with lansoprazole, amoxicillin, and clarithromycin. The latter study had an eradication rate of 87%. Treatment for a 2-week period with metronidazole, clarithromycin, and omeprazole was successful in 93% of the children treated. Although concerns about the use of bismuth salts in children have been expressed, none of the potential side effects have been reported when used for the treatment of *H. pylori* in children. Table I summarizes the recommended eradication therapies for *H. pylori*-associated disease in children.

TABLE I Recommended Eradication Therapies for *Helicobacter pylori*-Associated Disease in Children

Medication	Dose mg/kg/day bid	Duration of treatment
Amoxicillin	50	14 days
Clarithromycin	15	14 days
Proton pump inhibitor	1	1 month
Amoxicillin	50	14 days
Metronidazole	20	14 days
Proton pump inhibitor	1	1 month
Clarithromycin	15	14 days
Metronidazole	20	14 days
Proton pump inhibitor	1	1 month

Unfortunately, attempts to cure large populations of individuals infected with *H. pylori* has led to the acquisition of antimicrobial resistance. A large number of *H. pylori* clinical isolates have acquired resistance to clarithromycin and metronidiazole. Drug therapies for eradication of *H. pylori* from the gastrointestinal tract continue to have a number of limitations, including poor compliance, adverse side effects, and antimicrobial resistance. Therefore, a number of laboratories began to investigate the possibility of developing a vaccine to prevent *H. pylori* infection. The results of this work have not only demonstrated the feasibility of a prophylactic vaccine, but have also demonstrated the feasibility of a therapeutic vaccine that is capable of curing patients already infected with *H. pylori*.

VACCINES FOR THE PREVENTION OF *H. PYLORI* INFECTION

In addition to its role in pathogenesis, there is potential for the immune response to play a role in controlling *Helicobacter* infection. The authors initially used the *H. felis*/mouse model to develop an oral immunization protocol that is capable of inducing protective immunity against infectious challenge. This protocol, which employs *Helicobacter* antigens in combination with the mucosal adjuvant cholera toxin (CT), has now also been applied to the *H. pylori* model by the authors' laboratory and other laboratories. Surprisingly, not only can

protective immunity be induced by prophylactic immunization of healthy mice, but in both mouse and ferret models of *Helicobacter* infection, immunization appears to be an effective immunotherapy to eradicate the organism from infected hosts.

One major question relating to *Helicobacter* immunity concerns why it is that both immunization and infection induce local immune responses but only the immunized animals are protected. It has been previously demonstrated that *Helicobacter*-specific antibodies are sufficient to provide passive protection in mice when applied directly to the stomach. Additionally, distinct differences in the antigenic specificities between mice that are infected with *Helicobacter* and those that are protectively immunized have been described. However, a recent study employing IgA knockout mice showed that they could be protected in the absence of secretory IgA. Additionally, when κMT knockout antibody-deficient mice were used to further investigate the role of antibodies in protective immunity, it was demonstrated that antibodies are not required for protection. Thus, it appears that the primary difference in the immune responses between protected and infected groups is probably in the T-cell response.

A role for T cells in protective immunity against *Helicobacter* is supported by experiments in which adoptively transferred $CD4^+$ T cells from *Helicobacter*-immunized mice, but not those from infected mice, were sufficient to provide protection from challenge. A role for the T_H2-type immune response was suggested by the finding that adoptively transferring a T_H2-cell line could reduce the bacterial load but using a T_H1-cell line could not. Furthermore, a switch from T_H1 to T_H2 cytokine production following successful therapeutic vaccination also supports a role for T_H2 cells in immune protection from *Helicobacter* infection. All of these data have been generated from immunization studies using CT or *Escherichia coli* heat-labile toxin (LT) as mucosal adjuvants for oral immunization and the observations supporting a role for T_H2-based immunity are consistent with the known activities of mucosal adjuvants such as CT or LT. Therefore, it appears that immunization to prevent and/or cure chronic *H. pylori* infection may be the ideal therapeutic approach.

VACCINE STUDIES IN HUMANS

Since greater than half of the world's population is chronically infected with *H. pylori*, an effective vaccine regimen is a realistic goal. Successful eradication of chronic *Helicobacter* infections in the three animal models discussed in this article is an encouraging finding. A therapeutic vaccine would be more cost-effective and result in fewer potential complications than the administration of multiple antimicrobial agents to affected populations. Recently, a clinical trial was performed on volunteers to test an oral therapeutic vaccine consisting of recombinant *H. pylori* urease apoenzyme in combination with the mucosal adjuvant *E. coli* LT. The results were encouraging as subjects receiving the test vaccine had significant reductions in gastric *H. pylori* density compared to those receiving placebo controls. Additionally, anti-*H. pylori* urease IgA-producing cells from the circulation were significantly elevated in those patients receiving the vaccine compared to controls. Although a significant number of patients experienced diarrhea due to the LT, and complete bacterial eradication was not observed in any of the subjects, this study was an important first step in suggesting that improved vaccine formulations and safer adjuvants may achieve satisfactory outcomes.

SUMMARY

There is now considerable evidence that *H. pylori* is a human pathogen. The strong association between *H. pylori* and gastroduodenal disease is well documented. A number of hypotheses have been suggested for the pathogenic mechanisms of *H. pylori*-induced gastroduodenal disease including the presence of bacterial virulence factors, the production of inflammatory mediators, dysregulation of acid secretion, and the host immune response. Although animal models in which protective immunity can be induced by oral vaccination have been developed, the technology has not been advanced for use in humans. At present, treatment with a combination of antimicrobial agents and proton pump inhibitors continues to be recommended for the treatment of *H. pylori*-associated peptic ulcer disease in children.

Acknowledgments

This work was supported by Grants DK-46461, AI-40701, and HL-37117 from the National Institutes of Health.

See Also the Following Articles

Gastritis • *Helicobacter pylori* • Proton Pump Inhibitors • Recurrent Abdominal Pain (RAP)

Further Reading

Appelmelk, B. J., Faller, G., Claeys, D., Kirchner, T., and Vandenbroucke-Grauls, C. M. (1998). Bugs on trial: The case of *Helicobacter pylori* and autoimmunity. *Immunol. Today* **19**, 296–299.

Boren, T., Falk, P., Roth, K. A., Larson, G., and Normark, S. (1993). Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* **262**, 1892–1895.

Chong, S. K., Lou, Q., *et al.* (1995). *Helicobacter pylori* infection in recurrent abdominal pain in childhood: Comparison of diagnostic tests and therapy. *Pediatrics* **96**.

Czinn, S. J., Cai, A., and Nedrud, J. G. (1993). Protection of germ-free mice from infection by *Helicobacter felis* after active oral or passive IgA immunization. *Vaccine* **11**, 637–642.

Elashoff, J. D., and Grossman, M. I. (1980). Trends in hospital admissions and death rates for peptic ulcer disease in the United States. *Gastroenterology* **78**, 280–285.

Forman, D., Webb, P., and Parsonnet, J. (1994). *H. pylori* and gastric cancer. *Lancet* **343**, 243–244.

Gold, B. D., Colletti, R. B., *et al.* (2000). *Helicobacter pylori* infection in children: Recommendations for diagnosis and treatment. *J. Pediatr. Gastroenterol. Nutr.* **31**, 490–497.

Lee, A., Fox, J. G., Otto, G., and Murphy, J. (1990). A small animal model of human *Helicobacter pylori* active chronic gastritis. *Gastroenterology* **99**, 1315–1323.

Marshall, B. J., Armstrong, J. A., and McGechie, D. B. (1985). Attempt to fulfill Koch's postulate for pyloric *Campylobacter*. *Med. J. Austr.* **142**, 436–439.

Michetti, P., Kreiss, C., Kotloff, K. L., Porta, N., Blanco, J.-L. Bachmann, D., Herranz, M., Saldinger, P. F., Corthesy-Theulaz, I., Losonsky, G., Nichols, R., Simon, J., Stolte, M., Ackerman, S., Monath, T. P., and Blum, A. L. (1999). Oral immunization with urease and *Escherichia coli* heat-labile enterotoxin is safe and immunogenic in *Helicobacter pylori*-infected adults. *Gastroenterology* **116**, 804–812.

Parsonnet, J., Friedman, G. D., Orentreich, N., and Vogelman, H. (1997). Risk of gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. *Gut* **1997**, 297–301.

Peterson, W. L. (1991). *Helicobacter pylori* and peptic ulcer disease. *N. Engl. J. Med.* **324**, 1043–1048.

Tomb, J. F., White, O., Kerlavage, A. R., *et al.* (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.

Valle, J., Kekki, M., Sipponen, P., Ihamaki, T., and Siurla, M. (1996). Long-term course and consequences of *Helicobacter pylori* gastritis. *Scand. J. Gastroenterol.* **31**, 546–550.

Xiang, Z., Censini, S., Bayeli, P. F., *et al.* (1995). Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun.* **63**, 94–98.

Gastro-colic Reflex

WILLIAM Y. CHEY AND KAE YOL LEE

Rochester Institute for Digestive Diseases and Sciences, New York

chemoreceptors Molecules on sensory neurons that recognize specific chemical substances in the extracellular milieu and signal changes in the concentration of the specific substance.

colonic motility Patterned activity of the musculature of the colon responsible for propulsion and mixing of the luminal contents.

enteric nervous system Division of the autonomic nervous system that behaves like an independent integrative nervous system within the wall of the digestive tract.

gastro-colic reflex Meal-induced change in motility of the large intestine.

irritable bowel syndrome Persistent symptoms of abdominal pain and alterations of bowel habit not explainable by any apparent physical or biochemical abnormality.

mechanoreceptors Molecules on sensory neurons that detect and signal mechanical energy changes such as contractile tension or stretch of a muscle.

As early as 1909, A. F. Hertz reported that ingestion of a meal is the most powerful of all stimuli for motor activity of the colon in humans. Hertz did a fluoroscopic examination of an individual who had eaten a supper of porridge containing bismuth oxychloride on the previous evening; fluoroscopy before breakfast showed the cecum and ascending colon and most of the transverse colon, and the terminal ileum was also faintly visible. By 45 minutes after breakfast, the ileum was no longer visible and the shadow of the whole colon, as far as the junction of the iliac with the pelvic colon, was clearly seen. The breakfast clearly acted as a powerful stimulant for forward propulsion of the contents of the colon and of the terminal ileum. Hertz believed that the effect of food entering the stomach on the terminal ileum and large intestine was a reflex action, which he called the "gastro-colic reflex."

Copyright 2004, Elsevier (USA). All rights reserved.

Chong, S. K., Lou, Q., *et al.* (1995). *Helicobacter pylori* infection in recurrent abdominal pain in childhood: Comparison of diagnostic tests and therapy. *Pediatrics* **96**.

Czinn, S. J., Cai, A., and Nedrud, J. G. (1993). Protection of germ-free mice from infection by *Helicobacter felis* after active oral or passive IgA immunization. *Vaccine* **11**, 637–642.

Elashoff, J. D., and Grossman, M. I. (1980). Trends in hospital admissions and death rates for peptic ulcer disease in the United States. *Gastroenterology* **78**, 280–285.

Forman, D., Webb, P., and Parsonnet, J. (1994). *H. pylori* and gastric cancer. *Lancet* **343**, 243–244.

Gold, B. D., Colletti, R. B., *et al.* (2000). *Helicobacter pylori* infection in children: Recommendations for diagnosis and treatment. *J. Pediatr. Gastroenterol. Nutr.* **31**, 490–497.

Lee, A., Fox, J. G., Otto, G., and Murphy, J. (1990). A small animal model of human *Helicobacter pylori* active chronic gastritis. *Gastroenterology* **99**, 1315–1323.

Marshall, B. J., Armstrong, J. A., and McGechie, D. B. (1985). Attempt to fulfill Koch's postulate for pyloric *Campylobacter*. *Med. J. Austr.* **142**, 436–439.

Michetti, P., Kreiss, C., Kotloff, K. L., Porta, N., Blanco, J.-L. Bachmann, D., Herranz, M., Saldinger, P. F., Corthesy-Theulaz, I., Losonsky, G., Nichols, R., Simon, J., Stolte, M., Ackerman, S., Monath, T. P., and Blum, A. L. (1999). Oral immunization with urease and *Escherichia coli* heat-labile enterotoxin is safe and immunogenic in *Helicobacter pylori*-infected adults. *Gastroenterology* **116**, 804–812.

Parsonnet, J., Friedman, G. D., Orentreich, N., and Vogelman, H. (1997). Risk of gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. *Gut* **1997**, 297–301.

Peterson, W. L. (1991). *Helicobacter pylori* and peptic ulcer disease. *N. Engl. J. Med.* **324**, 1043–1048.

Tomb, J. F., White, O., Kerlavage, A. R., *et al.* (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.

Valle, J., Kekki, M., Sipponen, P., Ihamaki, T., and Siurla, M. (1996). Long-term course and consequences of *Helicobacter pylori* gastritis. *Scand. J. Gastroenterol.* **31**, 546–550.

Xiang, Z., Censini, S., Bayeli, P. F., *et al.* (1995). Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun.* **63**, 94–98.

Gastro-colic Reflex

WILLIAM Y. CHEY AND KAE YOL LEE

Rochester Institute for Digestive Diseases and Sciences, New York

chemoreceptors Molecules on sensory neurons that recognize specific chemical substances in the extracellular milieu and signal changes in the concentration of the specific substance.

colonic motility Patterned activity of the musculature of the colon responsible for propulsion and mixing of the luminal contents.

enteric nervous system Division of the autonomic nervous system that behaves like an independent integrative nervous system within the wall of the digestive tract.

gastro-colic reflex Meal-induced change in motility of the large intestine.

irritable bowel syndrome Persistent symptoms of abdominal pain and alterations of bowel habit not explainable by any apparent physical or biochemical abnormality.

mechanoreceptors Molecules on sensory neurons that detect and signal mechanical energy changes such as contractile tension or stretch of a muscle.

As early as 1909, A. F. Hertz reported that ingestion of a meal is the most powerful of all stimuli for motor activity of the colon in humans. Hertz did a fluoroscopic examination of an individual who had eaten a supper of porridge containing bismuth oxychloride on the previous evening; fluoroscopy before breakfast showed the cecum and ascending colon and most of the transverse colon, and the terminal ileum was also faintly visible. By 45 minutes after breakfast, the ileum was no longer visible and the shadow of the whole colon, as far as the junction of the iliac with the pelvic colon, was clearly seen. The breakfast clearly acted as a powerful stimulant for forward propulsion of the contents of the colon and of the terminal ileum. Hertz believed that the effect of food entering the stomach on the terminal ileum and large intestine was a reflex action, which he called the "gastro-colic reflex."

Copyright 2004, Elsevier (USA). All rights reserved.

INTRODUCTION

As described by Hertz in 1909, there is a propulsive response of the bowels following ingestion of food (Fig. 1). Normally, the desire for defecation is felt following a meal, and most often after breakfast. Stimulation of the gastro-colic reflex by food entering an empty stomach after a fast of 12 hours is greater than that produced by food taken in the middle of the day or in the evening at a comparatively short interval after a noon-day meal. Investigators during the early 20th Century observed that under natural conditions, the passage of feces along the colon was mainly due to a rapid mass movement, repeated three or four times a day, in the direction of the anus along a considerable length of bowel. It was thought that these movements were produced by a powerful peristaltic propulsive wave. The propulsive wave was characterized by manometric studies of the human colon several decades later. Manometric studies found that a high-amplitude propagating contraction (HAPC) of the colonic musculature accounted for the mass movements of feces.

The increased colonic motility in response to a meal occurs even in patients who have had their stomach removed. Occurrence of meal-evoked movements in the colon of these individuals suggests that the gastro-colic reflex can be evoked both by entry of food into the stomach and by entry of food into the upper small intestine.

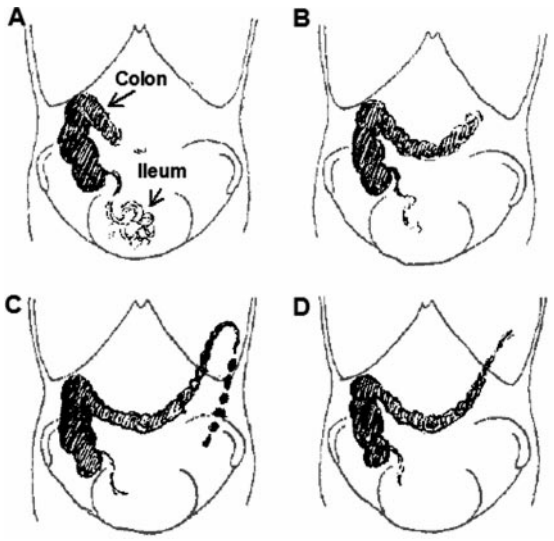


FIGURE 1 An early radiographic study by A. F. Hertz, with bismuth as the contrast material, illustrating the gastro-colic reflex. (A) Prior to the morning meal, contrast material ingested the previous evening was localized to the terminal ileum, ascending colon, and proximal region of the transverse colon. (B) Postprandial motility moved the contrast material into most of the transverse colon and emptied the terminal ileum 25 minutes after the morning meal. (C) The contrast material appears throughout the transverse colon and in the descending colon down to the recto-sigmoid colon 45 minutes after the meal. (D) By 60 minutes after the meal, defecation emptied the left colon.

CLINICAL SIGNIFICANCE

Exaggeration of the stimulating action of a meal on colonic motility appears to be the cause of the symptoms frequently observed in patients with the irritable bowel syndrome (IBS); these patients experience abdominal pain, diarrhea, and urgency of defecation following ingestion of a meal. Ingestion of a meal, particularly one rich in fat, results in powerful contractions of the left colon that propagate rapidly to the sigmoid colon. Motility studies show that the powerful contractions that propagate from the proximal colon to the distal colon, including the sigmoid colon, are HAPCs. The luminal pressure in the colon can increase to a level as high as 500 mmHg during the HAPCs in IBS patients (Fig. 2).

When HAPCs occur in the left colon, the IBS patients often experience sensations of urgency to defecate. These observations imply that the sensation of urgency followed by defecation after breakfast is due to a movement of colon contents caused by powerful propulsive HAPCs. On the contrary, in patients with severe constipation associated with thoracic spinal injury, multiple sclerosis, or diabetes mellitus, the colonic motor response to a meal is absent.

NEURONAL MECHANISMS

Understanding of detailed mechanisms of meal-induced changes in colonic motility is still in its infancy; nevertheless, the evidence suggests that neurohormonal mechanisms are the underlying factors responsible for the gastro-colic reflex. The nervous system extrinsic to the stomach and colon has been suggested to be involved in the mechanisms that evoke colonic motor response to ingestion of a meal. In patients with thoracic spinal injury and multiple sclerosis, the colonic motor response to eating is either absent or impaired. In diabetic patients with severe constipation, there is no postprandial increase in colonic motor activity. In diabetic patients, an autonomic neuropathy (i.e., malfunction of the extrinsic innervation of the colon) may be responsible for absence of the postprandial gastro-colic response.

Involvement of the extrinsic nervous system in the control of colonic motility has been demonstrated in experimental animal models. In the monkey, reversible cooling of the vagus nerves significantly decreases colonic motility both in fasting and in the postprandial state. Reversible vagal cooling also reduces spontaneously occurring colonic motility in the ferret. Moreover,

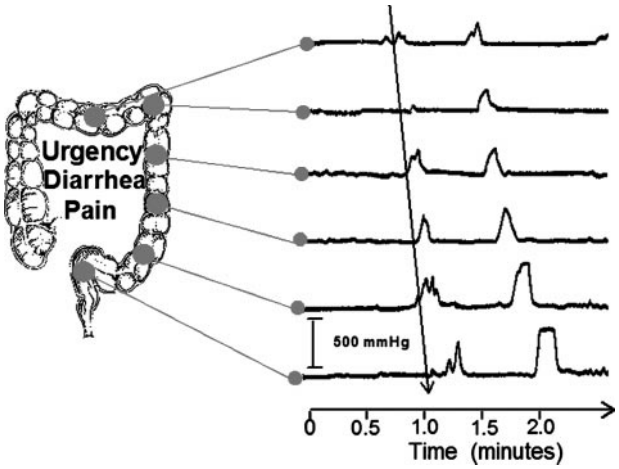


FIGURE 2 In patients with irritable bowel syndrome, high-amplitude propagating contractions occur after meals. Amplitudes of the contractions may exceed 500 mmHg when detected with manometric sensors positioned within the colonic lumen at six locations. High-amplitude propagating contractions occur twice on the manometric record. They are first recorded in the midtransverse colon and propagate rapidly to the rectosigmoid colon. The high-amplitude propagating contractions are associated with lower abdominal cramping pain, sensations of fecal urgency, and diarrhea in patients with irritable bowel syndrome. Reprinted from Chey *et al.* (2001). With permission from the American College of Gastroenterology.

in anesthetized ferrets, electric stimulation of either the cut central or peripheral end of a branch of the abdominal vagus nerve evokes a large-amplitude contraction of the colon. The response is not blocked by atropine or atropine plus adrenergic blockers, suggesting that neither acetylcholine nor norepinephrine is involved. Two separate vagal motor pathways to the colon influence colonic motility. One pathway releases the neurotransmitter acetylcholine and is referred to as cholinergic. The second pathway involves neither acetylcholine nor norepinephrine and is therefore referred to as a noncholinergic–nonadrenergic pathway.

John Wiley *et al.*, have demonstrated the presence of mechanoreceptors in the human stomach that respond to gastric distension and that appear to be the starting point for a reflex increase in rectosigmoid motility. The response to gastric distension is abolished by atropine, indicating that it is mediated via a cholinergic pathway. A recent finding in IBS patients is that the colonic motor response with HAPCs to ingestion of a meal or to administration of the hormone cholecystokinin is abolished by a drug that blocks the 5-hydroxytryptamine subtype 3 (5-HT3) receptor for serotonin. Thus, both cholinergic and serotonergic receptors appear to be involved in gastro-colonic responses, probably via the enteric nervous system.

HORMONAL MECHANISMS

In as few as 5 minutes after ingested liquids and solids enter the stomach, the liquefied component of the contents leaves the stomach and enters the duodenum. Thus, the colonic motor response that occurs in a half hour after a meal is probably due not only to distension of the stomach with food and signaling of gastric distension by mechanoreceptors, but also to gastric contents entering the upper small intestine. Of the nutrients that enter the duodenum, fat and its digestive products are the most potent stimulants for triggering increases in colonic motor activity, particularly that of the left colon, to produce mass movements that lead to defecation. The action of fat is mediated via chemoreceptors in the upper small intestinal mucosa and is partially atropine sensitive. Atropine blocks the muscarinic type receptor for acetylcholine, and the action of atropine in this case suggests the involvement of neural pathways that release the neurotransmitter acetylcholine.

Cholecystokinin (CCK) is an important gut hormone that is released by fat. Although intravenous administration of CCK has long been known to stimulate contractile activity of the human sigmoid colon and rectum, it was initially unknown whether fat-induced colonic motility was mediated by release of CCK as a hormonal signal. This was clarified by observations that fat-induced elevation of CCK in the blood was incapable of increasing rectosigmoid motility in healthy volunteers. Moreover, a drug that blocks receptors for CCK does not affect postprandial colonic motility in healthy subjects or in patients with IBS. That release of CCK as a hormonal signal in normal conditions cannot account for the increased rectosigmoid motility that occurs after a fat-containing meal is now clear. Aside from CCK, other hormones are released in response to normal and fat-rich meals. These meal-stimulated hormones include peptide YY and motilin. Whether these hormones are involved in the signaling of the gastro-colic reflex remains as an open question.

See Also the Following Articles

Autonomic Innervation • Colonic Motility • Enteric Nervous System • Irritable Bowel Syndrome • Parasympathetic Innervation • Postprandial Motility • Power Propulsion • Sensory Innervation • Sympathetic Innervation • Vagus Nerve

Further Reading

Bassotti, G., and Gaburri, M. (1988). Manometric investigation of high amplitude contractile activity of the human colon. *Am. J. Physiol.* **255**, G660–G664.

Battle, W. M., Snape, W. J., Jr., Alavi, A., Cohen, S., and Braunstein, S. (1980). Colonic dysfunction in diabetes mellitus. *Gastroenterology* **79**, 1217–1221.

Chey, W. Y., Jin, H. O., Lee, M. H., Sun, S. W., and Lee, K. Y. (2001). Colonic motility abnormality in patients with irritable bowel syndrome exhibiting abdominal pain and diarrhea. *Am. J. Gastroenterol.* **96**, 1499–1506.

Collman, P. I., Grundy, D., and Scratcherd, T. (1984). Vagal control of colonic motility in the anesthetized ferret: Evidence for a noncholinergic excitatory innervation. *J. Physiol.* (*Lond.*) **348**, 35–42.

Cook, I. J., Furukawa, Y., Panagopoulos, V., *et al.* (2000). Relationship between special patterns of colonic pressure and individual movements of content. *Am. J. Physiol.* **278**, G329–G341.

Dinoso, V. P., Meshkinpour, H., Lorber, S. H., Gutierrez, J. G., and Chey, W. Y. (1973). Motor responses of the sigmoid colon and rectum to exogenous cholecystokinin and secretin. *Gastroenterology* **45**, 364–367.

Esser, M. J., Cowles, V. E., Robinson, J. C., Schulte, W. J., Gleysteen, J. J., and Condon, R. E. (1989). Effects of vagal cryointerruption on colonic contractions in monkeys. *Surgery* **106**, 139–145.

Glick, M. C., Meshkinpour, H., Haldeman, S., Bhatia, N. N., and Bradley, W. E. (1982). Colonic dysfunction in multiple sclerosis. *Gastroenterology* **83**, 1002–1007.

Glick, M. C., Meshkinpour, H., Haldeman, S., Hoeler, F., Downey, N., and Bradley, W. E. (1984). Colonic dysfunction in patients with thoracic spinal cord injury. *Gastroenterology* **86**, 287–294.

Hertz, A. F. (1909). "Constipation and Allied Intestinal Disorders." p. 18. Oxford.

Hertz, A. F., and Newton, A. (1913). The normal movements of the colon in man. *J. Physiol.* (*Lond.*) **47**, 57–65.

Holdstock, D. J., and Misiewicz, J. J. (1970). Factors controlling colonic pressure and transit after meals in patients with total gastrectomy, pernicious anemia and duodenal ulcer. *Gut* **11**, 100–110.

Lee, K. Y., Kallas, F. G., Sun, S. W., and Chey, W. Y. (1999). Serotonin antagonist, ondansetron, inhibits colonic contraction and alleviates abdominal pain induced by cholecystokinin in humans. *Gastroenterology* **116**, G4453.

Niederau, C., Faber, S., and Karaus, M. (1992). Cholecystokinin's role in regulation of colonic motility in health and in irritable bowel syndrome. *Gastroenterology* **102**, 1889–1898.

Renny, A., Snape, W. J., Jr., Sun, E. A., London, R., and Cohen, S. (1983). Role of cholecystokinin in the gastrocolic response to a fat meal. *Gastroenterology* **85**, 17–21.

Wiley, J., Tatum, D., Keinath, R., and Owyang, C. (1988). Participation of gastric mechanoreceptors and intestinal chemoreceptors in the gastrocolic response. *Gastroenterology* **94**, 1144–1149.

Gastroenteritis

SUZANNE M. MATSUI

Veterans Affairs Palo Alto Health Care System and Stanford University School of Medicine

Norovirus Recently adopted name to designate the genus of the viral family, Caliciviridae; previously called Norwalk-like viruses or small round structured viruses.

reverse transcription and polymerase chain reaction Method used to amplify viral ribonucleic acid.

Sapovirus Recently adopted name to designate the genus of the viral family, Caliciviridae; previously called Sapporo-like viruses or classic caliciviruses.

virus-like particle(s) Spontaneously assemble when the viral capsid protein is expressed *in vitro*.

Acute gastroenteritis is a common, generally self-limiting disease characterized by rapid onset of watery diarrhea and associated symptoms such as nausea, vomiting, fever, and/or abdominal pain. In some cases, nausea and vomiting may be the predominant symptoms. It is estimated that this illness occurs at an overall rate of 1.2 to 1.5 episodes per person per year in the United States and at higher rates among children under the age of 3 years (1.3 to 2.3 episodes per child per year). Although gastroenteritis may be caused by many infectious agents, viruses are the most common cause in industrialized countries. Rotaviruses are the leading cause of severe diarrhea in childhood; noroviruses are a diverse group of viruses responsible for a large proportion of outbreaks of gastroenteritis in adults, and sapoviruses, astroviruses, and enteric adenoviruses are other viral agents of gastroenteritis.

INTRODUCTION

Gastrointestinal viruses were identified by electron microscopy in the early 1970s, and since then, much has

Copyright 2004, Elsevier (USA). All rights reserved.

Battle, W. M., Snape, W. J., Jr., Alavi, A., Cohen, S., and Braunstein, S. (1980). Colonic dysfunction in diabetes mellitus. *Gastroenterology* **79**, 1217–1221.

Chey, W. Y., Jin, H. O., Lee, M. H., Sun, S. W., and Lee, K. Y. (2001). Colonic motility abnormality in patients with irritable bowel syndrome exhibiting abdominal pain and diarrhea. *Am. J. Gastroenterol.* **96**, 1499–1506.

Collman, P. I., Grundy, D., and Scratcherd, T. (1984). Vagal control of colonic motility in the anesthetized ferret: Evidence for a noncholinergic excitatory innervation. *J. Physiol.* (*Lond.*) **348**, 35–42.

Cook, I. J., Furukawa, Y., Panagopoulos, V., *et al.* (2000). Relationship between special patterns of colonic pressure and individual movements of content. *Am. J. Physiol.* **278**, G329–G341.

Dinoso, V. P., Meshkinpour, H., Lorber, S. H., Gutierrez, J. G., and Chey, W. Y. (1973). Motor responses of the sigmoid colon and rectum to exogenous cholecystokinin and secretin. *Gastroenterology* **45**, 364–367.

Esser, M. J., Cowles, V. E., Robinson, J. C., Schulte, W. J., Gleysteen, J. J., and Condon, R. E. (1989). Effects of vagal cryointerruption on colonic contractions in monkeys. *Surgery* **106**, 139–145.

Glick, M. C., Meshkinpour, H., Haldeman, S., Bhatia, N. N., and Bradley, W. E. (1982). Colonic dysfunction in multiple sclerosis. *Gastroenterology* **83**, 1002–1007.

Glick, M. C., Meshkinpour, H., Haldeman, S., Hoeler, F., Downey, N., and Bradley, W. E. (1984). Colonic dysfunction in patients with thoracic spinal cord injury. *Gastroenterology* **86**, 287–294.

Hertz, A. F. (1909). "Constipation and Allied Intestinal Disorders." p. 18. Oxford.

Hertz, A. F., and Newton, A. (1913). The normal movements of the colon in man. *J. Physiol.* (*Lond.*) **47**, 57–65.

Holdstock, D. J., and Misiewicz, J. J. (1970). Factors controlling colonic pressure and transit after meals in patients with total gastrectomy, pernicious anemia and duodenal ulcer. *Gut* **11**, 100–110.

Lee, K. Y., Kallas, F. G., Sun, S. W., and Chey, W. Y. (1999). Serotonin antagonist, ondansetron, inhibits colonic contraction and alleviates abdominal pain induced by cholecystokinin in humans. *Gastroenterology* **116**, G4453.

Niederau, C., Faber, S., and Karaus, M. (1992). Cholecystokinin's role in regulation of colonic motility in health and in irritable bowel syndrome. *Gastroenterology* **102**, 1889–1898.

Renny, A., Snape, W. J., Jr., Sun, E. A., London, R., and Cohen, S. (1983). Role of cholecystokinin in the gastrocolic response to a fat meal. *Gastroenterology* **85**, 17–21.

Wiley, J., Tatum, D., Keinath, R., and Owyang, C. (1988). Participation of gastric mechanoreceptors and intestinal chemoreceptors in the gastrocolic response. *Gastroenterology* **94**, 1144–1149.

Gastroenteritis

SUZANNE M. MATSUI

Veterans Affairs Palo Alto Health Care System and Stanford University School of Medicine

Norovirus Recently adopted name to designate the genus of the viral family, Caliciviridae; previously called Norwalk-like viruses or small round structured viruses.

reverse transcription and polymerase chain reaction Method used to amplify viral ribonucleic acid.

Sapovirus Recently adopted name to designate the genus of the viral family, Caliciviridae; previously called Sapporo-like viruses or classic caliciviruses.

virus-like particle(s) Spontaneously assemble when the viral capsid protein is expressed *in vitro*.

Acute gastroenteritis is a common, generally self-limiting disease characterized by rapid onset of watery diarrhea and associated symptoms such as nausea, vomiting, fever, and/or abdominal pain. In some cases, nausea and vomiting may be the predominant symptoms. It is estimated that this illness occurs at an overall rate of 1.2 to 1.5 episodes per person per year in the United States and at higher rates among children under the age of 3 years (1.3 to 2.3 episodes per child per year). Although gastroenteritis may be caused by many infectious agents, viruses are the most common cause in industrialized countries. Rotaviruses are the leading cause of severe diarrhea in childhood; noroviruses are a diverse group of viruses responsible for a large proportion of outbreaks of gastroenteritis in adults, and sapoviruses, astroviruses, and enteric adenoviruses are other viral agents of gastroenteritis.

INTRODUCTION

Gastrointestinal viruses were identified by electron microscopy in the early 1970s, and since then, much has

Copyright 2004, Elsevier (USA). All rights reserved.

been learned about their important role in causing epidemic and sporadic gastroenteritis. The noroviruses have been especially challenging to study, because they cannot be cultivated in cell culture, infected individuals may not shed large quantities of the virus, and there is no small animal model that can be infected. Molecular characterization of the viral genome has greatly improved diagnostic capability and epidemiological surveillance. Noroviruses are now considered to be a major cause of acute gastroenteritis globally. Although viral gastroenteritis is generally self-limiting, with a benign course, dehydration may be a complication requiring prompt medical attention, particularly in infants and the elderly.

VIRAL PATHOGENS

The viral pathogens that cause acute gastroenteritis represent four distinct viral families. Rotaviruses are members of the family Reoviridae and have a segmented, double-stranded RNA genome enclosed in a triple-layered capsid. Noroviruses and sapoviruses constitute two genera of the family Caliciviridae, with a positive-sense RNA genome inside a capsid composed of a single structural protein. Astroviruses, of the family Astroviridae, are also nonenveloped viruses with a positive-sense RNA genome, but their capsid is composed of several structural protein subunits. Enteric adenoviruses represent a unique subgroup of the family Adenoviridae and have a double-stranded DNA genome enclosed in a characteristic icosahedral capsid. Each of these viruses also can be distinguished by electron microscopy by their characteristic morphologic features.

Rotaviruses

Group A rotaviruses are the most important cause of severe dehydrating diarrhea in children. On a global scale, it is estimated that there are 125 million cases of rotavirus diarrhea, including 8 million cases of severe diarrhea and 600,000 deaths. In the United States, it is estimated that there are 1 million cases of rotavirus diarrhea annually, with 65,000 requiring hospitalization and less than 20 deaths.

Children between the ages of 3 months and 3 years are most susceptible to infection. Because immunity to reinfection is not complete, rotavirus disease can also occur in otherwise healthy adults, as well as in patients who are elderly or immunosuppressed. Infection in children living in temperate climates usually occurs in the cooler months. In the United States and Europe, a wave of illness has been observed to spread from the Southwest to Northeast, starting in the autumn. In tropical climates and among adults, there does not appear to be seasonal variation.

Infection is transmitted primarily through the fecal–oral route in a highly efficient manner that owes to the large amount of viruses shed in feces. Food- and waterborne transmission is not common, but has also been reported. The incubation period ranges from 1 to 3 days and is followed by the abrupt onset of fever, malaise, vomiting, and watery diarrhea. Duration of illness varies from 3 to 8 days. Severely immunocompromised children may have a protracted course. Dehydration, with associated mild elevation of blood urea nitrogen (BUN) and mild metabolic acidosis, may require treatment with oral rehydration solutions.

If specific diagnosis of rotavirus infection is needed, enzyme immunoassays, with >90% sensitivity and specificity, are commercially available. Other diagnostic methods, including immune electron microscopy, nucleic acid hybridization with or without reverse transcriptase and polymerase chain reaction (RT-PCR), RNA electrophoresis, and cell culture isolation, are not as widely available.

Caliciviruses

Norovirus

Norovirus is the new name for a genus of the family *Caliciviridae* that includes a genetically diverse, but related, group of nonenveloped, single-stranded RNA viruses that are a major cause of acute gastroenteritis in humans. This group of viruses had previously been called "Norwalk-like viruses" (NLVs), after the prototype Norwalk virus that was isolated from a 1968 outbreak in Norwalk, Ohio, or "small round structured viruses" (SRSVs), a name based on their ultrastructural appearance. It is estimated that 23 million people become ill with noroviruses each year in the United States.

In 1972, the Norwalk virus was identified by immune electron microscopy, providing the first proof that a virus was the cause of a syndrome that had been called "winter vomiting disease." It was not until the early 1990s that the viral genome was cloned and sequenced, and the virus could be classified as a calicivirus. Sequence information also provided a basis for diagnosis by RT-PCR. The discovery that expression of the capsid protein spontaneously yields recombinant virus-like particles (VLPs) further enhanced studies of virus structure and immunity.

Between the early 1970s and the early 1990s, human volunteer studies were the main research tool to study viral pathogenesis. Norwalk virus infection was associated with broadening and blunting of villi in the

jejunum, crypt cell hyperplasia, cytoplasmic vacuoles, and polymorphonuclear and mononuclear cell infiltrates in the lamina propria. Intestinal brush border enzyme activity was decreased. Delayed gastric emptying was also documented and may have contributed to the symptoms of nausea and vomiting. Both structural and functional abnormalities returned to normal after resolution of the infection.

Immunity to infection with noroviruses is not completely understood. Studies have shown that only about half the people who are exposed to noroviruses develop illness, although asymptomatic infection can be documented in an additional 30%. Short-term immunity (a few weeks to a few months) can be demonstrated in infected individuals who mount an immune response, but the level of serum antibody does not predict whether the individual will be protected from reinfection or illness when challenged with the same norovirus in the long term. Some studies find no evidence for long-term immunity; those individuals who become ill with the initial exposure to the virus paradoxically become ill with repeated challenges several months to years later. Recent studies have suggested the possible role of blood group antigens in viral pathogenesis and determining susceptibility to infection. Persons belonging to blood group O may be most susceptible to severe norovirus infections. Recombinant Norwalk VLPs also show promise in further investigations of viral immune response and vaccine development.

Recent epidemics of gastroenteritis on cruise ships have sharply focused attention on the noroviruses, but it is well known that outbreaks can occur in many different settings and can be associated with the ingestion of contaminated food or water (less common). Raw or inadequately cooked shellfish pose a higher than average risk because viruses can be concentrated in these filter feeders. Uncooked or previously cooked foods that require extensive preparation and handling by food workers have also been associated with outbreaks. In most outbreaks, primary attack rates are 50% or more and secondary attack rates are 30% or more. Transmission occurs primarily through the fecal–oral route, but transmission through the airborne route and by fomites may also contribute. It has been estimated that infection can occur with the ingestion of as few as 10 viral particles.

Typically, after an incubation period of 12 to 48 hours, patients will experience the acute onset of nausea, abdominal cramping, diarrhea, and vomiting, frequently associated with headache, body aches, and low-grade fever. The symptoms last for 12 to 60 hours, after which there is complete resolution of the illness, although viral shedding may persist for up to 2 more weeks. Children tend to have more vomiting and less diarrhea; the reverse is observed for most adults. Severe dehydration is rarely encountered, but elderly patients with severe chronic diseases should be monitored for this potential complication and treated as clinically indicated.

The most widely used assay at present for diagnosing norovirus infection is RT-PCR, which can detect 10^2 to 10^4 viral particles/ml of stool. Because of the high level of sensitivity of this assay, care must be taken to avoid contamination of samples. The genetic diversity of the noroviruses has made it difficult to design primer sets that are capable of detecting all strains of virus, but current assays based on sequences from conserved regions of the genome encoding the viral RNA-dependent RNA polymerase are capable of detecting multiple strains. Sequence analysis of successfully amplified segments can help to determine the genogroup or genetic cluster of a particular norovirus, trace outbreaks to their source vehicle(s), and establish links between outbreaks.

Sapovirus

Sapporo virus, a calicivirus with classic morphology, is the prototype of the newly named genus *Sapovirus*. The genome organization of this group of viruses is more like those of animal caliciviruses, such as rabbit hemorrhagic disease virus, than that of the noroviruses. *Sapovirus* infection has different epidemiologic features, compared to noroviruses. Sapoviruses cause acute gastroenteritis primarily in infants and young children, but infection involving elderly adults has also been reported. *Sapovirus* outbreaks have also been reported to occur in institutional settings, although not as frequently as noroviruses.

Seroprevalence studies indicate that nearly all children have acquired *Sapovirus* antibodies by the age of 12 years, and these antibodies seem to confer long-term resistance to reinfection. Infection is most likely transmitted through the fecal–oral route. Oysters and cold foods have also been implicated as vectors of transmission. Following an incubation period of 24 to 72 hours, patients develop vomiting and diarrhea acutely, with associated abdominal pain, fever, and respiratory symptoms. The symptoms, which resemble those of a mild rotavirus infection, last for 24 to 48 hours. The infection can be diagnosed by direct electron microscopy, given the large quantities of viruses shed in feces and their distinctive appearance. There is no commercially available enzyme immunoassay, and tests based on viral nucleic acid sequence are not as far along as for the noroviruses.

Astroviruses

Human astroviruses primarily cause acute gastroenteritis in children, but illness has been reported to occur among elderly patients residing in institutions and in immunocompromised individuals (HIV-infected individuals and bone marrow transplant recipients). Most adults who are otherwise healthy do not tend to become infected or to develop symptoms. Similar to the sapoviruses, astrovirus antibodies are acquired early in childhood, with antibody prevalence rates in the range of 70% by school age. Astrovirus antibodies are also detected in gamma globulin pools from the United States and Japan.

The medical importance of astrovirus infection has been demonstrated by epidemiological surveys that show astrovirus to be second only to rotavirus as a cause of viral gastroenteritis among children attending outpatient clinics in Thailand. An Australian study has shown that among children hospitalized for diarrhea, rotaviruses are the most common etiology (as expected), followed by astroviruses. A typical case of astrovirus gastroenteritis in an immunocompetent child consists of a mild, watery diarrhea, vomiting, fever, and abdominal pain, which begin after an incubation period of 1 to 4 days. The symptoms last 2 to 3 days. In some immunocompromised patients, protracted and severe illness may occur.

Normal individuals are protected from repeated astrovirus infections by serum antibodies and CD4+ T cells in the gut lamina propria that recognize astrovirus antigens in an human leukocyte antigen (HLA)-restricted manner. When CD4+ T cells are depleted by the use of chemotherapeutic agents, such as fludarabine, patients may develop a prolonged course of astrovirus infection. In such patients, treatment with immunoglobulin (intravenously and/or orally) has been attempted in a small number of cases and has shown variable success rates. This type of treatment requires further study in controlled clinical trials.

Because astroviruses are shed in large quantities during infection and because at least 10% of the viruses display a distinctive surface star, diagnosis can be made by direct electron microscopy. Because astroviruses have been erroneously identified, the caveat is that accurate identification of astroviruses may require an experienced microscopist; a monoclonal antibody that can detect all eight serotypes of the human astrovirus is available.

Enteric Adenoviruses

Enteric adenoviruses are members of the Adenoviridae, a large family of DNA viruses that are associated with a wide spectrum of illnesses. Enteric adenoviruses that consistently cause gastrointestinal illness in infants and young children have been assigned to subgroup F, serotypes 40 and 41. Nonenteric types of adenoviruses have been identified in the stools of HIV-infected patients with acute or chronic diarrhea, but it has not been clearly demonstrated that these types of adenoviruses have an etiologic role.

Enteric adenovirus infections have a worldwide distribution, primarily infecting children under the age of 2 years. They are, in general, a less common cause of acute gastroenteritis, compared to astroviruses or noroviruses. However, in one study from Guatemala, enteric adenovirus infection was found to be more common than rotavirus infection (14 vs. 4.7%, respectively). There is no seasonal variation associated with enteric adenovirus infection.

After an incubation period of about 7 days, patients develop watery diarrhea and vomiting (less commonly). Respiratory symptoms and low-grade fever may also develop during the course of illness. Infection can range from mild to severe, but the majority of cases are mild. Viral shedding in stool lasts for up to 2 weeks, substantially longer than for rotavirus. The clinical course may be complicated by prolonged lactose intoleration and malabsorption. Whether adenovirus infection has any relationship to the development of celiac disease has not been resolved. The preferred method to detect the presence of enteric adenovirus in stool is the enzyme immunoassay, which is based on monoclonal antibodies to each type of enteric adenovirus. The assay is highly sensitive and specific and is commercially available.

TREATMENT AND PREVENTION

There are no specific antiviral therapies at the present time. Treatment is aimed at rehydration and supportive care, including early refeeding. Nonspecific treatment such as bismuth subsalicylate may decrease the severity of abdominal cramping in norovirus infection. For patients who develop dehydration, oral rehydration therapy, as recommended by the World Health Organization, is usually successful. In rare cases of severe vomiting or shock, parenteral hydration may be required. Once rehydration is accomplished, resumption of a regular diet is encouraged, even though some degree of infection-associated lactase deficiency may be present.

Prevention of foodborne infections begins with good handwashing and safe practices for food preparation. Public education about the risks associated with eating certain foods and transmission of viruses is essential.

Improved methods to decontaminate food or for rapid detection of gastroenteric viruses in food should also be helpful. Given the magnitude of the problem, a major approach to prevention of rotavirus illness has focused on vaccine development. Although several live attenuated vaccine candidates have yielded promising results, no vaccine is currently available for widespread use.

See Also the Following Articles

Diarrhea • Diarrhea, Pediatric • Eosinophilic Gastroenteritis • Food Poisoning • Rotavirus

Further Reading

Avery, M. E., and Snyder, J. D. (1990). Oral therapy for acute diarrhea. The underused simple solution [see comments]. *N. Engl. J. Med.* **323**, 891.

Becker, K. M., Moe, C. L., Southwick, K. L., and MacCormack, J. N. (2000). Transmission of Norwalk virus during a football game. *N. Engl. J. Med.* **343**, 1223–1227.

Franco, M. A., and Greenberg, H. B. (2001). Challenges for rotavirus vaccines. *Virology* **281**, 153.

Grohmann, G. S., Glass, R. I., Pereira, H. G., *et al.* (1993). Enteric viruses and diarrhea in HIV-infected patients. Enteric Opportunistic Infections Working Group. *N. Engl. J. Med.* **329**, 14.

Hutson, A. M., Atmar, R. L., Graham, D. Y., and Estes, M. K. (2002). Norwalk virus infection and disease is associated with ABO histo-blood group type. *J. Infect. Dis.* **185**, 1335–1337.

Molberg, Ø., Nilsen, E. M., Sollid, L. M., *et al.* (1998). CD4+ T cells with specific reactivity against astrovirus isolated from normal human intestine. *Gastroenterology* **114**, 115.

Parashar, U. D., Quiroz, E. S., Mounts, A. W., Monroe, S. S., Fankhauser, R. L., Ando, T., Noel, J. S., Bulens, S. N., Beard, R. S., Li, J.-F., Bresee, J. S., and Glass, R. I. (2001). "Norwalk-like viruses:" Public health consequences and outbreak management. *MMWR Morb. Mortal. Wkly. Rep. CDC Recommend. Rep.* **50**(No. RR-9), 1–17.

Gastroenterostomy

MALCOLM K. ROBINSON
Harvard University

anastomosis A connection created by surgical means between two normally distinct organs or structures that allows the free flow of contents between the two structures.

duodenum The first portion of the small intestine, which lies between the stomach and the jejunum.

gastroduodenostomy Surgical creation of an artificial passage between the stomach and the duodenum. This is one of the two types of gastroenterostomy.

gastroenterostomy Surgical creation of an artificial passage between the stomach and the small intestine.

gastrojejunostomy Surgical creation of an artificial passage between the stomach and the jejunum. This is one of the two types of gastroenterostomy.

jejunum The second portion of the small intestine, which immediately follows the duodenum.

stomach The part of the gastrointestinal tract that lies between the esophagus and the duodenum; it serves as a reservoir for swallowed food and is where digestion of food is initiated before it is passed into the duodenum.

A gastroenterostomy is a surgical procedure in which an artificial passage or connection is created between the stomach and a portion of the small intestine. The term comes from the Greek for "gaster," meaning stomach, "enteron," meaning intestine, and "stomoun," meaning to provide with an opening. The connection, which is frequently referred to as an "anastomosis," is held together by stitches or surgical staples and allows the free flow of contents between the two structures.

TYPES OF GASTROENTEROSTOMIES

There are two types of gastroenterostomy, which are named according to the portion of small intestine to which the stomach is attached. These are gastroduodenostomy, in which the stomach is connected to the duodenum, and gastrojejunostomy, in which the

Copyright 2004, Elsevier (USA). All rights reserved.

Improved methods to decontaminate food or for rapid detection of gastroenteric viruses in food should also be helpful. Given the magnitude of the problem, a major approach to prevention of rotavirus illness has focused on vaccine development. Although several live attenuated vaccine candidates have yielded promising results, no vaccine is currently available for widespread use.

See Also the Following Articles

Diarrhea • Diarrhea, Pediatric • Eosinophilic Gastroenteritis • Food Poisoning • Rotavirus

Further Reading

Avery, M. E., and Snyder, J. D. (1990). Oral therapy for acute diarrhea. The underused simple solution [see comments]. *N. Engl. J. Med.* **323**, 891.

Becker, K. M., Moe, C. L., Southwick, K. L., and MacCormack, J. N. (2000). Transmission of Norwalk virus during a football game. *N. Engl. J. Med.* **343**, 1223–1227.

Franco, M. A., and Greenberg, H. B. (2001). Challenges for rotavirus vaccines. *Virology* **281**, 153.

Grohmann, G. S., Glass, R. I., Pereira, H. G., *et al.* (1993). Enteric viruses and diarrhea in HIV-infected patients. Enteric Opportunistic Infections Working Group. *N. Engl. J. Med.* **329**, 14.

Hutson, A. M., Atmar, R. L., Graham, D. Y., and Estes, M. K. (2002). Norwalk virus infection and disease is associated with ABO histo-blood group type. *J. Infect. Dis.* **185**, 1335–1337.

Molberg, Ø., Nilsen, E. M., Sollid, L. M., *et al.* (1998). CD4+ T cells with specific reactivity against astrovirus isolated from normal human intestine. *Gastroenterology* **114**, 115.

Parashar, U. D., Quiroz, E. S., Mounts, A. W., Monroe, S. S., Fankhauser, R. L., Ando, T., Noel, J. S., Bulens, S. N., Beard, R. S., Li, J.-F., Bresee, J. S., and Glass, R. I. (2001). "Norwalk-like viruses:" Public health consequences and outbreak management. *MMWR Morb. Mortal. Wkly. Rep. CDC Recommend. Rep.* **50**(No. RR-9), 1–17.

Gastroenterostomy

MALCOLM K. ROBINSON
Harvard University

anastomosis A connection created by surgical means between two normally distinct organs or structures that allows the free flow of contents between the two structures.

duodenum The first portion of the small intestine, which lies between the stomach and the jejunum.

gastroduodenostomy Surgical creation of an artificial passage between the stomach and the duodenum. This is one of the two types of gastroenterostomy.

gastroenterostomy Surgical creation of an artificial passage between the stomach and the small intestine.

gastrojejunostomy Surgical creation of an artificial passage between the stomach and the jejunum. This is one of the two types of gastroenterostomy.

jejunum The second portion of the small intestine, which immediately follows the duodenum.

stomach The part of the gastrointestinal tract that lies between the esophagus and the duodenum; it serves as a reservoir for swallowed food and is where digestion of food is initiated before it is passed into the duodenum.

A gastroenterostomy is a surgical procedure in which an artificial passage or connection is created between the stomach and a portion of the small intestine. The term comes from the Greek for "gaster," meaning stomach, "enteron," meaning intestine, and "stomoun," meaning to provide with an opening. The connection, which is frequently referred to as an "anastomosis," is held together by stitches or surgical staples and allows the free flow of contents between the two structures.

TYPES OF GASTROENTEROSTOMIES

There are two types of gastroenterostomy, which are named according to the portion of small intestine to which the stomach is attached. These are gastroduodenostomy, in which the stomach is connected to the duodenum, and gastrojejunostomy, in which the

Copyright 2004, Elsevier (USA). All rights reserved.

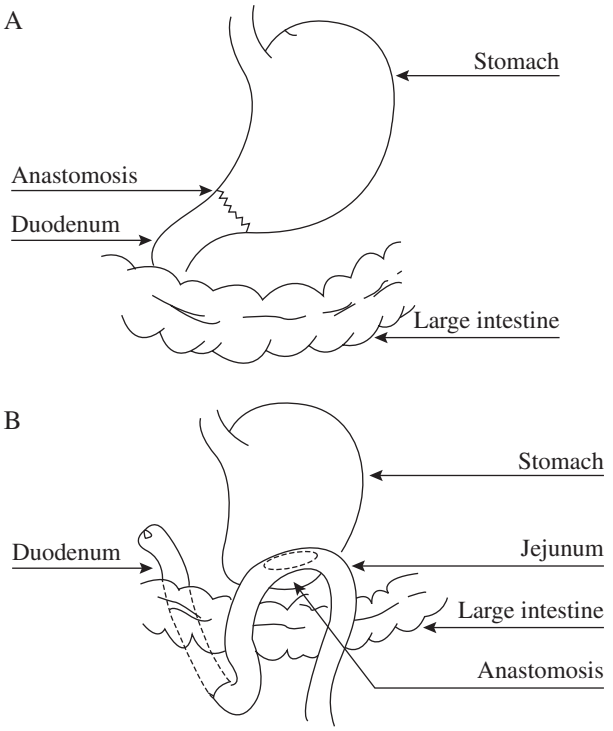


FIGURE 1 (A) Gastroduodenostomy. (B) Gastrojejunostomy.

stomach is connected to the jejunum. Gastrojejunostomy is the most frequently used type of gastroenterostomy (see Figs. 1 and 2).

SURGICAL INDICATIONS FOR GASTROENTEROSTOMY

There are two main indications or situations in which a surgeon may use a gastroenterostomy. The first reason is to reestablish continuity between the stomach and the small intestine following partial removal of the stomach. For example, a patient may have a large benign (i.e., non-cancer-containing) mass of the stomach near the duodenum. The surgeon may choose to remove the part of the stomach with the mass plus a piece of the duodenum to ensure that the entire mass has been removed. The surgeon must now reconnect the remaining part of the stomach to the small intestine so that food and juices can pass from the stomach to the intestine as before. This is accomplished using a gastroenterostomy. The surgeon may choose to connect the stomach to the remaining duodenum (i.e., a gastroduodenostomy; see Fig. 1A) or to the side of the jejunum (i.e., a gastrojejunostomy; see Fig. 1B).

A gastroenterostomy may also be used to circumvent an area of the stomach that is blocked. This may occur for a variety of reasons. For example, a patient may have a cancer that blocks the passage of food from the stomach into the duodenum. If the cancer has grown too large to be removed safely, one option is for a surgeon to connect the stomach to the side of the jejunum (i.e., a gastrojejunostomy) without the necessity of removing the mass or any part of the intestinal tract. This allows food to travel directly from the stomach to the jejunum, bypassing the blockage in the duodenum (see Fig. 2).

There are a large number of techniques for constructing gastroenterostomies and the versatility of this procedure makes it useful in a wide variety of circumstances.

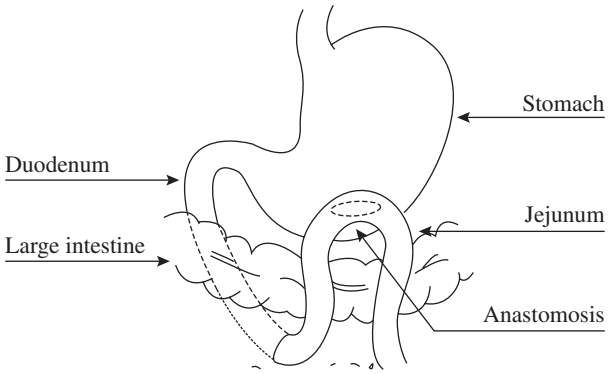


FIGURE 2 Gastrojejunostomy.

See Also the Following Articles

Gastric Surgery • Gastrostomy

Further Reading

Chassin, J. L. (1994). Gastrojejunostomy. *In* "Operative Strategy in General Surgery: An Expositive Atlas" (J. L. Chasin, ed.), 2nd Ed., pp. 194–200. Springer-Verlag, New York.

Gastroesophageal Reflux Disease (GERD)

J. PATRICK WARING
Digestive Healthcare of Georgia, Atlanta

Barrett's esophagus A condition that develops when the normal squamous tissue in the esophagus is repopulated by intestinal epithelium.
fundoplication Surgical procedure in which the fundus of the stomach is mobilized and wrapped around the distal esophagus.
gastroesophageal reflux disease Mucosal injury or persistent symptoms that develop in response to reflux of gastric contents into the esophagus.
proton pump inhibitors Medications that block the production of gastric acid secretion by inhibiting the H^+,K^+-ATPase pump.

Gastroesophageal reflux disease (GERD) is a very common problem. As many as 7% of adult Americans have daily heartburn. Patients with GERD may present with a variety of symptoms, ranging from typical symptoms, such as heartburn or regurgitation, to atypical symptoms, such as chest pain, asthma, laryngitis, or chronic cough. GERD may cause complications, such as esophagitis, esophageal ulceration, esophageal strictures, or Barrett's esophagus. Most patients with GERD will have a chronic problem requiring long-term therapy. Seventy-five percent of patients with chronic heartburn have moderate to severe symptoms. Many patients with GERD have an impaired quality of life, as assessed by the Gastrointestinal Symptom Rating Scale, the Health-Related Quality of Life scale, and the Psychological General Well-Being scale. Interestingly, there is little correlation between quality of life scoring and severity of esophagitis. Even patients without esophagitis have a markedly impaired quality of life. Quality of life scores improve to the normal range with successful medical or surgical therapy. The goals of modern medical therapy for GER are threefold: first, eliminate symptoms; second, heal injured esophageal mucosa; and third, manage and/or prevent complications.

PATHOPHYSIOLOGY

Gastroesophageal reflux disease (GERD) occurs as a consequence of gastric contents, predominantly acid, refluxing into the distal esophagus. The complicated pathophysiology includes competence of the lower esophageal sphincter (LES), the ability to clear refluxed gastric acid from the esophagus, esophageal mucosal defenses, and abnormalities of gastric function. The most important aspect of the pathophysiology of GERD is the competency of the anti-reflux barrier. Factors contributing to its integrity include the LES pressure, the presence or absence of a hiatal hernia, and the occurrence of transient lower esophageal sphincter relaxations. LES pressure less than 6 mm Hg, LES length less than 2 cm, and intra-abdominal length less than 1 cm are strongly associated with GERD. The crural diaphragm contributes substantially to the integrity of the LES. When patients have a hiatal hernia, the crural diaphragm and the LES are displaced from each other and consequently the sphincter is less competent. The most common pathophysiologic event at the time of a reflux episode is a transient lower esophageal sphincter relaxation. Poor esophageal clearance may lead to prolongation of the acid reflux events. Abnormalities of gastric function, such as hyperacidity, delayed gastric emptying, or duodenogastric reflux, may contribute to the severity of GERD by making the refluxate more noxious. It is important to remember that severe discomfort from acid reflux does not require severe injury to the esophageal mucosa. Acid exposure in the distal esophagus causes destruction of the tight junctions between the epithelial cells. This leads to penetration of the acid deeper into the epithelial lining. Sensory neurons are only two to three layers deep and when exposed to the acid may create a painful stimulus that may last for several hours. These histopathologic findings are found on electron microscopy in patients with and without esophagitis. Attempts to withdraw gastroesophageal reflux (GER) therapy are frequently unsuccessful because medical therapy offers no permanent resolution of pathophysiologic abnormalities.

SYMPTOMS

Most physicians are well acquainted with the typical manifestations of gastroesophageal reflux disease, such as heartburn and regurgitation. Heartburn is

 Copyright 2004, Elsevier (USA). All rights reserved.

defined as a burning retrosternal discomfort. Regurgitation is the effortless movement of food or fluid from the stomach to the esophagus. At least 10% of patients with GERD have what are known as extra-esophageal manifestations. The more common complaints include chest pain, laryngitis, chronic cough, and asthma. Other complaints that are postulated to be associated with GERD, although not as well proven, include laryngeal cancer, otitis media, globus sensation, vocal cord granulomas, halitosis, and dental erosions. Interestingly, these extra-esophageal manifestations may be present with or without typical reflux symptoms. As such, management of the extra-esophageal manifestations of GERD can be rather challenging.

As many as 30% of patients undergoing cardiac evaluation for chest pain have no cardiac lesions. Unexplained chest pain has been linked to microvascular angina and esophageal, pulmonary, musculoskeletal, and psychological causes. For years, it was thought that esophageal spasm or another esophageal motility disorder was the most common cause of unexplained chest pain. However, evidence of gastroesophageal reflux is seen in 50–70% of patients with unexplained chest pain. Esophageal pain from GERD may or may not respond to antacids or nitroglycerin. Neither a rapid response nor the lack of a response to symptomatic therapy is a reliable indicator of the cause of the chest pain. Patients with GERD-related chest pain frequently have typical reflux symptoms as well. It is important to have the patient describe exactly what he or she is feeling. Some patients use the words heartburn, indigestion, and chest pain interchangeably. Most patients are able to distinguish between the burning sensation and pressure sensations very nicely. Others may be unable to tell whether they have one or two painful sensations. Patients should be asked about heartburn, a separate burning sensation in the midchest. Some patients may experience this for years prior to developing the chest pain. Many patients do not have heartburn and chest pain at the same time. This does not exclude gastroesophageal reflux as the cause of the chest discomfort.

GERD and asthma commonly coexist. GERD may worsen asthma either by direct acid injury to the lungs or by a vagally mediated reflex bronchospasm. Asthma may worsen GERD by altering dynamics at the LES or promoting reflux episodes by coughing and wheezing. Efforts to identify asthmatics that will respond to antireflux therapy may be very difficult. Since many intrinsic and extrinsic factors can exacerbate bronchospasm, the cause and effect relationship between GERD and asthma is not obvious. Furthermore, GERD and asthma may occur in the same patient with no apparent association.

Acid and peptic injury to laryngeal strictures can lead to persistent symptoms of throat discomfort, throat clearing, hoarseness, voice loss, and chronic cough. The pathophysiologic abnormality in this situation may actually be at the upper esophageal sphincter. Gastric acid must travel all the way up the esophagus, past the upper esophageal sphincter, and onto the vocal cords. The majority of these patients actually have mild reflux, but even small amounts of acidic gastric juice may be quite injurious to the vocal cords. More controversial symptoms that may be reflux related are chronic sore throat and neck pain. True dysphagia is rare, but an uncomfortable sensation during swallowing is not. This should be distinguished from the globus sensation, the feeling that something is stuck in the throat, when there is obviously nothing there. The most common laryngeal finding in these patients is posterior laryngitis; however, it is not specific. Other causes of laryngeal irritation, such as voice abuse, heavy smoking, endotracheal intubation, and exposure to toxic fumes, should be considered. Approximately half of patients with posterior laryngitis will not have typical symptoms of heartburn to suggest the presence of GERD. Studies from the late 1980s suggested that laryngeal cancer occurs in patients with reflux disease in the absence of smoking or alcohol. However, there have been only limited data published since then. At this point, it does not seem likely that reflux predisposes patients to laryngeal carcinoma.

GERD is the third most common cause of chronic unexplained cough following postnasal drip and asthma, occurring in 10 to 21% of patients. Typical symptoms, such as heartburn, are frequently absent. Ambulatory pH monitoring may be the only abnormal test. It should be considered in patients refractory to antihistamine–decongestant therapy, in patients without typical reflux symptoms, and in patients who fail an empiric trial of anti-reflux therapy.

COMPLICATIONS

Barrett's esophagus is a premalignant condition in which specialized intestinal metaplasia replaces the normal squamous esophageal mucosa. This occurs almost always in response to chronic gastroesophageal reflux. Barrett's esophagus is the only known risk factor for esophageal adenocarcinoma. The reported incidence of esophageal adenocarcinoma for these patients has ranged from 0.2% to as high as 2.1% per year. Over the past two decades, the incidence of adenocarcinoma of the esophagus and gastric cardia has increased at a rate exceeding that of any other cancer in several developed countries, including the United States. Patients

with Barrett's esophagus are 30 to 40 times more likely to develop adenocarcinoma of the esophagus than the general population. Nearly all studies on Barrett's esophagus show a strong correlation between that condition and the presence of chronic reflux symptoms for greater than 5 years. The incidence of Barrett's esophagus in patients with chronic GERD symptoms is thought to be approximately 10–12%, although several studies suggest rates near 20%, particularly if male Caucasians with chronic reflux symptoms are targeted.

Patients with chronic GERD symptoms, such as heartburn or regurgitation, should have an endoscopy to look for evidence of columnar epithelium. Identifying patients with Barrett's esophagus is the major role of endoscopy in GERD. Patients with columnar epithelium should have a biopsy to look for evidence of specialized intestinal metaplasia. These patients are at greatest risk for the development of esophageal adenocarcinoma. Patients with Barrett's esophagus should be enrolled in a surveillance program to look for signs of dysplasia.

GERD is the most common cause of a benign esophageal stricture. Historically, nearly 10% of patients with GERD seeking medical attention have esophageal strictures, although this number is probably much smaller now. This condition is related to chronic fibrosis and scarring in response to prolonged esophageal acid exposure. Patients often complain of heartburn and regurgitation, as well as dysphagia. Reflux strictures are almost always located at the gastroesophageal junction. Rarely, the strictures may be long and have a ringed appearance. Strictures in patients with Barrett's esophagus usually occur at the squamocolumnar junction. Pill-induced esophageal injury may be confused with GERD. Up to 20% of patients with suspected reflux strictures may actually have pill-induced strictures. All pills have the potential to produce esophageal injury, but tetracycline, doxycycline, potassium chloride, quinidine, nonsteroidal anti-inflammatory drugs (NSAIDs), and alendronate are the most recognized for causing esophageal strictures. Medications may also be a cofactor with GERD, leading to esophageal stricture formation. Patients who have GERD and a stricture are far more likely to be taking aspirin or NSAIDs than patients who have GERD but no stricture.

DIAGNOSTIC TESTING

There is no ideal diagnostic test for gastroesophageal reflux disease. Most patients with GERD will have a normal endoscopy, esophageal manometry, or upper gastrointestinal X ray. History alone is good, but not perfect, with an accuracy of approximately 80%. Other problems, such as cardiac, biliary, gastric, or psychological abnormalities, may cause symptoms that are confused with chronic heartburn. Many physicians are offering patients an empiric trial of therapy prior to diagnostic testing if the history is consistent with GERD. In fact, a response to empiric therapy could be considered a positive diagnostic test. In one study, the sensitivity of a 7-day course of omeprazole 40 mg daily was 27%, but increased to 83% with a 7-day course of omeprazole 40 mg bid. An empiric trial of therapy may reduce diagnostic testing costs by 53–65% compared to endoscopy or pH monitoring. Indeed, recent guidelines have advocated empiric therapy as the initial approach for patients with chronic heartburn symptoms.

The role of endoscopy in patients with GERD has changed over the past few years. There appears to be little difference in symptom severity or treatment outcomes between patients with esophagitis and those without esophagitis. Consequently, the value of early endoscopy has been questioned. The degree of esophagitis no longer appears to be helpful in guiding therapy. Endoscopy should be considered in two situations. In patients with bothersome symptoms, such as weight loss, intestinal bleeding, nausea, vomiting, dysphagia, or odynophagia, an endoscopy should be performed to rule out more serious causes of these problems. Endoscopy is also indicated for patients with long-standing reflux symptoms to exclude Barrett's esophagus. Current guidelines recommend an endoscopy in anyone who has had heartburn at least twice a week for at least 5 years.

For many patients, ambulatory esophageal pH monitoring of the esophagus is the only test that will be abnormal. Here, a catheter with an antimony pH electrode is passed nasally into the distal esophagus, 5 cm above the lower esophageal sphincter. The pH can be recorded every few seconds over a prolonged period of time. Some investigators have demonstrated that an accurate reading may be obtained in as little as 3 h; however, the standard length of time remains 24 h. A reflux episode is defined as beginning when the esophageal pH drops below 4.0. Patients at the author's clinic are asked to eat normally and be active during the study period. There have been several "normal" values published, but in general, a study is considered abnormal if the pH in the distal esophagus is <4 at least 4.2% of the time. It is important to review the pH tracing to exclude artifacts. This may include drifting of the pH value to acidic levels, rather than an abrupt drop of a pH value that occurs with a reflux episode. Additionally, lead-positioning problems may create a deceptively abnormal study; for example, when the pH value is always in the range of 1–2, the pH probe is likely in the stomach.

The pH value may abruptly become 0 or 9 if there are reference electrode problems. Most computerized programs allow the physician to exclude these artifacts and obtain an accurate study.

Ambulatory esophageal pH monitoring should be performed in two major circumstances. The first is when the physician does not know whether the patient has GERD, but must determine this. For example, for patients with non-erosive or atypical reflux who are considering anti-reflux surgery, it is imperative to document an abnormal amount of acid exposure in the distal esophagus prior to anti-reflux surgery. In general, patients with erosive esophagitis considering anti-reflux surgery do not need a pH study. Second, it may be useful to follow a patient's response to therapy. Many patients with atypical reflux symptoms, such as chest pain, asthma, laryngitis, or chronic cough, will have an incomplete response to treatment. Documenting whether or not there is adequate acid suppression with medical therapy may help determine whether the refractory symptoms are related to continued acid reflux.

There are a variety of ways to alter this study to obtain more information. It can be performed on or off medical therapy. The pH study should be performed "off-therapy" simply to document whether or not acid reflux is present, for example, in a patient considering anti-reflux surgery. Occasionally, a patient will present with such unusual symptoms that they are probably not related to acid reflux. A normal pH test off-therapy would be helpful to exclude reflux as a possibility. The study should be performed "on-therapy" in patients with an incomplete response to treatment. It is reasonable to offer patients double-dose proton pump inhibitor (PPI) therapy for a couple of months before investigating with a pH study. This is more likely to occur in patients with atypical GERD symptoms. This may help determine whether the refractory symptoms are related to continued acid reflux or to some other medical problem. Placing the pH probe 1 cm above the upper esophageal sphincter is an intriguing way to look for problems in patients with laryngeal symptoms. Several papers suggest that this is far more sensitive for identifying reflux into the hypopharynx than traditional ambulatory esophageal pH monitoring. However, there are no established normal values. Gastric pH monitoring may be useful for identifying patients who will respond to advancing treatment to very high doses of PPIs or anti-reflux surgery. Many patients with refractory GERD symptoms simply do not demonstrate a suppression of gastric acid with PPI therapy. An attachable pH probe has just recently been developed. Preliminary data suggest that it is well tolerated and at least as accurate as traditional pH monitoring. This technology could cause a revision in the role of ambulatory esophageal pH monitoring. If a safe, simple, accurate, and well-tolerated pH test becomes available, clinicians may be inclined to use it earlier in the course of the patient's illness.

One theoretical advantage of ambulatory pH monitoring is the ability to correlate symptoms with reflux episodes. Several authors have created scoring systems, such as the symptom index, the symptom sensitivity index, or the symptom association probability. The last system may be the most accurate. It calculates the probability of a reflux event being associated with a given symptom by creating a 2×2 table comparing 2 min intervals with and without symptoms versus reflux episodes. The number of reflux-associated symptomatic periods should be greater than that expected by chance. Unfortunately, there are no clinical trials to prove that any of the symptom association scores actually predict a true cause and effect relationship. As with most aspects of ambulatory pH monitoring, symptom association often requires a good deal of clinical judgment.

The barium swallow is of limited benefit in patients with GERD, because of its poor sensitivity and specificity. It is very useful in the evaluation of dysphagia, but in general is not helpful in patients with GERD.

Provocative tests, such as the acid perfusion test or tensilon challenge, were popular in the 1980s for the evaluation of unexplained chest pain, although they are of limited usefulness now that ambulatory pH monitoring is available.

MEDICAL THERAPY

The selection of a particular medical regimen depends on the severity of disease, the effectiveness of the therapy, and the cost and convenience of the medical regimen. Historically, physicians have used the step-up approach, which involves using the least effective regimen and intensifying therapy only if necessary. However, in patients with frequent or severe symptoms or those with signs of GERD complications, beginning the regimen with intensive PPI therapy is a reasonable choice. Unlike healing of peptic ulcers, healing of esophagitis requires more aggressive acid suppression. Patients with atypical symptoms of GERD also may require more aggressive therapy.

Five PPIs, omeprazole, lansoprazole, rabeprazole, pantoprazole, and esomeprazole, are currently available in the United States. These medications profoundly inhibit the secretion of acid into the gastric lumen, but rarely eliminate gastric acid secretion. All five PPIs have comparable ability to control symptoms and

heal esophagitis. Their effectiveness is dose-dependent. Most studies have shown that the PPIs heal esophagitis in 85–96% of patients. PPIs are far superior to all other forms of medical therapy. Long-term safety and effectiveness have been demonstrated.

PPI therapy improves heartburn symptoms in 60–80% of patients without esophagitis. Again, PPIs are vastly superior to other medical therapies for control of heartburn symptoms for patients with and without esophagitis. PPI therapy can normalize Health-Related Quality of Life, Gastrointestinal Symptom Rating Scale, and the Psychological General Well-Being scores.

Four histamine-2 receptor antagonists (H2RAs), cimetidine, ranitidine, famotidine, and nizatidine, are available in the United States in prescription and over-the-counter formulations. H2RAs inhibit gastric acid secretion to a lesser extent than PPIs. H2RAs provide successful symptomatic relief in 60–70% of patients who are treated for 6–8 weeks. Only 25–45% of patients with esophagitis will remain free of mucosal injury after 1 year of H2RAs, perhaps because of developing tolerance to their effects. One study found that only 41% of patients had improvement in heartburn symptoms after 8 weeks with ranitidine 150 mg bid. Patients were then randomized to continue the standard dose (150 mg) bid or to use a higher dose (300 mg bid). Interestingly, complete heartburn resolution was observed in less than 20% of the patients after an additional 8 weeks in either group. This finding suggests that standard-dose H2RAs are moderately successful at improving heartburn symptoms, but the common practice of doubling the dose of ranitidine is not particularly helpful. With the exception of patients with mild disease, standard doses of H2RAs probably have a limited role in the management of GER complications.

Metoclopramide, a prokinetic agent, may improve symptoms of GER but does not heal esophagitis. It also does not improve esophageal peristalsis or acid clearance. Furthermore, it also has frequent side effects. Metoclopramide has virtually no role as the sole therapy in the management of GER. Cisapride and domperidone are not available in the United States. Antacids are very useful in symptomatic relief in patients with mild intermittent heartburn. They are no better than a placebo for the healing of esophagitis. They may be used with other medical therapy as an adjunct for breakthrough GER symptoms. There are conflicting data regarding the effectiveness of sucralfate therapy in GER. Considering that there are other more superior and more convenient treatment options, sucralfate has a limited role, if any, in the modern medical therapy of GER.

Lifestyle modifications may help patients with mild or intermittent symptoms of GERD. However, lifestyle modifications alone will be effective in less than 30% of patients with esophagitis. Consequently, it is not appropriate as the sole initial therapy for patients with frequent or severe symptoms of GERD. Diet alteration is perhaps the most useful of the lifestyle modifications. A variety of foods, such as chocolate, peppermint, raw onions, and fatty foods, decrease LES pressure. Certain liquids, such as coffee, colas, and beer, increase gastric acid secretion. The effect of coffee is greater in patients with LES dysfunction. Coffee and citrus juices are directly irritating to the esophageal mucosa. Patients are advised to avoid eating 2–3 h before bedtime, to elevate the head of the bed 6–10 in. during sleep, and to stop smoking cigarettes.

Weight loss is commonly associated with improvement in symptoms of GER, although there are no control studies regarding this observation. In fact, obesity has not been shown to be associated with LES or esophageal acid exposure. It is possible that the observed improvement with weight loss may be due to the dietary modifications implemented to produce the weight loss. Medications, such as theophylline and calcium channel blockers, can inhibit the LES pressure. Theophylline worsens esophageal acid exposure in asthmatics with GERD. If possible, these medications should be stopped or another medication should be prescribed.

It seems easy to lump the extra-esophageal symptoms into one category, but these symptoms can be quite different. Patients with chest pain usually have esophageal symptoms such as heartburn, regurgitation, or dysphagia. They tend to respond over a 2- to 4-week period with PPI therapy. Patients who fail PPI therapy often improve with an empiric trial of a low-dose tricyclic antidepressant. Asthmatics also frequently have other esophageal symptoms. It may be more difficult to determine whether gastroesophageal reflux is the cause of these symptoms, because asthma often improves or worsens indepent of GERD. Three to 4 months of therapy is probably indicated before deciding whether or not the asthma is truly reflux-related. Patients with reflux-related chronic cough usually respond promptly; however, profound acid suppression is often required. Patients with reflux-related laryngitis often require therapy with high-dose PPIs for 2 months or more before they experience improvement.

Intensive medical therapy is a very important adjunct to dilation in the management of reflux strictures, especially in those who also have erosive esophagitis. Two studies have shown that patients who received omeprazole have greater relief of dysphagia and required fewer re-dilations than patients who received ranitidine. Successful anti-reflux surgery decreases the need for further dilation as well. Patients

who already present with dysphagia are at greater risk for pill-induced injury by worsening an existing esophageal stricture.

SURGICAL MANAGEMENT

Since its introduction approximately 10 years ago, laparoscopic surgery for gastroesophageal reflux disease has become one of the more common operations performed in the United States. It is interesting that there is still no consensus among physicians as to the proper role of this procedure in the management of GERD. In most published series, the average patient having anti-reflux surgery has nearly complete relief of symptoms and a markedly improved quality of life. In fact, most comparative studies demonstrate the superiority of surgical therapy compared to medical therapy. However, individual complications have prompted some gastroenterologists to be cautious about recommending anti-reflux surgery.

The classic operation is a Nissen fundoplication. Here the fundus is fully mobilized; specifically; the short gastric blood vessels are disconnected or "taken down" from the fundus. A 1.0 to 1.5 cm plication is made using the fundus to wrap around the distal esophagus. Three sutures are usually placed, securing the wrap to the stomach and the esophagus. The hiatal hernia is reduced and the diaphragmatic crurae are closed. Many surgeons will sew the fundoplication to the diaphragm to prevent it from moving postoperatively. This is often performed over a large dilator to prevent the wrap from being too tight. However, there have been reports of perforations, which have led some experienced surgeons to abandon this part of the procedure.

The success rate with anti-reflux surgery, defined by Visick I or II, is over 90%. In one study, 100 patients were followed for an average of 21 months. Ninety-six percent of patients had a good outcome and 71% were completely free of symptoms. Only 3 patients were back on medical therapy. Two patients experienced worsening of symptoms after the procedure. Dysphagia was present in only 2 patients up to a year after surgery. Clinically significant complications occurred in 4 patients. The study was limited to patients with positive 24 h pH studies and "typical" symptoms of GERD. This study is noteworthy because these patients are ideal candidates for anti-reflux surgery. Additionally, the surgeons in this study were highly skilled and very experienced. Consequently, these are probably the best results that one could hope for.

The opinions of the author regarding the long-term durability of laparoscopic anti-reflux surgery come from a handful of studies with at least 5 years of follow-up. Long-term data from centers of excellence suggest a 96% satisfaction rate between 5 and 8 years after surgery. However, 6% of patients have symptoms of chest pain, 6% have heartburn, 6% have regurgitation, 20% have abdominal bloating, 12% have diarrhea, 27% have some difficulty swallowing, and 7% require an esophageal dilation. The same results may not apply to less experienced centers. A recently published study from the Veterans Administration (VA) showed that 62% of patients having surgery required medications for heartburn after 10–13 years. Another study showed only a 57% satisfaction rate, with 67% of the patients developing new symptoms, such as difficulty swallowing or bloating. Additionally, 6.7% of the patients needed a repeat operation rate in less than 1 year.

There are two large published studies comparing medical and surgical therapy: the VA Cooperative Study and the Nordic study. The VA study enrolled patients with complicated GERD in the mid-1980s. It showed that surgical therapy was markedly superior to medical therapy of the era, H2-receptor blockers and metoclopramide. The Nordic trial enrolled 310 patients with erosive esophagitis; 155 patients were randomized to continuous omeprazole therapy and 155 patients were randomized to open anti-reflux surgery. At both 3 and 5 years of follow-up, there were far more treatment failures in patients who were randomized to omeprazole treatment. However, the protocol also allowed dose adjustment to either 40 or 60 mg daily in case of symptom recurrence in patients randomized to omeprazole therapy. If this fact is considered, the failure rates still remained in favor of surgery, although the difference was not quite statistically significant. Quality of life assessment revealed values within normal ranges in both therapy arms during the 5 years of follow-up.

Numerous studies show that medical or surgical therapy improves but does not always relieve the symptoms of chest pain, asthma, chronic cough, and hoarseness. In one study, heartburn was relieved by fundoplication in 93% of patients compared to only 56% of patients with atypical symptoms. Success with medical therapy was the only factor that predicted a successful surgical outcome. The success of anti-reflux surgery will parallel that of intensive medical treatment with high-dose PPIs. The only possible exception is in asthmatics. There are no comparative trials, but several studies on anti-reflux surgery in asthmatics are pretty impressive. Seven of nine patients who required daily oral corticosteroids for asthma were able to discontinue this treatment entirely after successful anti-reflux surgery. In patients with a partial response to medical therapy, repeated ambulatory pH monitoring may be of benefit in separating the true medical failures from

those patients with refractory atypical symptoms due to other causes.

The approach to postfundoplication symptoms is not difficult. Postoperative symptoms can be divided into two categories: recurrent or new. Recurrent symptoms after surgery suggest that the surgery simply did not work. New postoperative symptoms suggest a complication at the time of surgery. On the other hand, the surgical management of these problems is very complex. The success rates for repeat operations are in the range of 75–85%, with a morbidity rate in the range of 20–30% and a mortality rate of over 1%. Conservative management, if successful, is always preferable to a repeat operation. The decision to redo a fundoplication should not be taken lightly and the surgery should probably be performed at a center with vast experience in these matters.

There are two reasons for a patient to have the same symptoms postoperatively: either the surgery was ineffective or the original diagnosis was incorrect. If the surgery simply did not work, the postoperative symptoms would be the same. Alternatively, there is no reason to believe that a patient will improve if the preoperative symptoms were not related to reflux in the first place. An incorrect original diagnosis is much more likely in patients who had no response to medical therapy preoperatively, especially those with atypical reflux symptoms. The initial approach to a patient with the same postoperative symptoms would be an empiric trial with a PPI for 1 or 2 months and this is frequently helpful. If the patient cannot stop the PPI without relapse of symptoms, it may be reasonable to simply continue the medical therapy. One may consider a surgical revision if chronic PPI therapy is required. However, given the high morbidity and mortality associated with second operations, this course of action should be discouraged. If there is no improvement with this empiric therapy, further evaluation will be needed. These patients should have assessment of the fundoplication with a barium swallow and/or an endoscopy. Certainly, if a second surgery is contemplated, a pH study must be performed to document the presence of reflux, especially if the patient has atypical GERD or if the fundoplication appears satisfactory.

The major new postoperative symptom is dysphagia. Dysphagia in the early postoperative period is very common. In fact, most patients are on soft diets for the first few weeks. Patients who are unable to drink liquids after 2 weeks or who are unable to eat solid foods after 6 weeks should be evaluated. The most important predictive factor in the management of postoperative dysphagia is the integrity of the fundoplication. A barium swallow and an endoscopy are complementary in the assessment of the integrity of the wrap. A barium swallow is better for examining the length of the fundoplication, whereas an endoscopy is better for assessing the location and orientation of the fundoplication. If the fundoplication is short (1–2 cm), parallel to the diaphragm, and at the top of the stomach, dilation therapy followed by watchful waiting will help the great majority of patients. The use of large dilators, i.e., 46 to 54 French, is very safe. There are conflicting reports on the use of the 3.0 cm pneumatic balloons. If the fundoplication is not intact, if it is too long, if it is twisted, or if gastric folds are seen above or alongside the wrap, surgical revision is usually necessary.

Surgical therapy is associated with serious, albeit infrequent complications. There is little doubt that successful anti-reflux surgery corrects the underlying pathophysiology, improves symptoms, normalizes quality of life scores, reduces health care utilization, and appears as good as or better than medical therapy in large comparative trials. Given the safety of PPI therapy, one must wonder if even the small risk of surgery is worthwhile in patients with uncomplicated GERD that responds to medical therapy. The decision to have anti-reflux surgery must be individualized. All patients undergoing long-term medical therapy for GERD should receive advice on the safety and wisdom of staying on that therapy, as well as information on anti-reflux surgery. Fundoplication should be considered in three circumstances: (1) It should be considered for individuals who are intolerant of proton pump inhibitor therapy because of side effects. However, this circumstance will be less common now that there are five PPIs available. (2) Fundoplication may be warranted for patients who are poorly responsive to proton pump inhibitor therapy. This outcome is probably not that common, given the effectiveness of the currently available PPIs. Such an outcome will be more common in patients with atypical GERD. The gastroenterologist should be as certain as possible that the patient not only has GERD, but that the patient's symptoms are reflux related. (3) Fundoplication may also be considered when a patient desires a permanent solution in order to be free of the need to take medications. These patients must be warned about potential suboptimal results, including the frequent need for medication within a few years of having the procedure, as well as the small, but real possibility of becoming worse after the operation. Even in experienced hands, 1–2% of patients are worse after the procedure.

Gastroesophageal reflux is a common, chronic problem that has great influence on the quality of life of patients. Excellent medical and surgical therapies are available. Diagnostic testing is largely reserved for

excluding complications such as Barrett's esophagus and evaluating those few patients who do poorly with therapy.

See Also the Following Articles

Antacids • Barrett's Esophagus • Chest Pain, Non-Cardiac • Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD • Esophageal Strictures • Esophageal Surgery • Gastric Acid Secretion • H2-Receptor Antagonists • Proton Pump Inhibitors • Sphincters

Further Reading

Achem, S. R., Kolts, B. E., MacMath, T., Richter, J., Mohr, D., Burton, L., and Castell, D. O. (1997). Effects of omeprazole versus placebo in treatment of noncardiac chest pain and gastroesophageal reflux. *Digest. Dis. Sci.* **42**, 2138–2145.

Bammer, T. Hinder, R. A., Klaus, A., and Klingler, P. J. (2001). Five- to eight-year outcome of the first laparoscopic Nissen fundoplications. *J. Gastrointest. Surg.* **5**, 42–48.

Carlson, M. A., and Frantzides, C. T. (2001). Complications and results of primary minimally invasive antireflux procedures: A review of 10,735 reported cases. *J. Am. Coll. Surg.* **193**, 428–439.

DeVault, K. R., and Castell, D. O. (1999). Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am. J. Gastroenterol.* **94**, 1434–1442.

Harding, S. M., Richter, J. E., Guzzo, M. R., *et al.* (1996). Asthma and gastroesophageal reflux: Acid suppression therapy improves asthma outcome. *Am. J. Med.* **100**, 395–405.

Hunter, J. G., Trus, T. L., Branum, G. D., Waring, J. P., and Wood, W. C. (1996). A physiologic approach to laparoscopic fundoplication for gastroesophageal reflux disease. *Ann. Surg.* **223**, 673–685.

Kahrilas, P. J., Fennerty, M. B., and Joelsson, B (1999). High- versus standard-dose ranitidine for control of heartburn in poorly responsive acid reflux disease: A prospective, controlled trial. *Am. J. Gastroenterol.* **94**, 92–97.

Lundell, L., Miettinen, P., Myrvold, H. E. *et al.*, (2001). Continued (5-year) follow-up of a randomized clinical study comparing antireflux surgery and omeprazole in gastroesophageal reflux disease. *J. Am. Coll. Surg.* **192**, 172–179.

Ours, T. M., Kavuru, M. S., Schilz, R. J., and Richter, J. E. (1999). A prospective evaluation of esophageal testing and a double-blind, randomized study of omeprazole in a diagnostic and therapeutic algorithm for chronic cough. *Am. J. Gastroenterol.* **94**, 3131–3138.

Peters, J. H., DeMeester, T. R., Crookes, P., *et al.* (1998). The treatment of gastroesophageal reflux disease with laparoscopic Nissen fundoplication: Prospective evaluation of 100 patients with "typical" symptoms. *Ann. Surg.* **228**, 40–50.

Richter, J. E., Kahrilas, P. J., Johanson, J., *et al.* (2001). Efficacy and safety of esomeprazole compared with omeprazole in GERD patients with erosive esophagitis: A randomized controlled trial. *Am. J. Gastroenterol* **96**, 656–665.

Sampliner, R. E. (2002). Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. *Am. J. Gastroenterol.* **97**, 1888–1895.

Spechler, S. J. (1992). Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. *N. Engl. J. Med.* **326**, 786–792.

Spechler, S. J., Lee, E., Ahnen, D., *et al.* (2001). Long-term outcome of medical and surgical therapies for gastroesophageal reflux disease: Follow-up of a randomized controlled trial. *J. Am. Med. Assoc.* **285**, 2331–2338.

Waring, J. P. (2001). Surgical and endoscopic treatment of gastroesophageal reflux disease. *Gastrointest. Clin. North Am.* **31**, S89–S109.

Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric

COLIN D. RUDOLPH
Medical College of Wisconsin and Children's Hospital of Wisconsin, Milwaukee

Barrett's esophagus Intestinal metaplasia of the esophageal mucosal lining.
esophageal stricture A narrowing of the esophageal lumen.
esophageal web A membrane obstructing the esophageal lumen.
gastroesophageal reflux The passage of stomach contents into the esophagus.
gastroesophageal reflux disease The disease that occurs when gastroesophageal reflux results in symptoms or signs, such as chest pain, esophagitis, chronic cough, or recurrent pneumonia.
histamine-2-receptor antagonists Medications that decrease gastric acid secretion by blocking the histamine-2 receptor.
Plummer-Vinson syndrome A disorder characterized by dysphagia, iron deficiency anemia, and a proximal esophageal web.
proton pump inhibitors Medications that decrease gastric acid secretion by inactivating the gastric proton pump that secretes acid from the parietal cells.
Schatzki ring A fibrous ring of tissue at the esophago-gastric mucosal junction that can cause dysphagia.
transient lower esophageal sphincter relaxation An episode of relaxation of the lower esophageal sphincter that is not associated with a swallow event.

Gastroesophageal reflux (GER) is a physiological event that occurs in normal infants, children, and adults. In the child and adult, refluxed gastric contents enter the esophagus and return to the stomach following transient episodes of relaxation of the lower esophageal sphincter. In the infant, the refluxed material often passes through the esophagus and mouth, resulting in regurgitation or vomiting of gastric contents. Gastroesophageal reflux can result in gastroesophageal reflux disease (GERD) when normal protective mechanisms fail. Inadequate esophageal clearance of the refluxed material results in esophageal inflammation and complications, such as stricture formation or Barrett's esophagus. Abnormalities of airway protective mechanisms may result in respiratory symptoms, such as pneumonia, apnea, laryngitis, and exacerbations of reactive airway disease. The first challenge in diagnosis and management of GERD is to determine whether GER is causing a specific symptom or sign since a variety of different pathogenic factors other than GER may present similarly in each case. Once the diagnosis is established, appropriate therapies need to be selected to balance risks of the disease versus the risks associated with each therapeutic option.

EPIDEMIOLOGY

During infancy, recurrent regurgitation or vomiting of gastric contents is frequent, occurring at least once per day in 50% of infants in the first 3 months of life, in 67% of infants who are 4 months old, and in 5% of infants who are 10 to 12 months old. Very few infants develop symptoms or signs of gastroesophageal reflux disease (GERD), which include feeding difficulties, irritability, hematemasis, anemia, failure to thrive, or respiratory disorders, such as apnea, stridor, chronic cough, or recurrent pneumonias. Little is known about the prevalence of GERD in children and adolescents. Preschool children may have problems, such as intermittent vomiting or vague complaints of abdominal pain. Older children and adolescents have presenting symptoms more similar to those of adults, such as complaints of heartburn. The incidence of supraesophageal complications of GERD in childhood is not known.

Although one study suggests that adults with GERD symptoms recall symptoms of gastroesophageal reflux (GER) during childhood more frequently than adults without GERD symptoms, there are very limited data regarding the natural history of GERD in infants and children. One retrospective study reports that in Finnish children diagnosed with GERD but without severe esophagitis, there was no progression of disease severity for 2 years after diagnosis but no longer term follow-up studies are available. Despite this lack of data, it is clear

 Copyright 2004, Elsevier (USA). All rights reserved.

that medical therapy for possible GER is frequently prescribed in adolescents. Pharmaceutical industry data from the United States indicate that 2% of adolescents and 3% of children are prescribed some form of antireflux therapy each year.

Families with multiple affected individuals with GERD have been described and it appears that there may be a genetic predisposition for GERD but no clear pattern of inheritance or genetic locus has yet been identified. A genetic locus associated with GERD was proposed to exist on chromosome 13q14 but these findings have not been replicated.

PATHOPHYSIOLOGY OF GASTROESOPHAGEAL REFLUX DISEASE

As in adults, most episodes of GER occur during transient lower esophageal sphincter relaxation (TLESR) episodes. These episodes are stimulated by gastric distension and allow venting of the stomach. Following the TLESR, secondary peristaltic contractions propel the refluxed gastric contents back into the stomach. Subsequent swallowing of saliva further neutralizes residual acid, returning the esophageal pH to normal and preventing inflammation. In normal infants, children, and adults, episodes of GER occur several times following meals and the esophageal pH returns to normal within minutes. Prolonged exposure to acid results in esophageal inflammation that may lead to scarring (stricture formation) or mucosal metaplasia (Barrett's esophagus).

Most episodes of GER do not reach the pharynx in normal adults or children. As refluxed material enters the upper esophagus, distension of the esophagus results in an initial reflex increase of the upper esophageal sphincter pressure, which tends to prevent most episodes of GER from reaching the pharynx. However, if the upper esophagus is rapidly distended, the upper esophageal sphincter relaxes and allows refluxed material to enter the pharynx. Simultaneously, airway protection mechanisms are invoked, including apnea and laryngeal closure to protect the airway from the refluxed gastric contents. Refluxed gastric contents can be vomited out of the mouth or may be reswallowed, clearing the pharynx before breathing resumes. Defects in esophageal clearance or airway protective mechanisms result in the diverse presentations of GERD outlined below (see Approach to the Gastroesophageal Reflux Disease Symptoms in Infants and Children).

DIAGNOSTIC EVALUATION FOR GASTROESOPHAGEAL REFLUX DISEASE

No single test can reliably differentiate whether GER is the cause of a specific symptom and sign. All of the tests have limitations and must be interpreted in the context of the specific clinical scenario.

Clinical History and Physical Exam

In infants and children, the clinical history and examination alone are often adequate to allow a diagnosis of uncomplicated GER. If warning signs (see Table I) suggest potential complications of GER, further diagnostic testing is indicated.

Radiographic Upper Gastrointestinal Series

Radiographic upper gastrointestinal series are used to evaluate the anatomy of the esophagus and upper gastrointestinal (GI) tract. This test is useful to exclude disorders that may present in a fashion similar to GER, such as an obstruction due to an esophageal stricture, achalasia causing dysphagia, aspiration during swallowing or a tracheoesophageal fistula causing recurrent pneumonia, or an anatomic disorder causing vomiting, such as pyloric stenosis or a malrotation. Findings of GER noted during the upper GI series are not a reliable predictor for GERD but often the evaluation of possible GER symptoms requires an upper GI series to exclude other disorders.

Esophageal pH Monitoring

Esophageal pH monitoring records the pH in the esophagus for 24 h using one or more pH electrodes.

TABLE I Warning Signals in the Vomiting Infant

Bilious vomiting
GI bleeding: hematemesis, hematochezia
Forceful vomiting
Onset of vomiting after 6 months of life
Failure to thrive
Diarrhea
Constipation
Fever
Lethargy
Hepatosplenomegaly
Bulging fontanelle
Macrocephaly/microcephaly
Seizures
Abdominal tenderness, distension
Genetic disorders

Traditional systems connect the sensor sites to an external recording device via a transnasal catheter. Recently, a system using telemetry connections was introduced such that in adults and older children the catheter may not be required. Normal values for the percentage of the 24 h recording time with esophageal pH being less than 4 (reflux index) vary depending on the site of the sensor in the esophagus and the age of the patient. Prolonged acid exposure correlates with an increased risk of esophagitis and it may be help determine whether GER is exacerbating asthma symptoms. Routine esophageal pH monitoring does not exclude GER as the cause of awake apnea in infants, recurrent pneumonia, or other supra-esophageal complications of GER. Esophageal pH monitoring may help clarify whether there is a temporal association between GER and the occurrence of a particular symptom, such as apnea, chest pain, or irritability. Calculation of the "symptom index" (the number of episodes of a symptom associated with GER divided by the number of episodes of a symptom) may help to determine the likelihood that GER is causing a specific symptom. Esophageal pH monitoring may be most useful to determine whether adequate GER therapy is being administered in a patient with persistent symptoms or signs despite aggressive medical therapy.

Esophagastroduodenoscopy and Biopsy

Esophagastroduodenoscopy (EGD) and biopsy involve the use of flexible fiber endoscopes to visualize and biopsy the esophageal mucosa. Small endoscopes with diameters as small as 5 mm allow endoscopy in infants and children with minimal morbidity. Sedation or anesthesia is usually necessary. Endoscopy and biopsy are useful to determine whether esophagitis is present and to confirm that GER is the most likely cause of esophagitis. Other causes of esophagitis, such as infection, Crohn's disease, eosinophilic esophagitis, or pill esophagitis, can also be identified. The stomach and small intestine can also be visualized during the same procedure so that EGD is useful for the evaluation of patients with symptoms including abdominal or chest pain, hematemasis, dysphagia, or odynophagia. EGD will also identify complications of GER, such as stricture or Barrett's esophagus.

Nuclear Scintigraphy

Nuclear scintigraphy is performed by the oral ingestion or instillation of technetium-labeled formula or food into the stomach. The areas of interest, the stomach, esophagus, and lungs, are scanned for evidence of GER and aspiration. The predictive value of this test for any complication of GER has not been established.

Esophageal Impedance Monitoring

Esophageal impedance monitoring measures both acid and nonacid episodes of GER. This new technology measures the electrical impedance between sequential sites in the esophagus. GER episodes are detected by changes in impedance. The use of this method remains experimental but it may be particularly useful for the diagnosis of airway-related complications of GER.

TREATMENT OPTIONS FOR GASTROESOPHAGEAL REFLUX DISEASE IN INFANTS AND CHILDREN

Treatment options for GERD are listed in Table II.

Conservative Therapy

Dietary changes may be useful in the treatment of infant GER. The number of episodes of vomiting and the amount of GER can be reduced by thickening infant

TABLE II Treatment of Gastroesophageal Reflux in Infants and Children

Conservative therapy
Infants
Thickening of formula with rice cereal or commerically available anti-reflux formula
Consider time-limited trial of protein hydrolysate formula
Rarely, consider prone positioning
Children
Sleep with head of bed elevated
Avoid caffeine, chocolate, spicy foods
Lose weight if obese
Pharmacologic Therapy
Antacids (use only for symptomatic relief in children and adolescents)
Magnesium hydroxide
Aluminum hydroxide
H2 blockers
Cimetidine
Ranitidine
Nazitidine
Famotidine
Proton pump inhibitors
Omeprazole
Lanzoprazole
Pantoprazole
Rabeprazole
Esomeprazole
Prokinetics
Surgical therapy

formula with up to 1 tablespoon of infant cereal per ounce. This very thick formula can be difficult to suck so that it is necessary to slit the nipple to allow easier flow. New formulas with starch solutions that thicken on contact with gastric acid are available. These formulas are easier for the infant to ingest and are effective in decreasing GER. Since cow's milk allergy can cause vomiting that is indistinguishable from GER in the infant, it is reasonable to consider a 2-week trial of a protein hydrolysate formula as treatment for an infant with recurrent vomiting. If there is a marked reduction in GER symptoms, it is reasonable to suspect that allergy may have caused the infant's vomiting.

Prone positioning decreases GER but infants that sleep in the prone position are at substantially increased risk of sudden infant death syndrome. Therefore, prone positioning during sleep is not recommended unless there is a substantial risk of possible aspiration with GER such as occurs in a child with an upper aerodigestive tract anomaly prior to surgical correction.

In children and adolescents, there is some evidence that avoidance of caffeine, chocolate, and spicy foods that provoke symptoms is useful for the treatment of GERD. There is also limited evidence that weight loss if obese and the avoidance of tobacco and alcohol may reduce GER. It is not known whether lifestyle changes have an additive benefit in patients receiving pharmacological therapy.

Pharmacologic Therapy

The medications currently available for the treatment of GERD in children are classified into those that decrease gastric acid secretion (anti-secretory agents) and those that decrease the number of GER episodes by improving gastrointestinal and/or esophageal motility (prokinetics). The anti-secretory agents include histamine-2 receptor antagonists (H2RAs) and proton pump inhibitors (PPIs). Both provide effective treatment for esophagitis but the PPIs are far more potent and result in more rapid healing. Long-term use of PPIs appears to be safe in infants and children. Hyperplastic and glandular gastric nodules and polyps may develop in children but this does not appear to be associated with an increased risk of gastric dysplasia or cancer. No effective prokinetic agents for the treatment of GERD are available in the United States. Metoclopramide does not decrease GER. Bethanechol is inadequately studied and has substantial side effects. Sucralfate has been used for the treatment of esophagitis in adults but there are inadequate data to support its use in children.

Surgical Therapy

As in adults, several approaches are available for surgical therapy of GERD in children. The Nissen fundoplication and modifications of this procedure are the most commonly performed operations. All of these procedures alter the anatomy at the diaphragmatic hiatus and increase the length of the intra-abdominal esophagus. Significant complications of the surgery include breakdown of the wrap, small bowel obstruction, gas bloat syndrome, dumping syndrome, infection, atelectasis or pneumonia, perforation, persistent esophageal stricture, and esophageal obstruction. These procedures are now frequently performed laparascopically, even in premature infants, which decreases the immediate postoperative morbidity. Alteration of the integrity of the fundoplication is more likely to occur in children who have neurological disease or in those with robust coughing (cystic fibrosis, asthma) and/or retching. Surgical therapy is usually not indicated for the treatment of infants with GERD unless the risks of complications of GERD outweigh the risks of surgery, since a combination of medical therapy and maturation often prevents the need for surgery. Surgical therapy may be more effective than medical therapy for the treatment of airway complications of GER, such as asthma or recurrent aspiration. In children and adolescents, growth and development are unlikely to alleviate symptoms. The decision to consider surgical therapy versus prolonged medical therapy requires balancing the risks and benefits of surgery with those of medical therapy. Recent studies indicate that despite surgical therapy, a large proportion of adults with GERD still receive PPIs when evaluated 5 years after surgery, indicating that surgery often does not lead to complete lifelong symptom resolution of GERD.

APPROACH TO GASTROESOPHAGEAL REFLUX DISEASE SYMPTOMS IN CHILDREN

The evaluation of each symptom presentation of GERD varies because the differential diagnosis, relative utility of different diagnostic tests, and consequences of making an incorrect diagnosis differ depending on the presenting symptom. The discussion below provides a brief overview of each symptom presentation and outlines symptom-directed diagnostic and therapeutic approaches.

Uncomplicated Gastroesophageal Reflux in an Infant

Normal infants have frequent episodes of GER with effortless, painless regurgitation. Simple "spitting up" must be distinguished from true vomiting, as each has different clinical implications. A large number of disorders can present with vomiting in infants but in the absence of warning symptoms and signs (Table I), parental education, reassurance, and anticipatory guidance are all that is required for management of simple GERD. No treatment is necessary if there are no symptoms in an otherwise well-appearing child but therapies including a time-limited trial of a hypoallergenic formula or use of thickened formulas may be used if desired. Recurrent GER generally decreases in frequency over the first year of life and resolves by 12 to 18 months of age. If symptoms worsen or do not improve by 18 to 24 months of age or are unresponsive to therapy much earlier, further evaluation including an upper GI series may be indicated.

Inadequate Weight Gain in an Infant

Attempts to curb the vomiting associated with uncomplicated infant GER by limiting the quantity of formula ingested can cause iatrogenic weight loss. If an infant is ingesting adequate calories without restrictions and weight loss persists, it is important that other causes of failure to thrive be considered before it is assumed that GERD is the cause. Vomiting due to allergy is indistinguishable from that due to GER so a time-limited (2 weeks) trial of administration of a hypoallergenic formula may be considered. Rarely, infants are unable to gain adequate weight due to problems of vomiting resulting from gastroesophageal reflux. If no other cause of vomiting or poor weight gain is identified, alternative management strategies include the administration of higher caloric density feedings (24 to 30 cal/oz), nasogastric tube feedings, or nasojejunal tube feeding. Pharmacologic therapy with a prokinetic would potentially be useful but no prokinetic agents that aid in management of GER are available. If the infant has concomitant delayed gastric emptying, metoclopramide or domperidone therapy may be helpful.

Excessive Irritability in an Infant

GER is rarely the cause of excessive irritability or colic in infants. Normal infants fuss or cry for an average of 2 h daily, with many normal infants crying for up to 6 h per day. Crying peaks at 6 weeks of age. There is no strong evidence demonstrating a relationship between excessive crying and GER; however, GERD is frequently invoked as a cause of irritability based on the frequency of vomiting in infants and extrapolations from the adult experience with heartburn. No good controlled studies demonstrate a response of irritability to GER therapy. In infants with severe irritability, other causes, such as neurologic disorders, urinary tract infection, allergy, or small intestinal disorders, need to be ruled out before a trial of empiric therapy for GERD is considered. A clear relationship between GER and episodes of irritability can be established by obtaining an esophageal pH monitoring study and determining the symptom index (see Diagnostic Evaluation for Gastroesophageal Reflux Disease). A trial of therapy of GER can be considered if other causes of irritability are excluded and episodes of irritability are associated with GER documented by vomiting or by esophageal pH monitoring.

Heartburn or Chest Pain in the Older Child

GER may cause symptoms of heartburn or chest pain in children. Some children have difficulty localizing pain and complain of "tummy" pain. In children with intermittent symptoms, initial empiric therapy with antacids, H2RAs, or PPIs can be considered. If symptoms recur following cessation of therapy or do not improve with empiric therapy, then diagnostic upper endoscopy with biopsy should be considered. Other disorders that cause esophagitis, including eosinophilic esophagitis or inflammatory bowel disease, and infectious esophagitis can present with similar symptoms. Some patients without esophagitis experience pain during physiologic reflux episodes. In these patients, it is reasonable to continue treatment with either H2RAs or PPIs to relieve symptoms despite the absence of esophagitis.

Feeding and Swallowing Difficulties

Feeding difficulties are associated with the presence of GER in infants and children but it is unclear whether GER (even in the presence of esophagitis) actually causes feeding problems in infants. Adults with relatively severe esophagitis rarely complain of significant dysphagia unless they develop stricture. No study has demonstrated that treatment of GER improves feeding in infants. Because a large variety of disorders may contribute to feeding difficulties or dysphagia in the infant and older child, empirical therapy for GER is generally not recommended. However, if there are other signs or symptoms suggestive of GERD, then a time-limited course of medical therapy can be considered. Usually, a diagnostic evaluation with a radiographic contrast study should first be performed to identify anatomic abnormalities, such as strictures or vascular rings,

and motility disorders, such as achalasia. Upper endoscopy with biopsy is also usually performed to rule out other causes of esophagitis.

Other Esophageal Symptoms and Signs

Esophagitis causes other symptoms including hematemasis and atypical seizure-like movements with torsion of the neck known as Sandifer's syndrome, esophageal stricture, and Barrett's esophagus. When erosive esophagitis is demonstrated during endoscopic evaluation, treatment with H2RAs or PPIs is usually indicated. PPI therapy is highly effective for treatment of esophagitis resulting from GER. Biopsy is mandatory to rule out other causes of esophagitis. Typical biopsies in GERD reveal thickening of the basal cell layer (>25% of mucosal thickness), increases in papillary height (>50% of mucosal thickness), and infiltration with neutrophils (>1 per HPF) or eosinophils (>1 and <15 per HPF). Barrett's esophagus is presumed to result from esophagitis. It is characterized by intestinal metaplasia of the mucosa and is associated with an increased risk of adenocarcinoma of the esophagus. Other causes of esophagitis include eosinophilic esophagitis, allergy, inflammatory bowel disease, infections, caustic esophagitis, and pill esophagitis, all of which are characterized by different findings on biopsy. The best approach to management of erosive esophagitis and Barrett's esophagus in children remains controversial. Due to the long-term risk of adenocarcinoma, many experts recommend that children with this diagnosis should undergo surveillance esophagoscopy and biopsy every 3 to 5 years. Treatment of GER-induced esophagitis should focus on reducing the acid exposure of the esophagus. No studies show a benefit of surgical anti-reflux therapy compared to long-term PPI therapy for prevention of GER complications including stricture or adenocarcinoma. The relative long-term risks and benefits of medical versus surgical therapeutic approaches are unclear.

Apnea

Cessation of breathing when gastric contents reflux into the pharynx is a normal physiologic protective mechanism that prevents aspiration. This reflex response appears to be somewhat more robust in the infant than in the adult such that laryngeal stimulation is more apt to cause cough in the child or adult but will cause apnea in early infancy. This relationship has led to a widely held belief that GER is a common cause of apnea of prematurity and of acute life-threatening events (ALTEs) in infancy. There is no good evidence to support a relationship between GER and the central apnea observed in premature infants. ALTEs are episodes in infants characterized by a combination of apnea, change in color (cyanosis, pallor, rubor, plethora), change in muscle tone (limpness, stiffness), and choking and gagging that requires intervention by the caretaker. GER is only one of many potential causes of an ALTE and those due to GER generally occur while the infant is awake, supine, and usually within 1 h of feeding. Prior to invoking GER as the cause of ALTE, it is essential to rule out other potential causes including anatomic airway obstructive disorders, central apnea, cardiac, central nervous system and infectious disorders, and possible intentional suffocation observed with Munchausen-by-proxy syndrome. Definitive demonstration that apnea is caused by GER in an individual patient is often difficult because the episodes occur sporadically. Thus, even during 24 h of monitoring, an episode often does not occur. Furthermore, it is impossible to determine whether GER is causing the episode or whether it results from the respiratory effort associated with obstructive apnea. Therefore, if esophageal pH monitoring is being performed to document a relationship between the apneic episodes and GER, it is useful only if combined with simultaneous recording of heart rate, chest wall impedance, nasal airflow, and oxygen saturation to detect obstructive apnea. Esophageal impedance recording may be more useful because it will detect acid and nonacid GER episodes, both of which could cause awake apneic episodes. Generally, it is reasonable to initiate therapy for presumed GER-related apnea without any diagnostic testing if there is a characteristic history combined with appropriate testing to exclude other causes of ALTEs. Conservative medical therapy for GER and possibly the administration of antisecretory agents decreases the frequency of episodes. Surgical therapy is usually not required since the frequency of episodes almost always decreases as the infant matures and the risk of morbidity from surgery generally exceeds the risk from the awake apneic episodes.

Asthma

Although the prevalence of GER in children with asthma is approximately 50%, GER exacerbates but does not cause asthma. Animal studies suggest that esophageal acid exposure may increase the reactivity of the airway through vagal reflex mechanisms. Microaspiration of refluxed gastric contents may also alter airway reactivity. Vigorous anti-reflux therapy (high-dose proton pump inhibitors) may decrease the severity of symptoms in some patients, especially those with coexisting symptoms of esophagitis. In patients without any symptoms suggestive of GER, esophageal pH

monitoring should be considered in those with asthma and recurrent pneumonia, those with nocturnal asthma more than once a week, those requiring either continuous oral corticosteroids or high-dose inhaled corticosteroids (more than two courses per year of corticosteroids), and those with persistent asthma unable to wean medical management. If esophageal pH monitoring demonstrates an increased frequency or duration of esophageal acid exposure, a trial of prolonged medical therapy for GER should be considered. In patients who respond to therapy, a decision to consider surgical management must balance the risks of long-term medical therapy versus surgical therapy. Some studies suggest that surgical therapy may be more effective than medical therapy, but these studies did not treat patients with high doses of PPIs.

Recurrent Pneumonia

In children with abnormal airway protective mechanisms, such as those with neurologic disease or with laryngeal anatomic defects, GER may cause recurrent pneumonia. Pulmonary fibrosis and chronic lung disease can result. Unfortunately, it is difficult to prove that GER is the cause of chronic lung disease. Normal esophageal pH monitoring does not exclude GER as a cause of recurrent aspiration. Bronchoalveolar lavage with large numbers of lipid-laden macrophages suggests that aspiration may be present, but it may be difficult to determine whether aspiration occurs during swallowing or only during episodes of GER. Regardless, treatment of GER is often indicated once other causes of recurrent pneumonia are ruled out, such as an immunodeficiency, cystic fibrosis, or anatomic defects, such as a tracheoesophageal fistula. Medical therapy may be tried in children with only moderate pulmonary disease, but in children with more severe pulmonary disease, antireflux surgery should be considered.

Hoarseness and Recurrent Croup

Hoarseness and recurrent croup have been associated with GER. The approach to diagnosis and treatment of GER-related laryngeal disease is at present undefined but a trial of aggressive anti-reflux therapy with a high-dose proton pump inhibitor is reasonable. If symptoms resolve and then recur with the cessation of therapy, GER is likely responsible. Decisions regarding long-term therapy must balance the risks of long term medical or surgical therapy versus the severity of underlying symptoms.

Other disorders, such as sinusitus, otitis media, and dental erosions, have been suggested to be related to GER but there is little evidence to support these contentions at this time.

CONGENITAL ESOPHAGEAL OBSTRUCTIVE LESIONS

Congenital Esophageal Stenosis and Web Diaphragms

Congenital esophageal stenosis and web diaphragms rarely occur (i.e., 1 in 25,000 live births) and usually present with dysphagia when feeds are advanced to puréed or solid consistencies. Diagnosis is made by radiographic contrast studies or endoscopy. Webs and fibromuscular stenosis usually can be successfully dilated using endoscopy- and/or fluoroscopy-guided balloon or bougie-type dilators. However, dilation of a stenosis resulting from ectopic cartilage (i.e., tracheobronchial remnant), which on endoscopy is more irregular and firmer than the fibromuscular type, seems somewhat more likely to result in perforation, leading some experts to recommend primary surgical resection of these lesions.

Acquired Esophageal Webs

Acquired esophageal webs are very uncommon in children. Plummer-Vinson syndrome, characterized by dysphagia, iron deficiency anemia, and an esophageal web in the upper esophagus, has only rarely been reported in childhood. Treatment of the iron deficiency and dilation of the web is therapeutic. An association with celiac disease has been reported and probably resulted from the severe iron deficiency sometimes observed in celiac disease.

Schatzki Rings

Schatzki rings are fibrous mucosal thickenings at the esophageal–gastric junction and are also very uncommon in children, presenting with progressive dysphagia of solids or acute food impaction. Although uncommon, this lesion can be seen in children with esophagitis and is treated with dilation.

Esophageal Duplications

Esophageal duplications are rare lesions that present as a tubular mass in the posterior mediastinum. Secretions cause distension with pressure on contiguous structures that can cause dysphagia or respiratory symptoms. Simple surgical excision is curative.

Congenital Esophageal Diverticulae

Congenital esophageal diverticulae are rare lesions that occur through the muscle layer located just above the cricopharyngeus muscle. They may present with symptoms of choking, coughing, and dysphagia in an older child. More commonly, acquired diverticulae without a muscle layer occur in neonates after traumatic efforts at airway intubation. Perforation results in infection and edema, which may cause obstruction. Antibiotic therapy and the cessation of oral feedings usually allow resolution of perforation if it is recognized early.

Vascular Rings

Vascular rings may also present with dysphagia or airway obstruction. Diagnosis is usually made by radiographic contrast esophagrams. The most common anomaly is that of an aberrant right subclavian artery, which crosses behind the upper esophagus and is seen in approximately 1% of the population. It usually is asymptomatic but in some children it may cause obstruction (dysphagia lusorum). If this occurs, other causes of dysphagia should be ruled out before surgical therapy is considered. The other most common vascular obstruction of the esophagus is a double aortic arch, which usually results in dysphagia and respiratory symptoms.

See Also the Following Articles

Barrett's Esophagus • Esophagus, Development • H2-Receptor Antagonists • Neonatal Tracheoesophageal Anomalies • Proton Pump Inhibitors • Webs

Further Reading

Buckley, K., Buonomo, C., Husain, K., and Nurko, S. (1998). Schatzki ring in children and young adults: Clinical and radiologic findings. *Pediatr. Radiol.* **28**, 884–886.

Kawahara, H., Dent, J., Davidson, G., and Okada, A. (2001). Relationship between straining, transient lower esophageal sphincter relaxation, and gastroesophageal reflux in children. *Am. J. Gastroenterol.* **96**, 2019–2025.

Mansell, N. J., Jani, P., and Bailey, C. M. (1999). Plummer-Vinson syndrome: A rare presentation in a child. *J. Laryngol. Otol.* **113**, 475–476.

Orenstein, S. R., Shalaby, T. M., Finch, R., Pfuetzer, R. H., DeVandry, S., Chensny, L. J., Bannada, M. M., and Whitcomb, D. C. (2002). Autosomal dominant infantile gastroesophageal reflux disease: Exclusion of a 13q14 locus in five well characterized families. *Am. J. Gastroenterol.* **97**, 2725–2732.

Randolph, J. G. (1986). Esophageal atresia and congenital stenosis. *In* "Pediatric Surgery" (K. J. Welch, J. G. Randolph, M. M. Ravitch, J. A. O'Neill, and M. I. Rowe, eds.), pp. 682–697. Year Book, Chicago.

Rudolph, C. D., Mazur, L. J., Liptak, G. S., Baker, R. D., Boyle, J. T., Colletti, R. B., Gerson, W. T., and Werlin, S. L. (2001). Guidelines for evaluation and treatment of gastroesophageal reflux in infants and children: Recommendations of the North American Society for Pediatric Gastroenterology and Nutrition. *J. Pediatr. Gastroenterol. Nutr.* **32**, S1–S31.

Ruuska, T., Karikoski, R., and Laippala, P. (2002). The natural course of gastroesophageal reflux disease in children. *Scand. J. Gastroenterol.* **37**, 638–641.

Sood, A. R., Strachan, D. R., and Fenwick, J. D. (1999). A case of Plummer-Vinson syndrome in childhood. *J. Pediatr. Surg.* **34**, 1570–1572.

Vasudevan, S. A., Kerendi, F., Lee, H., and Ricketts, R. R. (2002). Management of congenital esophageal stenosis. *J. Pediatr. Surg.* **37**, 1024–1026.

Waring, J. P., Feiler, M. J., Hunter, J. G., Smith, C. D., and Gold, B. D. (2002). Childhood gastroesophageal reflux symptoms in adult patients. *J. Pediatr. Gastroenterol. Nutr.* **35**, 334–338.

Gastrointestinal Matrix, Organization and Significance

RICHARD C. BATES
Beth Israel Deaconess Medical Center and Harvard University

basement membrane A specialized, sheet-like, extracellular matrix structure that separates the connective tissue from the epithelia, blood vessels, muscle fibers, and nerves.

lamina propria The connective tissue compartment located just below the epithelium consisting of interstitial matrix as well as various cell types (fibroblasts and immune cells).

mucosa The most highly differentiated layer of the gastrointestinal tract, comprising the epithelium, basement membrane, lamina propria, and layer of smooth muscle known as muscularis mucosae. Tissue specialization and surface shape are correlated with functional differentiation along the tract.

restitution The process that repairs damage to the epithelium above the lamina propria, involving epithelial cell migration and spreading across the site of injury in order to resurface it.

The mucosal layers of the stomach, intestine, and colon are well-differentiated tissues that display variations in cell type and structure, reflecting the functional specialization of the gastrointestinal tract. However, these mucosa share a common anatomy comprising epithelial cells, underlying connective tissue regions rich in extracellular matrix components and mesenchymal cells, and also smooth muscle. The biochemical components of the matrix form supramolecular complexes that provide physical support for the overlying epithelium and interact, via specific domains, with cellular receptors capable of influencing signal transduction pathways and gene transcription. Thus, the gastrointestinal matrix forms a regulatory network that directs such fundamental cellular processes as proliferation, differentiation, and migration to influence organ function.

INTRODUCTION

The extracellular matrix (ECM) network supporting the gastrointestinal tract is a rich milieu of structural proteins, proteoglycans, growth factors, and matricellular proteins. This stromal compartment also contains cells of mesenchymal origin, such as fibroblasts and immune cells, that are responsible for the maintenance of mucosal integrity. This article describes the major structural components of the matrix, which in general tend to be multifunction, multidomain molecules whose interactions are mediated through specific modules. Localization of these proteins into distinct subepithelial structures and/or regions suggests that functional specialization accompanies differentiation in the gastric mucosa. In addition to providing architectural support, it is now known that the matrix itself promotes signaling that is critical for cellular function, primarily via the integrin family of adhesion receptors. Indeed, matrix-derived signals direct such fundamental processes as proliferation, differentiation, maturation, and survival—which in turn affect tissue homeostasis, organ topography, and, ultimately, gastrointestinal function.

STRUCTURE AND COMPOSITION

Basement Membrane

Along the gastrointestinal tract, a simple columnar epithelium resides upon a basement membrane that also provides an interface with the lamina propria beneath it (Fig. 1). At the electron microscope level, the basement membrane appears as a continuous sheet of electron-dense material and is composed primarily of laminins, type IV collagen, nidogen, and heparin sulfate proteoglycans (HSPGs) integrated into a common structure, instead of simply being layered. Additional ECM components, such as fibronectin and tenascin, as well as associated proteins, such as SPARC (secreted protein, acidic and rich in cysteine), are also present in the basement membranes of the gastrointestinal tract. It is now appreciated that the interactions between components of the basement membrane are regulated through multiple, mostly domain-specific mechanisms.

Laminins, Nidogen, and Type IV Collagen

Laminins are a major component of the basement membrane. All laminins are large heterotrimeric glycoproteins composed of α, β, and γ subunits that

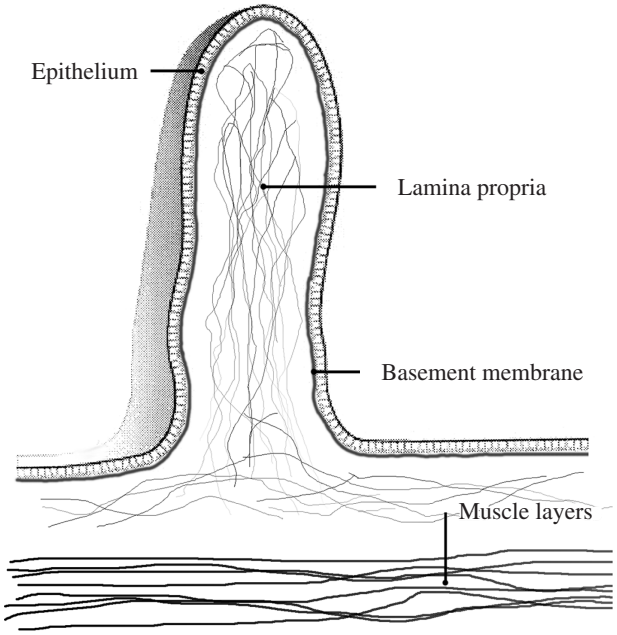


FIGURE 1 Schematic overview of an intestinal villus demonstrating the ultrastructural arrangement of the gastrointestinal mucosa layer. A columnar epithelium sits upon a basement membrane, which also provides an interface with the underlying loose connective tissue, the lamina propria. A smooth muscle layer defines the extent of the mucosa proper.

associate to form a cross-like structure. Five distinct α, three β, and two γ chains associate to form at least 11 currently known laminin isoforms, of which laminins-1, -2, -5, and -10 are most abundantly expressed in the intestinal matrix. Furthermore, these laminin proteins are often expressed in a developmental- and tissue-specific manner, suggesting a functional diversity of the individual variants. For example, laminin-1 and -2 show a complementary pattern of expression along the crypt–villus axis in the adult human intestine: laminin-1 presents a decreasing gradient of intensity from the tip of the villus to the crypt mouth, whereas laminin-2 is exclusively present in the basement membrane that surrounds the bottom of the crypts. At the C-terminal region of the heterotrimer, all laminins possess a large globular domain that contains modules via which laminin molecules interact with cells as well as other matrix molecules. In addition to self-assembly, laminins do associate with other laminin isoforms and other ECM molecules. Indeed, the ECM glycoprotein nidogen functions as a laminin–collagen IV bridge that is essential for the structural integrity of the basement membrane. Nidogen forms a link between the laminin and type IV collagen networks and, similarly, mediates the formation of a tertiary complex between laminin and the core protein of the proteoglycans. Because nidogen appears to play such an important role in basement membrane formation, it is not surprising that isoforms of this protein have now been discovered.

Collagens, which represent the most abundant proteins of the body and function as extracellular building blocks, are encoded by a family of at least 30 genes. Each gene encodes a precursor chain, known as a pro α chain; these chains assemble together in a tight triple-helix composed of three chains both as homotrimers and as heterotrimers. Of this family, type IV collagen is the major structural component of the basement membrane. It is composed of three of six genetically distinct α (IV) chains. The distribution of the α (IV) chains varies with tissues: α1 (IV) and α2 (IV) chains are widely distributed, whereas α3–6 (IV) chains are restricted to certain basement membranes. The presence of these multiple chains that give rise to at least five distinct type IV isoforms adds to the complexity of the basement membrane organization. Not surprisingly, during development and in the adult organs, type IV collagen expression parallels that of laminin-1 in the gut.

Proteoglycans

Proteoglycans are large proteins that carry sulfated polysaccharides (glycosaminoglycans) as side chains and are ubiquitously expressed in virtually all mammalian tissues. Heparin sulfated proteoglycans (HSPGs) are typical constituents of basement membranes, with the most prominent HSPG in the gastrointestinal tract being perlecan, which consists of a 400 to 500 kDa core protein with three heparin side chains. Incorporation of proteoglycans into the basement membrane involves heparin binding to laminin and type IV collagen, as well as core protein binding to nidogen.

Fibronectin and Tenascin

Additional ECM components fibronectin and tenascin are also present in the basement membranes of gastrointestinal tissues. Fibronectin is a large, multidomain adhesive glycoprotein that is an essential constituent of many extracellular matrices found throughout the body and also occurs as a soluble form in plasma. The amino acid sequence of fibronectin reveals three types of internally homologous repeats separated by short connecting sequences. There are 12 type I, 2 type II, and 15 type III modules and homologous domains are also found in other proteins, especially the type III repeat. Fibronectin is synthesized by cells as a soluble protomer and then converted into an insoluble form that becomes incorporated into the fibrillar meshwork

of the lamina propria (see below) or included into the basement membrane itself.

Tenascin-C (Tenascin-cytoactin) is the prototype member of a family of glycoproteins that arise from alternative splicing of a single gene. Their molecular structure is that of a hexameric oligomer with a striking symmetrical appearance. Six polypeptide chains emanate from a central core where the amino-termini are linked via a tenascin assembly domain; each arm contains a row of epidermal growth factor-like as well as fibronectin type III repeats and the terminal knob is composed of a globular domain resembling the carboxy-terminal region of the β and γ chains of fibrinogen. During embryogenesis, tenascin is broadly expressed, but its expression in adult tissues is far more restricted.

Interestingly, tenascin and fibronectin exhibit opposing gradients of expression in the basement membrane of the crypt–villus axis, consistent with a proposed fibronectin-antagonizing role for tenascin. Tenascin shows a gradient of increasing immunoreactivity toward the tip of the villus and it has been suggested that tenascin promotes the sloughing of intestinal epithelia into the gut lumen, which occurs at the top of the villi in the small intestines and on the surface of the colonic mucosa. Conversely, fibronectin reactivity is strongest in the lower regions of the crypts, with decreased staining intensity in basement membranes toward the surface of the villus.

SPARC

SPARC is a multifunctional glycoprotein that belongs to the group of matricellular proteins that mediate cell–matrix interactions, but do not serve primarily structural roles, as opposed to fibronectin, laminin, and the collagens. Other members of this group of proteins include thrombospondin and tenascin, as outlined above. SPARC modulates cellular interactions with the ECM by binding to structural matrix proteins, in particular collagen, as well as exerting a counteradhesive effect on cells. It has a broad tissue distribution, but in the adult appears to be most highly expressed in remodeling tissues and those with high rates of epithelial turnover, such as the gut, bone, and tissues undergoing repair.

Lamina Propria

The lamina propria constitutes the layer of loose connective tissue and interstitial matrix located just below the epithelium. In the stomach, the lamina propria tends to be relatively inconspicuous, filling the interstitial spaces between the tubular gastric glands. It is more expansive in the small intestine, where it occupies the cores of the villi and envelops the crypts. In the colon, it again becomes more retracted based on morphology, restricted largely to the regions between crypts. The structural composition of this zone differs fundamentally from that of the basement membrane in terms of its high content of fibrillar collagens and proteoglycans. However, as shown in Fig. 2, there is some overlap in terms of content with basement membrane-associated proteins. Notably, in addition to their reciprocal expression pattern seen in the basement membrane, both tenascin and fibronectin are found as part of the meshwork of the lamina propria interstitial matrix. Functionally, this region provides the structural support for the lymphatics and vasculature. Furthermore, the lamina propria of the stomach and intestine is also particularly cell-rich, including fibroblasts, lymphocytes, macrophages, plasma cells, and mast cells. This large proportion of cells with immune function provides an effective secondary line of defense against potential invading microorganisms and aggregations of lymphoid nodules within the lamina propria of the small intestine give rise to the specialized areas known as Peyer's patches. The lamina propria extends to the thin layer of smooth muscle, or muscularis mucosae, which together with the epithelium and basement membrane constitutes the mucosa of the gastrointestinal tract.

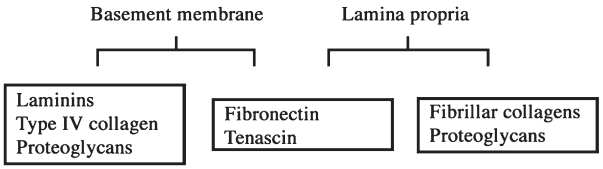


FIGURE 2 Summary of the major extracellular matrix protein components and their distribution within the basement membrane and lamina propria.

DERIVATION OF THE MATRIX

Both basement membrane assembly and lamina propria synthesis result from a complex interplay between epithelial and mesenchymal cells. Clearly, basement membrane assembly must occur in the extracellular environment at the interface of these two layers. With regard to the laminin family of proteins, evidence suggests that individual epithelial or mesenchymal cells are capable of producing several laminin chains and isoforms simultaneously, before depositing them together into the adjacent basement membrane. For example, epithelial cells produce laminin-1 and mesenchymal cells produce laminin-2. The mesenchymal compartment is thought to be the principal source of endogenous type IV collagen. Similarly, nidogen, whose crucial

role as a intermolecular cross-linker is required for basement membrane assembly, has been shown to be a mesenchymal product. In contrast, HSPG molecules located in the basement membrane of the developing intestine appear to be produced by the epithelial cells exclusively. Fibronectin is produced by epithelial cells, but is also synthesized by cells within the lamina propria, such as fibroblasts. Taken together, these findings reinforce the concept that epithelial–mesenchymal interactions play an important role in the formation of a complete basement membrane.

SIGNALING FROM THE MATRIX

Far from merely providing a static architectural role, it is now established that the matrix composition of the gastrointestinal tract defines the necessary microenvironment essential for multiple cellular functions including differentiation, proliferation, migration, and even survival. The best characterized ECM adhesion receptors are the integrins, which constitute a diverse family of integral membrane glycoproteins. Integrins function as noncovalently linked αβ heterodimers and, in addition to their ability to support cell adhesion, they activate signal transduction pathways via their cytoplasmic domains. The dual ability of these receptors to promote adhesion and influence signal pathways and gene expression in response to matrix proteins has established them as the principal mediators of specific signaling from the matrix. As demonstrated by the complementary gradients of laminin-1 and -2 expression, and the opposing expression profiles of fibronectin and tenascin in the basement membrane, specific regional differences in the biochemical composition of the matrix would be expected to influence epithelial cell behavior. Additionally, changes in integrin receptor expression by the epithelial cells themselves dictate their functional responses to the underlying matrix.

The complexity of integrin expression and function is highlighted by studies of the laminin-binding integrins in the gastrointestinal tract. Expression is complex and exhibits variations, for example, during development and even along the crypt–villus axis in the adult intestine. Here, it has been shown that two potential laminin-binding receptors exhibit complementary staining patterns along the crypt–villus axis, with α2β1 predominant in the crypts and α3β1 more highly expressed along the villus. Although α2β1 expression appears to correlate with that of laminin-2 in the crypts, this integrin actually has a higher affinity for laminin-1 binding than laminin-2. Since the crypt basement membrane is devoid of laminin-1, and α2β1 can also bind to collagen, it appears that this integrin is actually functioning as a collagen receptor in the intestine. Additional complexity arises when it is considered that α3β1 is further capable of binding multiple ligands including fibronectin, type IV collagen, and nidogen; thus, any of these components could influence gene expression in the epithelial cells by activating this receptor. Furthermore, other integrins such as α6β1, α7β1, and α6β4, all of which function exclusively as laminin receptors, are also co-expressed by these same cells and appear to have more uniform distributions. This apparent redundancy suggests that these different receptors must serve distinct signaling functions when binding to the same matrix protein.

MATRIX REGULATION OF RESTITUTION AND MUCOSAL HEALING

The mucosal epithelium of the gastrointestinal tract serves as an important barrier to a wide variety of noxious and potentially harmful agents. It is essential, therefore, that the epithelial layer is rapidly resealed following injury, for example, as the result of gastric ulceration, bacterial infection, inflammatory bowel conditions, ischemia, or radiation. This is accomplished by a process known as restitution, whereby viable epithelial cells migrate from regions proximal to the site of injury to cover the denuded surface. This response reestablishes epithelial continuity more rapidly than can be achieved by cell proliferation and helps prevent deeper mucosal damage. In order to do so, the epithelial cells temporarily and reversibly lose their polarized appearance and assume a more squamous morphology, best described not as dedifferentiation but rather as redifferentiation toward a motile phenotype. Restitution is subject to strict regulatory controls, directed in part by soluble mediators such as growth factors and bioactive peptides, as well as nonsoluble factors, including the extracellular matrix itself. The molecular architecture of the matrix, with its adhesive protein constituents, provides a physical substrate permitting traction and the generation of forces needed for migration and thus directly influences the rate of cellular motility. Furthermore, the matrix can also directly and specifically modulate signal transduction pathways within the epithelial cells themselves, at the level of protein activation, synthesis, and localization. The same adhesive components, including the collagens and fibronectin, have been shown to influence cell morphogenesis and differentiation, mediated largely through integrin activation and signaling. These signaling pathways direct the cytoskeletal changes and adhesive responses seen in migrating cells. Moreover, it is emerging that matrix-dependent motility-related

signaling is likely to be involved in cross talk with signals provided by other stimuli, such as growth factors, adding yet another level of complexity in the regulation of this critical homeostatic process.

See Also the Following Articles

Epithelium, Proliferation of • Epithelium, Repair of

Further Reading

Basson, M. D. (2001). *In vitro* evidence for matrix regulation of intestinal epithelial biology during mucosal healing. *Life Sci.* **69**, 3005–3018.

Beaulieu, J.-F. (1999). Integrins and human intestinal cell functions. *Front. Biosci.* **4**, D310–321.

Honenester, E., and Engel, J. (2002). Domain structure and organisation in extracellular matrix proteins. *Matrix Biol.* **21**, 115–128.

Jones, F. S., and Jones, P. L. (2000). The tenascin family of ECM glycoproteins: Structure, function and regulation during embryonic development and tissue remodeling. *Dev. Dyn.* **218**, 235–259.

Lotz, M. M., Nusrat, A., Madara, J. L., Ezzell, R., Wewer, U. M., and Mercurio, A. M. (1997). Intestinal epithelial restitution: Involvement of specific laminin isoforms and integrin laminin receptors in wound closure of a transformed model epithelium. *Am. J. Pathol.* **150**, 747–760.

Lussier, C., Basora, N., Bouatrouss, Y., and Beaulieu, J.-F. (2000). Integrins as mediators of epithelial cell–matrix interactions in the human small intestinal mucosa. *Microsc. Res. Tech.* **51**, 169–178.

Simon-Assmann, P., Kedinger, M., De Arcangelis, A., Rousseau, V., and Simo, P. (1995). Extracellular matrix components in intestinal development. *Experientia* **51**, 883–900.

Simon-Assmann, P., Lefebvre, O., Bellissent-Waydelich, A., Orian-Rosseau, V., and De Arcangelis, A. (1998). The laminins: Role in intestinal morphogenesis and differentiation. *Ann. N. Y. Acad. Sci.* **859**, 46–64.

Gastrointestinal Tract Anatomy, Overview

JOHN M. RUSSO* AND JERROLD R. TURNER†

**University of Chicago Children's Hospital and †University of Chicago*

celiac disease Autoimmune disorder affecting the small intestine; caused by abnormal responses to certain grains, leads to small intestinal mucosal damage and nutrient malabsorption.

gastroesophageal reflux Retrograde movement of gastric contents into the esophagus, resulting in esophageal damage and inflammation (esophagitis).

pernicious anemia Form of anemia (red blood cell deficiency) caused by vitamin B_{12} deficiency; occurs secondary to an absence of gastric intrinsic factor, which is necessary for B_{12} absorption in the terminal ileum.

volvulus Twisting of intestinal loops, with subsequent obstruction of the vascular supply and/or lumen.

The gastrointestinal tract is a tubular conduit; it is responsible for processing food into an absorbable form, absorption of nutrients and electrolytes from the lumen, delivery of these nutrients to the body, and excretion of waste products. With the largest surface area of any organ in the body, the gastrointestinal tract is also exposed to a diverse assortment of foreign materials. Sampling of these materials and generation of appropriately directed immune responses, e.g., to pathogens but not to food products, are an essential part of gastrointestinal function.

ESOPHAGUS

The esophagus is a 20- to 25-cm-long tube (see Fig. 1) that transports food from the oral cavity to the stomach. It is composed of three primary segments; cervical, thoracic, and abdominal. Accordingly, the vascular supply is also segmental. The upper esophagus is supplied by the thyroid arteries, the midesophagus is supplied by branches of the aorta, and the lower esophagus is supplied primarily by the left gastric artery. Venous drainage is similar, with the upper third of the esophagus draining into the superior vena cava, the middle third draining into the azygous veins, and the lower

Copyright 2004, Elsevier (USA). All rights reserved.

signaling is likely to be involved in cross talk with signals provided by other stimuli, such as growth factors, adding yet another level of complexity in the regulation of this critical homeostatic process.

See Also the Following Articles

Epithelium, Proliferation of • Epithelium, Repair of

Further Reading

Basson, M. D. (2001). *In vitro* evidence for matrix regulation of intestinal epithelial biology during mucosal healing. *Life Sci.* **69**, 3005–3018.

Beaulieu, J.-F. (1999). Integrins and human intestinal cell functions. *Front. Biosci.* **4**, D310–321.

Honenester, E., and Engel, J. (2002). Domain structure and organisation in extracellular matrix proteins. *Matrix Biol.* **21**, 115–128.

Jones, F. S., and Jones, P. L. (2000). The tenascin family of ECM glycoproteins: Structure, function and regulation during embryonic development and tissue remodeling. *Dev. Dyn.* **218**, 235–259.

Lotz, M. M., Nusrat, A., Madara, J. L., Ezzell, R., Wewer, U. M., and Mercurio, A. M. (1997). Intestinal epithelial restitution: Involvement of specific laminin isoforms and integrin laminin receptors in wound closure of a transformed model epithelium. *Am. J. Pathol.* **150**, 747–760.

Lussier, C., Basora, N., Bouatrouss, Y., and Beaulieu, J.-F. (2000). Integrins as mediators of epithelial cell–matrix interactions in the human small intestinal mucosa. *Microsc. Res. Tech.* **51**, 169–178.

Simon-Assmann, P., Kedinger, M., De Arcangelis, A., Rousseau, V., and Simo, P. (1995). Extracellular matrix components in intestinal development. *Experientia* **51**, 883–900.

Simon-Assmann, P., Lefebvre, O., Bellissent-Waydelich, A., Orian-Rosseau, V., and De Arcangelis, A. (1998). The laminins: Role in intestinal morphogenesis and differentiation. *Ann. N. Y. Acad. Sci.* **859**, 46–64.

Gastrointestinal Tract Anatomy, Overview

JOHN M. RUSSO* AND JERROLD R. TURNER†

**University of Chicago Children's Hospital and †University of Chicago*

celiac disease Autoimmune disorder affecting the small intestine; caused by abnormal responses to certain grains, leads to small intestinal mucosal damage and nutrient malabsorption.

gastroesophageal reflux Retrograde movement of gastric contents into the esophagus, resulting in esophageal damage and inflammation (esophagitis).

pernicious anemia Form of anemia (red blood cell deficiency) caused by vitamin B_{12} deficiency; occurs secondary to an absence of gastric intrinsic factor, which is necessary for B_{12} absorption in the terminal ileum.

volvulus Twisting of intestinal loops, with subsequent obstruction of the vascular supply and/or lumen.

The gastrointestinal tract is a tubular conduit; it is responsible for processing food into an absorbable form, absorption of nutrients and electrolytes from the lumen, delivery of these nutrients to the body, and excretion of waste products. With the largest surface area of any organ in the body, the gastrointestinal tract is also exposed to a diverse assortment of foreign materials. Sampling of these materials and generation of appropriately directed immune responses, e.g., to pathogens but not to food products, are an essential part of gastrointestinal function.

ESOPHAGUS

The esophagus is a 20- to 25-cm-long tube (see Fig. 1) that transports food from the oral cavity to the stomach. It is composed of three primary segments; cervical, thoracic, and abdominal. Accordingly, the vascular supply is also segmental. The upper esophagus is supplied by the thyroid arteries, the midesophagus is supplied by branches of the aorta, and the lower esophagus is supplied primarily by the left gastric artery. Venous drainage is similar, with the upper third of the esophagus draining into the superior vena cava, the middle third draining into the azygous veins, and the lower

Copyright 2004, Elsevier (USA). All rights reserved.

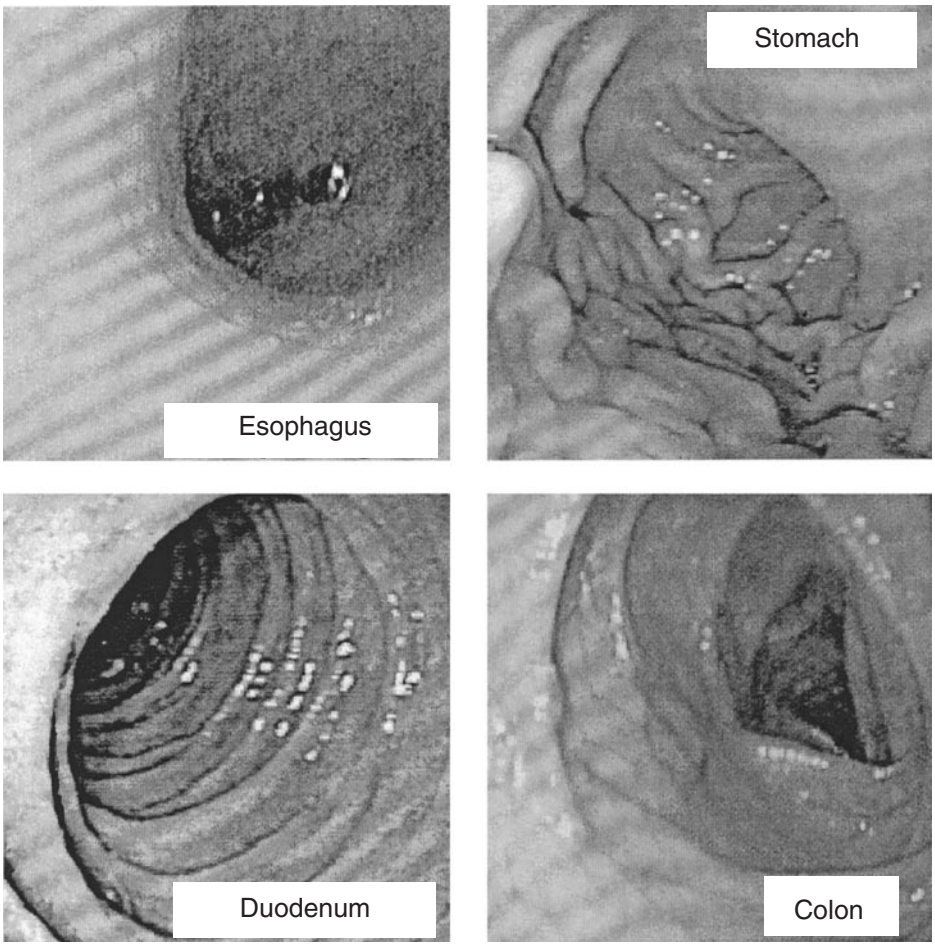


FIGURE 1 Endoscopic survey of the gastrointestinal tract. Endoscopic views of the esophagus, stomach, duodenum, and colon. The squamous mucosa of the esophagus is smooth and pink, without folds. In contrast, the gastric rugae are prominent. These stretch and flatten when the stomach is distended. In the duodenum, the plicae circularis are circumferential. The villi are appreciated as a velvety mucosal surface. The colonic mucosal folds are not circumferential and are tethered at sites corresponding to the teniae coli, resulting in a triangular profile.

third draining into the portal vein (via the gastric veins). With this rich vasculature, ischemia of the esophagus is uncommon. Unfortunately, this also makes the vascular network accessible to esophageal tumors and vulnerable to early local metastasis.

The esophagus begins in the pharynx, at the cricoid cartilage, and passes through the posterior mediastinum behind the aortic arch and left main stem bronchus. The esophagus then courses anterior to the aorta as it passes through the diaphragm, to the gastroesophageal junction. From the lumen outward, the esophageal wall is composed of the mucosa, submucosa, muscularis propria, and adventitia, respectively.

The principal esophageal function, transport of substances from the oral cavity to the stomach, is accomplished by coordinated contraction of the muscular layers. Like the entire gastrointestinal tract, the muscularis propria is composed of inner circular and outer longitudinal layers. Unlike the remainder of the gastrointestinal tract, the muscle layers include both skeletal (voluntary) and smooth (involuntary) muscle. Muscle in the upper third of the esophagus is composed primarily of skeletal muscle, while muscle in the middle third of the esophagus is a mixture of skeletal and smooth muscle, and that in the distal third is entirely smooth muscle. Thus, when food or liquid is ingested, a swallow is initiated voluntarily by coordinated relaxation and contraction of upper esophageal skeletal muscle. This continues involuntarily, as food is propelled via waves of contraction into the distal esophagus

and stomach. Auerbach's plexus lies between the circular and longitudinal muscles and generates the neural signals that control these contractions.

Entry into and exit from the esophagus is restricted by contractile muscle bundles termed the upper and lower esophageal sphincters, respectively. At rest, both sphincters are contracted, or closed; the upper esophageal sphincter closes to prevent air entry and the lower esophageal sphincter closes to prevent reflux of gastric contents. Relaxation of each sphincter must occur in coordination with a swallow to allow the appropriate passage of food and liquids. Failure of lower esophageal sphincter relaxation prevents passage of food from the esophagus into the stomach, resulting in the disease known as achalasia. In contrast, insufficient contraction of the lower esophageal sphincter results in gastroesophageal reflux and esophagitis. Commonly used substances, such as caffeine, alcohol, and nicotine, can prevent complete closure of the lower esophageal sphincter and cause or aggravate gastroesophageal reflux disease (GERD).

The esophageal lining is normally pink, moist, and covered by stratified squamous epithelium, much like the skin (Fig. 2). This ends abruptly at the Z-line, which marks the transition to the columnar epithelium of the stomach. The submucosa of the esophagus contains modified salivary glands that secrete mucus (for lubrication), growth factors (to augment epithelial cell growth and aid in repair), and, in the distal esophagus, bicarbonate (to neutralize acids).

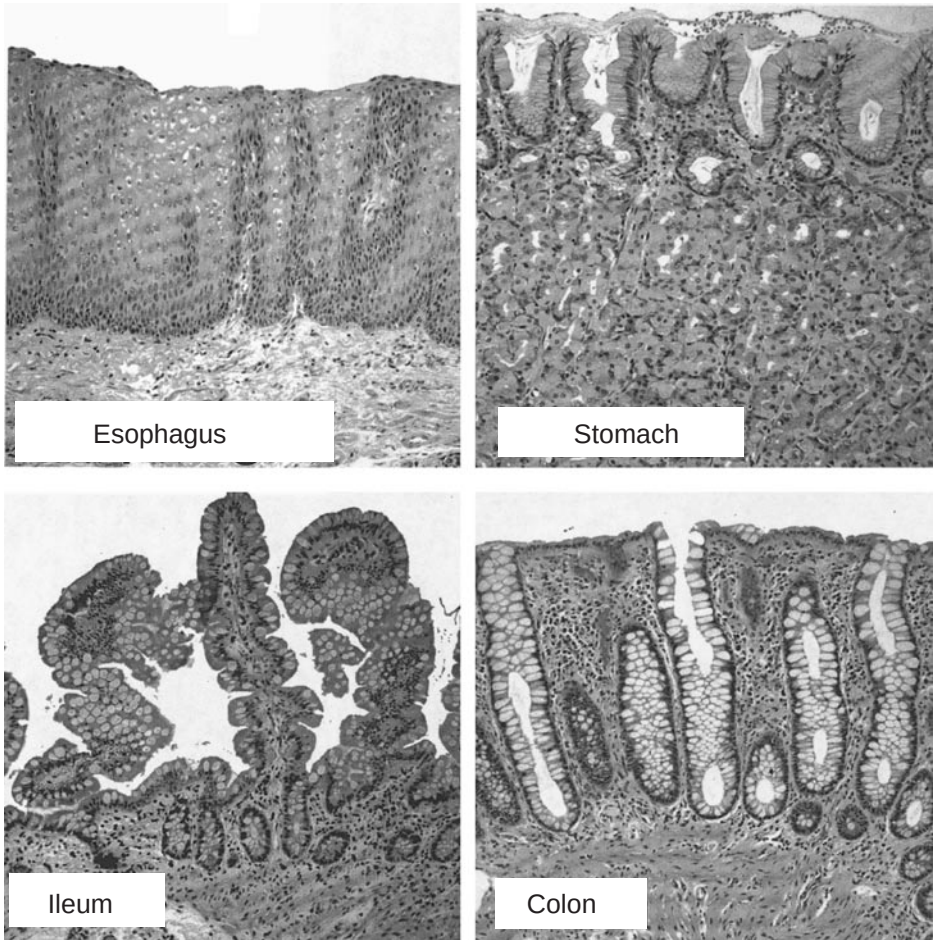


FIGURE 2 Histology of the esophagus, stomach, ileum, and colon. The esophagus is lined by a multilayered squamous epithelium, with the basal proliferative zone represented by the darker region at the bottom of the epithelium. The gastric body is lined superficially by mucus cells. Specialized parietal and chief cells are seen in deeper layers of the mucosa, in the glandular compartment. The ileum of the small intestine is distinguished by the presence of villi. These are absent in the colonic mucosa, which is dominated by crypts that are normally arranged in a uniform array.

STOMACH

The stomach is a distensible saccular portion of the gastrointestinal tract between the esophagus and the small intestine. Food enters the stomach through the gastroesophageal junction. While in the stomach, food is mixed with gastric acid and digestive enzymes to form chyme. Thus, the stomach serves, in part, as a reservoir, holding ingested food and chyme and delivering the latter to the small intestine at a controlled rate. As chyme exits the stomach, the narrow pyloric channel acts as a sieve, preventing large food particles from entering the duodenum until they are adequately processed.

The stomach is located in the left upper quadrant of the abdomen. In general, the stomach is J-shaped, but there is considerable variation based on the volume of the gastric contents. Sites within the stomach are commonly described based on two curvatures; the lesser curvature of the stomach forms the right upper border and the greater curvature forms the left lower border. The stomach receives arterial circulation from branches of the celiac artery.

The mucosal surface of the stomach forms prominent folds, or rugae, in an empty state (see Fig. 1). The gastric wall consists of four layers: mucosa, submucosa, muscularis propria, and serosa. The mucosa contains the secretory glands. The submucosa is mainly connective tissue containing lymphocytes, plasma cells, and neurovascular elements. The muscularis propria contains three layers: longitudinal fibers, circular fibers, and oblique fibers. The circular fibers course around the body of the stomach and thicken at the exit from the stomach, forming the pyloric sphincter.

The stomach can be divided into cardia, fundus, body, antrum, and pylorus. The cardia is a small section of the stomach located next to the gastroesophageal (GE) junction, just left of the midline. The Z-line marks the abrupt mucosal transition from the esophagus to the cardia. The fundus is a dome-shaped region projecting upward and to the left of the GE junction. The body is the largest section, beginning below the fundus and extending to the incisura angulus, a notch in the lesser curvature. The glands of the body and fundus are composed of parietal, chief, mucus, and endocrine cells (Fig. 2). Parietal cells secrete acid as well as intrinsic factor, which is necessary for the absorption of vitamin B_{12} in the terminal ileum. Lack of intrinsic factor causes an inability to absorb vitamin B_{12} in the ileum, resulting in pernicious anemia. The antrum extends from the incisura angulus and its border with the body to the pylorus. The pylorus contains the pyloric sphincter, a thick ring of muscle that regulates release of gastric contents into the duodenum. Hypertrophy of this muscle results in the inability of food to pass, resulting in projectile vomiting, and is most commonly seen in infants. The antrum, similar to the cardia, contains mainly mucus-secreting cells, but also includes endocrine and gastrin-secreting cells. The vagus nerve is one source that stimulates this secretion.

SMALL INTESTINE

The small intestine receives the contents of the stomach and is the primary site of nutrient absorption. It consists of three parts, duodenum, jejunum, and ileum, for a total length of about 6 m in the adult. The celiac trunk provides the arterial supply to the proximal duodenum and the superior mesenteric artery supplies the distal duodenum, jejunum, and ileum.

From the lumen outward, the small intestinal wall consists of four layers: mucosa, submucosa, muscularis propria, and serosa. Like the other sections of the gastrointestinal tract, the muscularis propria contains outer longitudinal and inner circular muscle layers, with the ganglion cells of Auerbach's plexus between the layers. The submucosa contains lymphatics, connective tissue, and Meissner's plexus. The submucosa also contains Brunner's glands in the duodenum and Peyer's patches in the ileum.

The duodenum is 25–30 cm in length and makes a C-shape to curve around the head of the pancreas. At the proximal end of the duodenum is the duodenal bulb, which has a smooth, featureless mucosal surface. The duodenum can be distinguished histologically by the presence of Brunner's glands, which secrete mucus. Distal to the duodenal bulb, the small intestinal mucosa is thrown into circular folds, termed plicae circulares, which decrease in number distally (Fig. 1). The ampulla of Vater, which drains secretions from the pancreatic and biliary ducts, is located in the second portion of the duodenum. Ampullary obstruction, as occurs with impacted gallstones or tumors, can result in pancreatitis.

The ligament of Treitz fixes the junction of the duodenum to the jejunum and marks the division of the upper and lower gastrointestinal tracts. Lymphoid follicles, present throughout the small intestinal mucosa, are most prominent in the ileum, where they form aggregates known as Peyer's patches. Intestinal infection may result in significant hyperplasia of these lymphoid nodules. The small intestine ends at the terminal ileum, where the ileocecal valve, made of two mucosal folds, forms the entrance into the colon.

The epithelium of the small intestine, consisting of crypts and villi (Fig. 2), is specialized to maximize

absorption. Villi are fingerlike projections into the lumen, where absorption occurs. They are covered by absorptive enterocytes (intestinal cells) and amplify their surface area further with a microvillus brush border. Blunting of villi, as occurs in celiac disease, results in significant malabsorption. Similarly, small intestinal infections can cause malabsorption and diarrhea. The brush border contains digestive enzymes and the transporters and ion channels necessary for efficient nutrient absorption. The crypts contain Paneth and endocrine cells as well as stem cells that differentiate as they migrate toward the villus. The majority of nutrient absorption takes place in the duodenum and jejunum. However, as noted previously, the ileum is essential for absorption of vitamin B_{12} and bile salts.

COLON

The colon begins at the ileocecal valve and continues through the rectum. It is approximately 1–1.5 m in length and is divided into cecum, ascending colon, transverse colon, descending colon, sigmoid colon, and rectum. Although the small intestine is mobile, the colon is relatively fixed. Significant landmarks include the hepatic flexure (adjacent to the liver), where the ascending colon joins the transverse colon, and the splenic flexure (adjacent to the spleen), where the transverse colon joins the descending colon. The main functions of the colon are absorption of water and electrolytes and storage of waste products before excretion.

The layers of the colonic wall are similar to those of the small intestine. Unlike the small intestine, however, the longitudinal muscle is organized into three separate bands, the teniae coli. These run from the appendix to the rectum. The outpouchings between the teniae are called haustra (Fig. 1). The superior mesenteric artery supplies the colon from the cecum to the proximal transverse colon. The inferior mesenteric artery supplies the remainder of the colon, excluding the rectum, which is supplied by rectal, or hemorrhoidal, vessels.

The cecum is the first section of the colon and is dilated relative to the remainder of the colon. It contains the appendix, a small, blind pouch enriched in mucosal lymphoid tissue. Inflammation of this tissue, resulting from retained material, results in appendicitis, a surgical emergency. The cecum is relatively mobile and therefore is more susceptible to volvulus, or twisting upon itself. Because it is dilated, a mass lesion or tumor in the cecum may go undetected for a long period of time.

The ascending colon extends from the cecum upward toward the liver, along the right side of the abdomen. At the hepatic flexure, it joins the transverse colon, which can be mobile as it drapes across the abdominal cavity toward the splenic flexure. Here the transverse colon joins the descending colon that courses along the left abdominal cavity to join the sigmoid colon in the lower abdomen. The sigmoid colon is S-shaped. It is the narrowest region of the colon, thus mass lesions in this area are generally symptomatic early.

The epithelium of the colon, unlike that of the small intestine, lacks villi. The mucosa consists of crypts containing goblet cells and absorptive enterocytes. Because the small intestine, rather than the colon, is responsible for the bulk of fluid absorption, the diarrhea associated with colonic disease is not as dramatic as that seen with small intestinal disease.

During development, ganglion cells of both Auerbach's and Meissner's plexi migrate in a proximal to distal direction. Failure of this migration results in Hirschprung's disease. The distal, aganglionic, segment of colon lacks inhibitory input and is tonically contracted, without normal peristaltic contractions. This results in obstruction and requires surgical removal of the aganglionic segment.

See Also the Following Articles

Biliary Tract, Anatomy • Colon, Anatomy • Duodenum, Anatomy • Esophagus, Anatomy • Liver, Anatomy • Pancreas, Anatomy • Peritoneum, Anatomy and Development • Rectum, Anatomy • Salivary Glands, Anatomy • Small Intestine, Anatomy • Stomach Anatomy

Further Reading

Antonioli, D. A., and Madara, J. L. (1998). Functional anatomy of the gastrointestinal tract. *In* "Pathology of the Gastrointestinal Tract" (S.-C. Ming and H. Goldman, eds.), 2nd Ed., pp. 13–34. Williams and Wilkins, Baltimore.

Keljo, D. J., and Gariepy, C. E. (2002). Anatomy, histology, embryology, and developmental anomalies of the small and large intestine. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, M. H. Sleisenger, and B. F. Scharschmidt, eds.), 7th Ed., pp. 1643–1647. W. B. Saunders, Philadelphia.

Redel, C. A. (2002). Anatomy, histology, embryology, and developmental anomalies of the stomach and duodenum. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, M. H. Sleisenger, and B. F. Scharschmidt, eds.), 7th Ed., pp. 675–682. W. B. Saunders, Philadelphia.

Gastrostomy

ZARA COOPER AND MALCOLM K. ROBINSON
Harvard Medical School

abdomen Portion of the body that lies between the chest and the pelvis and contains several organs, including the stomach, intestines, liver, spleen, pancreas, kidneys, and bladder.
gastrostomy Artificial opening in the stomach wall, usually created by surgical means.
gastrostomy tube Flexible, rubberlike tube that is inserted through a gastrostomy into the stomach.
stomach Part of the gastrointestinal tract that lies between the esophagus and first part of the small intestine; serves as a reservoir for swallowed food and is where food digestion is initiated. Although the terms "stomach" and "abdomen" are frequently used interchangeably, they are distinct entities and should not be confused.
volvulus Twisting or knotting of the stomach or intestines; obstructs passage of fluid and food.

A gastrostomy is a surgical procedure that creates an artificial opening in the stomach wall. Insertion of flexible tubing into the opening provides a means to introduce nutrients directly into the stomach, to drain stomach fluids, and to repair gastric volvulus.

GASTROSTOMY AND GASTROSTOMY TUBES

The medical procedure known as gastrostomy involves cutting a hole cut into the stomach wall with surgical instruments or a needle. An artificial opening is thus obtained. A flexible, rubberlike tube known as a "gastrostomy tube" (g-tube) can be inserted through this opening such that one end of the tube lies within the stomach cavity. This end is usually enlarged and shaped like a mushroom to prevent the tube from coming out of the stomach inadvertently. The other end of the g-tube is narrower and is pulled through another surgical opening that is made in the wall of the abdomen. This procedure thus creates a conduit between the stomach and the outside of a person's abdomen (see Fig. 1). Gastrostomy and insertion of a gastrostomy tube go hand in hand, and thus these terms are frequently used interchangeably.

MEDICAL INDICATIONS FOR GASTROSTOMY

There are three major medical indications or reasons for insertion of a gastrostomy tube. First, gastrostomy tubes are used in people who have a normal stomach and intestines but are unable to take food and medications by mouth for a prolonged period of time (e.g., 6 or more weeks). This is the most common reason for a gastrostomy. Insertion of a g-tube for this indication may be appropriate, for example, in persons who are in the intensive care unit and sedated, in those with dementia who refuse to eat, and in patients who have cancer of the throat. In such people, liquid food and medications are infused via the g-tube directly into the stomach for

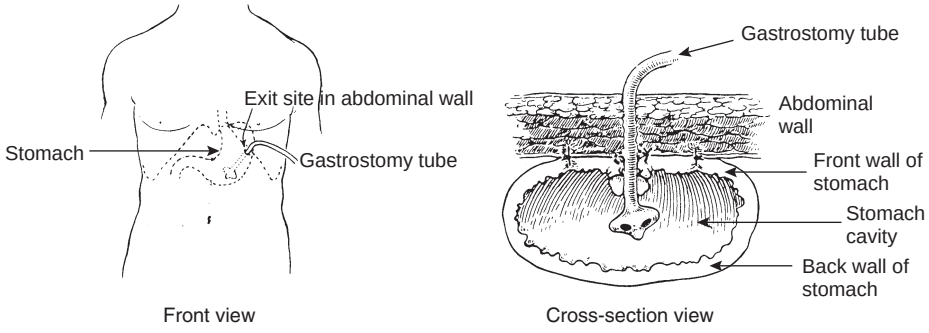


FIGURE 1 Gastrostomy tube.

 Copyright 2004, Elsevier (USA). All rights reserved.

digestion and subsequent passage into the intestines, where further digestion and absorption occur.

Second, gastrostomy tubes are used to drain fluid from the stomachs of those patients whose stomachs cannot propel food and secretions into the intestines. This may occur if the stomach or intestines are blocked (e.g., by a cancerous mass) or if the stomach and intestines stop contracting. Buildup of fluid and food in the stomach due to blockage or poor contraction eventually leads to vomiting. This may be avoided by draining fluid from the stomach via a g-tube. Fluid can accumulate in the stomach under these abnormal conditions even without eating. This is because fluid is constantly secreted into the stomach and intestines regardless of food consumption.

Finally, g-tube insertion may be indicated as part of treatment for a gastric volvulus. In the treatment of this condition, the stomach is untwisted and is then "fixed" in an untwisted position by inserting a g-tube through the abdominal wall and into the stomach via a gastrostomy. Recurrent volvulus is prevented because the stomach is no longer mobile enough to twist on itself. Volvulus of the stomach is a rare condition, and therefore this is an uncommon indication for g-tube insertion.

See Also the Following Articles

Gastric Surgery • Gastric Volvulus • Gastroenterostomy • Volvulus

Further Reading

Sax, H. C. (1998). Nutritional support in the critically ill. In "Current Surgical Therapy" (J. L. Cameron, ed.), 6th Ed. pp. 1143–1145. Mosby, New York.

Zollinger, R. M., and Zollinger, R. M., Jr. (1993). Gastrostomy. *In* "Atlas of Surgical Operations," 7th Ed., p. 30. Macmillan, New York.

Genetic Counseling and Testing

ERIC ROSENTHAL
University of California, San Diego

genetic counselor Professional with master's-level training in medical genetics, counseling, and the psychosocial/legal issues associated with inherited disorders and genetic testing. Most practicing genetic counselors are board certified by the American Board of Genetic Counseling.

genetic testing Analysis of chromosomes, genes, and/or gene products (e.g., proteins or enzymes) to determine whether a genetic alteration related to a specific disease or condition is present in an individual.

hereditary cancer syndrome Collection of clinical features, including cancers, attributable to an alteration in a single gene that can be passed from parents to their children.

Genetic counseling is a communications process dealing with the medical and psychosocial problems associated with the presence of, or the risk for, a genetic disorder. Genetic counseling is an essential element in the management of patients with hereditary cancer syndromes, particularly when genetic testing is offered. Genetic counseling and testing for hereditary cancer syndromes are best provided by genetic counselors, genetics nurses, and physicians with specialized expertise and experience in dealing with the unique issues and challenges associated with this rapidly evolving area of medicine.

INTRODUCTION

Recent years have seen an explosion of knowledge concerning the genetic basis of hereditary cancer syndromes (HCSs). An ever-increasing number of genetic tests are available to identify gene alterations that can increase lifetime risks for gastrointestinal tumors to as high as 100%. Utilization of these tests can identify patients who will benefit from increased surveillance, prophylactic surgery, chemoprevention, lifestyle and diet modifications, or other targeted interventions with the potential to reduce morbidity and mortality from cancer. However, the decision to utilize genetic testing should not be made without careful

Copyright 2004, Elsevier (USA). All rights reserved.

digestion and subsequent passage into the intestines, where further digestion and absorption occur.

Second, gastrostomy tubes are used to drain fluid from the stomachs of those patients whose stomachs cannot propel food and secretions into the intestines. This may occur if the stomach or intestines are blocked (e.g., by a cancerous mass) or if the stomach and intestines stop contracting. Buildup of fluid and food in the stomach due to blockage or poor contraction eventually leads to vomiting. This may be avoided by draining fluid from the stomach via a g-tube. Fluid can accumulate in the stomach under these abnormal conditions even without eating. This is because fluid is constantly secreted into the stomach and intestines regardless of food consumption.

Finally, g-tube insertion may be indicated as part of treatment for a gastric volvulus. In the treatment of this condition, the stomach is untwisted and is then "fixed" in an untwisted position by inserting a g-tube through the abdominal wall and into the stomach via a gastrostomy. Recurrent volvulus is prevented because the stomach is no longer mobile enough to twist on itself. Volvulus of the stomach is a rare condition, and therefore this is an uncommon indication for g-tube insertion.

See Also the Following Articles

Gastric Surgery • Gastric Volvulus • Gastroenterostomy • Volvulus

Further Reading

Sax, H. C. (1998). Nutritional support in the critically ill. In "Current Surgical Therapy" (J. L. Cameron, ed.), 6th Ed. pp. 1143–1145. Mosby, New York.

Zollinger, R. M., and Zollinger, R. M., Jr. (1993). Gastrostomy. *In* "Atlas of Surgical Operations," 7th Ed., p. 30. Macmillan, New York.

Genetic Counseling and Testing

ERIC ROSENTHAL
University of California, San Diego

genetic counselor Professional with master's-level training in medical genetics, counseling, and the psychosocial/legal issues associated with inherited disorders and genetic testing. Most practicing genetic counselors are board certified by the American Board of Genetic Counseling.

genetic testing Analysis of chromosomes, genes, and/or gene products (e.g., proteins or enzymes) to determine whether a genetic alteration related to a specific disease or condition is present in an individual.

hereditary cancer syndrome Collection of clinical features, including cancers, attributable to an alteration in a single gene that can be passed from parents to their children.

Genetic counseling is a communications process dealing with the medical and psychosocial problems associated with the presence of, or the risk for, a genetic disorder. Genetic counseling is an essential element in the management of patients with hereditary cancer syndromes, particularly when genetic testing is offered. Genetic counseling and testing for hereditary cancer syndromes are best provided by genetic counselors, genetics nurses, and physicians with specialized expertise and experience in dealing with the unique issues and challenges associated with this rapidly evolving area of medicine.

INTRODUCTION

Recent years have seen an explosion of knowledge concerning the genetic basis of hereditary cancer syndromes (HCSs). An ever-increasing number of genetic tests are available to identify gene alterations that can increase lifetime risks for gastrointestinal tumors to as high as 100%. Utilization of these tests can identify patients who will benefit from increased surveillance, prophylactic surgery, chemoprevention, lifestyle and diet modifications, or other targeted interventions with the potential to reduce morbidity and mortality from cancer. However, the decision to utilize genetic testing should not be made without careful

Copyright 2004, Elsevier (USA). All rights reserved.

consideration. This process is time consuming and requires the participation of professionals with specialized knowledge and skills.

HEREDITARY CANCER SYNDROMES

A partial list of HCSs with gastrointestinal (GI) involvement is presented in Table I. HCSs are caused by mutations in genes that play a role in preventing cells from becoming malignant. These genes are often involved in DNA repair and/or the regulation of cell division. When a person inherits an altered version of one of these genes, from either parent, certain tissues in their body are more susceptible to the random events that cause cells to become cancerous. Lifetime risks for various malignancies vary dramatically depending on the gene involved. The exact form of the mutation within the gene may also have an effect. Some mutations confer a risk for a specific cancer that is only slightly increased over that for the average population. Other mutations in the same gene may generate lifetime risks for certain cancers that approach 100%, as is the case for colon cancer and certain adenomatous polyposis *APC* gene mutations.

HCSs are usually inherited in a dominant fashion, which means that an individual is affected when they inherit an altered gene from only one of their parents. Each child of an affected parent has a 50% chance of inheriting the syndrome. New mutations arise periodically, which accounts for affected individuals who seem to be the first in their family to display signs of the syndrome. Other explanations, such as inaccuracies in the family history, or misattributed parentage, should also be considered in these cases.

RECOGNIZING SIGNS OF A HEREDITARY CANCER SYNDROME

The presence of a hereditary cancer syndrome is suspected when one or more of the following factors present in a family:

1. Cancer at unusually young ages.
2. Similar cancers, or cancers known to be related, occurring in several relatives, across more than one generation.
3. Individuals with multiple primary cancers in the same or different organs.

TABLE I Representative Hereditary Cancer Syndromes with GI Involvement

Name	Major features	Incidence	Gene(s) responsible
Hereditary nonpolyposis colon cancer	Colorectal cancer, endometrial cancer, ovarian cancer, gastric cancer, transitional cell cancer of the ureter and renal pelvis, other cancers, sebaceous adenomas	Estimates vary; ~5% of colorectal cancer	*MLH1*, *MSH2*, *MSH6*, *PMS1*, *PMS2*, others?
Familial adenomatous polyposis	>100 adenomatous colorectal polyps, colorectal cancer, cancer of the ampulla of Vater, duodenal cancer, epidermoid cysts, desmoid tumors, osteomas, congenital hypertrophy of the retinal pigment epithelium	1 : 10,000; ~1% of colorectal cancer	*APC*
Cowden syndrome, or PTEN hamartoma tumor syndrome	Hamartous polyps of the colon and stomach, breast cancer, papillary or follicular thyroid cancer, mucocutaneous lesions, endometrial fibroids, macrocephaly	1 : 300,000	*PTEN*
Peutz–Jeghers syndrome	Hamartous gastrointestinal polyps; cancer of the colon, stomach, small intestine, and pancreas; ovarian sex cord tumors; adenoma malignum of the cervix; pigmented skin lesions	1 : 200,000	*LKB1*, others not yet identified
Familial atypical mole–malignant melanoma	Dysplastic nevi, melanoma, pancreatic cancer	Unknown	*p16* (*CDKN2*), others not yet identified

4. The presence of rare or unusual cancers.
5. The presence of precancerous lesions (i.e., adenomatous polyps) at unusually young ages or in unusually large numbers.
6. The presence of other clinical findings consistent with known HCSs (i.e., desmoid tumors and congenital hypertrophy of the retinal pigment epithelium in familial adenomatous polyposis).

ELEMENTS OF A GENETIC COUNSELING ASSESSMENT FOR A HEREDITARY CANCER SYNDROME

The genetic counseling process should begin with an exploration of the patient's concerns, goals, and cancer risk perception. Many patients grossly over- or underestimate their cancer risks and this can impact their compliance with screening recommendations. Even when no testing is performed, genetic counseling can have a significant impact by providing reassurance or raising awareness.

The value of the risk assessment is dependent on the accuracy and completeness of the medical and family history information provided by the patient, and whenever possible, verified by medical records. Construction of at least a three-generation pedigree is the standard of care, including information about types of primary cancers, age of diagnosis for all primary cancers, and current age or age at death for all unaffected as well as affected relatives. Information about clinical features other than malignancy is essential, so patients must be queried about surgeries, hospitalizations, unusual skin features, and chronic use of medications. Patients evaluated for GI cancer syndromes should always be asked about the results of sigmoidoscopies, colonoscopies, or other examinations of the colon. Ethnic background is also important, because certain conditions are more common in some groups.

Family history information is used to predict the likelihood that the pattern of cancers in a family is due to a HCS, and the probability that genetic testing will find a mutation in one of the known genes. These analyses are based on data that are being updated at a rapid pace, which is one of the reasons why professionals engaging in genetic counseling for HCSs must devote a heroic effort to keep up with current information.

MAKING DECISIONS ABOUT GENETIC TESTING

A positive genetic test result, indicating that a patient has a high risk for certain malignancies, can provide valuable information to guide management. A negative test result can reduce anxiety and obviate the need for expensive and uncomfortable screening in family members whose risk is not elevated. Nevertheless, genetic testing is not feasible for many patients referred for genetic counseling. The family history may not be consistent with any of the known HCSs. Many patients lack insurance coverage for genetic testing, especially those in Medicare and Medicaid programs. Clinical testing is not yet available for genes, i.e., the PMS1 and PMS2 genes involved in hereditary non-polyposis colon cancer. Testing through research protocols is sometimes possible, but this approach requires that patients understand the limitations of investigational testing.

Patients may decline testing because of concerns that positive results will adversely impact their access to health and life insurance, or employment. State and federal laws provide some protection, but the situation remains unsettled, and these are legitimate concerns that often require lengthy discussion. Patients who wish to have their test results remain confidential may chose to pay laboratory costs "out of pocket" and request that no record be placed in their medical charts.

The value of genetic testing is equivocal in situations in which results will not alter medical management. For example, there are no proved effective interventions to reduce mortality in people with a hereditary risk for pancreatic cancer. Some individuals who have already had a colectomy following a diagnosis of colon cancer have little interest in establishing that they inherited a risk for this malignancy.

LIMITATIONS TO GENETIC TESTING TECHNOLOGY

It is imperative that patients and health professionals understand certain limitations and caveats of current genetic testing, including the following considerations:

1. Many genes responsible for HCSs have not yet been identified.
2. The testing technology used in clinical laboratories is not able to detect all of the clinically significant alterations in those genes that have been identified.
3. The testing often identifies gene changes of uncertain clinical significance, i.e., amino acid substitutions that may or may not affect protein function. The affect of these "variants" on cancer risk is often unknown.
4. Genetic test results can provide information only about the probability that an individual will develop certain cancers. There is enormous variability in the

clinical manifestations of HCSs between families and between members of the same family.

5. It is vital that patients understand that even those individuals who do not have a mutation in one of the genes that cause a HCS still have same risk for malignancy as others in the general population.

GENETIC TESTING AND THE FAMILY

Genetic testing is distinguished from other areas of medicine in the extent to which family involvement impacts care of the patient. This is immediately apparent when patients have difficulty providing a complete and accurate family history. Medical issues are not discussed in some families, and the absence of an accessible "family historian" can stymie efforts to gather the necessary information. It is not uncommon for some family members to regard the investigation of genetic cancer risks as "a can of worms." They may be uncooperative or actively hostile toward a relative seeking detailed health information, and they are unlikely to permit the release of necessary medical records over which they have control. Professionals providing genetic counseling for HCSs must be prepared to help patients deal with this issue.

The importance of family involvement is even more pronounced when the patient seeking information is an "at-risk" family member who has not yet manifested clinical features of the suspected syndrome. A negative genetic test result in such an individual is of limited value unless the family history of cancers has already been traced to a gene mutation in a relative who has had cancer or other signs of the syndrome. Otherwise, the patient must be counseled that the family history could be due to a mutation in a different gene (including those that have not yet been discovered), or a mutation that is not identifiable with current technology.

Following testing, there may be issues about communicating results within the family. Occasionally, patients refuse to share information about hereditary risks with relatives, and some affected parents may even balk at telling their own children that they are at a 50% risk of having inherited a HCS. The physician's moral and legal responsibilities in such situations are still under debate. It is best to discuss plans for sharing information within the family prior to initiating genetic testing.

See Also the Following Articles

Cancer, Overview • Familial Adenomatous Polyposis (FAP) • Familial Risk of Gastrointestinal Cancers • Lynch Syndrome/ Hereditary Non-Polyposis Colorectal (HNPCC)

Further Reading

Cummings, S. (2000). The genetic testing process: How much counseling is needed? *J. Clin. Oncol.* **18**, 60s–64s.

Fearnhead, N. S., Britton, M. P., and Bodmer, W. F. (2001). The ABC of APC. *Hum. Mol. Genet.* **10**, 721–733.

Hampel, H., and Peltomaki, P. (2000). Hereditary colorectal cancer: Risk assessment and management. *Clin. Genet.* **57**, 89–97.

Ivanovich, J. L., Read, T. E., Ciske, M. S., Kodner, I. J., and Whelan, A. J. (1999). A practical approach to familial and hereditary colorectal cancer. *Am. J. Med.* **107**, 68–77.

Lindor, N. M., Greene, M. H. and The Mayo Familial Cancer Program (1998). The concise handbook of family cancer syndromes. *J. Natl. Cancer Inst.* **90**, 1039–1071.

Stopfer, J. E. (2000). Genetic counseling and clinical cancer genetics services. *Semin. Surg. Oncol.* **18**, 347–357.

Vasen, H. F. A., Gruis, N. A., Frants, R. R., van der Velden, P. A., Hille, E. T. M., and Bergman, W. (2000). Risk of developing pancreatic cancer in families with familial atypical multiple mole melanoma associated with a specific 19 deletion of p16. *Int. J. Cancer* **87**, 809–811.

Giardiasis

FRANCES D. GILLIN AND LARS ECKMANN
University of California, San Diego

excystation Emergence of trophozoites from cysts.
trophozoites Metabolically active, proliferating forms of protozoa.

Giardiasis, caused by the protozoan parasite *Giardia lamblia*, is one of the most important waterborne diarrheal diseases, worldwide. Although its life cycle can be reproduced *in vitro* and its small (~12 Mb) genome has been sequenced, neither the basic biology of *G. lamblia* nor the pathophysiology of giardiasis is completely understood.

PARASITE BIOLOGY

Since *Giardia lamblia* neither invades nor secretes any known toxin, its adaptations to survival in hostile environments and ability to differentiate are key virulence factors. The dormant ovoid cyst form of *G. lamblia*, which is responsible for transmission of giardiasis, persists in fresh cold water for weeks and is relatively resistant to common disinfectants. Infection is initiated by ingestion of as few as 10 cysts in fecally contaminated water or occasionally food. Exposure of the quadrinucleate cysts to gastric acid initiates excystation. In the small intestine, the parasite emerges and divides into two binucleate trophozoites that attach to the apical surface of the epithelium. In humans, *G. lamblia* generally colonize below the entrance of the common bile duct. When trophozoites are carried downstream by the flow of intestinal fluid, they differentiate into infectious cysts that are passed in the feces. Trophozoites are ~10 μm long and shaped like a half pear, with an attachment disc on their ventral surface. Trophozoites also use their eight flagella to "swim" in the lumenal fluid and remain in the small intestine. *G. lamblia* lacks many *de novo* biosynthetic pathways and scavenges most of its purines, pyrimidines, amino acids, and lipids from the host. This parasite is amitochondriate, with a microaerophilic metabolism, using glucose and arginine for energy. *Giardia* belongs to one of the earliest lineages to branch from the eukaryotic line of descent, which makes it a valuable model for elucidating the biological innovations associated with the evolutionary appearance of nucleated cells.

EPIDEMIOLOGY, DISEASE, AND TREATMENT

In the United States and other developed countries, hikers and campers, families with children in day care, and travelers are at particular risk for giardiasis. Infections are much more common in most developing countries. In some areas, nearly 100% of children may be infected before the age of 2 years. Both the duration and the severity of giardiasis are highly variable. Approximately half of infected people are asymptomatic and the infection frequently resolves spontaneously within ~2 weeks. However, *G. lamblia* can cause severe and protracted diarrhea, which may lead to villus blunting and malabsorption, resulting in malnutrition and weight loss, especially in children. Importantly, trophozoites do not invade or secrete any known toxin or cause a mucosal inflammatory response. Trophozoites are covered with a dense layer of a cysteine-rich protein that is immunogenic, but also highly variable. Since the genome encodes >150 of these variable surface proteins, they are poor vaccine candidates. Moreover, antigenic variation is likely important for chronicity of infections and susceptibility to re-infection. A crude veterinary vaccine is available for cats, but the relevant antigens are not known. Repeated infections appear to confer some immune protection in humans. In endemic areas, adults are less frequently infected than children or than visiting adults from non-endemic areas. Experimental studies suggest that both B-cell-independent and B-cell-dependent host defenses, especially secretory immunoglobulin A, play important roles in controlling and clearing *G. lamblia*. Treatment with 5-nitroimidazoles, such as metronidazole (Flagyl), is effective in >85% of cases. *G. lamblia* is sensitive to these drugs because of its anaerobic metabolic pathways, which are not present in the host. Metronidazole is activated by

 Copyright 2004, Elsevier (USA). All rights reserved.

these pathways to a toxic form that likely acts by binding to and damaging parasite DNA. Clinically relevant drug resistance has not been documented.

See Also the Following Articles

Diarrhea • Diarrhea, Infectious • Immunodeficiency • Malabsorption • Parasitic Diseases, Overview • Traveler's Diarrhea

Further Reading

Adam, R. D. (2001). Biology of *Giardia lamblia*. *Clin. Microbiol. Rev.* **14**, 447–475.

Faubert, G. (2000). Immune response to *Giardia duodenalis*. *Clin. Microbiol. Rev.* **13**, 35–54.

Gardner, T. B., and Hill, D. R. (2001). Treatment of giardiasis. *Clin. Microbiol. Rev.* **14**, 114–128.

Glial Cells (Enteric)

ANNE RÜHL

Technical University Munich

capacitative Ca^{2+} entry Influx of extracellular Ca^{2+} across the cellular plasma membrane following the depletion of intracellular Ca^{2+} stores.

class II major histocompatibility complex Highly polymorphic cell surface molecules primarily expressed on B lymphocytes, macrophages, and dendritic cells; they are involved in peptide antigen presentation to $CD4^+$ T lymphocytes (helper T cells), which stimulates humoral immunity and macrophage activation.

Crohn's disease Chronic inflammatory bowel disease; may involve all parts of the gastrointestinal tract.

cytokines Relatively low-molecular-weight proteins produced by leukocytes and other cell types with a broad spectrum of functional activities, mainly regulating inflammatory and immune responses.

intercellular adhesion molecules Assist leukocytes in interacting with their environments through adherence; they are changed on the surfaces of endothelial cells and leukocytes during inflammation and help recruit leukocytes to sites of inflammation.

intermediate filament Part of the cytoskeleton that constitutes a link between the nucleus and the cell surface.

lymphocytic infiltration Depending on the underlying immunopathology, lymphocytes—i.e., a subpopulation of leukocytes—may be the predominant cell type in inflamed tissue; the presence of lymphocytes is suggestive of viral infection and autoimmune and/or chronic inflammation.

nestin Intermediate filament protein characteristic of undifferentiated cells during cortical and enteric nervous system development.

neural crest Ectodermal embryonic structure that arises from the neural primordium lateral borders when they form the neural tube; source of virtually all neurons and glial cells of the peripheral nervous system.

T cells Thymus-derived immune cells that are essential for cell-mediated immune responses.

Glial cells found in the enteric nervous system differ from glial cells found in other compartments of the central and peripheral nervous system. Although within the population of enteric glial cells there is a relative homogeneity, there is evidence to suggest some subsets have diverse functions. Enteric glia serve as a supportive metabolic and structural framework, but also play a role in information transfer and inflammation.

STRUCTURE, GROWTH, AND DEVELOPMENT

Morphology

The enteric nervous system (ENS) has unique structural and functional properties that have led to the concept of a "little brain" in the gut, serving as an independent center of integrative neural activity. It is organized as two ganglionated plexuses, the myenteric plexus and the submucous plexus, which contain no connective tissue or blood vessels but consist exclusively of densely packed neurons, glial cells, and their

Copyright 2004, Elsevier (USA). All rights reserved.

these pathways to a toxic form that likely acts by binding to and damaging parasite DNA. Clinically relevant drug resistance has not been documented.

See Also the Following Articles

Diarrhea • Diarrhea, Infectious • Immunodeficiency • Malabsorption • Parasitic Diseases, Overview • Traveler's Diarrhea

Further Reading

Adam, R. D. (2001). Biology of *Giardia lamblia*. *Clin. Microbiol. Rev.* **14**, 447–475.

Faubert, G. (2000). Immune response to *Giardia duodenalis*. *Clin. Microbiol. Rev.* **13**, 35–54.

Gardner, T. B., and Hill, D. R. (2001). Treatment of giardiasis. *Clin. Microbiol. Rev.* **14**, 114–128.

Glial Cells (Enteric)

ANNE RÜHL

Technical University Munich

capacitative Ca^{2+} entry Influx of extracellular Ca^{2+} across the cellular plasma membrane following the depletion of intracellular Ca^{2+} stores.

class II major histocompatibility complex Highly polymorphic cell surface molecules primarily expressed on B lymphocytes, macrophages, and dendritic cells; they are involved in peptide antigen presentation to $CD4^+$ T lymphocytes (helper T cells), which stimulates humoral immunity and macrophage activation.

Crohn's disease Chronic inflammatory bowel disease; may involve all parts of the gastrointestinal tract.

cytokines Relatively low-molecular-weight proteins produced by leukocytes and other cell types with a broad spectrum of functional activities, mainly regulating inflammatory and immune responses.

intercellular adhesion molecules Assist leukocytes in interacting with their environments through adherence; they are changed on the surfaces of endothelial cells and leukocytes during inflammation and help recruit leukocytes to sites of inflammation.

intermediate filament Part of the cytoskeleton that constitutes a link between the nucleus and the cell surface.

lymphocytic infiltration Depending on the underlying immunopathology, lymphocytes—i.e., a subpopulation of leukocytes—may be the predominant cell type in inflamed tissue; the presence of lymphocytes is suggestive of viral infection and autoimmune and/or chronic inflammation.

nestin Intermediate filament protein characteristic of undifferentiated cells during cortical and enteric nervous system development.

neural crest Ectodermal embryonic structure that arises from the neural primordium lateral borders when they form the neural tube; source of virtually all neurons and glial cells of the peripheral nervous system.

T cells Thymus-derived immune cells that are essential for cell-mediated immune responses.

Glial cells found in the enteric nervous system differ from glial cells found in other compartments of the central and peripheral nervous system. Although within the population of enteric glial cells there is a relative homogeneity, there is evidence to suggest some subsets have diverse functions. Enteric glia serve as a supportive metabolic and structural framework, but also play a role in information transfer and inflammation.

STRUCTURE, GROWTH, AND DEVELOPMENT

Morphology

The enteric nervous system (ENS) has unique structural and functional properties that have led to the concept of a "little brain" in the gut, serving as an independent center of integrative neural activity. It is organized as two ganglionated plexuses, the myenteric plexus and the submucous plexus, which contain no connective tissue or blood vessels but consist exclusively of densely packed neurons, glial cells, and their

Copyright 2004, Elsevier (USA). All rights reserved.

processes. Enteric glial cells have unique ultrastructural features clearly different from satellite cells of the autonomic ganglia and Schwann cells of peripheral nerves. They do not form basal laminae, they ensheathe axons not individually but in groups, they are extensively coupled by gap junctions, and they make numerous synaptoid contacts with vesicle-containing nerve varicosities.

Enteric glial cells are recognized as being similar to central nervous system astrocytes. This is confirmed by their expression of glial fibrillary acidic protein and vimentin, the calcium-binding protein S-100, glutamine synthetase, and Ran-2. Additionally, enteric glial cells transiently express the intermediate filament nestin during development. Primarily, enteric glial cells have been regarded as a relatively homogeneous cell population. Nevertheless, there is evidence to suggest morphological and functional heterogeneity among these cells, even though distinct subsets have not yet been consistently identified in the mature ENS.

Development and Differentiation

Enteric glia, like enteric neurons, develop from multipotential precursor cells in the neural crest. Enteric neurons evolve from vagal neural crest cells colonizing the entire length of the gut in a proximodistal direction and from sacral neural crest cells migrating in a distal–proximal direction, and most enteric glia also originate from these initial waves of crest-derived cells. However, enteric glia may also arise from glial precursors that enter the gut along with the extrinsic innervation later in development.

During the course of ontogeny of the ENS, nestin-immunoreactive neural crest cells emerge and give rise to both enteric glia and neurons. These cells proliferate in response to two neurotrophic factors, glial cell line-derived neurotrophic factor (GDNF; most likely produced by mesenchymal cells) and neurturin, but not to persephin or endothelin-3. Later on, nestin-positive precursors are replaced by separate lineages of neuronal and glial progenitors and the specified progenitors lose nestin, which is replaced by type-specific cytosolic and intermediate filament proteins. At this stage, GDNF acts as a neurotrophic factor and maintains cells in the neuronal lineage, but does not promote the development of glia. The factor required for further glial development may be a neuregulin, such as glial growth factor (GGF). Glial development in the ENS is also promoted by neurotrophin-3 (NT-3), through its high-affinity receptor TrkC, and the neuropoietic cytokines ciliary neurotrophic factor (CNTF) and leukemia inhibitory factor (LIF), through the α (CNTFRα) and β subunits (gp130 and LIFRβ) of the tripartite CNTF receptor, respectively. In contrast, neural growth factor (NGF) appears not to affect the development of enteric glia.

Postnatally, fibroblast growth factor and GGF seem to increase the division rate of enteric glia. In addition, high GDNF levels in the adult intestine imply an important role for this neurotrophic factor in the maintenance of the adult ENS. Finally, neurotrophins have been implicated in intestinal inflammation and repair processes and it is conceivable that enteric glia may mediate such effects. The high-affinity neurotrophin receptors (NTRs) TrkA (for NGF) and TrkB (for brain-derived neurotrophic factor and the neurotropins 4 and 5) have been found on enteric glial cells from all sections of the adult human gut, and $p75^{NTR}$ receptor immunoreactivity has been demonstrated on enteric glial cells of adult human and rat gut.

Proliferation

Cultured enteric glial cells have been used in *in vitro* studies to understand the functional properties of these cells. Enteric glial cells proliferate in response to compounds that increase intracellular cyclic adenosine monophosphate (cAMP), such as cholera toxin, dibutyryl cAMP, 8-bromo-cAMP, and forskolin, as well as to laminin and fibronectin, but not to epidermal growth factor. Furthermore, it has been suggested that enteric neurons inhibit the proliferation of enteric glia and Schwann cells and that enteric glial cells may depend to some degree on neuronal signals for terminal differentiation. The molecular nature of these signals is as yet unknown.

Inflammatory mediators appear to modulate the proliferation rate of enteric glial cells. Reactive proliferation of enteric glia is documented in animal models of intestinal inflammation. The identity of the inflammatory mediators is unclear. Nevertheless, evidence from *in vitro* studies suggests that the proinflammatory cytokine interleukin-1β (IL-1β) suppresses enteric glial cell proliferation, in contrast to the immunosuppressive cytokine IL-10.

FUNCTION

Electrophysiological Properties and Neuroglial Interactions

Historically, enteric glia have been viewed as a supportive framework—both structurally and metabolically—of the neuronal elements in the ENS. Evidence for a functional role of enteric glial cells in information transfer is now emerging. Enteric glia have

cell-to-cell coupling, suggesting the formation of enteric glial networks in which neuroligand signal transduction systems are linked to increases in cytosolic Ca^{2+} levels. Enteric glial cells respond to endothelins, vasopressin, adenosine and uridine triphosphate (ATP, UTP), epinephrine, serotonin, histamine, and prostaglandins. The presence of purinoceptors of the P_{2Y2} subtype has been reported, and it has been proposed that enteric glia possess an endothelin (ET_B) receptor linked to phospholipase C as well as intracellular inositol 1,4,5-trisphosphate (IP_3) receptors. IP_3-sensitive Ca^{2+} stores appear to predominate in enteric glia. Depletion of internal Ca^{2+} stores induces capacitative Ca^{2+} entry via a Ca^{2+} channel that is not voltage gated, but may be lanthanum sensitive. Aside from capacitative Ca^{2+} entry, there is evidence from patch-clamp studies that enteric glial cells possess a number of voltage-gated ion channels, including inward-rectifier and delayed-rectifier K^+ channels as well as voltage-activated Na^+ channels. Enteric glia also display large "passive" octanol-sensitive ionic currents, consistent with coupling by gap junctions among these cells.

Enteric Glia and Inflammation

Astrocytes in the central nervous system participate in inflammatory and immune responses. Because of the marked similarities between astrocytes and enteric glial cells, enteric glia are believed also to be involved in inflammatory processes. There is emerging evidence that enteric glia respond to inflammation and are specifically activated in the course of intestinal inflammation. The proliferation rate of enteric glial cells is enhanced in animal models of intestinal inflammation, and there is evidence for the expression of major histocompatibility complex class II (MHC II) antigen on enteric glia in patients with Crohn's disease, which is associated with lymphocytic infiltration of the ENS. This suggests a role for enteric glial cells as antigen-presenting cells that attract immune cells to the ENS. This concept is strengthened by the established capacity of cultured enteric glial cells for phagocytosis and cytokine-inducible MHC II and intercellular adhesion molecule (ICAM-1) expression, which is fully functional for antigen-specific T cell activation. Additionally, enteric glial cells appear to possess functional receptors for IL-1β and IL-6; activation of enteric glial cells by IL-1β via the IL-1 receptor has been shown to stimulate the synthesis and release of IL-6, which displays feedback inhibition of its own secretion. Enteric glial cells appear also to produce IL-1β. Little is presently known about the regulatory mechanisms underlying the production of IL-1β. Evidence suggests that tumor necrosis factor α is not synthesized by enteric glial cells and does not affect cytokine production.

Support for the concept that enteric glial cells produce and are regulated by cytokines is growing. Enteric glial cells appear to possess neurotransmitter receptors, and a preliminary report about stimulated release of IL-6 from enteric glial cells by the neurotransmitter noradrenaline has been published. Thus, enteric glial cells may be an integral part of the inflammatory process in the intestine through the generation of cytokines that regulate the inflammatory response. Simultaneously, enteric glia may serve as an essential link between the nervous and immune systems in the gut by enabling neurotransmitters to modulate inflammation via the release of proinflammatory cytokines from the enteric glia.

A study in adult transgenic mice has implicated enteric glia in maintaining the integrity of bowel function, because ablation of these cells in the small intestine is followed by fulminant jejuno-ileitis and partial degeneration of enteric nerves. This suggests that enteric glia possess factors that are immunosuppressive and antiinflammatory under physiological conditions.

Overall, the cross-talk occurring among the various cell populations in the gut is an area of growing research, and future studies that further examine the role of enteric glia are likely to yield important information in this regard.

See Also the Following Articles

Colitis, Ulcerative • Crohn's Disease • Enteric Nervous System

Further Reading

Burns, A. J., and Le Douarin, N. M. (1998). The sacral neural crest contributes neurons and glia to the post-umbilical gut: Spatiotemporal analysis of the development of the enteric nervous system. *Development* **125**, 4335–4347.

Bush, T. G., Savidge, T. C., Freeman, T. C., Cox, H. J., Campbell, E. A., Mucke, L., Johnson, M. H., and Sofroniew, M. W. (1998). Fulminant jejuno-ileitis following ablation of enteric glia in adult transgenic mice. *Cell* **93**, 189–201.

Gabella, G. (1981). Ultrastructure of the nerve plexuses of the mammalian intestine: The enteric glial cells. *Neuroscience* **6**, 425–436.

Gershon, M. D., and Bursztajn, S. (1978). Properties of the enteric nervous system: Limitation of access of intravascular macromolecules to the myenteric plexus and muscularis externa. *J. Comp. Neurol.* **180**, 467–488.

Gershon, M. D., and Rothman, T. P. (1991). Enteric glia. *Glia* **4**, 195–204.

Giaroni, C., De Ponti, F., Cosentino, M., Lecchini, S., and Frigo, G. (1999). Plasticity in the enteric nervous system. *Gastroenterology* **117**, 1438–1458.

Jessen, K. R., and Mirsky, R. (1983). Astrocyte-like glia in the peripheral nervous system: An immunohistochemical study of enteric glia. *J. Neurosci.* **3**, 2206–2218.

Rühl, A., Franzke, S., Collins, S. M., and Stremmel, W. (2001). Interleukin-6 expression and regulation in rat enteric glial cells. *Am. J. Physiol. (Gastrointest. Liver Physiol.)* **280**, G1163–G1171.

Sarosi, G. A., Barnhart, D. C., Turner, D. J., and Mulholland, M. W. (1998). Capacitative Ca^{2+} entry in enteric glia induced by thapsigargin and extracellular ATP. *Am. J. Physiol. (Gastrointest. Liver Physiol.)* **275**, G550–G555.

Glucose-Dependent Insulinotropic Polypeptide (GIP)

CHI-CHUAN TSENG AND M. MICHAEL WOLFE
Boston University School of Medicine

Glucose-dependent insulinotropic polypeptide (GIP), originally named "gastric inhibitory peptide" for its ability to inhibit acid secretion in dog, is a 42-amino-acid polypeptide and is structurally related to the secretin family of gastrointestinal regulatory polypeptides. GIP is expressed in cells of the upper small intestine and has been shown to be released into the circulation after ingestion of carbohydrate and fat. The full range of biological activity of GIP has yet to be completely elucidated.

STRUCTURE

Glucose-dependent insulinotropic polypeptide (GIP) was first isolated by Brown *et al.* from porcine small intestine in 1969 and was originally named "gastric inhibitory peptide" on the basis of its ability to inhibit acid secretion in dog. Its primary structure was first described in 1971 and revised in 1981. GIP contains 42 amino acid residues and is structurally related to the secretin family of gastrointestinal regulatory polypeptides, which includes secretin, glucagon, glucagon-like peptides (GLP-1 and GLP-2), vasoactive intestinal peptide (VIP), peptide histidine isoleucine, growth hormone-releasing factor, and pituitary adenylate cyclase-activating polypeptide. The amino acid sequences of porcine, human, bovine and rat GIP peptides are shown in Fig. 1.

The human GIP gene, which is located on chromosome 17q, spans approximately 10 kb; the rat GIP gene is slightly smaller, spanning approximately 8.2 kb (Fig. 2). Immediately upstream of the GIP gene is a promoter believed to control tissue-specific and glucose-regulated expression of GIP mRNA. Both the human and rat GIP promoters contain a TATA-box, located 28 and 27 bp upstream of the putative transcriptional start site, respectively. The GIP promoter also contains consensus sequences for activator protein-1 and the cyclic AMP (cAMP)-responsive element, and these elements may be involved in the regulation of GIP gene expression by exogenous factors such as glucose and lipids.

Sequence analysis of the GIP cDNA clones revealed a 459 bp open reading frame encoding a 153-amino-acid polypeptide in the human cDNA and a 432 bp open reading frame encoding a 144-amino-acid polypeptide in the rat cDNA (Fig. 3). The predicted amino acid sequences indicate that both human and rat GIP are derived by proteolytic processing of a preprohormone. Recently, GIP(3–42) was purified from the upper part of the porcine intestine and was shown to have antibacterial activity. It is presumed that this peptide is also derived by proteolytic processing of prepro-GIP.

LOCALIZATION AND DEVELOPMENT

Immunoreactive GIP (IRGIP) cells have been localized to the upper small intestine of ruminants, humans, pigs,

Copyright 2004, Elsevier (USA). All rights reserved.

Jessen, K. R., and Mirsky, R. (1983). Astrocyte-like glia in the peripheral nervous system: An immunohistochemical study of enteric glia. *J. Neurosci.* **3**, 2206–2218.

Rühl, A., Franzke, S., Collins, S. M., and Stremmel, W. (2001). Interleukin-6 expression and regulation in rat enteric glial cells. *Am. J. Physiol. (Gastrointest. Liver Physiol.)* **280**, G1163–G1171.

Sarosi, G. A., Barnhart, D. C., Turner, D. J., and Mulholland, M. W. (1998). Capacitative Ca^{2+} entry in enteric glia induced by thapsigargin and extracellular ATP. *Am. J. Physiol. (Gastrointest. Liver Physiol.)* **275**, G550–G555.

Glucose-Dependent Insulinotropic Polypeptide (GIP)

CHI-CHUAN TSENG AND M. MICHAEL WOLFE

Boston University School of Medicine

Glucose-dependent insulinotropic polypeptide (GIP), originally named "gastric inhibitory peptide" for its ability to inhibit acid secretion in dog, is a 42-amino-acid polypeptide and is structurally related to the secretin family of gastrointestinal regulatory polypeptides. GIP is expressed in cells of the upper small intestine and has been shown to be released into the circulation after ingestion of carbohydrate and fat. The full range of biological activity of GIP has yet to be completely elucidated.

STRUCTURE

Glucose-dependent insulinotropic polypeptide (GIP) was first isolated by Brown *et al.* from porcine small intestine in 1969 and was originally named "gastric inhibitory peptide" on the basis of its ability to inhibit acid secretion in dog. Its primary structure was first described in 1971 and revised in 1981. GIP contains 42 amino acid residues and is structurally related to the secretin family of gastrointestinal regulatory polypeptides, which includes secretin, glucagon, glucagon-like peptides (GLP-1 and GLP-2), vasoactive intestinal peptide (VIP), peptide histidine isoleucine, growth hormone-releasing factor, and pituitary adenylate cyclase-activating polypeptide. The amino acid sequences of porcine, human, bovine and rat GIP peptides are shown in Fig. 1.

The human GIP gene, which is located on chromosome 17q, spans approximately 10 kb; the rat GIP gene is slightly smaller, spanning approximately 8.2 kb (Fig. 2). Immediately upstream of the GIP gene is a promoter believed to control tissue-specific and glucose-regulated expression of GIP mRNA. Both the human and rat GIP promoters contain a TATA-box, located 28 and 27 bp upstream of the putative transcriptional start site, respectively. The GIP promoter also contains consensus sequences for activator protein-1 and the cyclic AMP (cAMP)-responsive element, and these elements may be involved in the regulation of GIP gene expression by exogenous factors such as glucose and lipids.

Sequence analysis of the GIP cDNA clones revealed a 459 bp open reading frame encoding a 153-amino-acid polypeptide in the human cDNA and a 432 bp open reading frame encoding a 144-amino-acid polypeptide in the rat cDNA (Fig. 3). The predicted amino acid sequences indicate that both human and rat GIP are derived by proteolytic processing of a preprohormone. Recently, GIP(3–42) was purified from the upper part of the porcine intestine and was shown to have antibacterial activity. It is presumed that this peptide is also derived by proteolytic processing of prepro-GIP.

LOCALIZATION AND DEVELOPMENT

Immunoreactive GIP (IRGIP) cells have been localized to the upper small intestine of ruminants, humans, pigs,

Copyright 2004, Elsevier (USA). All rights reserved.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PORCINE:	Tyr -	Ala -	Glu-	Gly -	Thr-	Phe-	Ile -	Ser-	Asp-	Tyr-	Ser-	Ile -	Ala-	Met-
HUMAN:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BOVINE:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RODENT:	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
PORCINE:	Asp-	Lys-	Ile -	Arg-	Gln-	Gln-	Asp-	Phe-	Val -	Asn-	Trp-	Leu	Leu-	Ala-
HUMAN:	.	.	.	His	.	.	.	.	.	.	.	.	.	.
BOVINE:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RODENT:	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	29	30	31	32	33	34	35	36	37	38	39	40	41	42
PORCINE:	Gln-	Lys-	Gly-	Lys-	Lys-	Ser-	Asp-	Trp-	Lys-	His-	Asn-	Ile -	Thr-	Gln
HUMAN:	.	.	.	.	.	Asn	.	.	.	.	.	.	.	.
BOVINE:	.	.	.	.	.	.	.	.	Ile	.	.	.	.	.
RODENT:	.	.	.	.	.	Asn	.	.	.	.	.	Leu	.	.

FIGURE 1 Amino acid sequences of porcine, human, bovine, and rat gastric inhibitory polypeptide (GIP). Adapted from Greeley, G. (1999). "Gastrointestinal Endocrinology." With permission from Humana Press.

dogs, and rats. In the gastrointestinal tract of dog and human, IRGIP is present in cells predominantly in the midzone of the duodenal villi and to a lesser extent in the jejunum. Other studies have detected some IRGIP cells in the terminal ileum of rats and humans, as well as GIP transcripts in the ileal mucosa of rats. Surprisingly, expression of the GIP gene was detected in human and mouse stomach, as well as in the rat submandibular

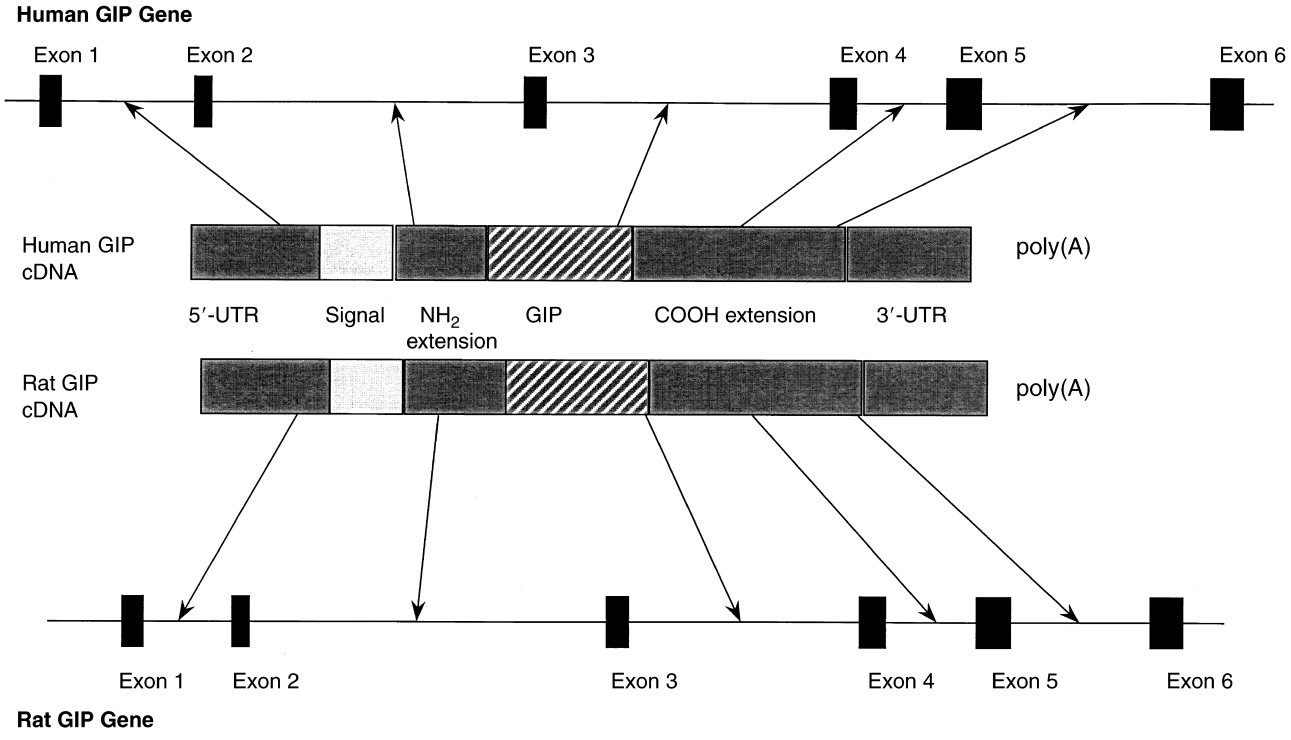


FIGURE 2 Spatial organization of the human and rat GIP gene exons. The location of the exons within the genomic sequence are indicated by solid bars, the location of introns are indicated by arrows. The regions within the cDNAs encoding the various domains of the preprohormone as well as the 5′ and 3′ untranslated sequences are also shown. Adapted from Greeley, G. (1999). "Gastrointestinal Endocrinology." With permission from Humana Press.

SIGNAL PEPTIDE:

Rat: **Met-Val-Ala**-Leu-**Lys-Thr**-Cys-Ser-**Leu-Leu-Leu**-Val-Leu-**Leu-Phe-Leu-Ala-Val-Gly-Leu-Gly**

Human: **Met-Val-Ala**-Thr-**Lys-Thr**-Phe-Ala-**Leu-Leu-Leu**-Leu-Ser-**Leu-Phe-Leu-Ala-Val-Gly-Leu-Gly**

N-TERMINAL PEPTIDE:

Rat: **Glu-Lys**-Glu-**Glu**-Val-Glu-**Phe**-Arg----------**Ser-His-Ala-Lys**-Phe-Ala-Gly-**Pro**-Arg-**Pro-Arg-Gly**-Gln-**Arg**

Human: **Glu-Lys**-Lys-**Glu**-Gly-His-**Phe**-Ser-Ala-Leu-Pro-Ser-Leu-Pro-Val-Gly-**Ser-His-Ala-Lys**-Val-Ser-Ser-**Pro**-Gln-**Pro-Arg-Gly**-Pro-**Arg**

MATURE GIP:

Rat: **Tyr-Ala-Glu-Gly-Thr-Phe-Ile-Ser-Asp-Tyr-Ser-Ile-Ala-Met-Asp-Lys-Ile**-Arg-**Gln-Gln-Asp-Phe-Val-Asn-Trp-Leu-Leu-Ala-Gln-Lys-**

Human: **Tyr-Ala-Glu-Gly-Thr-Phe-Ile-Ser-Asp-Tyr-Ser-Ile-Ala-Met-Asp-Lys-Ile**-His-**Gln-Gln-Asp-Phe-Val-Asn-Trp-Leu-Leu-Ala-Gln-Lys-**

Rat: **Gly-Lys-Lys-Asn-Asp-Trp-Lys-His-Asn**-Leu-**Thr-Gln**

Human: **Gly-Lys-Lys-Asn-Asp-Trp-Lys-His-Asn**--Ile--**Thr-Gln**

C-TERMINAL PEPTIDE:

Rat: **Arg-Glu-Ala-Arg-Ala-Leu-Glu-Leu-Ala**-Gly-**Gln**-Ser-Gln-**Arg**-Asn-**Glu-Glu**-Lys-Glu-Ala-Gln-----Gly-**Ser-Ser**-Leu-Pro-**Lys**-Ser-Leu-**Ser**

Human: **Arg-Glu-Ala-Arg-Ala-Leu-Glu-Leu-Ala**-Ser-**Gln**-Ala-Asn-**Arg**-Lys-**Glu-Glu**-Glu-Ala-Val-Glu-Pro-Gln-**Ser-Ser**-Pro-Ala-**Lys**-Asn-Pro-**Ser**

Rat: **Asp-Glu-Asp**-Val-**Leu-Arg-Asp-Leu-Leu-Ile-Gln-Glu-Leu-Leu-Ala**-Trp-Met--Ala--**Asp-Gln**-Ala-Glu-**Leu-Cys-Arg-Leu-Arg-Ser**-Gln

Human: **Asp-Glu-Asp**-Leu-**Leu-Arg-Asp-Leu-Leu-Ile-Gln-Glu-Leu-Leu-Ala**-Cys-Leu-Leu-**Asp-Gln**-Thr-Asn-**Leu-Cys-Arg-Leu-Arg-Ser**-Arg

FIGURE 3 Comparison of the predicted amino acid sequences of the rat and the human GIP precursors. Dashes in the NH2-and COOH-terminal peptides indicate gaps introduced to maximize alignment between the two sequences. Homologous amino acid residues are designated by boldface lettering. Adapted from Greeley, G. (1999). "Gastrointestinal Endocrinology." With permission from Humana Press.

salivary gland. Ultrastructural studies of human IRGIP cells have demonstrated a characteristic appearance of the K cell: intracellular secretory granules having a small electron-dense core surrounded by a concentric electron-lucent halo. In the dog, however, IRGIP cells identified in the duodenum contained uniformly electron-dense secretory granules consistent with the cell type recognized as the I cell of the endocrine cell classification.

During fetal development, immunoreactive GIP cells appeared during the 14th week of gestation in humans, initially in the duodenal region and then emerging distally with age. An ontogenic study in rats reported that GIP mRNA was present between days 18 and 20 of embryonic development, and mRNA levels markedly increased between days 3 and 5 of postnatal life, followed by a gradual increase toward adult levels. GIP transcripts in the submandibular gland, however, were not detected until day 10 postnatally.

METABOLISM

The kidney is the major site of GIP clearance. Several recent studies have demonstrated that intact GIP(1−42) was hydrolyzed into GIP(3−42) by a ubiquitous enzyme dipeptidyl peptidase IV by cleavage at the penultimate alanine residue. This observation was significant since GIP(3−42) was previously shown to be biologically inactive, yet likely immunoreactive with most GIP antisera. Using a specific GIP antiserum against the N-terminus of the intact GIP, GIP(3−42) was found to account for $73.8 \pm 2.9\%$ of total GIP measured in healthy individuals during the fasting state. An hour after a mixed meal, GIP(1−42) and GIP(3−42) constitute $35.7 \pm 4.2\%$ and $58.1 \pm 2.7\%$, respectively, of total immunoreactive GIP detected in the serum. It was concluded that dipeptidyl peptidase IV is the primary enzyme responsible for GIP inactivation and that the true biological half-life for GIP is more likely approximately 2−7 min. Interestingly, the half-life of GIP appears to be shorter in the diabetic than in the normal individuals, a mechanism that may partly account for the diminished GIP effect observed in many diabetic patients.

GIP RECEPTOR

The GIP receptor is a member of the secretin−VIP family of G-protein-coupled receptors with seven

transmembrane-spanning domains. The gene encoding the human GIP receptor (GIPR) spans approximately 13.8 kb and has been assigned to chromosome bands 19q13.2−q13.3. The human GIPR gene consists of 14 exons, with 13 containing protein-coding sequences and one, exon 1, containing 5′-untranslated sequences. The rat GIPR gene is approximately 10.2 kb in length and contains an additional intronic sequence, resulting in 15 exons. When spliced, these exons produce mature transcripts that contain a 1389 bp open reading frame (ORF), encoding a 466-amino-acid protein, and a 1365 bp ORF encoding a 455 bp protein in humans and rats, respectively. Northern analysis has demonstrated GIPR transcripts of approximately 5.5 kb in human and rat insulinoma cell lines. In addition, GIPR transcripts that are 3.8 kb in length have been detected in rat pancreatic islets and a hamster islet tumor cell line (HIT-T15). RNA encoding the GIPR has been identified not only in the rat pancreas and stomach, but also in adipose tissue, heart, pituitary, adrenal cortex, cerebral cortex, bone, hippocampus, and olfactory bulb. *In situ* hybridization has also detected GIPR-specific transcripts in some major blood vessels and within some of the cells lining bronchioles in the lung of rat embryos. The wide- spread distribution of GIP receptor suggests much broader physiological function than has been reported previously.

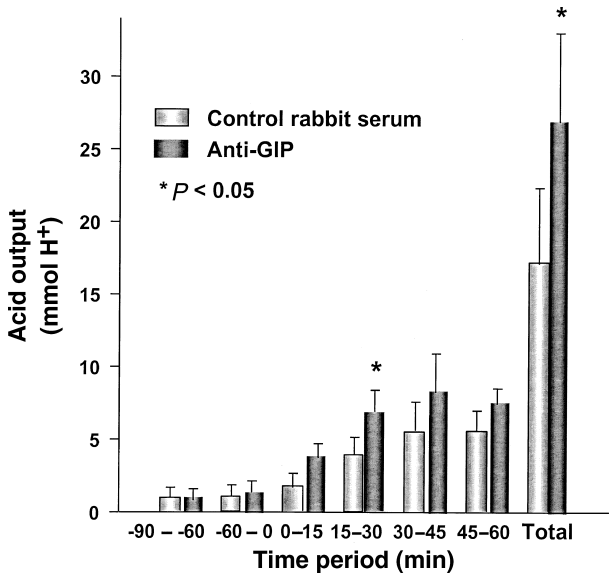


FIGURE 4 Gastric acid secretion compared before and after the intravenous infusion of control rabbit serum or anti-GIP antiserum in dogs. A 10% peptone meal was administered at time 0 minutes. No differences were detected in basal acid output, however the 60-minute meal-stimulated gastric acid secretory response after the infusion of control serum (17.2 ± 5.3 mmol H^+) was significantly less than after the infusion of the GIP antiserum (27.2 ± 6.1 mmol H^+). Data from Greeley, G. (1999). "Gastrointestinal Endocrinology." With permission from Humana Press.

BIOLOGICAL FUNCTIONS

GIP as Enterogastrone

Earlier observation had demonstrated that gastric acid secretion and emptying could be inhibited by the intravenous infusion of a crude extract from small intestinal mucosa. The substance in the extract was later named "enterogastrone," representing a hormone located in the mucosa of the proximal small intestine that was released into the circulation upon the ingestion of fat. Although the full spectrum of biological activity of GIP has yet to be completely elucidated, GIP appears to fulfill the requirements of an enterogastrone. Purified GIP was shown to be a potent inhibitor of gastric acid and pepsin secretion in dogs with denervated pouches. These effects were, however, much attenuated in the intact, innervated stomach, suggesting that cholinergic innervation might antagonize the inhibitory properties of GIP in the stomach. Using a different approach, Wolfe *et al*. infused GIP antiserum into conscious dogs prepared with gastric fistulas. With the binding of 98% of plasma GIP, peptone meal-stimulated gastric acid output increased from 17.2 to 27.2 mmol/h (Fig. 4) and the integrated gastrin response increased from 2.27 to 3.71 ng · min/ml. Moreover, recent studies have demonstrated that GIP stimulated somatostatin release from the isolated perfused rat stomach and inhibited carbachol-stimulated gastrin release. These results suggest that GIP inhibition of rat antral gastrin cells appears to be mediated by the local release of antral somatostatin. Therefore, although the precise physiological relationship between GIP and gastric function has not been determined, it is possible that the effects are species-specific and that GIP may act synergistically with other candidate enterogastrones to inhibit antral gastrin release and acid secretion under physiological conditions.

GIP as Incretin

The existence of a chemical stimulant of the endocrine pancreas had been suggested earlier and the term "incretin" was proposed as the insulin-stimulatory substance released from small intestine following food. Although several gastrointestinal regulatory peptides have been proposed as incretins, only GIP and GLP-1 appear to fulfill the requirements to be considered

physiological stimulants of postprandial insulin release. Many recent studies have demonstrated that, when administered in physiological doses and in the presence of glucose, GIP is a potent stimulator of insulin release by pancreatic islet beta cells. Subsequent studies have also demonstrated the insulinotropic properties of GIP in several experimental models, including the isolated perfused pancreas, isolated pancreatic islet preparations, and *in vivo*. GIP was found to increase intracellular cAMP levels and calcium influx that paralleled the stimulation of insulin release. Furthermore, GIP-stimulated insulin release appeared to be dependent on serum glucose concentration in both the human and rat, with a concentration of 4.4– 5.5 m*M* essential for its insulinotropic action. The observed glucose-dependency of GIP-stimulated insulin secretion appeared to provide an important safeguard against hypoglycemia by preventing the inappropriate stimulation of insulin release during a high-fat, low-carbohydrate meal.

GIP Stimulates GLP-1 Secretion

GLP-1 is a product of intestinal posttranslational processing of preproglucagon and is synthesized in endocrine L cells of the distal small intestine and colon. In humans, GLP-1 infusion has been found to increase plasma insulin concentrations and decrease glucose levels and it has thus been proposed as one of the incretins. Despite all the available data, controversy exists regarding the relative physiological importance of GIP and GLP-1 in stimulating insulin secretion. Recently, studies have demonstrated that GIP stimulated GLP-1 release *in vitro* in rat mucosal cells, but similar results have not been observed in human studies. However, using immunohistochemical studies, immunoreactive GIP and GLP-1 cells were found to be adjacent to each other in the human small intestine, suggesting the existence of a paracrine interaction between K and L cells in this area.

GIP in Adipose Tissue

As stated above, the finding that the GIP receptor is widely distributed in peripheral organs suggests that GIP may have functions in addition to the well-documented enterogastrone and insulinotropic actions. GIP receptor mRNA is expressed in adipose tissue and GIP has been shown to increase lipoprotein lipase activity in a concentration-dependent manner in cultured 3T3-L1 preadipocytes and in explants of rat adipose tissue, suggesting that GIP may play a role in the clearance of chylomicron triglyceride after feeding. However, exogenous porcine GIP and endogenous GIP released by a test meal was without effect on the elimination rate of intravenously injected soy oil in human.

GIP also appears to have action on lipid metabolism within adipocytes. GIP has been reported to stimulate fatty acid synthesis in adipose tissue and to enhance insulin-stimulated incorporation of fatty acids into triglycerides in rat epididymal fat pads. In addition, rat adipocytes incubated with GIP developed increased insulin receptor affinity and showed an increased sensitivity of insulin-stimulated glucose transport. Although still controversial, these results suggest that GIP may play a role in fat metabolism.

GIP in Brain

Although GIP mRNA has not been detected in the brain, GIP receptor mRNA was identified in many regions of the brain, including the olfactory bulb, cerebral cortex, hippocampus, the mammillary bodies, and part of the inferior colliculus. Pharmacological doses of GIP injected into the third ventricle resulted in lower plasma levels of follicle-stimulating hormone without altering levels of luteinizing hormone. GIP also stimulated growth hormone secretion, but was without effect on prolactin and thyroid-stimulating hormone. It remains to be determined whether circulating GIP can gain access to areas in the brain where the receptor is located.

Miscellaneous GIP Biological Properties

In addition to the above putative biological properties of GIP, other actions have been reported in sites where the receptor appears to be present. In the gastrointestinal tract, GIP has been observed to lower esophageal sphincter pressure, to decrease motility of the small intestine, and to reduce intestinal water and electrolyte transport. Receptor mRNA was also detected in the endothelium of major blood vessels where GIP has been reported to increase mesenteric and portal blood flow. Moreover, specific binding sites for GIP have been detected in skeletal muscle and bone, suggesting that GIP may play a role in exercise-induced improvement of peripheral glucose utilization and the function of osteoblasts. Finally, recent studies have shown that GIP may function physiologically as a growth factor for insulin-producing islet beta cells.

REGULATION OF EXPRESSION

The release of GIP into the circulation has been demonstrated primarily after the ingestion of two major nutrient stimuli, carbohydrate and fat. However, the release of GIP in response to oral carbohydrates differs in both magnitude and timing from that following fat

ingestion. Following oral glucose, GIP release is rapid, preceding insulin release, and reaches a peak in peripheral venous blood in approximately 15–30 min, returning to basal values by 3 h in the human and dog. In the rat, plasma GIP levels following a glucose meal do not increase significantly until 60 min and the increment found in rats is much smaller in magnitude when compared to human and dogs. Moreover, only glucose, galactose, and sucrose stimulated GIP release into the circulation, whereas no effect was observed in response to small intestinal perfusion with fructose, mannose, or lactose.

Intralumenal lipid also provides a potent stimulus for the release of GIP into the circulation in humans, dogs, and rats. However, after the ingestion of 100 ml of a triglyceride suspension, the peripheral blood GIP response does not reach a peak until 120–150 min and does not return to baseline by 3 h. In addition to the prolonged response, the amount of GIP released in response to fat is greater than that which follows glucose ingestion. These differences may be due, at least in part, to the effects of lipid-containing meals on gastric emptying. Furthermore, long-chain, but not medium-chain, fatty acids have been found to stimulate GIP release.

In addition to glucose- and fat-stimulated GIP release, some studies have reported the release of GIP after the administration of protein and specific amino acids. An amino acid mixture containing arginine, histidine, isoleucine, leucine, lysine, and threonine provided a potent stimulus for both GIP and insulin release in humans. In contrast, a mixture known to stimulate cholecystokinin release, containing methionine, phenylalanine, tryptophan, and valine, produced only small increases in serum GIP and insulin concentrations. Following intragastric infusion of 10% peptone, a 10-fold increase in serum GIP concentrations was observed in dogs within 15 min. However, the release of GIP was not seen in rats until 120 min after ingestion of a peptone meal. This increment in rats was attenuated by omeprazole, suggesting that protein stimulated GIP release by enhancing gastric acid secretion. Furthermore, some investigators have failed to detect an increase in GIP release in humans after the ingestion of protein meals consisting of cod or steak. These different responses to protein meals may thus be due to interspecies variation.

Prior to the release of any peptide, several intracellular events must occur, among them the stimulation or inhibition of expression of the gene encoding that protein. Because hormonal gene expression can be regulated at several different steps along the biosynthetic pathway, nutrient-regulated GIP gene expression has recently been examined in rats. In animals that ingested

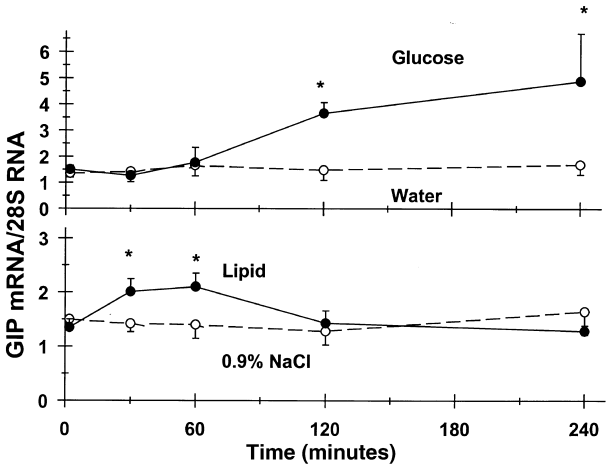


FIGURE 5 Steady state levels of duodenal GIP mRNA in glucose- (top) and Lipomul-fed (bottom) rats. GIP mRNA concentrations are compared to water-fed rats and are expressed as the mean (± SE) ratio of GIP mRNA to 28S ribosomal RNA to correct for gel loading. Each time point consists of 4–6 rats for each group. *$P < 0.05$. Data from Greeley, G. (1999). "Gastrointestinal Endocrinology." With permission from Humana Press.

a 10% glucose solution orally, duodenal GIP mRNA increased steadily over a 4 h period, at which time a fourfold increase in GIP mRNA levels in the small intestine was detected (Fig. 5). In contrast to the effects of glucose, duodenal GIP mRNA following intestinal perfusion with a 20% lipid emulsion was increased, but to a lesser degree when compared with oral glucose ingestion. Moreover, whereas an increase in duodenal GIP mRNA occurred primarily during the latter portion of the meal, the increase in lipid-stimulated GIP gene expression occurred at earlier time points, between 30 and 60 min after the perfusion, and GIP mRNA levels appeared to decline thereafter (Fig. 5). These studies indicate that both glucose- and lipid-containing meals increase GIP gene expression at the pretranslational level. In contrast to glucose and lipid, peptone infusion increased duodenal and plasma GIP concentrations, but failed to induce a significant change in duodenal GIP mRNA levels, suggesting that peptone-stimulated GIP expression might occur only at the posttranslational phase.

CONTRIBUTION TO DISEASES

The precise contribution of GIP to diabetes mellitus is unclear as no definite and consistent abnormalities in GIP release in diabetic patients have been found. Although newly diagnosed type I diabetic patients have been reported to have decreased circulating concentrations of GIP, enhanced and normal GIP release into the

circulation has also been reported in patients with type 2 diabetes mellitus. Furthermore, studies in patients with type 2 diabetes have demonstrated that although the incretin activity of GLP-1 is normal, the insulinotropic properties of GIP appear to be greatly diminished. Although the mechanisms remain unknown, a decrease in GIP receptor number on the pancreatic islet beta cells, a shorter GIP half-life, and chronic desensitization of the GIP receptor due to elevated serum GIP levels might account for impaired insulin secretion seen in diabetic patients.

Recently, corticotropin-independent, food-dependent nodular adrenal hyperplasia has been identified as a rare cause of Cushing's syndrome. In these patients, the serum cortisol levels were significantly elevated after the ingestion of glucose or a mixed meal, and elevated postprandial cortisol levels were not suppressed by a large dose of dexamethasone. Further investigation has demonstrated aberrant expression of the GIP receptor on the steroid-producing adrenal tumors and the infusion of GIP significantly increased plasma cortisol levels. It was concluded that this type of nodular adrenal hyperplasia and Cushing's syndrome may be food-dependent and may result from abnormal responsiveness of GIP receptors on the adrenal cells to the physiological secretion of GIP. Whether GIP plays any significant role in regulating adrenal gland secretion under physiological conditions is currently unknown.

Although GIP is one of the candidate enterogastrones, no evidence has suggested that GIP plays a role in the pathogenesis of duodenal ulcer, as no significant changes in serum GIP levels were detected in a group of duodenal ulcer patients. Furthermore, fasting and food-stimulated GIP levels were significantly elevated in duodenal ulcer patients, indicating that hyposecretion of GIP did not account for hypersecretion of gastric acid seen in this group of patients.

See Also the Following Articles

Gastric Acid Secretion • Secretin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Bollag, R. J., Zhong, Q., Phillips, P., Min, L., Zhong, L., Cameron, R., Mulloy, A. L., Rasmussen, H., Qin, F., Ding, K. H., and Isales, C. M. (2000). Osteoblast-derived cells express functional glucose-dependent insulinotropic peptide receptors. *Endocrinology* **141**, 1228–1235.

Brubaker, P. L. (1991). Regulation of intestinal proglucagon-derived peptide secretion by intestinal regulatory peptides. *Endocrinology* **128**, 3175–3182.

Deacon, C. F., Nauck, M. A., Meier, J., Hucking, K., and Holst, J. J. (2000). Degradation of endogenous and exogenous gastric inhibitory polypeptide in healthy and in type 2 diabetic subjects as revealed using a new assay for the intact peptide. *J. Clin. Endocrinol. Metab.* **85**, 3575–3581.

Kieffer, T. J., McIntosh, C. H. S., and Pederson, R. A. (1995). Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 *in vitro* and *in vivo* by dipeptidyl peptidase IV. *Endocrinology* **136**, 3585–3596.

Lacroix, A., Bolté, E., Tremblay, J., Dupré, J., Poitras, P., Fournier, H., Garon, J., Garrel, D., Bayard, F., Taillefer, R., Flanagan, R. J., and Hamet, P. (1992). Gastric inhibitory polypeptide-dependent cortisol hypersecretion–A new cause of Cushing's syndrome. *N. Engl. J. Med.* **327**, 974–980.

Mortensen, K., Petersen, L. L., and Orskov, C. (2000). Colocalization of GLP-1 and GIP in human and porcine intestine. *Ann. N. Y. Acad. Sci.* **921**, 469–472.

Nauck, M. A., Helmesaat, M. M., Ørskov, C., Holst, J. J., Ebert, R., and Creutzfeldt, W. (1993). Preserved incretin activity of GLP-1 but not of synthetic human GIP in patients with type 2 diabetes mellitus. *J. Clin. Invest.* **30**, 301–307.

Trumper, A., Trumper, K., Trusheim, H., Arnold, R., Goke, B., and Horsch, D. (2001). Glucose-dependent insulinotropic polypeptide is growth factor for beta (INS-1) cells by pleiotropic signaling. *Mol. Endocrinol.* **15**, 1559–1570.

Tseng, C. C., Boylan, M. O., Jarboe, L. A., Usdin, T. B., and Wolfe, M. M. (1996). Chronic desensitization of the glucose-dependent insulinotropic polypeptide receptor in diabetic rats. *Am. J. Physiol.* **270**, E661–E666.

Tseng, C. C., Jarboe, L. A., Landau, S. B., Williams, E. K., and Wolfe, M. M. (1993). Glucose-dependent insulinotropic peptide: Structure of the precursor and tissue-specific expression in rat. *Proc. Natl. Acad. Sci. USA* **90**, 1992–1996.

Tseng, C.-C., Jarboe, L. A., and Wolfe, M. M. (1994). Regulation of glucose-dependent insulinotropic peptide gene expression by a glucose meal. *Am. J. Physiol.* **266**, G887–G891.

Usdin, T. B., Mezey, É., Button, D. C., Brownstein, M. J., and Bonner, T. I. (1993). Gastric inhibitory polypeptide receptor, a member of the secretin–vasoactive intestinal peptide receptor family, is widely distributed in peripheral organs and the brain. *Endocrinology* **133**, 2861–2870.

Wolfe, M. M., Boylan, M. O., Kieffer, T. J., and Tseng, C.-C. (1998). Glucose-dependent insulinotropic polypeptide: Incretin vs enterogastrone. *In* "Gastrointestinal Endocrinology" (J. H. Greeley, Jr., eds.), pp. 439–466. Humana Press, Totowa, NJ.

Wolfe, M. M., Zhao, K. B., Glazier, K. D., Jarboe, L. A., and Tseng, C.-C. (2000). Regulation of glucose-dependent insulinotropic polypeptide release by protein in the rat. *Am. J. Physiol.* **279**, G561–G566.

Yeung, C. M., Wong, C. K., Chung, S. K., Chung, S. S., and Chow, B. K. (1999). Glucose-dependent insulinotropic polypeptide gene expression in the stomach: Revealed by a transgenic mouse study, *in situ* hybridization and immunohistochemical staining. *Mol. Cell. Endocrinol.* **154**, 161–170.

Glycogen Storage Disease

DAVID A. WEINSTEIN AND JOSEPH I. WOLFSDORF
Children's Hospital Boston and Harvard Medical School

gluconeogenesis Formation of new glucose from noncarbohydrate substrates, including various amino acids, lactate, pyruvate, and glycerol.

glycogen Complex, hydrated polymer of glucose with a very large molecular weight (ranging over several million Daltons), consisting of many glucose molecules joined together to form a compact, highly branched spherical structure with a large number of exposed terminal glucose molecules that are accessible to the enzymes involved in glycogen breakdown (glycogenolysis).

glycogenolysis Intracellular breakdown of glycogen to glucose.

glycogen storage diseases (glycogenoses) Several inherited diseases caused by abnormalities of the enzymes that regulate glycogen synthesis and degradation.

Glucose in excess of cellular energy needs is stored in both liver and muscle as glycogen. The formation of glycogen (glycogenesis) enables storage of glucose without the osmotic consequences of free glucose molecules. Defects in the enzymatic reactions involved in converting glucose to glycogen result in various types of glycogen storage diseases (glycogenoses). Hypoglycemia is the primary manifestation of the hepatic glycogenoses; weakness and muscle cramps are the predominant features of the muscle glycogenoses.

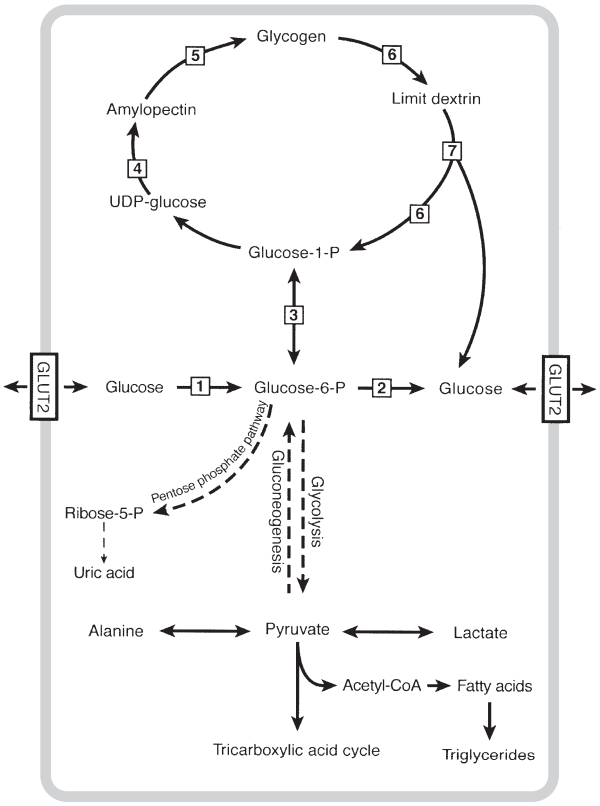


FIGURE 1 Simplified scheme of glycogen synthesis and breakdown, indicating the enzymes at each step: 1, hexokinase/glucokinase; 2, glucose-6-phosphatase; 3, phosphoglucomutase; 4, glycogen synthase; 5, branching enzyme; 6, glycogen phosphorylase; 7, debranching enzyme. UDP-glucose, uridine diphosphoglucose. Reproduced with permission from Wolfsdorf *et al.* (1999).

INTRODUCTION

The average well-fed man on a diet rich in carbohydrate stores about 70 g of glycogen in his liver and 200 g in his muscles. The liver is freely permeable to glucose, which is rapidly phosphorylated by glucokinase to form glucose-6-phosphate, which is then converted to glucose-1-phosphate, the starting point for glycogen synthesis (Fig. 1). Hepatic glycogen synthase catalyzes the formation of α-1,4-linkages that elongate the chains of glucose molecules. A branching enzyme leads to formation of α-1,6-linkages at branch points, on average, every 10 glucose units along the chain. In the intervals between meals and during the overnight fast, a cascade of enzymatic reactions activates hepatic glycogen phosphorylase, the rate-limiting enzyme in glycogenolysis, which removes glucose from the outer branches of glycogen, leading, ultimately, to the formation of glucose-6-phosphate. Glucose-6-phosphatase catalyzes the terminal reaction of glycogenolysis and gluconeogenesis, the hydrolysis of glucose-6-phosphate, allowing glucose to be released from the liver into the systemic circulation, a process that is critically important for the maintenance of glucose homeostasis. Although glycogen storage diseases involve both the

 Copyright 2004, Elsevier (USA). All rights reserved.

muscles and the liver, only the glycogenoses with hepatic involvement are discussed in this article.

GLYCOGEN STORAGE DISEASE TYPE 0

Glycogen storage disease type 0 (GSD0) is caused by a deficiency of hepatic glycogen synthase, leading to a marked decrease in glycogen stores. The inability to store glucose as glycogen causes postprandial hyperglycemia, and increased conversion of dietary glucose to lactate resulting in postprandial hyperlactatemia.

Clinical Features

Individuals with GSD0 are usually asymptomatic early in infancy. Fasting ketotic hypoglycemia develops on cessation of nighttime feeding, and weaning from overnight feeds is difficult. During gastrointestinal illness or periods of poor feeding for any reason, children may become lethargic; hypoglycemia usually is an incidental finding. Childhood growth may be mildly delayed, but most children continue to develop normally. The physical examination is usually normal and there is no hepatomegaly. Hyperglycemia and hyperlactatemia occur after meals, particularly after breakfast, and may be confused with early diabetes.

Epidemiology

GSD0 is inherited in an autosomal recessive manner. It is caused by mutations in the *GYS2* gene located on chromosome 12p12.2. Only 18 cases have been described in the literature, but there is evidence that this disease may be underdiagnosed. GSD0 affects both genders equally and cases from Eastern Europe, Western Europe, North America, and South America have been described.

Diagnosis

Frequent measurements of blood glucose, lactate, and ketones in both the fed and fasting states demonstrate the unique biochemical pattern of postprandial hyperglycemia and hyperlactatemia alternating with fasting ketotic hypoglycemia. Despite the decrease in hepatic glycogen content, the glycemic response to glucagon is variable. A liver biopsy shows hepatocytes that contain only small amounts of glycogen (approximately 0.5% of wet liver weight) and moderate steatosis. The diagnosis can now be made noninvasively by mutational analysis.

Treatment

The goal of treatment is to prevent hypoglycemia. Fasting must be avoided. Frequent high-protein meals and snacks are given every 3–4 hours during the day. Uncooked cornstarch (0.5–1 g/kg) at bedtime prevents overnight hypoglycemia. Because surplus glucose cannot be stored as glycogen and is converted to lactic acid, the diet should contain an increased amount of protein to provide substrate for gluconeogenesis and decreased carbohydrate to minimize postprandial hyperglycemia and hyperlactatemia.

GLYCOGEN STORAGE DISEASE TYPE I

Glycogen storage disease type I (GSDI), also known as Von Gierke disease, is caused by deficient glucose-6-phosphatase activity. Glucose-6-phosphatase is a multicomponent enzyme system located in the membrane of the endoplasmic reticulum; the system consists of the enzyme, a calcium-binding protein, and three proteins that facilitate transport of glucose-6-phosphate (T1), phosphate (T2), and glucose (T3) across the endoplasmic reticulum membrane. More than 80% of patients with GSDI have mutations that cause deficient catalytic activity of the glucose-6-phosphatase system (type Ia GSD). Mutations in the *G6PT1* gene cause deficiency of the glucose-6-phosphate transporter (type Ib GSD), which also results in inability to convert glucose-6-phosphate to glucose.

Clinical Features

Symptomatic hypoglycemia may occur soon after birth, but most children do not present with symptoms until 3–6 months of age, when they begin to sleep through the night, or when an intercurrent illness disrupts the normal pattern of frequent feeding. GSDI is occasionally diagnosed in an asymptomatic infant after hepatomegaly and a protuberant abdomen are detected on a routine physical examination. It is more usual, however, for affected infants to present with tachypnea, seizures, lethargy, developmental delay, or failure to thrive. Impaired production of glucose via glycogenolysis and gluconeogenesis causes severe hypoglycemia and increased hepatic production of lactate, triglycerides, and uric acid. Activation of glucose counterregulation as a result of hypoglycemia increases the plasma levels of glucagon, epinephrine, and cortisol and can rapidly lead to severe lactic acidosis. Surgical stress, intercurrent illness, and administration of pharmacologic doses of glucocorticoids can, similarly, cause lactic acidosis without hypoglycemia. There is no effect on social and cognitive development unless the infant suffers cerebral damage from recurrent severe hypoglycemia.

Long-term complications may include hepatic adenomas, focal segmental glomerulosclerosis, renal tubular dysfunction, nephrocalcinosis, nephrolithiasis, gout, anemia, and osteoporosis. Poor metabolic control contributes to the pathogenesis of these complications. More than 60% of patients with GSDI develop hepatic adenomas, which are usually first detected during adolescence (occasionally seen in childhood) and may cause serious morbidity. Adenomas may remain stable or gradually enlarge over time. They may, however, undergo malignant transformation into hepatocellular carcinoma, or may hemorrhage into the abdomen, causing fatal hemoperitoneum. The serum alkaline phosphatase concentration and erythrocyte sedimentation rate are often increased in patients with adenomas. Ultrasonography is the preferred method for screening. When malignancy is suspected because of a change in the sonographic appearance, magnetic resonance imaging provides greater definition. Serum α-fetoprotein is normal in patients with adenomas, but may be increased in some cases of hepatocellular carcinoma. Large hepatic adenomas (>7 cm in diameter) have been associated with an unremitting, iron-resistant anemia, and a peptide (hepcidin) that blocks intestinal absorption of iron and macrophage recycling of iron is inappropriately expressed in the adenomas. In addition to all of these clinical features, patients with type Ib GSD have neutropenia, recurrent bacterial infections, and may develop inflammatory bowel disease.

Epidemiology

GSDI is inherited in an autosomal recessive manner. It is caused by mutations in the *G6PC* gene located on chromosome 17q21. More than 70 mutations have been identified worldwide. The overall incidence is estimated to be 1 in 200,000 births and the highest incidence appears to be in Ashkenazi Jews.

Diagnosis

The simplest means of determining the probable enzymatic defect in a child suspected of having a glycogenosis is to obtain serial blood measurements of metabolites (glucose, lactate, and ketones) during a fast of 3–4 hours duration, performed under close observation. In a child with GSDI, fasting rapidly leads to hypoglycemia and progressive lactic acidosis. Administration of glucagon (30 μg/kg) fails to elicit a glycemic response and may exacerbate the acidosis.

A liver biopsy and assay of hepatic glucose-6-phosphatase activity has, until recently, been the primary method of confirming the diagnosis. Differentiation between the types of GSDI requires an analysis of glucose-6-phosphatase activity in both intact and fully disrupted microsomes. In type Ib GSD, glucose-6-phosphatase activity is normal when the assay is performed on previously frozen tissue (freezing disrupts the microsomal membrane). The diagnosis can now be confirmed by mutational analysis.

Treatment

Treatment consists of providing a continuous dietary source of glucose to prevent blood glucose from falling below the threshold for glucose counterregulation (approximately 70 mg/dl). In infants, continuous glucose can be provided with frequent (every 2–3 hours) feeds during the day and continuous intragastric feeds at night via a nasogastric or gastrostomy tube. From about 6–12 months of age, uncooked cornstarch can be used as an alternative method of glucose delivery. Uncooked cornstarch is slowly absorbed into the circulation and allows the interval between feeds to be increased. The amount of cornstarch and the interval between feeds have to be individualized based on results of blood glucose monitoring and periodic metabolic evaluations. An estimate of glucose requirements can be obtained by calculating the basal glucose production rate using the following formula: $y = 0.0014x^3 - 0.214x^2 + 10.411x - 9.084$, where y = milligrams of glucose/minute and x = body weight (kilograms). If the patient has to fast for any reason or if gastrointestinal illness does not allow adequate intake of glucose, intravenous glucose must be administered at 125–150% of the estimated glucose production rate to ensure euglycemia and to minimize lactic acidosis.

Foods that contain galactose and fructose must be restricted because these sugars cannot be converted to glucose and large quantities may exacerbate the biochemical derangements. Oral citrate supplementation may prevent renal calcification in patients with hypercalciuria and hypocitraturia. If optimal dietary management fails to lower serum uric acid and triglycerides to acceptable levels, treatment with allopurinol and gemfibrozil, respectively, is indicated. Neutropenia in type Ib GSD responds well to granulocyte colony-stimulating factor (G-CSF) therapy, which is also efficacious for inflammatory bowel disease.

GLYCOGEN STORAGE DISEASE TYPE III

Glycogen storage disease type III (GSDIII), also known as Cori disease or Forbes disease, is caused by deficiency of glycogen debrancher enzyme (GDE). Release of glucose from glycogen stores requires the sequential

actions of both glycogen phosphorylase and GDE. After phosphorylase has acted on the outer branches of glycogen, GDE hydrolyzes the branch points. Lack of GDE activity causes glycogenolysis to stop when the outermost branch points are reached, resulting in the accumulation of an abnormal form of glycogen (limit dextrin) in affected tissues. The two main subtypes are defined by the location of the enzyme deficiency. Type IIIa GSD accounts for 85% patients in the United States, and results from lack of GDE activity in both liver and muscle. Only the liver is affected in type IIIb.

Clinical Features

Patients with hepatic involvement have hepatomegaly and ketotic hypoglycemia. In infancy and early childhood it may be indistinguishable from children with GSDI. Because the outer segments of glycogen are accessible and gluconeogenesis can occur normally, the hypoglycemia typically improves with age. In type IIIA, muscle weakness is usually minimal in childhood. Clinically significant myopathy affecting principally the large proximal muscles of the shoulders and hips usually becomes prominent in adulthood. Biochemical abnormalities include increases in aspartate transaminase (AST), alanine transaminase (ALT), and creatine kinase concentrations, and hyperlipidemia. Long-term complications include short stature, cardiomyopathy, and hepatic adenomas in 25% of patients.

Epidemiology

The incidence of GSDIII is about 1 in 100,000 live births and is unusually frequent in Israeli Jews of North African descent (prevalence, 1 in 5400; carrier prevalence, 1 in 35). GSDIII is inherited in an autosomal recessive manner and is caused by mutations in the debrancher gene on chromosome 1p21.

Diagnosis

Fasting causes ketotic hypoglycemia without the hyperlactatemia characteristic of GSDI. Glucagon does not elicit a glycemic response when given after an overnight fast, but does elicit a glycemic response when given 2 hours after a carbohydrate-rich meal. A definitive diagnosis is obtained by demonstrating abnormal glycogen (limit dextrin with short outer branches) in liver and/or muscle and deficient debranching enzyme activity. Definitive subtyping of GSDIII requires biopsies of both liver and muscle, because a normal serum creatine kinase concentration does not rule out muscle involvement. Measurement of enzyme activity in skin fibroblasts or lymphocytes can be used to screen for GSDIII, but results may not be definitive and cannot be used for subtyping.

Treatment

As in GSDI, the goal is to provide a continuous source of glucose to maintain blood glucose above 70 mg/dl. After infancy, uncooked cornstarch (1.75 g/kg at 6-hour intervals) can be used to maintain normoglycemia. Cornstarch has been shown to improve growth and ameliorate the biochemical abnormalities. Milk products and fruit should not be restricted because fructose and galactose can be normally converted to glucose. Amino acids can be used as a substrate for gluconeogenesis, and there is some evidence that a high-protein diet may be beneficial.

GLYCOGEN STORAGE DISEASE TYPE IV

Glycogen storage disease type IV (GSDIV), also known as amylopectinosis or Anderson disease, is caused by deficient branching enzyme (amylo-1,4 → 1,6-transglucosidase) activity, which causes accumulation in the liver of an abnormal unbranched glycogen molecule with long outer branches, resembling amylopectin. The abnormal glycogen acts as a foreign body that induces cirrhosis.

Clinical Features

GSDIV typically presents in early infancy with hepatosplenomegaly and failure to thrive. Because nonbranched glycogen is available for glycogenolysis, hypoglycemia is unusual in GSDIV until late in the disease, when cirrhosis is advanced. The usual clinical course is progressive liver cirrhosis with portal hypertension, esophageal varices, and ascites, culminating in death from liver failure by 5 years of age. Accumulation of amylopectin-like polysaccharide in cardiac muscle can result in a fatal cardiomyopathy. A childhood-onset variant presents as myopathy or cardiomyopathy, and a milder adult-onset variant may present with central and peripheral nervous system involvement resulting from accumulation of unbranched glycogen in neuronal tissue (polyglucosan body disease).

Epidemiology

GSDIV is inherited as an autosomal recessive trait and accounts for about 3% of all cases of GSD. Mutations in one glycogen-branching enzyme (GBE) gene, located on chromosome 3p14, are responsible for both the hepatic and the neuromuscular forms of the disease. All

three forms of GSDIV are caused by mutations in the same gene; significant retention of GBE activity accounts for mild disease.

Diagnosis

The diagnosis of GSDIV is established by demonstrating abnormal glycogen (an amylopectin-like abnormal polysaccharide with long outer chains) that stains with periodic acid Schiff reagent but is partially resistant to diastase digestion. Branching enzyme is deficient in liver, muscle, leukocytes, erythrocytes, and fibroblasts. In adult polyglucosan body disease, the branching enzyme deficiency can be detected only in leukocytes or in a nerve biopsy.

Treatment

There is no specific treatment for GSDIV. Transplantation has been an effective treatment for progressive liver failure.

GLYCOGEN STORAGE DISEASE TYPES VI AND IX

Glycogen storage disease types VI and IX [glycogen phosphorylase deficiency (Hers disease) and phosphorylase kinase deficiency, respectively] are considered together because both impair activity of liver phosphorylase, which catalyzes the rate-limiting step in glycogenolysis. GSDVI is caused by deficiency of glycogen phosphorylase, whereas GSDIX is caused by deficiency of phosphorylase b kinase, which results in failure to activate hepatic phosphorylase.

Clinical Features

Presentation is usually with hepatomegaly and varying degrees of growth retardation. Prolonged fasting or strenuous physical exercise can cause ketotic hypoglycemia. Mild hyperlipidemia and elevations in serum transaminase concentrations are common. Lactic acid and uric acid concentrations are normal and metabolic acidosis is rare. Motor development may be delayed as a consequence of muscular hypotonia and weakness. Uncommon clinical phenotypes have been described, including proximal renal tubular acidosis, neurologic abnormalities, and cirrhosis. A rare severe cardiac-specific phosphorylase kinase variant that can lead to cardiac failure has been reported.

Epidemiology

GSDVI is rare except in the Mennonite community (estimated disease frequency 1 in 1000 in that population). It is inherited as an autosomal recessive trait. *PYGL*, the gene encoding glycogen phosphorylase, is located on chromosome 14q21–q22. GSDIX is inherited either as an X-linked or an autosomal recessive trait, occurs in approximately 1 in 100,000 births, and accounts for approximately 25% of all patients with GSD.

Diagnosis

Glycogen phosphorylase deficiency can be diagnosed by assaying activity of the enzyme in leukocytes and erythrocytes. A definitive diagnosis requires demonstration of the enzyme defect in a liver biopsy. Phosphorylase b kinase can also be assayed in leukocytes and erythrocytes. Because the enzyme has several isoenzymes, the diagnosis can be missed without studies of liver and muscle. Definitive diagnosis of phosphorylase b kinase deficiency, therefore, requires demonstration of the enzyme defect in affected tissues. Subtyping of the disease requires molecular genetic analyses. Mutation analysis is likely to become the standard method for diagnosing both GSDVI and the X-linked variant of GSDIX. It is unlikely, however, to be able to rule-out all forms of GSDIX, because multiple genes are involved in synthesizing the phosphorylase kinase protein.

Treatment

Most patients do not require specific treatment. They should routinely have a bedtime snack and avoid prolonged fasting. In the unusual patient with nocturnal hypoglycemia and ketosis, uncooked cornstarch (2 g/kg) at bedtime prevents hypoglycemia and ketosis. These disorders have an excellent prognosis.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Galactosemia • Hepatic Adenomas • Hereditary Fructose Intolerance • Liver Biopsy

Further Reading

Bao, Y., Kishnani, P., Wu, J. Y., and Chen, Y. T. (1996). Hepatic and neuromuscular forms of glycogen storage disease type IV caused by mutations in the same glycogen-branching enzyme gene. *J. Clin. Invest.* **97**, 941–948.

Chen, Y.-T. (2001). Glycogen storage diseases. *In* "The Metabolic and Molecular Bases of Inherited Disease" (C. Scriver, A. Beaudet, W. Sly, *et al.,* eds.), 8th Ed., Vol. 1, pp. 1521–1555. McGraw-Hill, New York.

Chou, J., and Mansfield, B. (1999). Molecular genetics of type I glycogen storage diseases. *Trends Endocrinol. Metab.* **10**, 104–113.

Elpeleg, O. N. (1999). The molecular background of glycogen metabolism disorders. *J. Pediatr. Endocrinol. Metab.* **12**, 363–379.

Gitzelmann, R., Spycher, M., Feil, G., *et al.* (1996). Liver glycogen synthase deficiency: A rarely diagnosed entity. *Eur. J. Pediatr.* **155**, 561–567.

Rake, J. P., ten Berge, A. M., Visser, G., *et al.* (2000). Glycogen storage disease type Ia: Recent experience with mutation analysis, a summary of mutations reported in the literature and a newly developed diagnostic flowchart. *Eur. J. Pediatr.* **159**, 322–330.

Wolfsdorf, J. I., Holm, I. A., and Weinstein, D. A. (1999). Glycogen storage diseases: Phenotypic, genetic, and biochemical characteristics, and therapy. *Endocrinol. Metab. Clin. North Am.* **28**, 801–823.

Growth Factors

RAYMOND J. PLAYFORD
Imperial College, London

lumenal surveillance peptides Molecules that are constantly present within the gut lumen; their predominant role is to stimulate repair at sites of injury, acting in a "surveillance" fashion.

mucosal integrity peptides Molecules that are constitutively expressed in the mucosa and which function to maintain normal mucosal integrity.

rapid response peptides Molecules produced rapidly at sites of injury and which function to stimulate the repair process.

redundancy State that occurs when two or more factors have similar overlapping functions; removal of one of the factors therefore has much less effect on the system than if it was the sole controlling agent.

restitution The process by which surviving cells at the edge of a wound migrate across the denuded area to reestablish epithelial continuity. This process is not dependent on cell proliferation.

Growth factors are generally considered to be peptides that act via specific receptors, triggering intracellular secondary messengers, resulting in cell proliferation. However, it is important to appreciate that when complex interactions are considered, such as the control of growth and differentiation of gut cells, ascribing somewhat arbitrarily an individual function to a molecule, e.g., to consider "epidermal growth factor" simply a stimulant of proliferation, can be misleading as it is now clear that such factors have multiple effects. For example, although they are often considered separately, the distinctions between cytokines and growth factors are sometimes blurred as the "cytokine" interleukin-8 has been shown to stimulate the migration of human colonic epithelial cells, a process normally associated with "growth factors," and peptides normally considered to be growth factors can influence immunological function. Similarly, molecules such as glutamine or butyrate, which have been generally considered simple energy providers, and vitamins, such as vitamins A and D, which at one time were thought to have limited biological functions, are now known to influence multiple other activities of the cell, such as development, differentiation, and proliferation. To complicate matters even further, peptides that stimulate repair are usually thought of as growth factors. Although this is true in the majority of cases, certain peptides appear to stimulate the repair process without influencing proliferation, the trefoil peptides being a notable example. These peptides are therefore not growth factors as such but may, nevertheless, play an important physiological role and offer potential as a therapeutic strategy. This article therefore focuses on peptide factors that influence gut proliferation and/or repair.

INTRODUCTION

Gut Mucosal Integrity

The gastrointestinal tract possesses the remarkable ability to remain intact despite being constantly bathed in acid and proteolytic enzymes that can digest virtually any form of food that is eaten. When a superficial mucosal injury occurs, such as following direct physical trauma or ingestion of noxious agents such as aspirin or alcohol, the gastrointestinal tract is rapidly healed. The

 Copyright 2004, Elsevier (USA). All rights reserved.

Gitzelmann, R., Spycher, M., Feil, G., *et al.* (1996). Liver glycogen synthase deficiency: A rarely diagnosed entity. *Eur. J. Pediatr.* **155**, 561–567.

Rake, J. P., ten Berge, A. M., Visser, G., *et al.* (2000). Glycogen storage disease type Ia: Recent experience with mutation analysis, a summary of mutations reported in the literature and a newly developed diagnostic flowchart. *Eur. J. Pediatr.* **159**, 322–330.

Wolfsdorf, J. I., Holm, I. A., and Weinstein, D. A. (1999). Glycogen storage diseases: Phenotypic, genetic, and biochemical characteristics, and therapy. *Endocrinol. Metab. Clin. North Am.* **28**, 801–823.

Growth Factors

RAYMOND J. PLAYFORD
Imperial College, London

lumenal surveillance peptides Molecules that are constantly present within the gut lumen; their predominant role is to stimulate repair at sites of injury, acting in a "surveillance" fashion.

mucosal integrity peptides Molecules that are constitutively expressed in the mucosa and which function to maintain normal mucosal integrity.

rapid response peptides Molecules produced rapidly at sites of injury and which function to stimulate the repair process.

redundancy State that occurs when two or more factors have similar overlapping functions; removal of one of the factors therefore has much less effect on the system than if it was the sole controlling agent.

restitution The process by which surviving cells at the edge of a wound migrate across the denuded area to reestablish epithelial continuity. This process is not dependent on cell proliferation.

Growth factors are generally considered to be peptides that act via specific receptors, triggering intracellular secondary messengers, resulting in cell proliferation. However, it is important to appreciate that when complex interactions are considered, such as the control of growth and differentiation of gut cells, ascribing somewhat arbitrarily an individual function to a molecule, e.g., to consider "epidermal growth factor" simply a stimulant of proliferation, can be misleading as it is now clear that such factors have multiple effects. For example, although they are often considered separately, the distinctions between cytokines and growth factors are sometimes blurred as the "cytokine" interleukin-8 has been shown to stimulate the migration of human colonic epithelial cells, a process normally associated with "growth factors," and peptides normally considered to be growth factors can influence immunological function. Similarly, molecules such as glutamine or butyrate, which have been generally considered simple energy providers, and vitamins, such as vitamins A and D, which at one time were thought to have limited biological functions, are now known to influence multiple other activities of the cell, such as development, differentiation, and proliferation. To complicate matters even further, peptides that stimulate repair are usually thought of as growth factors. Although this is true in the majority of cases, certain peptides appear to stimulate the repair process without influencing proliferation, the trefoil peptides being a notable example. These peptides are therefore not growth factors as such but may, nevertheless, play an important physiological role and offer potential as a therapeutic strategy. This article therefore focuses on peptide factors that influence gut proliferation and/or repair.

INTRODUCTION

Gut Mucosal Integrity

The gastrointestinal tract possesses the remarkable ability to remain intact despite being constantly bathed in acid and proteolytic enzymes that can digest virtually any form of food that is eaten. When a superficial mucosal injury occurs, such as following direct physical trauma or ingestion of noxious agents such as aspirin or alcohol, the gastrointestinal tract is rapidly healed. The

 Copyright 2004, Elsevier (USA). All rights reserved.

gut therefore possesses powerful mucosal defense and repair mechanisms. These mechanisms include a high rate of cellular turnover (second only to the hematopoietic system), an efficient mucosal blood flow, a continuous adherent mucus layer (which, in combination with gastric bicarbonate secretion, allows the formation of a pH gradient so that the apical surface of gastric epithelial cells is at a neutral pH, despite gastric juice having a pH of 2–3), and the presence of regulatory peptides that can directly stimulate repair and also influence all of the other protective factors.

Ulceration occurs when this balance is disrupted due to additional aggressive factors [mainly *Helicobacter pylori* colonization or ingestion of nonsteroidal anti-inflammatory drugs (NSAIDs)], decreased mucosal defense such as an abnormal mucus layer or reduced production of growth regulatory peptides, or more likely, when there is a combination of the two. This idea would explain why only a minority of patients colonized with *H. pylori* go on to develop ulceration, a finding that cannot be explained at present by differences in the toxigenicity of *H. pylori* strains alone. It is also important to note that aggressive factors and defense mechanisms are not independent since a weakened mucosal defense might be due, in part, to a hyperimmune response from the host damaging the mucosa, a bystander effect, thereby reducing the production of growth regulatory peptides.

Mechanisms of Healing

When an injury occurs, the healing process can be considered in three phases. There is an initial rapid response involving the migration of surviving cells from the wound edge to cover the denuded area. This begins to occur within the first hour following injury and is termed restitution. This is followed by a much slower increase in proliferation and differentiation that only begins 1 to 2 days after the injury has occurred. An additional factor that is often overlooked is the final stage of remodeling, in which the mucosa slowly re-establishes an essentially normal-looking mucosa. Failure to achieve this final stage can leave a weakened mucosal defense against further damage and emphasizes the importance of the quality of healing.

FUNCTIONS OF PEPTIDES IN MUCOSAL INTEGRITY

There are at least 30 different peptides produced within the gut that are involved in mucosal integrity and repair. Although there have been many papers examining the effect of individual peptides on gastrointestinal function, there are little data regarding their function as part of an integrated defense mechanism *in vivo*. Different aspects of their pathophysiological role can therefore be inferred from considering them in isolation, belonging to broad groups based on their overall role in maintaining gut growth or in stimulating repair, and by examining the distribution of their sites of production compared with their sites of action (based on receptor distribution). It is also important to remember that modulation of tissue mass is not simply dependent on the rate of proliferation, but is dependent on the equilibrium established among cell production, migration, and loss (including apoptosis). Peptide growth factors can influence all of these aspects; for example, epidermal growth factor (EGF) stimulates cell proliferation and migration and also influences crypt fission, a recently identified mechanism by which new crypts are produced. Other peptides have also been shown to influence the rate of programmed cell death (apoptosis) within the gut, acting, for example, via the Fas/Fas ligand (FasL) signaling system. Fas is a member of the tumor necrosis factor α/nerve growth factor receptor family and is expressed in various cells including gastrointestinal mucosal cells. Binding of its ligand, FasL, triggers apoptosis.

CONSIDERATION OF REPRESENTATIVE INDIVIDUAL GROWTH FACTORS

EGF Receptor Ligand Family

This group of polypeptides, with the common property of binding to the EGF receptor (also known as the c-erb-1 receptor), includes EGF itself, transforming growth factor-α (TGF-α), amphiregulin, betacellulin, and heparin-binding EGF. Factors that probably also bind to this receptor but are less well identified include mammary-derived growth factor II and human milk growth factor III (which might be the same molecule as EGF). EGF receptor activation requires receptor–ligand interaction followed by dimerization with a subsequent intracellular signaling cascade involving pathways such as tyrosine kinase activity. The realization that heterodimerization with other members of the erb receptor family may also occur adds to the complexity of the situation and is an area that is poorly understood.

Epidermal Growth Factor

EGF is a 53-amino-acid peptide produced by the adult salivary glands and the Brunner's glands of the duodenum. Although there is no doubt that systemically administered EGF has multiple effects on

TABLE I Actions of the Growth Factor EGF

Action	Effect	*Possible* secondary message
Proliferation	↑	
Acid secretion	↑	Protein kinase C, cyclic AMP
Bicarbonate	↑	Prostaglandins
NaCl and glucose uptake	↑	**Brush border area, Na^+–glucose co-transporter, lipids**
Chloride secretion	↑/↓	Phosphatidylinositol 3-kinase
Amylase secretion—pancreas	↑&↓	Cyclic AMP, phospholipase C
Mucus secretion	↑	Prostaglandins
Gastrointestinal blood flow	↑	β-adrenergic, nitric oxide, prostaglandins
Smooth muscle contraction longit	↑	Prostaglandins
Smooth muscle contraction circular	↑	(densensitizes) not prostaglandins
Gastric emptying	↓	
Restitution	↑	Cell migration, prostaglandins
Permeability	↑	
Muscosal protection	↑	**Proliferation, polyamines, mucus, trefoil peptides**

Note. Adapted from Uribe *et al*. (1997).

gastrointestinal and other structures (Table I), there is continuing controversy regarding the physiological role of lumenal EGF. One reason for this controversy is that EGF, along with other ingested products, is susceptible to digestion. When EGF_{1-53} comes into contact with adult acidic gastric juice, it is rapidly digested to an EGF_{1-49} form that only has 25% of the biological activity of the intact EGF molecule. Once EGF enters the small intestine, it is susceptible to further proteolytic digestion under fasting conditions, but may be preserved in the presence of ingested food proteins. Lumenal concentrations of EGF (and other peptide growth factors) are therefore not just modulated only by the amount of production and secretion but are also highly dependent on the intralumenal milieu, particularly with regard to pH and proteolytic activity.

Most recent studies examining the distribution of the EGF receptor in the normal adult human gastrointestinal tract have found it to be present only on basolateral membranes and not on the apical (lumenal) surfaces. The distribution of the EGF receptors might, however, vary between species; for example, autoradiographic studies have been reported to show apical receptors in the pig intestine. If EGF receptors are distributed only on the basolateral membranes of the normal adult human gut, the EGF present within the intestinal lumen is unlikely to exert any biological activity except at sites of injury. Evidence in favor of this role for EGF includes the finding that rats that have had their salivary glands removed do not develop spontaneous ulcers or atrophy of the gut, but do develop increased ulceration and diminished repair if artificial ulcers are induced. This has led to the suggestion that EGF acts as a "lumenal surveillance peptide" in the adult gut, having little activity under basal (nondamaged) conditions but being readily available to stimulate the repair process at sites of injury. It is important to note, however, that lumenal EGF might gain access to basolateral receptors in the immature neonate gut because of its increased permeability.

Transforming Growth Factor-α

TGF-α is a 50-amino-acid molecule that is present in the mucosa throughout the gastrointestinal tract. Systemic administration of TGF-α stimulates gastrointestinal growth and repair, inhibits acid secretion, stimulates mucosal restitution after injury, and increases gastric mucin levels.

Within the small intestine and colon, TGF-α expression occurs mainly in the upper (nonproliferative) zones, which suggests that its major physiological role may be to influence differentiation and cell migration rather than cell proliferation. TGF-α may therefore play a complementary role with TGF-α to control the balance between proliferation and differentiation in the intestinal epithelium. Up-regulation of TGF-α expression has been shown to occur in the gastrointestinal mucosa at sites of injury as well as in the liver following partial hepatectomy, supporting a role for TGF-α in mucosal growth and repair. Further evidence for this role comes from research on mice that have had the TGF-α gene "knocked out" using homologous recombination. They have a relatively normal phenotype under control conditions but an increased sensitivity to colonic, although not small intestinal, injury. These findings support the role of TGF-α in maintaining epithelial continuity but suggest that the relative importance of peptides such as this might vary from one region of the gut to another.

Taken together, most studies suggest that the major physiological role of TGF-α is to act as a "mucosal integrity peptide" maintaining normal epithelial function in the nondamaged mucosa.

The Transforming Growth Factor-β Family

This family of molecules is structurally distinct from TGF-β and, in most systems, actually inhibits proliferation. There are at least five different isoforms of TGF-β and their major site of expression in the normal gastrointestinal tract is in the superficial zones where they may function to inhibit proliferation once the cells have left the crypt region. TGF-β has many diverse functions, which include being a potent chemoattractant for neutrophils and stimulating epithelial cell migration at sites of wounds. It is therefore likely to be a key player in stimulating restitution.

Insulin-like Growth Factors and Their Binding Proteins

Insulin-like growth factor-I (IGF-I) and IGF-II promote cell proliferation and differentiation. They are similar in structure to pro-insulin and it is possible that they also exert insulin-like effects at high concentrations. The liver is a major site of IGF synthesis. IGF-I is known to promote protein accretion; i.e., it is an anabolic agent and is at least partly responsible for mediating the growth-promoting activity of growth hormone. IGF-II also has anabolic activity and has been shown to reduce the catabolic state in starved animals.

Most of the biological activity of IGF is dependent on it being in a free form rather than being bound to one of its binding proteins. The physiological relevance of growth regulatory factors such as these can therefore not be understood without also considering their relevant binding factors. Six IGF-binding proteins (IGFBPs) have been identified and cloned. It was initially thought that their main function was to act as carrier proteins, reducing the proteolytic digestion of IGF and limiting its biological activity. Additional roles for IGFBPs have been suggested as a result of the fact that the different IGFBPs show distinct patterns of distribution in different tissues and their levels are altered in response to hormone or nutrient status. Examples include the findings that administration of dexamethasone to rats increases hepatic production of IGFBP-1 and that malnutrition of neonatal rats resulted in reduced serum IGF-I and IGF-II but caused an elevation in serum IGFBP-2. The detailed functions of the IGFBPs are unclear although it is probable that one of the roles of secreted or soluble IGFBP is to inhibit IGF-mediated proliferation or amino acid uptake by limiting the availability of free IGF to bind to its receptors. Conversely, cell surface/matrix-associated IGFBPs may potentiate the actions of IGF by increasing local concentrations of IGF-I and IGF-II next to their receptors.

Platelet-Derived Growth Factor

Platelet-derived growth factor (PDGF) is an acid-stable molecule that was originally identified from platelets but is also synthesized and secreted by macrophages. It consists of two disulfide-linked polypeptides: chain A (14 kDa) and chain B (17 kDa). The dimer therefore exists in three isoforms (AA, AB, and BB), which bind to tyrosine kinase-type receptors. PDGF is a potent mitogen for fibroblasts and arterial smooth muscle cells and oral administration of exogenous PDGF has been shown to facilitate ulcer healing in animals.

Vascular Endothelial Growth Factor

Vascular endothelial growth factor (VEGF) is a homodimeric 34 to 42 kDa heparin-binding glycoprotein with potent angiogenic, mitogenic, and vascular permeability-enhancing factors and it is related to PDGF. Specific receptors for VEGF have been identified on the apical membranes of the human colonic cell line Caco-2 and also the human cell line H-4. Although VEGF bound to these cell lines, it did not induce a proliferative response. The pathophysiological role of VEGF is therefore unclear although its angiogenic activity may play an important role in the healing of conditions such as peptic ulceration. In addition, there is currently interest in the use of anti-angiogenic factors, such as blockers of VEGF, for the treatment of malignant disorders.

Hepatocyte Growth Factor

Hepatocyte growth factor (HGF), one of the heparin-binding growth factors, is a potent mitogen for gastric cells *in vitro* and also acts as a morphogen. Expression of both the HGF gene and its receptor, c-*met*, is increased after induction of injury in various animal models. The timing of the response varies according to the method used but suggests that this signaling system may well have relevance in both the early and late phases of ulcer repair. HGF mRNA is located in stromal cells between the regenerative glands and in the arterial vessels in the submucosa and *met* mRNA is located in the epithelial cells of the regenerative glands. This system therefore provides an example of epithelial–mesenchymal cross talk (Table II) and is

TABLE II Production versus Receptor Distribution

Peptide	Source of peptide	Receptor distribution/effects
Transforming growth factor-α	Epithelial cells	c-erb-B1 receptor on epithelial cells
Transforming growth factor-β	Epithelial cells	Suppresses proliferation of enterocytes, stimulates restitution
Hepatocyte growth factor	Cells of mesodermal origin, e.g., fibroblasts	c-*met* receptor on epithelial cells
Basic fibroblast growth factor	Fibroblasts, activated macrophages	Fibroblasts and epithelial cells, stimulate restitution and angiogenesis
Trefoil factor family peptides	Mucus cells, epithelial cells at sites of injury	Poorly defined receptor present on epithelial cells, stimulates restitution not proliferation; mucus stabilization

further complicated by the fact that HGF is secreted from stromal fibroblasts as a single-chain, biologically inactive precursor (pro-HGF). This precursor is converted to an active heterodimeric protein by a novel serine protease (HGF activator). This proteolytic process is probably essential for HGF to exert its mitogenic activity. Interestingly, HGF is converted to its active heterodimeric form only in injured tissue, suggesting that selective activation of HGF by the HGF activator is a mechanism by which HGF action is localized to damaged tissues.

HGF has also been implicated in mediating abnormal gastric mucosal growth, resulting in a condition called large-fold gastritis, in some patients whose gastrointestinal tracts are colonized with *H. pylori*. Eradication of *H. pylori* resulted in a reduction in fold thickness and a parallel fall in HGF levels, suggesting that HGF is important in this trophic response. The molecular link between *H. pylori*, HGF, and abnormal mucosal growth may be mediated by the cytokine interleukin-β (IL-β), since administration of an IL-1β antagonist reduces the production of HGF in these patients. This situation therefore provides some insight into the link between the presence of *H. pylori*, cytokine and growth factor production, and abnormal mucosal growth. Similar mechanisms may also come into play to explain the relationship between *H. pylori* colonization and the development of gastric cancer.

Basic Fibroblast Growth Factor

Basic fibroblast growth factor (bFGF) is normally present in the human gastric mucosa and is a potent stimulant of angiogenesis. It accelerates the healing of experimental gastric and duodenal ulcers in rats and in patients with NSAID-associated gastric ulcers. In ulcerated human gastric mucosa, immunoactive bFGF is unregulated in the granulation tissue, endothelial cells, mononuclear cells, and epithelial cells at the ulcer rim. Taken together, it seems likely that this peptide plays an important role in reestablishing mucosal homeostasis following injury.

Keratinocyte Growth Factor

Keratinocyte growth factor (KGF), a member of the fibroblast growth factor family (FGF7), is a potent stimulant of keratinocyte proliferation, binding to a splice variant of FGF receptor 2. KGF has been shown to ameliorate damage in an experimental model of colitis in rats but KGF mRNA expression was not increased after indomethacin- or acetic acid-induced gastric injury in rats and was ineffective in reducing indomethacin-induced gastric damage. Current data are therefore insufficient to judge whether KGF plays an important role in gastric ulcer healing.

Trefoil Peptides

The trefoil motif consists of a unique "three-loop" structure formed by intrachain disulfide bonds. Three members of this family have been identified in humans and various nomenclatures have been used to describe them in the past. The current accepted practice is to term them all as trefoil factor family members 1–3 (TFF1–3). TFF1 (previously pS2) and TFF2 (previously spasmolytic polypeptide) are present mainly in the mucus-producing cells of the stomach; TFF1 is produced by the superficial epithelial cells throughout the stomach and TFF2 is produced by the deeper glandular elements within the gastric antrum. In contrast, TFF3 (previously intestinal trefoil factor) is predominantly produced by goblet cells of the small and large intestine. All three peptides are also produced at sites of injury, such as those that occur with gastric ulceration or Crohn's disease. Recent evidence suggests that the trefoil peptides have two roles in the gastrointestinal tract: (1) under basal conditions, they may play a role in mucus stabilization and (2) when an acute injury occurs, their rapid up-regulation is important

in stimulating the repair process, particularly that of epithelial restitution, acting as an example of a rapid response peptide.

The trefoil peptides possess the unusual property of decreasing injury and stimulating repair by processes independent of proliferation. This aspect seems particularly attractive when considering using them for treating patients because the pro-mitogenic activity of some of the other peptides could potentially stimulate premalignant lesions elsewhere in the body.

Gut Hormones as Growth Factors

The bowel shows a remarkable ability to respond to changes in dietary intake. Cross-circulation experiments support the concept of circulating trophic factors influencing gut growth although the identity of such a factor(s) remains unclear. Gastrin probably plays a role as a trophic factor for mucosal growth within the stomach and there is currently much interest in the role of glucagon-like peptide-2 (GLP-2), as systemic infusion of GLP-2 resulted in a general trophic response within the gut. This has led to the possibility that GLP-2 ligands may be useful in stimulating growth of the bowel in conditions such as short bowel syndrome. In contrast, early enthusiasm for a major trophic role for the gut hormones peptide YY and cholecystokinin within the gastrointestinal tract has diminished due to the absent or weak response in gut growth when recombinant forms of the hormones have been infused.

REDUNDANCY AND COMPLEX INTERACTIONS

Genomic removal of growth factor expression (producing knockout mice) has provided important insights into the pathophysiological role of several peptides. At first sight, the finding that deletion of factors with potent biological activity such as EGF or TGF-α results in an essentially minimal alteration in mucosal homeostasis seems surprising. However, it has subsequently been shown that much more extreme phenotypes can be generated by cross-breeding to produce double or triple knockouts for EGF receptor ligands. When these animals are produced, much more serious phenotypical abnormalities are caused and deletion of the EGF receptor itself results in multiple problems in the gastrointestinal tract and elsewhere, with fatal consequences. The most likely explanation for these findings is that there is redundancy within the system, allowing mucosal homeostasis to be maintained even if an individual peptide is lost for whatever reason. A further important finding from these studies is that the phenotypes resulting from such experiments are often strain dependent. The fact that major variations can occur within a single species therefore requires that caution be exercised in extrapolating findings from mice to humans.

The cooperative effects of peptides involved in the healing process can be mediated via distinct receptors (e.g., trefoil peptides and EGF) or by the same receptor (e.g., EGF and TGF-α, which both bind to the EGF or c-erb-B1 receptor). In addition, a number of growth factors, such as EGF, are monomeric in solution yet have been shown to activate their cognate receptors by inducing dimerization. It is therefore important to realize that, *in vivo*, EGF may signal to the cell through the formation of receptor homodimers (where two c-erb-B1 receptors link together) or through the formation of heterodimers (where one c-erb-B1 receptor binds to another activated member of the c-erb-B receptor family, such as c-erb-B4, which binds heregulin). This ability to form heterodimers increases markedly the repertoire of ligand–receptor interactions and is an area that can be overlooked if the effects of adding a single growth factor to cell lines *in vitro* are being examined.

A much greater understanding of the physiological roles of peptides can be achieved by considering the epithelium not in isolation but as part of an integrated cross-talking mechanism involving the epithelium and its underlying mesenchyme (see Table II). Control of growth and differentiation of a cell is influenced by factors secreted into the local microenvironment and may be mediated in a paracrine (cell to cell) manner or indirectly by influencing the extracellular matrix composition. A further example of potential cross talk between various systems is demonstrated by the relationship between EGF and trefoil peptides. In the normal mucosa, lumenal EGF is unable to reach its receptor because of its basolateral distribution. However, at sites of injury, EGF can bind to its receptor and, in addition to direct proliferative stimulation, may induce expression of trefoil peptide production (as at least one member of the family has an EGF-responsive enhancer region) as well as that of its own receptor. When produced, the trefoil peptides can act in a synergistic manner with EGF to stimulate the repair process.

CLINICAL APPLICATIONS OF GROWTH FACTORS

General Considerations

Advances in recombinant peptide technology now allow virtually pure peptide to be produced at moderate

cost. This removes the concern of contamination with infectious agents (as seen in some patients treated with purified growth hormone who develop Creutzfeldt–Jakob disease) and the development of antibodies to purified peptides from other species (as seen with porcine insulin). The most widely prescribed recombinant peptide used by gastroenterologists is probably interferon-α for the treatment of infectious liver disease. The use of growth factors for hollow organ gastroenterology is, however, at a much earlier stage. One of the major areas of concern regarding the use of peptides such as EGF in the clinical setting is the worry that systemic administration of growth factors may act to promote tumor growth elsewhere in the body. Oral, as opposed to systemic, administration of growth factors provides one possible approach to this problem. However, for EGF and the trefoil peptides, the oral doses required to treat gastrointestinal damage may be up to 1000 times greater than when the peptide is given systemically, making oral therapy economically unrealistic. Stabilizing molecules against digestion in the small intestine is likely to be more difficult due to the many different proteolytic enzymes produced by the pancreas. Possible strategies include administering the peptide along with nonspecific serine protease inhibitors or administering the peptide in a site-specific release formulation to overcome these problems.

Specific Disease Considerations

Recombinant peptides are unlikely to be of value for the treatment of esophagitis or peptic ulceration due to the high efficacy of acid suppressants and *H. pylori* eradication regimens. There are, however, a number of serious gastrointestinal pathologies for which novel therapies might prove useful. Some examples are given below.

Short Bowel Syndrome

Patients usually acquire this condition as a result of massive intestinal resection for vascular insufficiency or following repeated operations for inflammatory bowel disease. Current therapeutic options consist of long-term parenteral (intravenous) feeding or, in a few selected cases, small bowel transplantation. However, both of these options are associated with high cost and morbidity. Systemic administration of individual growth factors such as EGF has been shown to stimulate bowel growth in animal models of short bowel syndrome and parenteral administration of trophic gut hormones such as GLP-II also show promise. Clinical trials are ongoing.

Chemotherapy-Induced Mucositis

In recent years, high-dose chemotherapy protocols are being increasingly used in an attempt to improve cancer cure rates. As a result of this escalation in dosing, toxic side effects on the bone marrow or gastrointestinal tract can be the factor limiting the dose or duration of treatment. The breakdown in mucosal integrity may range from oral stomatitis to massive intestinal ulceration. Strategies to protect these tissues and encourage their recovery may therefore facilitate the use of higher dosage with greater potential for cure. Several peptide growth factors, such as EGF, KGF, and TGF-β, are currently under examination, although it is important to realize that the great initial promise seen in animal models has yet to be confirmed in clinical trials. For example, in a phase I clinical study of patients undergoing chemotherapy, EGF had only a minor beneficial effect in reducing mouth ulceration. If peptides with growth stimulatory or inhibitory effects are to be used, the timing of administration is likely to be critical; growth-arresting factors might protect bone marrow or gut from the damaging effects of chemotherapy (which tends to affect areas with the highest cell turnover) if given prior to chemotherapy. In contrast, growth-stimulating factors might "rescue" the recovery of injured areas if administered following chemotherapy. This latter approach is already being used clinically as colony-stimulating growth factor is used to stimulate bone marrow recovery following chemotherapy.

Inflammatory Bowel Disease and Necrotizing Enterocolitis

The etiology of ulcerative colitis and Crohn's disease is unknown and current treatment of these severe, incapacitating conditions therefore must be on an empiric basis. Several peptide growth factors are at various stages of investigation. These include EGF, PDGF, TGF-α, IGF-I, KGF, and trefoil peptides. Most are still at a very early (animal model) stage and are unlikely to be in standard clinical usage for many years although a preliminary report of the use of EGF in the treatment of patients with left-sided colitis seems encouraging and provides proof of concept.

Necrotizing enterocolitis is a severe life-threatening illness of young children that causes severe ulceration of the small and large bowel. Its etiology is unclear and initial treatment consists of general supportive measures consisting of fluid replacement and antibiotic therapy, although intestinal resection is often required. A recent case report indicated that a continuous infusion of EGF had a remarkable restorative effect on gut histology and larger clinical trials are ongoing.

Liver Disease

The use of growth factors in the area of hepatology is at an early stage but may be of value in stimulating hepatic regeneration following major surgical resections or for reducing injury and stimulating repair of liver function following exposure to hepatotoxins. For example, EGF has been shown to reduce the amount of hepatotoxicity induced by the toxin carbon tetrachloride in rats. The mechanisms underlying this protective effect are unclear but may be related to the induction of free-radical scavengers as well as growth stimulatory effects.

Multiple-Organ Failure

Multiple-organ failure (MOF) is a severe, life-threatening condition that usually occurs as a result of major trauma, burns, or fulminant infections. Whatever the initiating event, once established, MOF has a high mortality (up to 80%). The pathophysiological mechanisms underlying this condition are unclear although important contributory factors probably include hypoxia, increased intestinal permeability, bacterial translocation, endotoxemia, and uncontrolled systemic inflammatory responses. Several factors, including EGF, are currently under examination as potential therapeutic agents to prevent or reduce the severity of this condition.

SUMMARY

It is now clear that multiple peptides are involved in maintaining gut mucosal homeostasis and in stimulating the repair process. Important information has been obtained by considering the individual factors in isolation but additional insight can be obtained by considering them as belonging to broad groups dependent on their pathophysiological role. Knockout mice studies have confirmed the importance of many of these peptides but also demonstrated that there is redundancy in the system. The advent of recombinant peptide technology offers the potential to use one or several (to elicit synergistic responses) peptides to treat a variety of gastrointestinal conditions. It is important to note, however, that although cell culture and animal models have shown proof of concept, the process of translating their use to standard clinical practice is still at a relatively early stage.

See Also the Following Articles

Colitis, Ulcerative • Crohn's Disease • Epithelial Barrier Function • Epithelium, Proliferation of • Epithelium, Repair of • Growth Hormone • *Helicobacter pylori* • Necrotizing Enterocolitis • Short Bowel Syndrome • Transforming Growth Factor-β (TGF-β)

Further Reading

Barnard, J. A., Beauchamp, R. D., Russell, W. E., Dubois, R. N., and Coffey, R. J. (1995). Epidermal growth factor-related peptides and their relevance to gastrointestinal pathophysiology. *Gastroenterology* **108**, 564–580.

Playford, R. J. (1995). Peptides and mucosal integrity. *Gut* **37**, 595–597.

Playford, R. J., Macdonald, C. E., and Johnson, W. S. (2000). Colostrum and milk-derived growth factors for the treatment of gastrointestinal disorders. *Am. J. Clin. Nutr.* **72**, 5–14.

Podolsky, D. K. (2000). Mechanisms of regulatory peptide action in the gastrointestinal tract: Trefoil peptides. *J. Gastroenterol.* **35**(Suppl.), 1269–1274.

Uribe, J. M., and Barret, K. E. (1997). Non-mitogenic actions of growth factors: An integrated view of their role in intestinal physiology and pathophysiology. *Gastroenterology* **112**, 255–268.

Growth Hormone

DOUGLAS W. WILMORE
Brigham and Women's Hospital and Harvard Medical School

acromegaly State of increased growth in adults due to the oversecretion of GH. The patient is unusually tall, with an enlarged jaw, and large hands and feet. Because this is most commonly related to a pituitary tumor, with tumor enlargement the patient experiences headaches and often visual disturbances.

binding proteins Large molecules that circulate in the bloodstream and attach to smaller molecules such as hormones. These proteins tend to prolong the life of the bound hormones by keeping them in the bloodstream.

endocrine Secretion of a hormone from a gland and transportation of the hormone in the bloodstream to distant sites, where the hormone exerts its action.

insulin resistance Glucose uptake, primarily by adipose tissue and skeletal muscle, is enhanced by the hormone insulin. A resistant state occurs when the same quantity of insulin fails to cause normal uptake of glucose by these tissues.

paracrine Secretion of a hormone from a cell, with the hormonal effect exerted on an adjacent cell.

pituitary gland Small endocrine organ located at the base of the brain; governs growth, metabolism, and reproduction through the secretion of a variety of hormones.

Growth hormone, a protein composed of 191 amino acids, is secreted by the somatotroph cells of the anterior pituitary gland. In children, this hormone is responsible for linear growth and enhancement of lean body mass through the stimulation of protein synthesis. In adults, growth hormone is directly or indirectly responsible for normalization of body composition, improved skeletal muscle and cardiac function, and maintenance of normal serum lipids. In all age groups, the hormone appears to have an effect on the central nervous system, improving cognitive function and life quality.

INTRODUCTION

Growth hormone (GH) elaboration increases in the first 10 years of life, achieves its maximum secretory levels at the time of puberty, and then secretion and blood levels gradually decrease with age. Deficiencies of GH in children result in stunted growth, and excesses due to pituitary tumors result in acromegaly (or gigantism in a child). More subtle are deficiencies that occur with aging, and these have only recently being diagnosed and reversed by exogenous administration of the hormone.

ENDOCRINOLOGY

GH, like some other pituitary hormones, is released from the somatotroph cells of the anterior pituitary gland in a pulsatile or episodic manner, with the greatest pulses coming at night. Secretion of GH is controlled by two antagonistic systems, growth hormone-releasing hormone (GHRH) and somatostatin. GHRH has a marked stimulatory effect whereas somatostatin, which is released from neurons in the median eminence of the brain, inhibits GH secretion. Both of these controlling factors are secreted into the hypophyseal portal system and transported to the anterior pituitary gland, where their interaction with the somatotroph cells accounts for the quantity of hormone secreted.

When GH is released into the bloodstream, about 50% of the hormone is attached to a binding protein. It is transported to various tissues, where it binds to two receptors, leading to homodimerization, which generates a biological response. Signal transduction then proceeds via tyrosine phosphorylation of Janus kinase (JAK) proteins, a common step for substances stimulating cells via the cytokine receptor superfamily, which includes the GH receptor. Once the target cell is stimulated, there are a variety of responses, but the most apparent is the generation of a second hormone, insulin-like growth factor-1 (IGF-1). Although this substance can be synthesized and secreted by multiple cell types, the greatest expression occurs in the liver. Here elaboration is clearly under GH control, and IGF-1 is released into the bloodstream and effects other tissues. In addition to the elaboration of IGF-1, GH also induces the synthesis of binding proteins, the most prominent one being insulin-like growth factor binding protein-3 (IGFBP-3). This binding protein greatly prolongs the half-life of IGF-1, and thus hormonal binding and transport within the bloodstream are regulatory steps in controlling anabolic action (Table I).

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Comparison of GH, IGF-1, and Insulin

Parameter	GH	IGF-1	Insulin
Molecular mass (Daltons)	22,000	7649	5734
Structure	Folded chain with disulfide bridges	One chain	Two chains
Origin	Anterior pituitary gland	Mostly liver; all tissues	Pancreatic beta cells
Secretion	Pulsatile	Constant, slow release	Pulsatile
Circulating form	About 50% bound	Mainly bound	Free
Production rate	0.2–1 mg/day in adults	10 mg/day	2 mg/day
Half-life	6–20 minutes	12–15 hours	10 minutes

IGF-1 actions are mediated via its receptor, although it can also be bound to the insulin receptor, but at a much lower affinity. IGF-1 stimulates cell proliferation and differentiation. In addition, it can cause hypoglycemia, stimulate protein synthesis, and enhance wound healing. IGF-1, although having a longer half-life than GH, is responsive to dietary perturbations. With fasting or reduction in carbohydrate ingestion, IGF-1 levels decrease, and then return to normal with refeeding. Thus, there appears to be an important regulatory interaction between substrate availability and this anabolic growth factor—e.g., if no nutrients are available, there is no need for a hormone to stimulate cell growth. In addition, IGF-1 elaboration is attenuated during states of selective nutrient deficiency, such as lack of zinc and manganese, and during diseases associated with inflammation. The interactions between GH and IGF-1 are shown in Fig. 1.

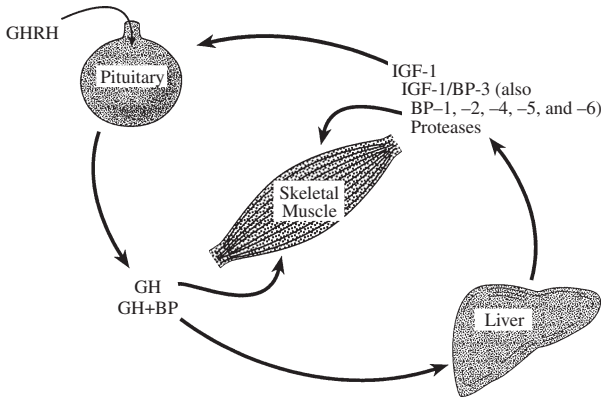


FIGURE 1 Growth hormone (GH) is secreted by the pituitary gland following stimulation by growth hormone-releasing hormone (GHRH). GH travels in the bloodstream, either in the free state or loosely bound to a binding protein (BP). It attaches to cell surface receptors and stimulates changes in metabolism. In the liver, it stimulates the production of insulin-like growth factor-1 (IGF-1) and its binding proteins. During disease states, there is also release of proteases that cleave IGF-1 from BP-3, and this reduces IGF-1 activity. Both GH and IGF-1 have mechanisms that control feedback to the brain, regulating GH elaboration.

PHYSIOLOGY

GH and IGF-1 (the latter working through both endocrine and paracrine mechanisms) work alone or in concert to exert a variety of metabolic effects on an organism (Fig. 2). IGF-1 causes hypoglycemia by decreasing glucose production and enhancing peripheral glucose clearance. In contrast, GH causes insulin resistance, thereby counteracting this undesirable effect of IGF-1. In addition to this antiinsulin effect, GH causes sodium and water retention (probably the most common side effect associated with the administration of GH) and stimulates the mobilization and utilization of adipose tissue. In the whole organism, this results in the loss of body fat and a lowering of blood lipids. This lipolytic effect is countered in part by the actions of IGF-1. Both hormones have pronounced effects on protein synthesis and epiphyseal (bone) growth. GH causes growth of all organs, and studies in both animals and humans have documented that excess GH treatment results in visceromegaly, with enlargement of the liver, kidneys, spleen, and gastrointestinal tract.

The net effect of GH, in conjunction with the secretion of IGF-1, is to produce a marked anabolic state. GH administered to animals or humans accelerates or normalizes growth. This is most easily observed in young children, but changes also occur in the adult. There is enhanced bone growth, a greater lean body mass (characterized by more muscle), and less fat (Fig. 3). If biochemical tests are performed, treated subjects are more resistant to insulin (more diabetic-like) and have lower lipid levels. Performance of cognitive and physical tasks is improved, especially when compared to GH-deficient subjects.

PATHOLOGIC CONDITIONS

GH Deficiency

Children

Congenital growth hormone deficiency can occur as a result of some type of anatomic developmental

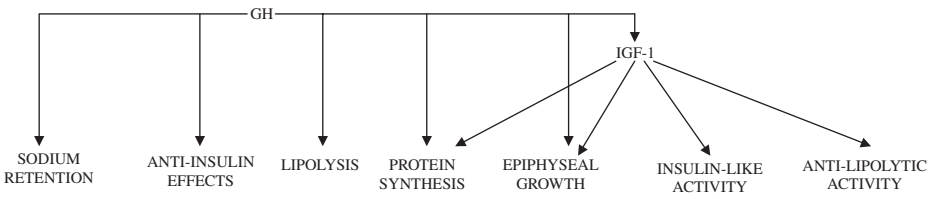


FIGURE 2 Multiple effects of growth hormone (GH) and insulin-like growth factor-1 (IGF-1).

abnormality or is associated with a genetic defect. Children may also acquire a growth hormone-deficient state as a result of radiation for tumors of the central nervous system or associated with resection of hypothalamic, pituitary, or other tumors in the adjacent areas of the brain. Trauma to the area and some infiltrating diseases have also been associated with the deficiency state.

Relative GH deficiency has been described in conjunction with chronic renal failure in children who fail to grow, and in those who receive steroid therapy as immunosuppressive therapy following organ transplantation. Turner's syndrome, a congenital abnormality in girls, is associated with short stature and is responsive to GH therapy.

Adults

The great majority of adults with GH deficiency have diseases associated with the pituitary gland (tumors, adenomas, infarction, or infiltrative diseases), most of which are treated surgically and/or with irradiation. However, GH deficiency is now being recognized in older adults, and when diagnosed, is often treated by replacement therapy.

GH Excess

GH excess is associated with pituitary tumors that secrete excessive amounts of GH. The result is enhanced growth (referred to as gigantism in the child or acromegaly in the adult). Such tumors are also associated with headaches, visual disturbances, abnormal sweating, and, eventually, weakness and fatigue. There is gradual and unexplained swelling of the feet and hands and thickening of the lips and tongue. Removal of the tumor (almost always a pituitary adenoma) necessitates GH replacement, but often resolves many of the unpleasant symptoms.

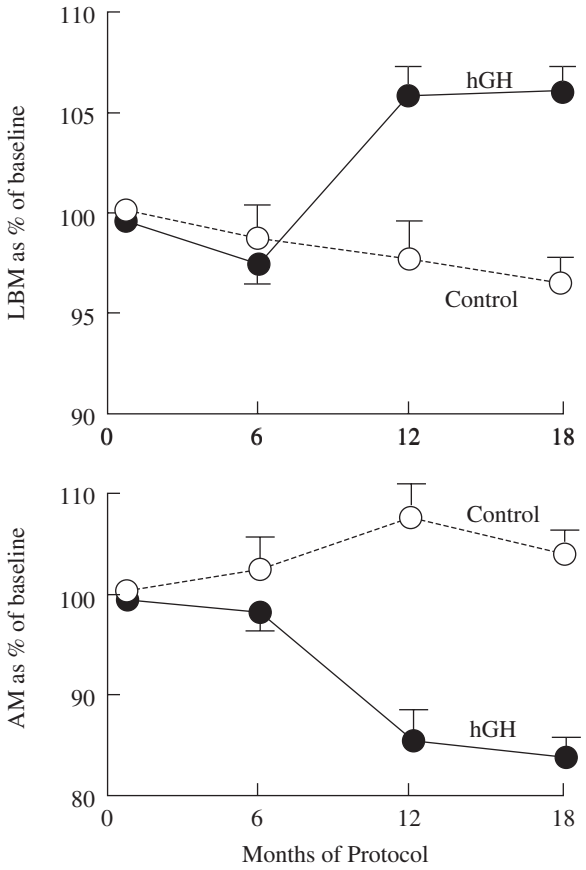


FIGURE 3 In adults, lean body mass (LBM) and adipose mass (AM) change with human growth hormone (hGH) administration, compared with controls who received placebo injections (each point is the mean ISEM for 16–21 individuals). Reproduced with permission from Cohn, L., Feller, A. G., Rudman, I. W., and Rudman, D. (1993). Further observations of the effects of human growth hormone in elderly hyposomatomedinemic men. *In* "Growth Hormone Replacement Therapy in Adults: Pros and Cons" (Z. Laron and O. Butenandt, eds.). Freund Publishing House, London.

GH THERAPY

GH was the first hormone to be produced by recombinant DNA technology and has been abundantly available for clinical use and study since the mid-1980s. Before that time, GH was extracted from pituitary glands obtained from cadavers, and only a limited quantity of GH was available for therapeutic and investigative use.

Testing for Deficiency

GH is rapidly secreted in response to hypoglycemia (rapidly falling blood glucose). Therefore, one standard test is to administer a small dose of insulin to induce hypoglycemia, and then draw serial blood samples for GH determination. An increase of less than 10 μg/liter from baseline is considered diagnostic of deficiency in children. In adults, the increase must be $<3-5$ μg/liter to diagnose the deficiency state (the value used depends on the method of blood testing). Other stimulation tests are available [e.g., administration of L-dihydroxyphenylalanine (L-DOPA), arginine, and GHRH] and similar guidelines are generally followed if these alternate stimulatory tests are used.

Results in Children

In cases of classic growth hormone deficiency, the administration of GH has resulted in a marked increase in linear growth (Fig. 4) that continues until puberty. In one study, GH-deficient children treated for an average of 8 years had a final height of 171.6 cm (5 ft. 7 in., boys) and 158.5 cm (5 ft. 2 in., girls). The average dose of GH given to children is 0.3 mg/kg/week. This is injected subcutaneously daily or every other day in divided doses. The recommended GH doses for children with renal failure and for Turner's syndrome are slightly higher (0.35 and 0.375 mg/kg/week, respectively) due to the resistance states associated with these disorders.

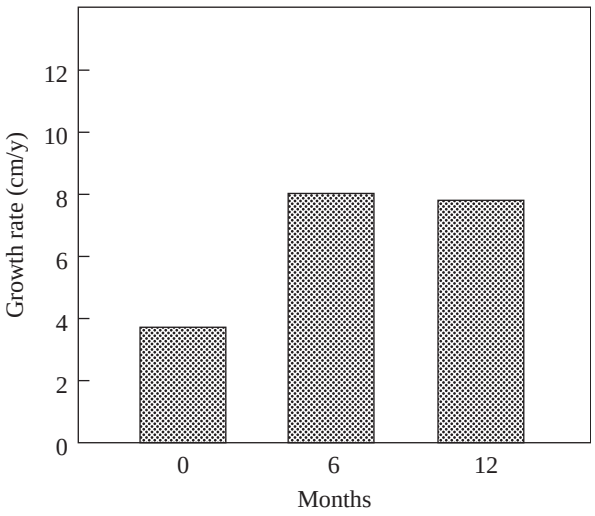


FIGURE 4 The growth rate in GH-deficient children after 6 and 12 months of treatment. Reproduced with permission from Takano, K., Shizume, K., Hizuka, N., *et al.* (1987). Therapeutic benefit of the treatment of GH-deficient children with growth hormone. *In* "Growth Hormone—Basic and Clinical Aspects" (O. Isaksson *et al.*, eds.). Elsevier Science Publ., New York.

Results in Adults

In adult individuals who have documented GH deficiency due to pituitary diseases or associated with aging, replacement doses range from 3 to 25 μg/kg/day (about 0.3−2 mg/day for the average adult). However, a recent consensus conference suggests that the starting dose should be 0.15−0.3 mg/day regardless of body weight. The dose is then adjusted to administer adequate quantities of GH so that IGF-1 levels are elevated to the midnormal range.

The responses to this therapy have been dramatic, especially when compared to the deficient state. In one study, treatment for 26 weeks resulted in a 13% decrease in subcutaneous fat mass and a 30% decrease in intra-abdominal fat mass. Muscle mass and strength were also characteristically increased.

Quality-of-life studies, which include energy levels, mood, and measures of emotional liability, all show general improvement with GH replacement. Long-term outcome studies are lacking, and it is not known at this time if GH therapy is associated with a reduction in cardiovascular risk or with an enhancement in cancer risk. Data from followup studies of patients receiving GH replacement therapy following excision of brain (pituitary) tumors show that, to date, concerns about GH stimulating tumor growth or recurrence are unfounded. Longer term studies of large groups of individuals are needed, however, to answer this question.

Other Uses

GH has been administered to small groups of patients with catabolic conditions, including respiratory failure, burn injury, recovery from surgical procedures, congestive heart failure, liver transplantation, and renal failure. These diseases process are associated with the accelerated breakdown of body tissues, particularly skeletal muscle, and this results in weakness, malnutrition, and poor wound healing, all of which prolong recovery. It is hoped that GH therapy will enhance protein synthesis and thus limit mortality, morbidity, and convalescence time.

The most consistent benefit of high-dose short-term GH therapy has been observed in children following thermal injury. Herndon and associates have shown that such therapy has accelerated wound healing and reduced hospitalization by approximately 2 weeks (Table II). In another large randomized trial in critically ill adults, high-dose growth hormone was associated with increased mortality. This observation has not been made in other studies, but enthusiasm for treatment of the critically ill has been tempered and few trials evaluating GH in this group are ongoing.

TABLE II GH Therapy in Thermally Injured Children

Factor	Control	GH
Wound healing rate (days/% burn)	0.8 ± 0.1	0.54 ± 0.04*
Length of hospitalization	46 ± 6	32 ± 4*

*$p < 0.05$

The anabolic effects of GH have also been utilized to support patients with HIV infection and associated wasting. Studies in such populations show that such GH therapy is effective in supporting lean tissue (muscle mass) during associated catabolic states.

TABLE III Side Effects of GH

Weight gain	Headaches
Dependent edema	Benign intracranial hypertension
Sensation of tightness	Hyperglycemia
Carpal tunnel syndrome	Hyperphosphatemia
Arthralgia	Cardiomegaly
Myalgia	Cardiac arrhythmias
Hypertension	

Side Effects of GH

The most common side effects of GH administration are fluid retention, swelling of the hands and feet, and the occurrence of joint pain and myalgesias. Hyperglycemia may also occur, especially in an individual prone to having diabetes. Other side effects are listed in Table III.

See Also the Following Articles

Growth Factors • Somastostatin

Further Reading

Juul A., and Jorgensen, J. O. L. (2000). "Growth Hormone in Adults," 2nd Ed. Cambridge University Press, Cambridge, U.K.

Smith, R. G., and Thorner, M. O. (eds.). (2000). "Human Growth Hormone. Research and Clinical Practice." Humana Press, Totowa, N.J.

Thapar, K., Kovaas, K., Scheithauer, B. W., and Lloyd, R. V. (eds.). (2001). "Diagnosis and Management of Pituitary Tumors." Humana Press, Totowa, N.J.

Torosian, M. H. (ed.) (1996). "Growth Hormone in Critical Illness—Research and Clinical Studies." R. G. Landes, Austin, Texas.

Guanylin

LEONARD RALPH FORTE
Harry S. Truman Memorial Veteran's Administration Hospital and University of Missouri

cyclic GMP Intracellular signaling molecule that is produced by the cell surface receptors for guanylin and uroguanylin in target cells. Responsible for mediating cellular responses to these peptide hormones.

epithelial cell balance Physiological process that controls the birth and death rates of cells to maintain the morphology and function of epithelia in organs of the body.

guanylin/uroguanylin Small peptide hormones produced in the intestine, where they act locally to regulate intestinal functions, including the secretion of electrolytes and fluid into the intestinal lumen.

intestine–kidney endocrine axis Hormonal connection between the intestine and kidney formed by enteric guanylin and/or uroguanylin peptides that are secreted into the bloodstream following the ingestion of dietary sodium chloride (i.e., table salt).

Guanylin and uroguanylin are small, heat-stable peptides that are produced throughout the body, but are most abundant in the intestinal tract, where they are secreted locally to regulate epithelial cell functions through the intracellular second-messenger molecule, cyclic 3′5′-guanosine monophosphate (cyclic GMP). Apical plasma membranes of epithelial cells lining the intestinal lumen express high levels of a guanylin/uroguanylin receptor. When the receptor is activated by uroguanylin/guanylin binding to an extracellular portion of this signaling molecule, cyclic GMP is produced to regulate cellular functions. Certain strains of enteric bacteria produce heat-stable enterotoxin peptides that are molecular mimics of uroguanylin. The toxin peptides secreted by some enteric microbes produce an unregulated stimulation of uroguanylin/guanylin receptors, causing a cholera-like disease, often called "traveler's" diarrhea. The high incidence of this disorder in people in developing nations may be linked to the markedly lower incidence of colon cancer found in the developing world, attributed to a protective effect of the guanylin peptide family of cyclic GMP-regulating hormones, as well as to their molecular mimics.

INTRODUCTION

Guanylin and uroguanylin, peptide hormones that are produced in all tissues of placental (i.e., eutherian) mammals, are found most abundantly within epithelial cells of the gastrointestinal tract. In a representative metatherian mammal, the opossum, and in some species of bony fish, three different guanylin-like peptides have also been identified. The biologically active forms of guanylin peptides are highly conserved throughout vertebrate evolution. Most of the guanylin peptides are 15–16 amino acids long and contain two intramolecular disulfide bonds formed between specific cysteine pairs, which provide for a heat- and acid-stable molecule that is biologically active (Fig. 1). Enteric bacteria produce heat-stable toxin (ST) peptides, which serve as molecular mimics of uroguanylin and induce a cholera-like form of diarrhea. The ST peptides secreted by diarrhea-inducing strains of *Escherichia coli*, *Vibrio cholerae*, and *Yersinia enterocolitica* have variable structures, but all of these ST molecules contain a highly conserved core peptide region consisting of 13 residues with three disulfide bonds that are essential for full biological activity. This bioactivity domain of ST peptides is most closely related to a similar region within the biologically active forms of uroguanylin. Bacterial ST peptides have a greater affinity for the intestinal receptors compared to that of uroguanylin, which contributes to the toxicity of STs *in vivo* (Fig. 1). Piscine (i.e., fugu) forms of uroguanylin and bioactive ST peptides secreted by enterotoxigenic strains of bacteria have the closest sequence similarity. A novel bioactive peptide, lymphoguanylin, has been identified in the opossum, and this peptide has only three cysteine residues and one intramolecular disulfide (Fig. 1).

In the intestinal tract, locally secreted guanylin and uroguanylin peptides participate in regulating the physiological process of fluid secretion during digestion. Biological actions have also been demonstrated in the kidney, as demonstrated by increases in urinary sodium chloride (salt), potassium, and water excretion following administration of either guanylin or uroguanylin. Because guanylin and uroguanylin were discovered only 10 years ago, it is likely that other physiological roles will be elucidated for these peptide hormones in the future. The widespread distribution of guanylin/uroguanylin-producing tissues and the broad distribution of their cognate cyclic GMP-producing receptor,

 Copyright 2004, Elsevier (USA). All rights reserved.

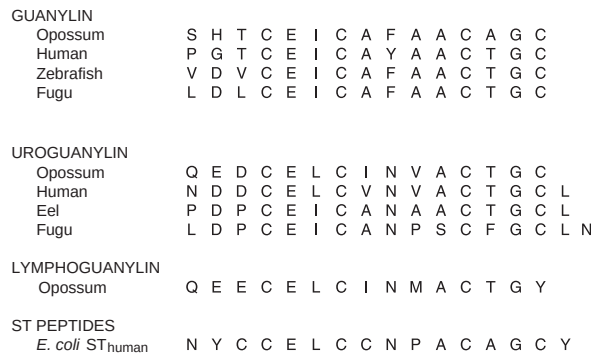


FIGURE 1 Primary structures of guanylin peptides. Amino acids are abbreviated using the single-letter code. Disulfide bonds connect the first to third and second to fourth cysteines in the linear peptide chain of guanylin and uroguanylin. Fugu (puffer fish) and zebrafish peptide sequences were derived from genome sequencing projects; other peptide sequences have been previously published. Heat-stable toxin (ST) peptides also have a third disulfide bond connecting their unique cysteine residues. *Escherichia coli* ST has three additional amino acids (i.e., NSS) on the N terminus, forming a 19-residue peptide.

clearly suggest that these peptide hormones influence cell and organ function throughout the body.

CELLULAR MECHANISM OF GUANYLIN AND UROGUANYLIN ACTION

Uroguanylin and guanylin bind to and activate a specific cell surface receptor that has an intrinsic catalytic activity that produces the intracellular second messenger, cyclic GMP. These receptor enzymes, guanylate cyclases, are localized to plasma membranes of various cells in the body. Thus, guanylin and uroguanylin bind to a receptor guanylate cyclase (R-GC) found on the surface of target cells and regulate cellular functions by stimulating the enzymatic portion of R-GCs to produce cyclic GMP from guanosine triphosphate (GTP). Cyclic GMP inside the cell then binds to specific receptor proteins, which can be either cyclic GMP-dependent protein kinases, cyclic GMP-regulated ion channels, or cyclic GMP-modulated phosphodiesterases.

One physiological mechanism that has been relatively well studied involves the robust stimulation of intestinal fluid and electrolyte secretion produced by guanylin, uroguanylin, and ST peptides mediated by intracellular cyclic GMP. In this example, cyclic GMP binds to and activates one or more protein kinases in certain specialized cells within the intestinal epithelium that are capable of secreting chloride and/or bicarbonate anions. Activated protein kinases catalyze the transfer of phosphate from adenosine triphosphate (ATP) to serine residues on the cystic fibrosis transmembrane conductance regulator (CFTR) protein, which is intimately involved in controlling the secretion of chloride and bicarbonate anions into the intestinal lumen. Phosphorylated CFTR is also associated with a reduction in the intestinal absorption of sodium into the body. The net physiological effect of either guanylin or uroguanylin is the stimulation of intestinal secretions containing the electrolytes sodium, chloride, and bicarbonate. The resulting fluid secretion permits the digestion and absorption of nutrients and other dietary factors.

REGULATION OF SODIUM CHLORIDE HOMEOSTASIS

Both guanylin and uroguanylin are produced in the intestine and then are released into the bloodstream when sodium chloride (salt) is ingested in food and drink. Secretion of these peptides into the bloodstream forms a novel endocrine axis linking the intestine with the kidney via guanylin and/or uroguanylin hormones. A fundamental response to the ingestion of salt in our diets is a marked increase in the excretion of urinary sodium chloride, which helps balance renal salt excretion and dietary salt intake, thus maintaining salt homeostasis in the body. Guanylin peptides stimulate the excretion of sodium, chloride, and potassium in the urine and also increase urinary volume. Kidney tubules express the R-GC type of receptor, but renal tubular cells also appear to have another form of receptor and signal transduction pathway that may be independent of R-GC and cyclic GMP. In fact, kidney cell responses to the peptides appear to be, at least in part, mediated by a G-protein-coupled receptor and pertussis toxin-sensitive mechanism of signaling. This novel signaling mechanism appears to regulate potassium channel activities in kidney cells and therefore may contribute to the observed effects of guanylin and uroguanylin to markedly increase urinary potassium excretion. High-salt diets also enhance the expression of both guanylin and uroguanylin messenger RNA transcripts in the intestine and lead to increases in the urinary excretion of uroguanylin. Both findings are consistent with a novel physiological role for guanylin peptides as regulators of salt balance in the body.

UROGUANYLIN, GUANYLIN, AND COLON CANCER

Tumors begin in the colorectal epithelium as small benign adenomas (polyps), which slowly grow over the course of many years. Many adenomas arise in the colonic mucosa secondary to mutations in the

adenomatous polyposis coli (APC) gene. Adenoma cells can acquire additional gene mutations that accumulate over time and culminate in the formation of malignant adenocarcinomas that may spread to other organs. In both polyp and adenocarcinoma cells, the expression of guanylin and uroguanylin genes is turned off, resulting in a hormone-deficient state within the tumors. However, the R-GC form of receptor for these peptides continues to be expressed at normal levels. In a mouse model of human colon cancer, the oral administration of uroguanylin had a profound effect of reducing both the number and the size of tumors that appeared within the intestinal tract, thus markedly reducing tumor burden. Moreover, when human colon cancer cells were treated with uroguanylin or related peptides, the cells responded by undergoing programmed cell death (apoptosis). Other agents that elevate intracellular cyclic GMP levels also elicit apoptosis of colon cancer cells. Based on these new findings, it may be postulated that a primary physiological role of the guanylin peptides is to regulate epithelial cell turnover by influencing cellular lifetimes. In this manner, guanylin and uroguanylin can participate in the complex cellular mechanisms that regulate birth and death rates of the epithelial cells that populate the intestinal mucosa in order to maintain both the morphology and the function of the intestinal epithelium.

The antitumor actions of uroguanylin in both animal and cell models of colon cancer have stimulated additional investigations of this peptide hormone for potential use in the prevention and/or therapy of colon cancer. In addition, chemical analogues of uroguanylin-like peptides have been synthesized; the analogues contain a radioactive atom that can be detected at the surface of the body using the noninvasive imaging method of single photon-emitting computed tomography (SPECT). Preclinical studies have demonstrated that these uroguanylin-like agents can be used to locate colon adenocarcinomas that have spread to other organs in the body. This diagnostic approach is very promising for the early detection of metastatic forms of colon tumors because the tumor cells have very high levels of uroguanylin/guanylin receptors. The radioactive forms of these uroguanylin-like peptides bind to cell surface receptors, and the macromolecular complexes are then internalized and concentrated within the tumors. SPECT imaging can then be used to detect and locate one or more secondary colorectal tumors that have spread to other organs. It is clear that discovery of the guanylin family of cyclic GMP-regulating peptides has provided new insights into colon tumor biology per se. Moreover, the uroguanylin-like peptides offer new therapeutic approaches to prevent, detect, or treat colon cancer.

See Also the Following Articles

Colorectal Adenocarcinoma • Small Intestine, Absorption and Secretion

Further Reading

Forte, L. R. (1999). Guanylin regulatory peptides: Structures, biological activities mediated by cyclic GMP and pathobiology. *Regul. Pept.* **81**, 25–39.

Forte, L. R., and Currie, M. G. (1995). Guanylin: A peptide regulator of epithelial transport. *FASEB J.* **9**, 643–650.

Forte, L. R., Fan, X., and Hamra, F. K. (1996). Salt and water homeostasis: Uroguanylin is a circulating peptide hormone with natriuretic activity. *Am. J. Kidney Dis.* **28**, 296–304.

Forte, L. R., and Hamra, F. K. (1996). Guanylin and uroguanylin: Intestinal peptide hormones that regulate epithelial transport. *News Physiol. Sci.* **11**, 17–24.

Forte, L. R., London, R. M., Freeman, R. H., and Krause, W. J. (2000). Guanylin peptides: Renal actions mediated by cyclic GMP. *Am. J. Physiol* **278**, F180–F191.

Forte, L. R., London, R. M., Krause, W. J., and Freeman, R. H. (2000). Mechanisms of guanylin action via cyclic GMP in the kidney. *Annu. Rev. Physiol.* **62**, 673–695.

Giannella, R. A. (1995). *Escherichia coli* heat-stable enterotoxin, guanylins, and their receptors: What are they and what do they do? *J. Lab. Clin. Med.* **125**, 173–181.

Krause, W. J., London, R. M., Freeman, R. H., and Forte, L. R. (1997). The guanylin and uroguanylin peptide hormones and their receptors. *Acta Anat.* **160**, 213–231.

Lohmann, S. M., Vaandrager, A. B., Smolenski, A., Walter, U., and De Jonge, H. R. (1997). Distinct and specific functions of cGMP-dependent protein kinases. *Trends Biochem. Sci.* **22**, 307–312.

Lucas, K. A., Pitari, G. M., Kazerounian, S., Ruiz-Stewart, I., Park, J., Schulz, S., Chepenik, K. P., and Waldman, S. A. (2000). Guanylyl cyclases and signaling by cyclic GMP. *Pharmacol. Rev.* **52**, 375–414.

Nakazato, M. (2001). Guanylin family: New intestinal peptides regulating electrolyte and water homeostasis. *J. Gastroenterol.* **36**, 219–225.

Vaandrager, A. B., and De Jonge, H. R. (1994). Effect of cGMP on intestinal transport. *Adv. Pharmacol.* **26**, 253–283.

Vaandrager, A. B. (2002). Structure and function of the heat-stable enterotoxin receptor/guanylyl cyclase C. *Mol. Cell. Biochem.* **230**, 73–83.

Wedel, B. J., and Garbers, D. L. (2001). The guanylyl cyclase family at y2k. *Annu. Rev. Physiol.* **63**, 215–233.

argyrophil (enterochromaffin-like) cells in the rat stomach. *Z. Zellforschung Mikroskop. Anat.* **122**, 460.
Popielski, L. (1919). β-Imidazolyläthylamin und die Organextrakte. I. β-Imidazolyäthylamin als mächtiger erreger der Magendrüsen. *Pfluegers Arch. Gesamte Physiol. Menschen Tiere* **178**, 214–259.
Prinz, C., Kajimura, M., Scott, D. R., Mercier, F., Helander, H. F., and Sachs, G. (1993). Histamine secretion from rat enterochromaffin-like cells. *Gastroenterology* **105**, 449.
Sachs, G., Zeng, N., and Prinz, C. (1997). Physiology of isolated gastric endocrine cells. *Annu. Rev. Physiol.* **59**, 243–256.

H2-Receptor Antagonists

MITCHELL L. SCHUBERT
Virginia Commonwealth University's Medical College of Virginia and McGuire Veterans Affairs Medical Center, Richmond

erosion A shallow ulcer confined to the mucosal layer.
gastric H^+,K^+-ATPase An ATP-hydrolyzing enzyme responsible for catalyzing the exchange of lumenal K^+ for cytoplasmic H^+ by parietal cells, bringing about gastric lumenal acidification.
gastroesophageal reflux Backflow of acidic gastric contents into the esophagus, producing a spectrum of clinical symptoms and manifestations ranging from mild episodic heartburn and regurgitation without macroscopic esophagitis to chronic inflammation and ulceration; severe cases may involve stricture and hemorrhage.
gynecomastia Excessive development of the male mammary glands.
Helicobacter pylori A gram-negative urease-producing bacterium that causes an active chronic gastritis and is an important etiological factor in the development of gastric and duodenal ulcers.
nonsteroidal anti-inflammatory drugs Medications that reduce inflammation by inhibiting the enzymatic activity of cyclooxygenase, a key enzyme in the biosynthetic pathway leading to the formation of prostaglandins. Examples include aspirin, ibuprofen, naproxen, and diclofenac.
peptic ulcer A loss of tissue, extending through the mucosa into the submucosa, in the esophagus, stomach, or duodenum, due to acidic gastric secretions.
pharmacokinetics The disposition of drugs within the body in relation to their absorption, distribution, metabolism, and elimination.

Histamine receptors have been classified into four major subclasses: H1, H2, H3, and H4. H1 receptors are widely distributed and mediate the actions of histamine in the allergic response and as a bronchial constrictor. H2 receptors are present primarily in the stomach, where they regulate gastric acid secretion, and to a lesser extent in the heart, central nervous system, reproductive system, and on lymphocytes. Histamine-2 (H2)-receptor antagonists are drugs that competitively block the ability of histamine to interact with H2-receptors. H3-receptors were originally identified as inhibitory autoreceptors on histamine-containing nerve terminals in the central nervous system. In the stomach, H3-receptors augment acid secretion by suppressing somatostatin secretion. In the antrum, suppression of somatostatin leads to stimulation of gastrin, the main hormonal stimulant of acid secretion, whereas in the fundus, it leads to stimulation of histamine and acid secretion. Most recently, an H4-receptor has been identified in intestine, spleen, thymus, and immune cells. The H4-receptor shares 40% homology and overlapping pharmacology with the H3 receptor; its physiologic function is not known. This article focuses on the role of blocking H2 receptors in the stomach. The identification of the H2 receptor on the basolateral membrane of the parietal cell by Black in 1972 and the subsequent development of safe and effective H2-receptor antagonists have revolutionized the treatment of gastrointestinal acid-related disorders.

INTRODUCTION

Until the development of histamine-2 (H2)-receptor antagonists, the only treatments available for patients with acid-related disorders were bed rest, bland diets, milk drips, antacids, and surgery. Ulcers and ulcer pain

Copyright 2004, Elsevier (USA). All rights reserved.

recurred and complications such as bleeding, perforation, and obstruction were frequent. Surgery, mainly vagotomy (to denervate the acid-producing area of the stomach) with antrectomy (to remove the hormone gastrin), became the "gold standard" for the treatment of recurrent, recalcitrant, or complicated disease. Unfortunately, surgery proved not to be definitive (1–10% recurrence rate) and produced its own set of problems, including gastric stasis, bile reflux gastritis and esophagitis, and gastric remnant carcinoma.

The introduction of cimetidine, the first commercially available H2-receptor antagonist, and the subsequent introduction of ranitidine, famotidine, nizatidine, and roxatidine revolutionized the management of acid-related diseases and improved the quality of life for a large number of patients. For the first time, peptic ulcer disease and gastroesophageal reflux disease could be treated and prevented pharmacologically.

MECHANISM OF ACTION

The main secretagogues active at the level of the parietal cell are acetylcholine, histamine, and gastrin (Fig. 1). Acetylcholine, released from intramural postganglionic nerve terminals, binds to parietal cell muscarinic subtype 3 (M3) receptors that are coupled to an increase in intracellular calcium. Histamine, released from enterochromaffin-like (ECL) cells, which reside in the basal half of the oxyntic gland in the vicinity of parietal cells, binds to parietal cell H2 receptors that are coupled to an increase in intracellular concentrations of adenosine 3′,5′-cyclic monophosphate (cAMP). Histamine-containing ECL cells are the predominant endocrine/paracrine cell type in the oxyntic mucosa and constitute 65–75% of the endocrine cells in rat and 30–35% of the endocrine cells in humans. Gastrin, released from antral G cells in the distal stomach, travels through the bloodstream to activate cholecystokinin subtype 2 (CCK_2; previously termed CCK-B) receptors present on both parietal and ECL cells that are coupled to an increase in intracellular calcium. The increase in parietal cell levels of cAMP and calcium synergistically activates the parietal cell proton pump, an H^+,K^+-ATPase, that catalyzes the exchange of lumenal K^+ for cytoplasmic H^+ and is responsible for gastric lumenal acidification (Fig. 1).

Gastrin, the main stimulant for acid secretion during ingestion of a meal, activates the parietal cell directly and, more importantly, indirectly by increasing the production and release of histamine from ECL cells (Fig. 1). In H2-receptor-deficient mice, gastrin is an ineffective secretagogue, confirming the notion that, although CCK_2 receptors are present on the parietal cell, gastrin stimulates acid secretion primarily by binding to CCK_2 receptors on ECL cells that are coupled to histamine release.

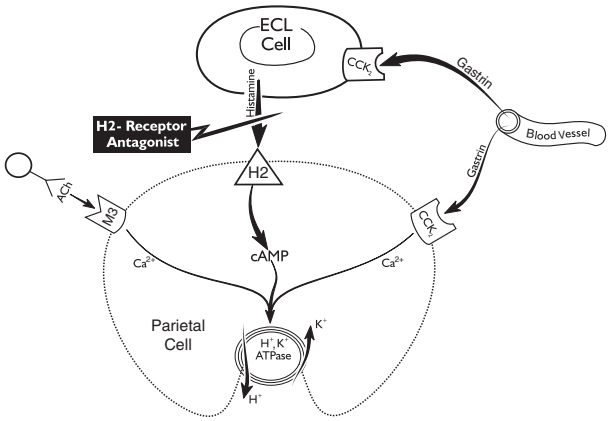


FIGURE 1 Model illustrating the action of the main secretagogues at the level of the parietal cell, the transduction pathways to which they are coupled, and the site of action of H2-receptor antagonists. Acetylcholine (ACh), released from intramural neurons, binds to M3 receptors that are coupled to an increase in intracellular calcium (Ca^{2+}). Histamine, released from enterochromaffin-like (ECL) cells, binds to H2 receptors that are coupled to an increase in cellular adenosine 3′,5′-cyclic monophosphate (cAMP). Gastrin, released from G cells in the antrum into the bloodstream, binds to CCK_2 receptors on the parietal cell that are coupled to an increase in Ca^{2+} and more importantly binds to CCK_2 receptors on ECL cells that are coupled to the release of histamine. The two transduction pathways in the parietal cell, i.e., cAMP and Ca^{2+}, converge on and activate the H^+,K^+-ATPase, the hydrogen pump of the parietal cell. H2-receptor antagonists competitively block the H2 receptor on the parietal cell, thus blocking both histamine-driven acid secretion and that elicited by gastrin, whose action is mediated primarily by the release of histamine from ECL cells.

The structure of H2-receptor antagonists is patterned after that of histamine (Fig. 2). The initial compounds (e.g., cimetidine) share histamine's imidazole ring. Subsequent compounds (ranitidine, famotidine, and nizatidine) use other five-membered rings (i.e., furan and thiazole rings). Roxatidine, which is available in Europe, differs from previous compounds in that the five-membered ring has been replaced with a phenyl ring. Unlike the other H2-receptor antagonists, roxatidine is a prodrug that is rapidly metabolized to yield the active compound.

H2-receptor antagonists competitively and selectively inhibit the binding of histamine to H2 receptors on the basolateral membrane of the parietal cell, thereby reducing intracellular concentrations of the signaling molecule cAMP and, as a result, the secretion of acid by the parietal cell. As discussed above, the H2-receptor

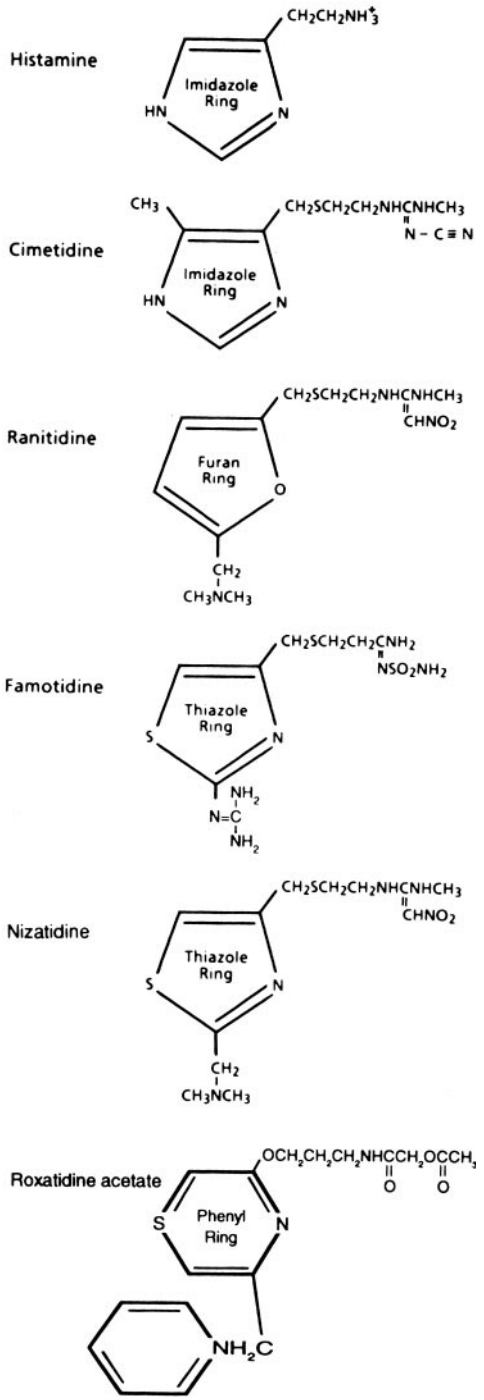


FIGURE 2 Chemical structure of histamine and H2-receptor antagonists.

antagonists are effective at blocking both histamine-driven acid secretion and that elicited by gastrin, whose action is mediated primarily by the release of histamine from ECL cells. The relative potencies of the H2-receptor antagonists clinically available are as follows: famotidine > ranitidine = nizatidine = roxatidine > cimetidine (Table I).

PHARMACOKINETICS

The absorption of H2-receptor antagonists from the gastrointestinal tract varies considerably; their mean bioavailability after oral administration is as follows (Table I): famotidine 45%, ranitidine 50%, cimetidine 60%, nizatidine 95%, and roxatidine 95%. Peak plasma concentrations occur 1–3.5 h after oral administration, although wide interindividual variations are observed. Concomitant administration of antacids or sucralfate may decrease the absorption of H2-receptor antagonists by up to 30%.

The H2-receptor antagonists are widely distributed throughout the body, with a mean volume of distribution of 1.5 liter/kg (Table I). Protein binding is low, in the range of 8% for roxatidine to 30% for nizatidine. Although all the H2-receptor antagonists cross the blood–brain barrier, cimetidine achieves the highest cerebrospinal fluid levels. All cross the placenta and are secreted into breast milk.

After oral administration, cimetidine, ranitidine, and famotidine are extensively metabolized by the liver. By contrast, 65% of nizatidine or roxatidine is excreted unchanged in the urine. Nizatidine produces an active metabolite, *N*-2-monodesmethylnizatidine, which exhibits 60% of the H2-receptor-blocking activities of the parent compound. After intravenous administration, 65–85% of each H2-receptor antagonist is excreted unchanged in the urine.

The elimination half-life varies considerably, averaging 1.5 h for cimetidine, ranitidine, and nizatidine; 4 h for famotidine; and up to 6 h for roxatidine (Table I). Because the half-life of H2-receptor antagonists is increased up to 10-fold in patients with renal insufficiency, it has been suggested that one-half the therapeutic dose be given to patients with creatinine clearances below 30 ml/min. No dosing adjustment is required for patients with hepatic dysfunction.

EFFICACY

Although a number of factors may contribute to the pathogenesis of peptic ulcer disease and reflux esophagitis, gastric acid secretion plays a pivotal role and reduction in acid secretion is the cornerstone of treatment for these diseases. Maximal healing rates for peptic ulcer are achieved when intralumenal pH is maintained at or above 3 for 16 h out of a 24 h period. In contrast, optimal healing of reflux esophagitis requires greater acid inhibition, a target pH at or above 4 for 16 h out

TABLE I Comparison of H_2-Receptor Antagonists

Characteristic	Cimetidine	Ranitidine	Famotidine	Nizatidine	Roxatidine
Absorption					
Bioavailability (%)	30–80	39–88	40–50	75–100	95
Time to peak concentration (h)	1–2	1–3	1–3.5	1–3	1–3
Distribution					
Volume (liter/kg)	0.8–2.1	1.0–1.9	1.1–1.3	1.2–1.6	0.9–1.9
Protein binding (%)	20	15	16	30	8
Elimination					
Serum half-life (h)	1.5–2.5	1.6–3.1	2.5–4.0	1.1–2.0	2.0–6.0
Relative potency	1	4–10	20–50	4–10	4–6

of a 24 h period. Treatment for as little as 2–7 days with H2-receptor antagonists may induce tachyphylaxis, i.e., progressive loss of efficacy, with time. The precise mechanism is not known.

Duodenal and Gastric Ulcers

Approximately 60–80% of duodenal and gastric ulcers can be healed by 4-week treatment and 80–90% by 6- to 8-week treatment with therapeutic doses of any of the H2-receptor antagonists. Ulcers can be successfully treated with either a twice-a-day dose or a single bedtime dose (cimetidine, 800 mg; ranitidine, 300 mg; famotidine, 20 mg; nizatidine, 300 mg; or roxatidine, 150 mg). Once healed, however, cure is not permanent and 80% of patients will have a relapse of their ulcer within 1 year. Maintenance therapy with one-half the standard therapeutic dose administered at bedtime (i.e., cimetidine, 400 mg; ranitidine, 150 mg; famotidine, 10 mg; nizatidine 150 mg; or roxatidine, 75 mg) can decrease the 1-year relapse rate to 10–20%.

In the vast majority of cases, duodenal and gastric ulcers are associated with either *Helicobacter pylori* infection of the stomach or the use of nonsteroidal anti-inflammatory drugs (NSAIDs); Zollinger-Ellison syndrome caused by a gastrin-secreting tumor (gastrinoma) is a rare cause of peptic ulcer disease. In patients infected with *H. pylori*, additional therapy aimed at eradicating the bacteria may accelerate the rate of ulcer healing over that achieved using an H2-receptor antagonist alone and prevents ulcer relapse. Ranitidine bismuth citrate, 400 mg twice a day for 4 weeks, in conjunction with clarithromycin, is an approved regimen for eradication of the bacteria.

NSAID-induced ulcers can be effectively treated with H2-receptor antagonists, especially if the NSAID can be discontinued. If the NSAID must be continued, H2-receptor antagonists achieve only a 50% healing rate at 8 weeks. For the prevention of NSAID-induced ulcers in high-risk patients, concomitant administration of H2-receptor antagonists is effective in preventing duodenal but not gastric ulcers.

Zollinger-Ellison syndrome is characterized by marked gastric acid hypersecretion. H2-receptor antagonists have been used successfully to control acid secretion but high and frequent dosing is required (e.g., ranitidine: 2700 mg per day in divided doses every 4–6 h) and this condition is now most often treated with proton pump inhibitors.

Reflux Esophagitis

Reflux esophagitis manifests as a spectrum of endoscopic and pathologic stages ranging in severity from reflux symptoms without evidence of esophageal damage to erosive and ulcerative esophagitis. Over-the-counter formulations of H2-receptor antagonists provide at best partial relief, with complete relief occurring in only 15% of patients. In patients with erosive or ulcerative esophagitis, standard doses of H2-receptor antagonists provide complete symptom relief and mucosal healing in only approximately 50–75% of patients at 8 weeks. Because healing of erosive esophagitis requires greater acid suppression than that required to heal duodenal or gastric ulcers, ranitidine 150 mg qid or comparable doses of another H2-receptor antagonist are recommended. After complete healing, symptomatic recurrences occur in 80% of patients. In contrast to duodenal and gastric ulcer, maintenance therapy to prevent recurrence requires the use of the same dose required to achieve healing.

Prophylaxis of Stress-Related Gastric Hemorrhage

Patients admitted to intensive care units, particularly those with respiratory failure, coagulopathy, extensive burns, head injuries, or extensive trauma, are at increased risk for stress-related gastric mucosal damage

and subsequent upper gastrointestinal bleeding. Despite controversy regarding the precise role of acid in the development of stress-related gastric damage, data show a strong relationship between mucosal damage, bleeding, and low pH. H2-receptor antagonists, administered continuously intravenously, are effective in preventing stress ulcer and stress ulcer bleeding.

Adjuvant Treatment of Cancer

H2-receptor antagonists, in particular, cimetidine, have been shown to improve the survival of patients with colorectal cancer, melanoma, and renal cell cancer. It has been proposed that cimetidine enhances the host immune response by increasing the cytotoxic activity of mononuclear cells and tumor-infiltrating lymphocytes and prevents cancer metastasis by inhibiting cancer cell adhesion to endothelial cells via down-regulation of the cell surface expression of E-selectin, a ligand on endothelial cells that binds tumor antigens.

ADVERSE EFFECTS

The H2-receptor antagonists are generally well tolerated with adverse effects observed in 1.5% of treated patients compared with 1.2% of placebo patients.

Central Nervous System

Cimetidine and ranitidine, but perhaps not nizatidine, cross the blood–brain barrier and may interact with H2 receptors present in the brain. A variety of central nervous system (CNS) side effects, ranging from headache, dizziness, anxiety, somnolence, and depression, to confusion and delirium, have been reported. It should be noted, however, that the overall frequency of CNS effects is very low, less than 0.2%. Advanced age and renal dysfunction may predispose to CNS side effects.

Endocrine

Cimetidine possesses weak anti-androgenic activity and may rarely cause gynecomastia, hyperprolactinemia, and impotence. Gynecomastia has been observed in up to 4% of patients with Zollinger-Ellison syndrome who were treated for prolonged periods with high doses of cimetidine. Prolactin levels may rise acutely but usually return to normal with continued use. Sperm counts may be depressed, but usually remain within normal limits. Although ranitidine, famotidine, and nizatidine do not possess anti-androgenic activity, nevertheless they may rarely cause gynecomastia.

Renal

Cimetidine has rarely (1 in 100,000 patients exposed) been associated with an idiosyncratic hypersensitivity interstitial nephritis, which may manifest as fever, rash, arthralgia, polymyositis, eosinophilia, and sterile pyuria. Renal function usually returns to normal once cimetidine is stopped. Cimetidine, but none of the other H2-receptor antagonists, has been reported to increase serum creatinine concentrations as a result of competition for renal tubular excretion between cimetidine and creatinine.

Hematologic

Thrombocytopenia, anemia, and leukopenia have been reported only rarely. In most circumstances, patients had serious underlying disease and/or were taking multiple drugs, making it difficult to establish a direct cause and effect. H2 receptors are present on the cell surface of suppressor T lymphocytes and H2-receptor antagonists have been reported to have mild immune-modulating effects; they have induced tumor regression in several studies and may cause false-negative immediate skin tests.

Pregnancy and Breast-Feeding

The H2-receptor antagonists cross the placenta and are also distributed in breast milk. Although data in humans are limited, these drugs have been assigned to Food and Drug Administration Pregnancy Category B and should be used in pregnancy only if clearly needed. Because of the potential for adverse reactions in nursing infants, caution should be exercised in administering these drugs to nursing mothers.

Other Effects

H2 receptors have been identified in the human atria and rare cases of dysrhythmia and atrioventricular conduction defects have been reported. Asymptomatic elevations in hepatic transaminases may occur. Studies suggest that cimetidine delays gastric emptying, whereas ranitidine and nizatidine accelerate it.

DRUG INTERACTIONS

Interference with Drug Absorption

H2-receptor antagonists may interfere with the absorption of other drugs or themselves by altering gastric emptying, binding directly to another drug, or altering intragastric pH. Cimetidine has been reported to delay gastric emptying, whereas ranitidine and

nizatidine accelerate it. Antacids and sucralfate bind to and inhibit the absorption of H2-receptor antagonists by 10–25%. Certain drugs that are weak bases, such as the antifungal medications ketoconazole and itraconazole, require an acid environment for optimal absorption. As the pH rises in response to H2-receptor antagonists, the absorption of these drugs decreases.

Cytochrome P450 Interactions

The liver is a major site for drug biotransformation. Many drugs are lipophilic and require conversion in the liver to more polar, aqueous-soluble compounds for elimination in bile or urine. Cytochrome P450, or the mixed-function oxidative microsomal enzyme system, refers to a group of heme-containing hepatic enzymes that catalyze the oxidative biotransformation of lipophilic substrates to more polar hydrophilic compounds. Cytochrome P450 is defined by its maximal absorption of light at 450 nm. Multiple different cytochrome P450 enzymes are expressed in a given individual and differences among individuals in their capacity to metabolize drugs may reflect genetically determined differences in the expression of these enzymes.

Imidazole-containing drugs, such as cimetidine, are capable of inhibiting the catalytic activity of one or more cytochrome P450 enzymes, in particular, CYP1A2 and CYP2C19. Ranitidine, which lacks the imidazole ring, interacts 5- to 10-fold less. Less information is available regarding the other H2-receptor antagonists, but they do not appear to have any significant interaction. Although cimetidine has the potential to interfere with the hepatic metabolism of many therapeutic drugs, only a few are of potential clinical significance because of their narrow therapeutic ranges: theophylline, warfarin, quinidine, phenytoin, cyclosporin, and tricyclic antidepressants. The clinical importance of these interactions depends on (1) how often cimetidine is taken, (2) the initial serum concentration of the therapeutic drug, (3) the individual susceptibility of the patient to the interaction, and (4) the individual sensitivity of the patient to the therapeutic drug.

Warfarin exists as a mixture of *S* and *R* isomers; (*S*)-warfarin is five times more potent as a vitamin K antagonist than (*R*)-warfarin. Cimetidine inhibits the metabolic clearance of the *R* isomer only; thus, it causes only a slight prolongation of the prothrombin time.

Many cardiac anti-arrhythmics, such as quinidine, lidocaine, and flecainide, are metabolized by cytochrome P450 and elevated levels may occur during co-administration with cimetidine. The elimination half-life of beta-blockers, such as propranolol and metoprolol, but not atenolol and nadolol, which are primarily cleared through the kidney, may be prolonged by cimetidine. Cimetidine may augment the blood pressure-lowering effect of nifedipine and the delay in atrioventricular conduction induced by verapamil.

Despite the potential for adverse interactions between cimetidine and various drugs metabolized by cytochrome P450, the reality is that serious complications only very rarely occur. Most problems can be avoided by monitoring the plasma levels of a drug with a narrow therapeutic window or by administering the H2-receptor antagonist in a single nighttime dose.

Interference with Drug Elimination

Cimetidine can interfere with the renal clearance of creatinine as well as cationic drugs that undergo active renal proximal tubular secretion. By competing for the transport mechanism, cimetidine may impair the elimination of procainamide and cyclosporin.

Alcohol Dehydrogenase Interaction

A portion of ingested alcohol is metabolized in the stomach by the enzyme alcohol dehydrogenase. Cimetidine, ranitidine, and nizatidine have been reported to inhibit gastric alcohol dehydrogenase. In addition, ranitidine and nizatidine have been reported to accelerate gastric emptying, thus shortening the exposure of alcohol to gastric alcohol dehydrogenase. Since the initial reports that cimetidine, ranitidine, and nizatidine, but not famotidine, increase blood alcohol levels, the validity of these findings and the clinical implications have been the subject of considerable controversy.

Acknowledgments

Thanks are extended to Mary Beatty for the artwork. This work was supported by the Veterans Administration Medical Research Fund.

See Also the Following Articles

Cholecystokinin (CCK) • Cytochrome P450 • Duodenal Ulcer • Enterochromaffin-like (ECL) Cells • Gastric Acid Secretion • Gastric H,K-ATPase • Gastric Ulcer • Gastroesophageal Reflux Disease (GERD) • *Helicobacter pyroli* • Histamine • NSAID-Induced Injury • Over-the-Counter Drugs • Pharmacology, Overview • Proton Pump Inhibitors

Further Reading

Farley, A., Wruble, L. D., and Humphries, T. J. (2000). Rabeprazole versus ranitidine for the treatment of erosive gastroesophageal reflux disease: A double-blind, randomized clinical trial. *Am. J. Gastroenterol.* **95**, 1894–1899.

Feldman, M., and Burton, M. E. (1990). Histamine2-receptor antagonists: Standard therapy for acid-peptic diseases. [first of two parts] *N. Engl. J. Med.* **323**, 1672–1680.

Feldman, M., and Burton, M. E. (1990). Histamine2-receptor antagonists: Standard therapy for acid-peptic diseases [second of two parts] *N. Engl. J. Med.* **323**, 1749–1755.

Feston, H. P. M., Schenk, E., Tan, G., Snel, P., Nelis, F., and The Dutch Reflux Study Group (1999). Omeprazole versus high-dose ranitidine in mild gastroesophageal reflux disease: Short- and long-term treatment. *Am. J. Gastroenterol.* **94**, 931–936.

Jensen, R. T. (1998). Management of the Zollinger-Ellison syndrome in patients with multiple endocrine neoplasia type 1. *J. Intern. Med.* **243**, 477–488.

Kobayashi, K., Matsumoto, S., Morishima, T., Kawabe, T., and Okamoto, T. (2000). Cimetidine inhibits cancer cell adhesion to endothelial cells and prevents metastasis by blocking E-selectin expression. *Cancer Res.* **60**, 3978–3984.

Kobayashi, T., Tonai, S., Ishihara, Y., Koga, R., Okabe, S., and Watanabe, T. (2000). Abnormal functional and morphological regulation of the gastric mucosa in histamine H2 receptor-deficient mice. *J. Clin. Invest.* **105**, 1741–1749.

Levitt, M. D. (1993). Review article: Lack of clinical significance of the interaction between H2-receptor antagonists and ethanol. *Aliment. Pharmacol. Ther.* **7**, 131–138.

Messori, A., Trippoli, S., Vaiani, M., Gorini, M., and Corrado, A. (2000). Bleeding and pneumonia in intensive care patients given ranitidine and sucralfate for prevention of stress ulcer: Meta-analysis of randomised controlled trials. *Br. Med. J.* **321**, 1103–1106.

Niopas, I., Toon, S., Aarons, L., and Rowland, M. (2000). The effect of cimetidine on the steady-state pharmacokinetics and pharmacodynamics of warfarin in humans. *Eur. J. Clin. Pharmacol.* **55**, 399–404.

Raychowdhury, R., Zhang, Z., Hockers, M., and Wang, T. C. (1999). Activation of human histidine decarboxylase gene promoter activity by gastrin is mediated by two distinct nuclear factors. *J. Biol. Chem.* **274**, 20961–20969.

Sachs, G., Zeng, N., and Prinz, C. (1997). Physiology of isolated gastric endocrine cells. *Annu. Rev. Physiol.* **59**, 243–256.

Shamburek, R. D., and Schubert, M. L. (1993). Pharmacology of gastric acid inhibition. *Balliere Clin. Gastroenterology* **7**, 23–54.

Vuyyuru, L., Harrington, L., Arimura, A., and Schubert, M. L. (1997). Reciprocal inhibitory paracrine pathways link histamine and somatostatin secretion in the fundus of the stomach. *Am. J. Physiol.* **273**, G106–G111.

Vuyyuru, L., Schubert, M. L., Harrington, L., Arimura, A., and Makhlouf, G. M. (1995). Dual inhibitory pathways link antral somatostatin and histamine secretion in human, dog, and rat stomach. *Gastroenterology* **109**, 1566–1574.

Yeomans, N. D., Tulassay, Z., Juhasz, L., Racz, I., Howard, J. M., van Rensburg, C. J., Swannell, A. J., and Haekey, C. J. (1998). A comparison of omeprazole with ranitidine for ulcers associated with nonsteroidal anti-inflammatory drugs. *N. Engl. J. Med.* **338**, 719–726.

Halitosis

KEN YAEGAKI
University of British Columbia, Vancouver

halitosis Abnormally foul or fetid breath/oral malodor or simply bad breath.
organoleptic measurement and score Scoring odor intensity by smelling a patient's breath. Observation conditions must ensure that breath air is not diluted with room air.
volatile sulfide compounds Compounds such as hydrogen sulfide, methyl mercaptan, and dimethyl sulfide that are the main cause of halitosis and that are produced from sulfur-containing amino acids by bacterial putrefaction.

Most cases (approximately 90%) of halitosis originate within the oral cavity. Oral bacteria degrade sulfur-containing amino acids to volatile sulfide compounds (VSCs). Numerous compounds exist in human breath, but only VSCs correlate well with organoleptic score. Approximately 10% of halitosis cases are related to conditions outside the mouth. In general, people are not able to assess their own malodor accurately and 30–90% of the patients visiting oral malodor clinics do not in fact exhibit bad breath.

INTRODUCTION

Halitosis can be classified into three main categories (Table I): genuine halitosis, pseudo-halitosis, and halitophobia. Genuine halitosis is subclassified as physiologic halitosis and pathologic halitosis. Physiologic halitosis is caused by physiologic factors, such as tongue coating. Pathologic halitosis is further subdivided into oral pathologic halitosis and extraoral pathologic halitosis. The principal causes of oral pathologic halitosis are periodontal conditions and progressive dental caries. Pseudo-halitosis and halitophobia are psychological phenomena in which patients complain of the presence of halitosis, but others do not perceive their breath to be offensive. Improvement in the condition of pseudo-halitosis can be obtained by explanation of examination results and simple treatment measures. In contrast, halitophobic patients cannot believe, despite evidence to the contrary, that they do not have halitosis. Halitophobic patients are thus considered to suffer from a social phobia.

TABLE I Classification of Halitosis

Genuine halitosis
Physiologic halitosis
Pathologic halitosis
Oral pathologic halitosis
Extraoral pathologic halitosis
Pseudo-halitosis
Halitophobia

EXTRAORAL PATHOLOGIC HALITOSIS

Gastrointestinal diseases, especially stomach conditions, are popularly believed to cause halitosis. Since the esophagus is collapsed, this does not happen unless function of the esophagus is abnormal. It has been demonstrated that gastric *Helicobacter pylori* infection may cause halitosis. Pyloric stenosis and extrinsic duodenal obstruction can also cause halitosis. Since malodorous compounds, such as volatile sulfide compounds (VSCs), are absorbed from the intestines into the bloodstream, the compounds diffuse into lung air. However, as thiols, such as H_2S and CH_3SH, immediately react irreversibly with blood components, only dimethyl sulfide is transported from the blood into the air in the lung.. Therefore, although hydrogen sulfide, methyl mercaptan, and ethanethiol are found in the breath of patients with liver conditions, *fetor hepaticus* is probably principally caused by dimethyl sulfide. Furthermore, C-2–C-4 aliphatic acids and lomonene have also been suggested as components of *fetor hepaticus*. Some metabolic disorders can result in halitosis. Trimethylaminuria causes a fishy odor in urine, sweat, and breath because an oxidative enzyme that converts trimethylamine to trimethylamineoxide is lacking. In patients with high blood cholesterol, isoprene with specific malodor can be detected in the breath utilizing a simple laboratory device. Diabetes elevates the acetone level in breath and is generally believed to be the most plausible reason for halitosis. However, acetone itself does not produce a fetid malodor; the foul breath of diabetics may be linked to their periodontal condition or other infections. Diabetes is an indirect rather than a direct reason for halitosis. Nasal, sinus, or pharyngeal conditions have also been assumed to be

 Copyright 2004, Elsevier (USA). All rights reserved.

frequent reasons for halitosis, but in fact this is seldom the case. Conditions of the bronchi and the lung, however, can cause halitosis. Aniline, *o*-toluidine, *n*-propanol, and methylethylketone have been found in the breath of patients with lung carcinoma. Some medications cause halitosis. Sulfur-containing drugs, such as cysteine, cysteamine, or disulfiram, result in foul breath. Psychopharmacologic, antirheumatic, or antihypertensive drugs also causes halitosis. Anticholinergics, antidepressants, analgesics, and numerous other drugs that suppress salivation may cause halitosis as they suppress the natural cleansing action of saliva in the mouth.

PHYSIOLOGIC HALITOSIS

The majority of halitosis cases are physiologic and evidence-based treatments of halitosis are necessary. To carry out a rational treatment, the halitosis must be classified and appropriate treatment needs (TN) determined. The TN for physiologic halitosis comprise instructions by the clinicians on oral hygiene and tooth brushing. The most important oral hygiene procedure is tongue cleaning. Physiological halitosis mainly originates from the dorso-posterior region of the tongue when covered with a coating containing desquamated epithelial cells, blood cells, and bacteria. Approximately 60% of total VSCs are produced from such tongue coatings. Tongue brushing has been found to be more effective and safer in reducing oral malodor than tongue scraping. A tongue brush with very soft and tiny bristles or a toothbrush for infants is recommended to clean the tongue. The clinician should instruct the patient that brushing must always be performed from the terminal sulcus to the front of the tongue to avoid brushing the tongue tonsil.

DISCRIMINATIVE DIAGNOSIS OF EXTRAORAL PATHOLOGIC HALITOSIS FROM OTHER HALITOSIS

To perform organoleptic measurements, a translucent tube (1 in. diameter, 4 in. length) is inserted through a privacy screen (a translucent tube is inserted at the center). The patient is asked to close his or her mouth and breathe normally through the nose for 1 min and then exhale his or her breath into the tube. For the first 1–2 s, the judge smells the patient's breath for oral malodor and gives a score (Table II). Then, the patient is instructed to brush the tongue and to rinse the mouth with 0.75–1.50% H_2O_2 or 0.12–0.20% chlorhexidine mouthwash. After these procedures that markedly reduce oral malodor are performed, organoleptic measurement is carried out again for odors originating in the lung.

TABLE II Organoleptic Score Scale

Category	Description
0: Absence of odor	Odor cannot be detected
1: Questionable odor	Odor is detectable, although the examiner could not recognize it as malodor
2: Slight malodor	Odor is deemed to exceed the threshold of malodor recognition
3: Moderate malodor	Malodor is definitely detected
4: Strong malodor	Strong malodor is detected, but can be tolerated by examiner
5: Severe malodor	Overwhelming malodor is detected and cannot be tolerated by examiner (examiner instinctively averts the nose)

See Also the Following Articles

Diabetes Mellitus • Duodenal Obstruction • *Helicobacter pylori* • Pyloric Stenosis

Further Reading

Hoshi, K., Yamano, Y., Miysunaga, A., Shimizu, S., Kagawa, J., and Ogiuchi, H. (2002). Gastrointestinal conditions and halitosis: Association of *Helicobacter pylori* infection. *Int. Dent. J.* **52**, 207–211.

Murata, T., Yamaga, T., Lida, H., Miyazaki, H., and Yaegaki, K. (2002). Classification and examination of halitosis. *Int. Dent. J.* **52**, 181–186.

Tangerman, A. (2002). Halitosis in medicine: A review. *Int. Dent. J.* **52**, 201–206.

Yaegaki, K., and Coil, J. M. (1999). Clinical dilemmas posed by psychosomatic halitosis patients. *Quintessence Int.* **30**, 302–306.

Yaegaki, K., and Coil, J. M. (2000). Genuine halitosis, pseudohalitosis and halitophobia: Classification, diagnosis and treatment. *Compendium* **21**, 880–890.

Yaegaki, K., Coil, J. M., Kamemizu, T., and Miyazaki, H. (2002). Tongue brushing and mouth rinsing as basic treatment measure for halitosis. *Int. Dent. J.* **52**, 192–196.

Hamartomatous Polyposis Syndromes

SHERRY C. HUANG AND JOHN M. CARETHERS
University of California, San Diego

bone morphogenetic protein receptor type 1A Mutated receptor in some patients with juvenile polyposis syndrome.

hamartoma Mature but disorganized normal tissue indigenous to the site of origin.

phosphatase and tensin homologue Tumor suppressor gene on chromosome 10; mutated in patients with Cowden disease and in some patients with juvenile polyposis syndrome and Bannayan–Riley–Ruvalcaba syndrome.

serine/threonine kinase 11 Gene that is mutated in patients with Peutz–Jeghers syndrome.

Hamartomatous polyposis syndromes are a group of clinically distinct disorders in which the predominant feature is multiple hamartomatous polyps in the gastrointestinal tract. These syndromes are transmitted in an autosomal dominant fashion to offspring, but sporadic forms do exist. Despite the nondysplastic histologic characteristics of the polyps, each hamartomatous polyposis syndrome carries specific cancer risks at different organ sites. Several mutated genes have recently been identified in the different polyposis syndromes. These genes may provide clues to the pathogenesis of the polyps and may play an important role in classifying the hamartomatous polyposis syndromes prior to cancer formation. Pathways involved in the hamartomatous syndromes include those of vascular endothelial growth factor, the transforming growth factor-β superfamily, and antagonizing the effects of Akt/protein kinase B.

INTRODUCTION

The autosomal dominant inherited hamartomatous polyposis syndromes consist of juvenile polyposis syndrome (JPS), Bannayan–Riley–Ruvalcaba syndrome (BRRS), Cowden syndrome (CS), Peutz–Jeghers syndrome (PJS), and hereditary mixed polyposis syndrome. Recently, these syndromes have been associated with germ-line mutations in certain tumor suppressor genes, but there remains some overlap in manifestations of a mutated phenotype. Because organ-specific cancer risks in each syndrome are different (see Table I), it is important to arrive at a molecular basis for classifying these syndromes. For instance, patients with JPS are predisposed to colorectal cancers with a 4- to 12-fold increase in incidence. Although in BRRS the risk for colorectal carcinoma is not known, carcinoma of the breast, endometrium, and thyroid has been well described in CS. As such, medical and surgical treatments of these syndromes differ based on the different disease phenotypes and subsequent malignant potential. Genotyping these syndromes has been important in classifying these diseases prior to cancer formation.

PEUTZ–JEGHERS SYNDROME

Patients with Peutz–Jeghers syndrome have associated characteristic mucocutaneous pigmentary spots along with intestinal hamartomatous polyps. Classically, pigmentary macules (1–5 mm) are often found in the perioral region and cross the vermilion border. The PJS polyps are distinct from those seen in the other hamartomatous polyposis syndromes in that the polyps demonstrate an arborizing pattern of growth of the muscularis mucosae, with extension into branching fronds of the polyp. These benign glands within the polyp are often surrounded by smooth muscle, which extends into the submucosa or muscularis propria (pseudoinvasion).

The incidence of PJS has been described as 1 in 120,000 births and occurs in all races and skin types. On biopsy, the mucocutaneous pigmentary macules are characterized by increased melanocytes at the epidermal–dermal junction. No malignant transformation has been ascribed to the hyperpigmentation associated with PJS. Young patients with PJS often present with symptoms of the small intestine (intussusception and abdominal pain), where the polyps occur most frequently. As the PJS patient ages, the morbidity shifts to malignancy. Patients with PJS have a 15-fold elevated relative risk for developing cancer, with 93% of patients developing cancer by age 64. The major sites of cancer are small intestine, stomach, pancreas, colon, esophagus, ovary, lung, uterus, and breast.

The genetic cause of PJS has been well described by many groups and appears to be a germ-line mutation in

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Cancer Risks and Screening in Hamartomatous Polyposis Syndromes

Syndrome	Cancer	Risk	Screening recommendations
Juvenile polyposis syndrome	Colon	Up to 68%	Initial colonoscopy in early teens (12–15 years) in asymptomatic patients; interval determined by symptoms, but every 3 years thereafter
	Gastric, duodenal	Rare	Upper GI endoscopy every 3 years to manage symptoms of polyps
Cowden syndrome	Colon	Minimal	As needed to manage symptoms of polyposis
	Thyroid	3–10%	Annual thyroid exam, start in teens
	Breast	25–50%	Breast exam, start at 25 years; annual mammograms, start at age 30
Peutz–Jeghers syndrome	Colon	35–40%	Initial colonoscopy in early teens in asymptomatic patients; interval determined by symptoms, but every 3 years thereafter
	Pancreas	35%	Endoscopic or abdominal ultrasound every year, start at age 30
	Stomach	25–30%	Upper GI endoscopy every 2 years, start at age 10
	Small bowel	10–15%	Annual hemoglobin; small bowel barium study every 2 years, start at age 10
	Esophagus	Rare	None
	Breast	50–60%	Annual breast exam; mammography every 2–3 years, start at age 25
	Ovarian	20%	Annual pap smear, annual pelvic ultrasound, start at age 20
	Uterine/cervical	8–10%	Annual pelvic exam, annual pelvic or vaginal ultrasound, start at age 20
	Testes	10%	Annual testicular exam, start at age 10; testicular ultrasound if feminizing features occur

the *STK11/LKB1* gene, located on chromosome 19p13.3. The protein product of *STK11/LKB1* is a tumor suppressor found in a variety of adult and fetal tissues. Genetic analysis of PJS polyps demonstrates that both alleles of *STK11/LKB1* are inactivated in most polyps. In adenomatous tissues or adenocarcinomas from patients with PJS, additional genetic lesions appear, suggesting the *STK11/LKB* may play a "gatekeeper" role for the development of malignancy in patients with PJS. Diagnosis is based on the finding of typical melanin pigment spots and characteristic gastrointestinal (GI) polyps. Management involves screening to prevent benign and malignant complications.

COWDEN SYNDROME

Cowden syndrome occurs in 1 of 200,000 individuals and is autosomal dominantly inherited. It is characterized by multiple hamartomatous polyps in the colon and throughout the GI tract. Juvenile-like polyps that contain neural elements are the most common, but a number of other polyps may occur, including lipomas, inflammatory polyps, ganglioneuromas, and lymphoid hyperplasia. The hallmark of CS is facial trichilemmomas, which most commonly occur around the mouth, nose, and eyes.

Two-thirds of patients with CS have thyroid adenomas and goiters, and as many as three-fourths will have fibrocystic breast disease and fibroadenomas. Excess risk of GI cancers has not yet been described, although the two most common cancers that develop in patients with CS are thyroid cancer (10%) and breast cancers (as many as 50% of patients). The genetic cause of CS has been well described. It arises from mutations of the *phosphatase and tensin homologue* (*PTEN*) gene on chromosome 10q23, which encodes a 403-amino-acid dually specific phosphatase. Other benign soft tissue and visceral tumors have been described, including hemangiomas, lipomas, lymphangiomas, neurofibromas, uterine leiomyomas, and meningiomas. Up to 85% of patients who meet the diagnostic criteria of CS have a mutation of the *PTEN* gene. Management centers on cancer surveillance and prevention.

BANNAYAN–RILEY–RUVALCABA SYNDROME

Bannayan–Riley–Ruvalcaba syndrome is also known as Bannayan–Zonana syndrome, Ruvalcaba–Myhre–Smith syndrome, and Riley–Smith syndrome. BRRS is a rare congenital syndrome with features of intestinal juvenile polyps, macrocephaly, subcutaneous and visceral lipomas and hemangiomas, cognitive and motor developmental delay, lipid storage myopathy, Hashimoto thyroiditis, and pigmentary spotting of the penis in males. The prevalence of BRRS is unknown. Approximately half of the patients with BRRS have evidence of germ-line mutations in *PTEN*, the same mutated gene that is associated with CS. BRRS and

CS have been described in the same family, suggesting that the two syndromes might be allelic. Analysis of patients with BRRS and CS for germ-line *PTEN* mutations indicates a similar mutational pattern. Some BRRS patients do not have germ-line mutations in *PTEN*. Rather, they have gross chromosome deletions or rearrangements involving the *PTEN* locus on chromosome 10, making the patient haploinsufficient. CS and BRRS are sometimes referred to as the *PTEN* hamartoma tumor syndrome (PHTS).

JUVENILE POLYPOSIS SYNDROME

Juvenile polyposis syndrome is a disease in which 10 or more juvenile polyps are found in the gastrointestinal tract without evidence of the extraintestinal manifestations that categorize the patient in the other hamartomatous polyposis syndromes. Juvenile polyps are hamartomas, which are mature but disorganized tissues indigenous to the anatomical site of origin. The cystic, inflammatory, nonneoplastic histological appearance of juvenile polyps initially led most clinicians to feel there was no risk of malignant transformation, but the association between cancer and JPS is well described. Several cases describing the presence of atypical, hyperplastic, or adenomatous and adenocarcinomatous changes have been reported in the literature, representing a potential progression of the histology toward cancer. The dysplasia has been described as presenting in two forms: the coexistence of adenomatous changes in a juvenile polyp, and coexisting adenomas lacking features of juvenile polyps found in patients with JPS.

Patients with JPS are predisposed to gastrointestinal malignancies, with a 16% risk of developing colorectal carcinoma in young patients and a cumulative risk of 68% by 60 years of life. It is noteworthy that the mean age at presentation is 34 years (range 15–59 years) and that the clinical outcome is usually poor, with a high percentage of poorly differentiated or mucinous cancers. It is possible that the cancer risk in this group is even higher given the large number of colectomies performed to treat unmanageable symptoms. Also associated with JPS are pancreatic, gastric, and duodenal cancers.

The *PTEN* gene has been implicated in a minority of patient with JPS, whereas approximately half of the patients with JPS have germ-line mutations in the *SMAD4* gene located on chromosome 18q21.1. This gene encodes a key intracellular signal transducer and transcriptional regulator for the transforming growth factor-β superfamily of ligands and receptors. Another one-quarter to one-half of JPS patients have mutations in the bone morphogenetic protein receptor type 1A (BMPR1A), a receptor for the bone morphogenetic protein ligand, which is upstream in the SMAD4 signaling cascade.

Management of JPS involves treatment of secondary symptoms arising from the polyposis (bleeding, malabsorption). Surveillance for cancer formation is critical.

See Also the Following Articles

Cancer, Overview • Familial Risk of Gastrointestinal Cancers

Further Reading

Carethers, J. M. (2002). Hamartomatous polyposis syndromes: Genetic pathways. *Curr. Opin. Gastroenterol.* **18**, 60–67.

Giardiello, F. M., Brensinger, J. D., Tersmette, A. C., *et al.* (2000). Very high risk of cancer in familial Peutz–Jeghers syndrome. *Gastroenterology* **119**, 1447–1453.

Giardiello, F. M., Hamilton, S. R., Kern, S. E., Offerhaus, G. J. A., Green, P. A., Celano, P., Krush, A. H., and Booker, S. V. (1991). Colorectal neoplasia in juvenile polyposis or juvenile polyps. *Arch. Dis. Child.* **66**, 971–975.

Goodman, Z. D., Yardley, J. H., and Milligan, F. D. (1979). Pathogenesis of colonic polyps in multiple juvenile polyposis. *Cancer* 43, 1906–1913.

Gruber, S. B., Entius, M. M., Petersen, G. M., *et al.* (1998). Pathogenesis of adenocarcinoma in Peutz–Jeghers syndrome. *Cancer Res.* **58**, 5267–5270.

Heiss, K. F., Schaffner, D., Ricketts, R. R., and Winn, K. (1993). Malignant risk in juvenile polyposis coli: Increasing documentation in the pediatric age group. *J. Pediatr. Surg.* **28**, 1188–1193.

Hemminki, A., Markie, D., Tomlinson, I., Axizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Aminoff, M., Hoglund, P., Jarvinen, H., Kristo, P., Pelin, K., Ridanpaa, M., Salovaara, R., Toro, T., Bodmer, W., Olschwang, S., Olden, A. S., Stratton, M. R., de la Chapell, A., and Aaltonen, L. A. (1998). A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature* **391**, 184–187.

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Jarvinen, H. J., Sistonene, P., Tomlinson, I. P. M., Houlston, R. S., Bevan, S., Mitros, F. A., Stone, E. M., and Aaltonen, L. A. (1998). Mutations in the *SMAD4/DPC4* gene in juvenile polyposis. *Science* **280**, 1086–1088.

Huang, S. C., Chen, C. R., Lavine, J. E., Taylor, S. F., Newbury, R. O., Pham, T. T. T., Ricciardiello, L., and Carethers, J. M. (2000). Genetic heterogeneity in familial juvenile polyposis. *Cancer Res.* **60**, 6882–6885.

Li, J., Yen, C., Liaw, D., *et al.* (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast and prostate cancer. *Science* **275**, 1943–1947.

Liaw, D., Marsh, D. J., Li, J., Dahia, P. L. M., Wang, S. I., Zheng, Z., Bose, S., Call, K. M., Tsou, H. C., Peacocke, M., Eng, C., and Parsons, R. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nature Genet.* **16**, 64–67.

Marsh, D. J., Coulon, V., Lunetta, K. L., Rocca-Serra, P., Dahia, P. L. M., Zheng, Z., Liaw, D., Caron, S., Duboue, B., Lin, A. Y., Richardson, A.-L., Bonnetblanc, J.-M., Gorlin, R. J., Hodgson, S. V., Huson, S., Lacombe, D., LePrat, F., Odent, S., Toulouse, C.,

Olopade, O. I., Sobol, H., Tishler, S., Woods, C. G., Robinson, B. G., Weber, H. C., Parson, R., Peacocke, M., Longy, M., and Eng, C. (1998). Mutations spectrum and genotype–phenotype analyses in Cowden disease and Bannayan–Zonana syndrome, two hamartoma syndromes with germline *PTEN* mutations. *Hum. Mol. Genet.* **7**, 505–517.

Nelen, M. R., Padberg, G. W., Peeters, E. A. J., Lin, A. Y., van den Helm, B., Frants, R. R., Coulon, V., Goldstein, A. M., van Reen, M. M. M., Easton, D. F., Eeles, R. A., Hodgson, S., Mulvihill, J. J., Murday, V. A., Tucker, M. A., Mariman, E. C. M., Starink, T. M., Ponder, B. A. J., Ropers, H. H., Kremer, H., Longy, M., and Eng, C. (1996). Localization of the gene for Cowden disease to chromosome 10q22–23. *Nat. Genet.* **13**, 114–116.

Zigman, A. F., Lavine, J. E., Jones, M. C., Boland, C. R., and Carethers, J. M. (1997). Localization of the Bannayan–Riley–Ruvalcaba syndrome to chromosome 10q23. *Gastroenterology* **113**, 1433–1437.

Haustra

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

taenia coli Three strips of longitudinal smooth muscle spaced equidistantly around the circumference of the colon.

Haustra appear as a prominent morphologic feature when the colon is viewed radiographically after instillation of barium contrast media in humans (Fig. 1). Haustra appear as sacculations that are separated from one another by constrictions of the colonic circular muscle coat; they are repeated more or less uniformly along the length of colon. The haustra are not permanently fixed structures, and sometimes appear to migrate along the length of the colon. At other times, they may be absent. In their absence (e.g., in certain disease states), the colonic wall appears on radiographs to have a smooth profile.

ANATOMY AND PHYSIOLOGY

The circular muscle coat of the ascending, transverse, and descending segments of the colon is overlaid on the surface with three thin strips of longitudinal muscle that are about 6 mm wide and spaced equidistantly around the circumference. These three strips of longitudinal muscle are the epiploic taenia, libera taenia, and mesocolic taenia; as a group, these strips are called the taenia coli. When the taenia coli reach the rectosigmoid region of the large intestine, they merge to envelop the entire circumference of the bowel. An

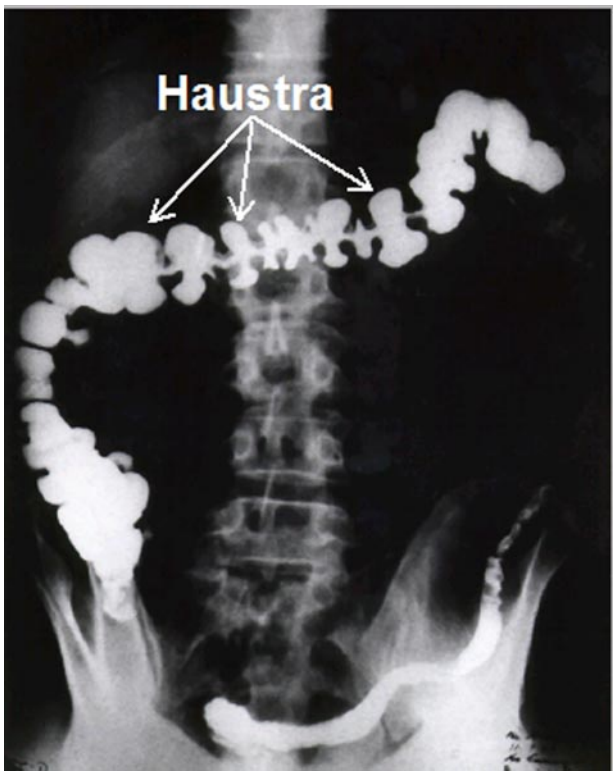


FIGURE 1 X-Ray film showing haustrations in the ascending and transverse colon of a human. Between the haustral "pouches" are segments of contracted circular muscle.

Copyright 2004, Elsevier (USA). All rights reserved.

Olopade, O. I., Sobol, H., Tishler, S., Woods, C. G., Robinson, B. G., Weber, H. C., Parson, R., Peacocke, M., Longy, M., and Eng, C. (1998). Mutations spectrum and genotype–phenotype analyses in Cowden disease and Bannayan–Zonana syndrome, two hamartoma syndromes with germline *PTEN* mutations. *Hum. Mol. Genet.* **7**, 505–517.

Nelen, M. R., Padberg, G. W., Peeters, E. A. J., Lin, A. Y., van den Helm, B., Frants, R. R., Coulon, V., Goldstein, A. M., van Reen, M. M. M., Easton, D. F., Eeles, R. A., Hodgson, S., Mulvihill, J. J., Murday, V. A., Tucker, M. A., Mariman, E. C. M., Starink, T. M., Ponder, B. A. J., Ropers, H. H., Kremer, H., Longy, M., and Eng, C. (1996). Localization of the gene for Cowden disease to chromosome 10q22–23. *Nat. Genet.* **13**, 114–116.

Zigman, A. F., Lavine, J. E., Jones, M. C., Boland, C. R., and Carethers, J. M. (1997). Localization of the Bannayan–Riley–Ruvalcaba syndrome to chromosome 10q23. *Gastroenterology* **113**, 1433–1437.

Haustra

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

taenia coli Three strips of longitudinal smooth muscle spaced equidistantly around the circumference of the colon.

Haustra appear as a prominent morphologic feature when the colon is viewed radiographically after instillation of barium contrast media in humans (Fig. 1). Haustra appear as sacculations that are separated from one another by constrictions of the colonic circular muscle coat; they are repeated more or less uniformly along the length of colon. The haustra are not permanently fixed structures, and sometimes appear to migrate along the length of the colon. At other times, they may be absent. In their absence (e.g., in certain disease states), the colonic wall appears on radiographs to have a smooth profile.

ANATOMY AND PHYSIOLOGY

The circular muscle coat of the ascending, transverse, and descending segments of the colon is overlaid on the surface with three thin strips of longitudinal muscle that are about 6 mm wide and spaced equidistantly around the circumference. These three strips of longitudinal muscle are the epiploic taenia, libera taenia, and mesocolic taenia; as a group, these strips are called the taenia coli. When the taenia coli reach the rectosigmoid region of the large intestine, they merge to envelop the entire circumference of the bowel. An

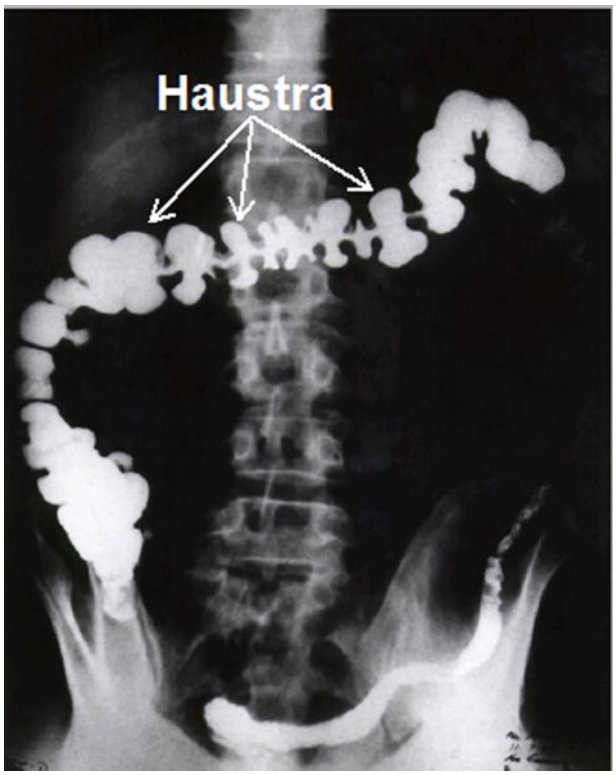


FIGURE 1 X-Ray film showing haustrations in the ascending and transverse colon of a human. Between the haustral "pouches" are segments of contracted circular muscle.

Copyright 2004, Elsevier (USA). All rights reserved.

earlier interpretation assumed that the taenia coli are shorter than the length of the colon and that this produces circumferential wall distortion, giving the sacculated appearance of haustra. Later discoveries of the role of the enteric nervous system in the control of the colonic musculature caused this concept to become outmoded.

A segmental pattern of colonic motility programmed by the enteric nervous system now accounts for the ultraslow forward movement of feces in the colon. The motility pattern is called "haustration." Ringlike contractions of the circular muscle divide the colon into chambers, in which contraction of the autogenic smooth muscle is maintained in a state of inhibition by ongoing activity of inhibitory motor neurons. The ringlike constrictions reflect autogenic muscle contractions in segments in which the inhibitory motor innervation has been inactivated by integrative functions in the colon's enteric neural network.

The neurally programmed haustrations are dynamic in that they form and reform at different sites. The most common pattern in the fasting individual is for the contracting segment to propel the contents in both directions into relaxed receiving segments. This mixes and compresses the semiliquid feces in the haustral pockets and probably facilitates absorption of water with minimal forward propulsion. Net forward propulsion occurs when sequential migration of the haustra occurs along the length of bowel. The contents of one haustral pocket are propelled into the next region, where a second pocket is formed, and from there to the next segment, where the same events occur. This results in slow forward progression and is believed to be a mechanism for compacting the feces in storage.

HAUSTRA AS MARKERS OF DISEASE STATES

The radiographic appearance of abnormal haustrations has been reported for patients diagnosed with irritable bowel syndrome. Major hallmarks of irritable bowel syndrome are abdominal pain and disordered defecation that cannot be ascribed to an organic or nonneurogenic disorder. Modern investigative methods now implicate neuropathy in the brain–gut axis as a factor in the etiology of the syndrome. Neuropathic alterations, especially in the neural networks of the enteric nervous system, are expected to underlie a motility component of the "irritable colon."

Variabilities in the shape, size, and spacing of colonic haustrations are reported for some cases of irritable bowel syndrome. One of the striking variations is a complete absence of haustra coincident with exaggerated narrowing of the colonic lumen that produces an appearance of shortening. The appearance of shortening, narrowing, and lack of haustration may also be observed in patients with advanced ulcerative colitis.

See Also the Following Articles

Brain–Gut Axis • Colonic Motility • Enteric Nervous System • Irritable Bowel Syndrome

Further Reading

Hasler, W. (1996). Embryology, anatomy, physiology of the colon, rectum, and anus. *In* "Gastroenterology and Hepatology" (M. Feldman, and C. Boland, eds.), Vol. 2, pp. 1–30. Current Medicine, Philadelphia.

Helicobacter pylori

ZHANNAT Z. NURGALIEVA*, KAREN J. GOODMAN†, AND DAVID Y. GRAHAM*,‡
*Baylor College of Medicine, †University of Texas, and ‡VA Medical Center, Houston

B-cell gastric lymphoma A monoclonal proliferation of cancerous B cells that have infiltrated the gastric glands.

CagA A protein produced by the cag pathogenicity island of *Helicobacter pylori* and encoded by cytotoxin-associated gene A. The protein induces the production of CagA antibodies and the $CagA^+$ strain of *H. pylori* is responsible for a more severe inflammatory response in the host than that produced by $CagA^-$ strains. $CagA^+$ strains are associated with a higher likelihood of peptic ulcer, atrophic changes in stomach, intestinal metaplasia, and gastric cancer.

***cag* pathogenicity island (*cag* PAI)** A group of approximately 40 genes present in the genomes of some *Helicobacter pylori* strains. Strains with the *cag* PAI cause more inflammation and the presence of this pathogenicity island is associated with an increased risk of developing gastric cancer or peptic ulcer disease.

gram-negative A characteristic that refers to bacteria that do not retain the violet stain used in Gram's method and therefore appear pink instead of blue after being stained.

gastritis Inflammation of the mucous membrane of the stomach.

outer inflammatory protein A An outer inflammatory protein of *Helicobacter pylori* that is involved in mucosal inflammation. The protein is considered a virulence factor of the bacterium and its presence is associated with increased mucosal inflammation.

salvage therapy Antimicrobial regimens designed to be used after failure of first-line therapies.

ulcer A break in the lining of the stomach, duodenum, or esophagus where hydrochloric acid and pepsin are present. Ulcers formed in the stomach are known as gastric or stomach ulcers and ulcers formed in the first part of the small intestine are known as duodenal ulcers.

urease An enzyme that breaks down urea to form ammonia and carbon dioxide.

***Helicobacter pylori* is a transmissible bacterial infection of the gastric mucosal surface. The infection results in progressive mucosal damage with eventual impairment of gastric function. *H. pylori* infection is now established as the most common cause of gastritis and is etiologically related to gastric ulcer disease, duodenal ulcer disease, primary gastric B-cell lymphoma, and gastric adenocarcinoma. This unique bacterium, which resides in the acidic environment of the human stomach, was identified in 1982. Since that time, extensive research by health scientists has resulted in considerable knowledge regarding its role in disease. There has also been success in developing simple methods to identify the infection and antibiotic regimens for its eradication. Nonetheless, there are still many avenues to explore with regard to the nature of the bacterium and its interactions with the host. Victory over *H. pylori* cannot be declared until it can be effectively prevented from causing human disease. Current gaps in knowledge regarding the natural history of the infection, its mode of transmission, pathogenic mechanisms, cofactors involved in disease pathways, and how to induce immunity present obstacles to prevention.**

MICROBIOLOGICAL PROPERTIES

Helicobacter pylori belongs to a family of gram-negative, spiral-shaped bacteria. It measures $0.6 \times 3.5\,\mu m$ in size. It is highly motile due to the presence of unipolar flagella that allow the organism to swim with a darting motion through the thick mucus lining the stomach cavity. The bacterium is microaerophilic and growth *in vitro* requires an enriched selective medium and an atmosphere with reduced oxygen and increased CO_2 concentrations. The optimum temperature for growth is 37°C. Growth on agar plates typically appears within 3 to 4 days as tiny, smooth, and translucent colonies of *H. pylori*. Initial growth may require a longer time and culture plates are typically kept for 2 weeks to improve the sensitivity of this method to diagnose infection.

BACTERIAL VIRULENCE FACTORS

Important characteristics of the bacterium that allow colonization of the gastric mucosa include motility, urease production, and the ability to adhere to the surface cells lining the stomach cavity. Motility allows the bacterium to escape the highly acidic surroundings in the stomach lumen and reach the more alkaline surface cells of the gastric mucosa. Urease produced by the bacterium hydrolyzes urea to liberate ammonia, which neutralizes gastric acid and provides a suitable microenvironment for *H. pylori* while protecting it from the acid in the stomach.

Colonization by the bacterium and production of toxins induce the production of pro-inflammatory

 Copyright 2004, Elsevier (USA). All rights reserved.

factors (cytokines) by the host. Evidence suggests that the severity of the host's inflammatory response is related to the presence of certain bacterial genetic factors including the presence of the *cag* pathogenicity island and the outer inflammatory protein, OipA. In Asia, more than 90% of *H. pylori* strains possess the *cag* pathogenicity island. In developed Western countries, the proportion is somewhat lower but is generally 60% or more. The presence of this pathogenicity island is associated with more aggressive clinical outcomes of *H. pylori* infection, including duodenal ulcer and gastric cancer. However, *H. pylori* strains lacking both the *cag* pathogenicity island and OipA have been found in patients with peptic ulcer and with gastric cancer, and therefore, there is no currently identifiable "safe" *H. pylori* infection.

NATURAL HISTORY

Acute *H. pylori* infection has been observed in a few cases of experimental and accidental inoculation as well as in prospective studies, especially in Japan. In these cases, *H. pylori* colonization led to inflammation of the gastric mucosa and transient reduction in the ability of the stomach to make acid (hypochlorhydria) accompanied by a broad spectrum of symptoms. Two voluntary inoculations produced distinct outcomes. The first resulted in gastritis accompanied by acute dyspeptic symptoms and detectable *H. pylori* on histologic examination of biopsies; the infection appeared to be eliminated spontaneously by 14 days. A 1-week course of tinidazole was also taken. No antibody response was detected. The second experimental inoculation resulted in a persistent infection, which resisted a series of trials using single antibiotics. In this case, anti-*H. pylori* immunoglobulin M (IgM) levels rose and fell within weeks following acute infection; IgG became detectable when IgM levels fell. Acute infection is recognized by the presence of new-onset upper gastrointestinal symptoms and either the characteristic histologic features or the presence of positive histology or urea breath test (see below) and negative serum IgG antibody levels. Several studies have followed IgG levels after *H. pylori* colonization is eliminated; titers generally decline following elimination, often reaching seronegative levels within 1 or 2 years. The immune response to *H. pylori* does not appear to confer immunity, given that previously infected individuals are susceptible to re-infection. Furthermore, co-infection with multiple strains, though uncommon, has been reported. Because spontaneous elimination is rarely observed when adults with preexisting *H. pylori* infection are followed over time, it has been assumed that infection generally persists once acquired. However, this assumption overlooks observations that suggest that exposure to *H. pylori* may sometimes result in brief, self-limiting infection, particularly in young children. Given that infection is not generally detected at onset, the proportion of acute infections that persist is not known. Furthermore, cases of infection detected in epidemiologic studies will tend to be persistent, particularly when measurement of antibodies is used for detection. Therefore, most of what is known about the epidemiology of *H. pylori* infection relates to persistent infection; little is known about the epidemiology of acute *H. pylori* infection.

EPIDEMIOLOGY

Acquisition of *H. pylori* infection typically occurs in childhood and in Western countries it is likely to result commonly from intrafamilial spread. Age, socioeconomic conditions (in particular, the stage of economic development of the geographic region, residential crowding, education level, and belonging to a socially marginalized group), and migration from a high-prevalence region are major determinants of the prevalence of the infection. Children from higher social strata in developed nations typically acquire the infection at a rate of less than 1% per year, except perhaps during early childhood, whereas in developing countries, the rate of acquisition among young children may exceed 10% per year. The prevalence of *H. pylori* infection in adults is believed to reflect the rate of acquisition when they were children and thus the social and economic conditions of their families during that time period. In developing countries, widespread acquisition of the infection occurs early in life, such that by adolescence or earlier, the prevalence is often in the range of 80–90% and remains constant across adult birth cohorts. In developed countries, the average rate of acquisition appears to have fallen as standards of living have improved. This change manifests in an increase in prevalence with age, which is believed to represent a progressive fall in acquisition in successive birth cohorts, rather than continued acquisition of infection during adulthood. Longitudinal studies in developed countries have shown that the prevalence of infection is declining in all age groups. It has been suggested that this apparent change in the epidemiology of *H. pylori* infection is possibly related to the widespread use of antibiotics for other infections.

TRANSMISSION

Because *H. pylori* is not easily isolated from extragastric secretions, the usual portals of entry and exit to and

from the human host are not known with certainty. Perinatal transmission from mother to infant does not appear to occur readily and blood-borne transmission is implausible. Iatrogenic transmission by gastroenterological procedures has been documented, but other specific modes of transmission have been neither confirmed definitively nor ruled out.

Given that residential crowding is a major risk factor for infection and that the infection clusters within families, person-to-person transmission is likely, although the relative importance of fecal–oral, oral–oral, and gastric–oral (for example, through vomitus) pathways is not clear. Studies from a number of areas, especially Andean Latin America and Kazakhstan, have suggested a role for waterborne transmission. Contradictory findings have been reported from other parts of the world, although investigators from various regions have reported the detection of *H. pylori* in water using molecular techniques such as polymerase chain reaction. The major route of transmission may well vary among populations, with the organism being opportunistic in the sense that any route that allows access to the stomach may lead to infection. Evidence consistent with each of these routes has been reported.

Studies in rural Andean children under 10 years of age suggest that *H. pylori* infection is most readily transmitted among siblings who are close in age, whereas researchers in Germany, where the average family size is smaller, believe that parent-to-child transmission may be more important. Studies of the *H. pylori* strains recovered from family members have generally shown them to be similar or identical to the strain from one of the parents, often the mother. Recently, *H. pylori* has been isolated from vomitus as well as from natural and cathartic stools, suggesting that transmission may be enhanced during episodes of acute gastroenteritis.

Studies investigating the possibility of zoonotic transmission have also had conflicting findings and recent data suggest that it is more likely that animals in close contact with humans may acquire human *H. pylori* strains rather than animals being a reservoir for human infection.

EFFECTS ON THE STOMACH

The stomach serves as a chamber for milling, disinfecting, and delivering components of ingested substances to the small intestine for absorption. Acid and pepsin secreted in the stomach help to denature proteins and prevent potential pathogens from colonizing the mucosa. The stomach can be divided into three major areas: one area, the body or corpus, where cells secrete acid, and two areas, the cardia and the antrum, the most proximal and distal parts of the stomach, respectively, which do not make acid. The corpus, the largest area of the stomach, contains glands that make acid and pepsin. The antrum contains endocrine cells that produce gastrin (G cells) and somatostatin (D cells), which help to control acid secretion in the corpus.

H. pylori-induced inflammation in the antrum can produce an imbalance between gastrin and somatostatin secretion, leading to a dysregulation of acid secretion, which tends to produce more acid for longer time periods. In contrast, inflammation in the corpus can inhibit acid-secreting cells. Therefore, the physiologic response to *H. pylori* infection depends on the site and severity of the inflammation. When the infection is primarily in the antrum (antral predominant gastritis), the individual tends to make excessive amounts of acid and is at increased risk of duodenal ulcer disease. Severe corpus gastritis is associated with low-level acid secretion and degeneration of the normal glandular architecture of the stomach, leading to gastric atrophy. Patients with severe corpus gastritis are at increased risk of gastric ulcer and gastric cancer.

H. PYLORI-ASSOCIATED DISEASES

Gastritis

Almost all *H. pylori*-infected individuals develop inflammation of the lining of the stomach (gastritis) (Fig. 1). Typically, this inflammation does not cause symptoms; thus, the infection has a long latent period. Acute infection may be accompanied by nonspecific transient abdominal pain, nausea, and vomiting. Chronic infection is characterized by infiltration of inflammatory cells within the gastric mucosa. Eradication of *H. pylori* generally results in healing of the gastric mucosa with reduction or disappearance of the inflammation followed by a normalization of gastric secretion. In some individuals, persistent *H. pylori* gastritis leads to a loss of the glandular components of the gastric mucosa, a process known as atrophy. In such cases, *H. pylori*-induced gastritis damages gastric structure and function, leading to reduced acid secretion (hypochlorhydria or achlorhydria), with consequences such as increased enteric infection or reduced absorption of iron and vitamin B12.

Peptic Ulcer Disease

The lifetime risk of developing peptic ulcer among *H. pylori*-infected individuals is approximately 1 in 6. Peptic ulcer disease is a chronic disease that can spontaneously heal and recur. The cure of *H. pylori* infection leads to healing of active ulcers and prevention of ulcer recurrence. Thus, discovery of effective

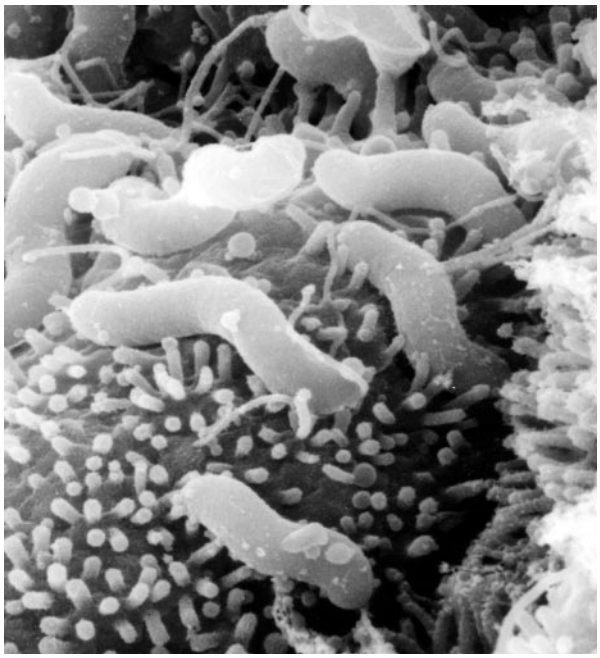


FIGURE 1 Scanning electron microscopic photograph of *Helicobacter pylori* attached to the lining of the human stomach. Courtesy of Professor Tatsuo Yamamoto, Niigata, Japan.

methods to treat *H. pylori* infection resulted in changing a major chronic disease into a curable condition.

Gastric Cancer

The World Health Organization's International Agency for Research on Cancer classified *H. pylori* as a definite carcinogen in 1994. Gastric cancer was known to be associated with atrophic gastritis before the discovery of its association with *H. pylori*. The idea that chronic *H. pylori*-induced gastritis could lead to cancer fit coherently into previous theories regarding the role of chronic inflammation in gastric carcinogenesis. The lifetime risk of gastric adenocarcinoma varies greatly across populations (e.g., 1−3% in the United States and 11−12% in Japan). However, not all populations with a high prevalence of *H. pylori* infection have a high rate of gastric cancer; the type of associated gastritis is the best predictor of the outcome. In the United States, where gastric cancer rates are lower, antral predominant gastritis and superficial gastritis without atrophy are the most common types. In contrast, in Japan, where gastric cancer rates are higher, *H. pylori*-associated gastritis is typically atrophic. The different cancer risks associated with *H. pylori* infection across populations most likely reflect differences in the distributions of cofactors, such as dietary patterns involved in carcinogenesis, as well as differences in the virulence of the dominant *H. pylori* strains. Current knowledge suggests that widespread prevention of *H. pylori* infection will yield dramatic declines in gastric cancer rates.

B-Cell Gastric Lymphoma

Continuous lymphoid infiltration of gastric mucosa can lead to the development of tumors, called mucosa-associated-lymphoid tissue lymphomas (MALT lymphomas), that arise in the lymphoid tissues of the stomach. MALT lymphomas associated with *H. pylori* infection are monoclonal proliferations of cancerous B cells that have infiltrated the gastric glands. The malignancy is typically low-grade. Eradication of *H. pylori* infection, if instituted early, leads to the remission of MALT lymphomas in most cases.

DIAGNOSIS

Patients with a potentially curable clinical manifestation of *H. pylori* should be tested. This group includes those patients with known peptic ulcer disease or low-grade gastric MALT lymphoma. Most clinicians agree that testing patients with dyspepsia is worthwhile because that population includes many patients with peptic ulcer disease. The controversy concerns how to target patients with latent *H. pylori* infection in order to prevent a clinical manifestation. Current recommendations by the Maastrich consensus conference also target those with a known increased risk of a clinical manifestation and include those with a family history of gastric cancer or peptic ulcer. Long-term use of drugs that suppress gastric acid secretion is now thought to promote the progression of *H. pylori* gastritis. Thus, many experts recommend testing patients for whom long-term antisecretory therapy is planned (e.g., those with gastroesophageal reflux disease). Cost-effectiveness research in the United States has shown that a population-based test and treat approach may be a worthwhile disease prevention effort, but the potential adverse effects associated with such an approach have not yet been adequately evaluated.

Active or current *H. pylori* infection can be diagnosed noninvasively using a urea breath test or a stool antigen test. If endoscopy is needed, it is possible to identify the infection in biopsies of the gastric mucosa by histology, rapid urease test, or culture. Current infection or past exposure to *H. pylori* can be diagnosed by detecting antibodies in blood or urine.

Testing for Active *H. pylori* Infection

Urea Breath Testing

The urea breath test is a noninvasive method that detects *H. pylori* urease activity. The organism produces

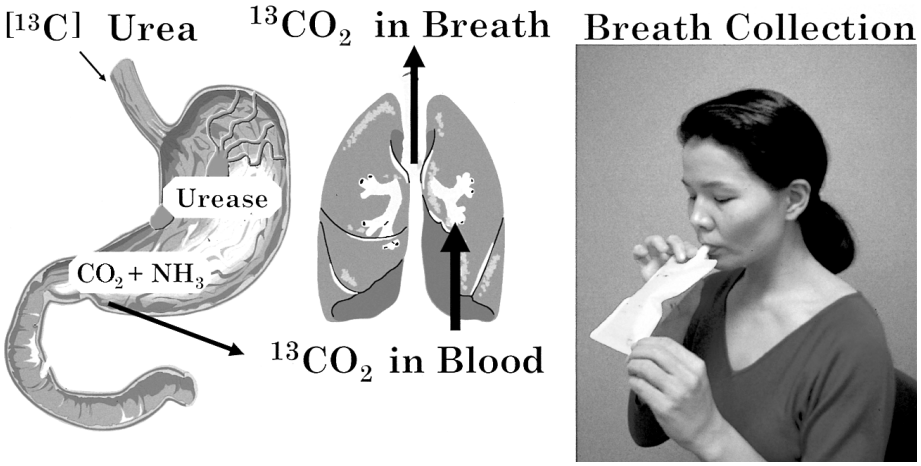


FIGURE 2 Urea breath test. This test is based on the measurement of labeled carbon dioxide produced by the enzyme urease. The steps include ingestion of labeled urea, hydrolysis of the urea into labeled CO_2, collection of a breath sample, and measurement of enrichment of the level of labeled CO_2.

the enzyme urease, which splits urea into carbon dioxide and ammonia. If one administers urea labeled with the nonradioactive carbon isotope, ^{13}C, or the radioactive isotope, ^{14}C, one can take breath samples before and after administering the labeled urea and determine whether urease is present in the stomach (Fig. 2) by measuring the excess labeled carbon in the expired breath using isotope-ratio spectrometry. Urea breath testing is considered the most accurate of the noninvasive methods for detecting active *H. pylori* infection. The use of drugs that temporarily reduce the number of *H. pylori* in the stomach can cause false-negative results. Thus, it is important to withhold proton pump inhibitors, antibiotics, and bismuth for at least 1 week prior to testing.

Stool Antigen Testing

The stool antigen test is based on the detection of *H. pylori* antigens in the stool. These antigens are detected using an enzyme immunoassay. Stool antigen testing is subject to the same limitations as the urea breath test, with the added need to refrigerate or freeze stool specimens that cannot be tested immediately. For initial diagnosis, it is comparable in accuracy to the urea breath test. It is less accurate when used to confirm successful eradication. Reliability is increased if testing is delayed until 6 to 8 weeks after treatment.

Histology

Biopsy of the gastric mucosa is easily and safely performed during examination of the stomach by endoscopy. Examination of the stained tissue under the microscope reveals both infiltration of the mucosa with inflammatory cells and the presence of the bacterium (Fig. 3). The accuracy is improved if special stains that target the bacteria are used. Biopsy-based methods (histology, culture, and rapid urease test) share the limitation that biopsy sampling may miss colonized areas of the stomach because the colonization may be patchy.

Culture

H. pylori can be cultured from gastric mucosal biopsies. Culture is not routinely available, but is particularly valuable for patients with treatment-resistant infections as one can test the recovered organism for susceptibility to different antibiotics and thus choose a therapy with an increased likelihood of success.

Rapid Urease Tests

The urea breath test detects one of the split products of ingested urea (labeled CO_2) and the rapid urease test detects the other product (ammonia). A biopsy of gastric tissue is placed into a medium containing urea and a pH indicator. When the bacterial urease splits the urea, the liberated ammonia will increase the pH; this is recognized by a color change in the test indicator. Rapid urease tests are fast, inexpensive, and easy to perform. A limitation is that this method requires a high density of bacteria in the specimen. Negative results could mean that the level of bacteria in the specimen obtained is low.

Testing for Current *H. pylori* or Past Exposure to the Infection

Antibody Testing

Detection of antibodies against *H. pylori* is widely used to check the infection status. A variety of serum,

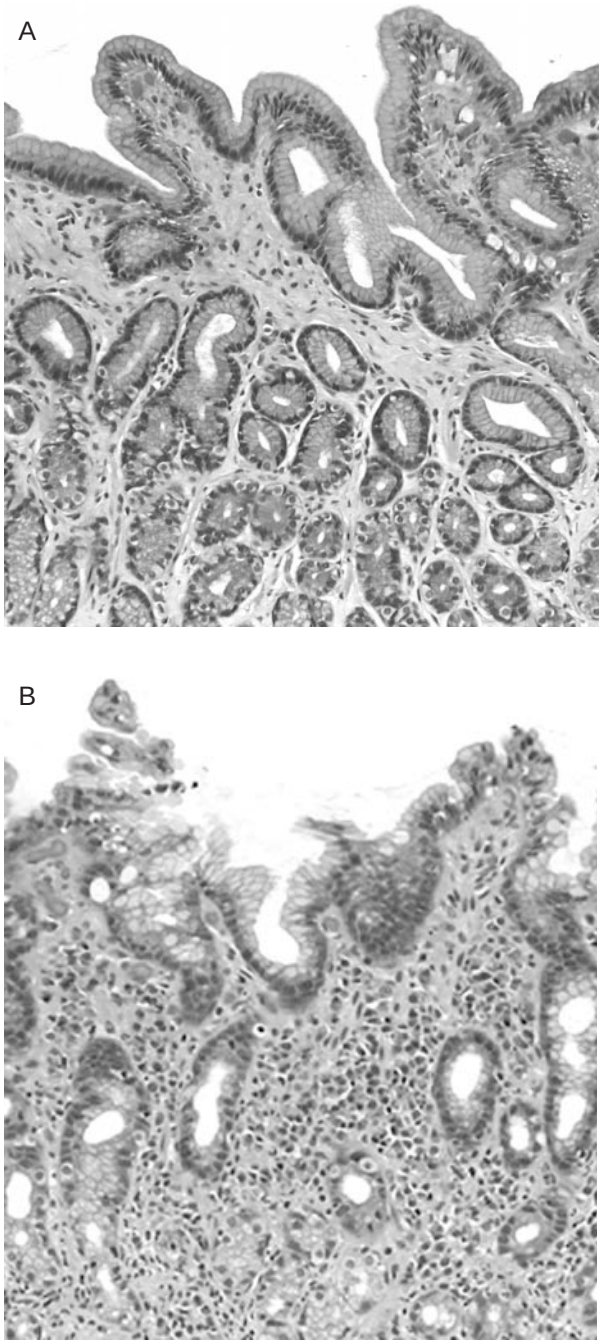


FIGURE 3 Photomicrographs of infected and normal stomachs. (A) A normal gastric mucosa essentially devoid of inflammatory cells. (B) Gastric mucosa with an inflammatory infiltrate made up of acute and chronic inflammatory cells.

whole-blood, and urine tests are available for use in the laboratory and in the office. The primary limitation is that antibody tests cannot distinguish a past history of the infection from an active infection. Antibody testing is useful for screening of patients with a high pretest probability of an active *H. pylori* infection, such as a patient with a peptic ulcer. The antibody level falls slowly after eradication of the infection and thus cannot be used to confirm successful therapy.

TREATMENT

The stomach is actually a very hostile place for antibiotics to work, as the contents are acidic and vary in volume. Nonetheless, successful combination therapies have been discovered. In general, therapy consists of a drug to inhibit acid secretion and two or more drugs to kill the bacterium. Two or more antibiotics are needed to prevent the development of antibiotic resistance. The antibiotics most widely used for anti-*H. pylori* therapy include clarithromycin, amoxicillin, metronidazole, tetracycline, and furazolidone. Clarithromycin is a member of the macrolide group of antibiotics that bind specifically to bacterial ribosomes and lead to bacterial death by disrupting protein synthesis. This drug is resistant to the acidic environment of the stomach, but is much more effective when acid secretion is suppressed. Amoxicillin is an acid-stable semisynthetic penicillin that is also a mainstay of *H. pylori* therapy and also requires suppression of acid secretion. Metronidazole is a pro-drug from the nitroimidazole group, which does not require inhibition of acid secretion. Tetracycline hydrochloride and oxytetracycline are both effective and neither requires inhibition of acid secretion. Bismuth salts including the bismuth subsalicylate and the colloid suspension of bismuth citrate are also commonly used. These drugs are thought to act topically in the stomach; they have the advantage of not requiring drugs to reduce acid secretion and not generating bacterial resistance.

Treatment Regimens

The most effective therapies are triple or quadruple therapies (based on the number of drugs used). Triple therapies typically consist of a combination of clarithromycin, amoxicillin, or metronidazole given twice a day along with a drug to reduce acid secretion. A commonly used triple or quadruple therapy consists of a bismuth salt, metronidazole, and either tetracycline or amoxicillin, with the optional addition of an antisecretory drug. The antisecretory drugs used are either histamine-2 receptor antagonists or proton pump inhibitors, with the latter being more effective for pH control (Table I). Rescue or salvage therapy is the name given to regimens designed to be used after failure of first-line therapies; such failures reflect the possibility of bacterial resistance to one or more of the drugs included in the

TABLE I Recommended Antibiotic Treatment Regimens for *Helicobacter pylori* Infection

Triple therapy
Proton pump inhibitor two times daily
Clarithromycin 500 mg two times daily
and
Amoxicillin 1000 mg two times daily
or
Metronidazole 500 mg two times daily
Quadruple therapy
Proton pump inhibitor two times daily
Bismuth four times daily
Tetracycline 500 mg four times daily
Metronidazole 500 mg three times daily

regimen. The growing problem of antibiotic resistance requires modification of the treatment approach in such cases. Culture of the organism and testing for antibiotic resistance are recommended so that a more effective new therapy can be selected.

Good compliance with the treatment regimen is considered critical for success. In general, therapies of longer duration (10 or 14 days) are more effective than those of shorter duration (7 days). Other predictors of treatment success include the prevalence of *H. pylori* strains in the local population that are resistant to antibiotics used in the treatment regimen and also the prevalence of *H. pylori* infection in children in the local population, which is an indicator of current levels of transmission. Currently, there is a need for more effective therapies in populations where antibiotic resistance and transmission levels are high.

Confirmation of Successful Therapy

Because overall current therapy for *H. pylori* is effective in only approximately 70% of cases, it is important to test for the presence of *H. pylori* posttherapy and confirm that the infection was cured. The availability of noninvasive diagnostic methods such as the urea breath test or stool antigen test enhances the feasibility of posttherapy testing and eliminates the need for posttherapy endoscopy to confirm cure of the infection. Avoiding endoscopy to verify elimination of *H. pylori* is particularly important in settings where the invasive procedure may serve as a source of re-infection. Testing should be withheld until after approximately 4 weeks of therapy cessation. With the stool antigen test, the waiting period should be longer, 6–8 weeks. The proton pump inhibitors should be stopped at least 1 week prior to testing.

PROSPECTS FOR PREVENTION

Prevention of *H. pylori* infection will require identification of the weak links in the chain of transmission that are amenable to modifications in behavior or in household sanitation. The prevalence of *H. pylori* infection "naturally" falls in association with improved standards of living, sanitation, and household hygiene. Thus, *H. pylori* infection is one of a whole host of infectious diseases that will decline and disappear as standards of living improve. For the short term, there is considerable interest in development of a vaccine to prevent infection. Work in experimental animals has provided proof of principle, but early human trials have been disappointing. As humans are the major reservoir of infection, cure of *H. pylori* infection reduces the pool of potential "transmitters." Studies are under way in developing countries to observe whether elimination of the infection in a village will lead to long-term eradication of the infection and whether it is necessary to treat the entire family of infected children to prevent re-infection. Clearly, there is still much work to be done.

Acknowledgments

This material is based upon work supported in part by the Office of Research and Development Medical Research Service Department of Veterans Affairs, by Public Health Service Grants DK-53659 and DK56338, and by funds from the Texas Gulf Coast Digestive Diseases Center. Dr. Nurgalieva is supported in part by a grant from the Foundation of Gastrointestinal Microbial Pathogens. Dr. Goodman is supported in part by Public Health Service Grant DK53664. Dr. Graham has received recent research support or honoraria for speaking engagements from AstraZeneca, Check Med, Enteric Products Inc, Meretek, Otsuka, SmithKline Diagnostics, TAP, Prometheus, Takeda, Wyeth-Ayerst, and Pharmacia. In addition, Dr. Graham is a consultant for Prometheus, Meretek, and Otsuka.

See Also the Following Articles

Atrophic Gastritis • Breath Tests • Duodenal Ulcer • Erosive and Hemorrhagic Gastritis (Gastropathy) • Functional (Non-Ulcer) Dyspepsia • Gastric Ulcer • Gastritis • Gastritis and *Helicobacter pylori*, Pediatric • H2-Receptor Antagonists • Lymphomas • Marginal Ulcer • Proton Pump Inhibitors • Stomach, Adenomas and Carcinomas of the

Further Reading

Breuer, T., Malaty, H. M., and Graham, D. Y. (1997). The epidemiology of *H. pylori* associated gastroduodenal disease. *In* "The Immunobiology of *H. pylori:* From Pathogenesis to Prevention" (P. B. Ernst, M. P. Michetti, and P. D. Smith, eds.), pp. 1–14. Lippincott-Raven, Philadelphia, PA.

Brown, L. M. (2000). *Helicobacter pylori:* Epidemiology and routes of transmission. *Epidemiol. Rev.* **22**, 283–297.

Ferrero, R. L., Kansau, I. N., and Labigne, A. (1997). Virulence factors produced by *H. pylori. In* "The Immunobiology of

H. pylori: From Pathogenesis to Prevention" (P. B. Ernst, P. Michetti, and P. D. Smith, eds.), pp. 29–45. Lippincott-Raven, Philadelphia, PA.
Fischbach, L. A., Goodman, K. J., Feldman, M., and Aragaki, C. (2002). Sources of variation of *Helicobacter pylori* treatment success in adults worldwide: A meta-analysis. *Int. J. Epidemiol.* **31**, 128–139.
Goodman, K. J., and Cockburn, M. (2001). The role of epidemiology in understanding the health effects of *Helicobacter pylori*. *Epidemiology* **12**, 266–271.
Goodman, K. J., and Correa, P. (1995). The transmission of *Helicobacter pylori*: A critical review of the evidence. *Int. J. Epidemiol.* **24**, 875–887.
Goodman, K. J., and Correa, P. (2000). Transmission of *Helicobacter pylori* among siblings. *Lancet* **29**, 358–362.
Graham, D. Y. (2000). Therapy of *Helicobacter pylori:* Current status and issues. *Gastroenterology* **118**(Suppl. 1), S2–S8.
Graham, D. Y., Genta, R. M., and Dixon, M. F. (1999). "Gastritis." Lippincott Williams & Wilkins, Philadelphia, PA.
Graham, K. S., and Graham, D. Y. (2002). "Contemporary Diagnosis and Management of *H. pylori*-Associated Gastrointestinal Diseases." Handbooks in Health Care Co., New Town, PA.
Parsonnet, J., Harris, R. A., Hack, H. M., and Owens, D. K. (1996). Modeling cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: A mandate for clinical trials. *Lancet* **348**, 150–154.
Parsonnet, J., Shmuely, H., and Haggerty, T. (1999). Fecal and oral shedding of *Helicobacter pylori* from healthy infected adults. *J. Am. Med. Assoc.* **282**, 2240–2245.
Torres, J., Perez-Perez, G., Goodman, K. J., Atherton, J. C., Gold, B. D., Harris, P. R., la Garza, A. M., Guarner, J., and Munoz, O. (2000). A comprehensive review of the natural history of *Helicobacter pylori* infection in children. *Arch. Med. Res.* **31**, 431–469.

Helminth Infections

RICHARD D. BUNGIRO AND MICHAEL CAPPELLO
Yale University

cestode A parasitic segmented flatworm of the class Cestoidea in the phylum Platyhelminthes; also known as a tapeworm.
definitive host The host in which a parasite reaches maturity.
helminth General term for a parasitic worm.
intermediate host An animal, vertebrate or invertebrate, that serves as a host for an intermediate developmental stage of a parasite. The intermediate host transmits the parasite by releasing a form infectious to the definitive host or by being consumed by the definitive host.
nematode A nonsegmented roundworm of the phylum Nematoda; may be free-living or parasitic.
trematode A parasitic nonsegmented flatworm of the class Trematoda in the phylum Platyhelminthes; also known as a fluke.

Helminths are worms that parasitize humans and cause varying degrees of pathology. The helminths that infect humans may be broadly classified as nematodes (also known as roundworms), trematodes (flukes), and cestodes (tapeworms). Helminths collectively infect billions of persons worldwide. They are exquisitely adapted to exploit their human hosts, gaining entry through the ingestion of water or food or by penetrating the skin (directly or with the help of an insect vector). As they thrive under conditions of poor sanitation and malnutrition, helminth infections are common in those regions of the globe most afflicted by poverty; however, they may be found to some degree in affluent industrialized nations (an excellent example is the pinworm, a frequent unwelcome guest in schools and day-care centers). Various species of helminths reside in or migrate through almost every tissue and organ and especially common sites of infection are the gastrointestinal and hepatobiliary systems. Many helminth infections are clinically silent for years to decades, but in some hosts severe, even fatal disease may occur.

NEMATODES

Nematodes are a highly diverse group of worms that are found in almost every terrestrial and aquatic habitat. Most are free-living (i.e., nonparasitic). Although

Copyright 2004, Elsevier (USA). All rights reserved.

H. pylori: From Pathogenesis to Prevention" (P. B. Ernst, P. Michetti, and P. D. Smith, eds.), pp. 29–45. Lippincott-Raven, Philadelphia, PA.
Fischbach, L. A., Goodman, K. J., Feldman, M., and Aragaki, C. (2002). Sources of variation of *Helicobacter pylori* treatment success in adults worldwide: A meta-analysis. *Int. J. Epidemiol.* **31**, 128–139.
Goodman, K. J., and Cockburn, M. (2001). The role of epidemiology in understanding the health effects of *Helicobacter pylori*. *Epidemiology* **12**, 266–271.
Goodman, K. J., and Correa, P. (1995). The transmission of *Helicobacter pylori*: A critical review of the evidence. *Int. J. Epidemiol.* **24**, 875–887.
Goodman, K. J., and Correa, P. (2000). Transmission of *Helicobacter pylori* among siblings. *Lancet* **29**, 358–362.
Graham, D. Y. (2000). Therapy of *Helicobacter pylori:* Current status and issues. *Gastroenterology* **118**(Suppl. 1), S2–S8.
Graham, D. Y., Genta, R. M., and Dixon, M. F. (1999). "Gastritis." Lippincott Williams & Wilkins, Philadelphia, PA.
Graham, K. S., and Graham, D. Y. (2002). "Contemporary Diagnosis and Management of *H. pylori*-Associated Gastrointestinal Diseases." Handbooks in Health Care Co., New Town, PA.
Parsonnet, J., Harris, R. A., Hack, H. M., and Owens, D. K. (1996). Modeling cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: A mandate for clinical trials. *Lancet* **348**, 150–154.
Parsonnet, J., Shmuely, H., and Haggerty, T. (1999). Fecal and oral shedding of *Helicobacter pylori* from healthy infected adults. *J. Am. Med. Assoc.* **282**, 2240–2245.
Torres, J., Perez-Perez, G., Goodman, K. J., Atherton, J. C., Gold, B. D., Harris, P. R., la Garza, A. M., Guarner, J., and Munoz, O. (2000). A comprehensive review of the natural history of *Helicobacter pylori* infection in children. *Arch. Med. Res.* **31**, 431–469.

Helminth Infections

RICHARD D. BUNGIRO AND MICHAEL CAPPELLO
Yale University

cestode A parasitic segmented flatworm of the class Cestoidea in the phylum Platyhelminthes; also known as a tapeworm.
definitive host The host in which a parasite reaches maturity.
helminth General term for a parasitic worm.
intermediate host An animal, vertebrate or invertebrate, that serves as a host for an intermediate developmental stage of a parasite. The intermediate host transmits the parasite by releasing a form infectious to the definitive host or by being consumed by the definitive host.
nematode A nonsegmented roundworm of the phylum Nematoda; may be free-living or parasitic.
trematode A parasitic nonsegmented flatworm of the class Trematoda in the phylum Platyhelminthes; also known as a fluke.

Helminths are worms that parasitize humans and cause varying degrees of pathology. The helminths that infect humans may be broadly classified as nematodes (also known as roundworms), trematodes (flukes), and cestodes (tapeworms). Helminths collectively infect billions of persons worldwide. They are exquisitely adapted to exploit their human hosts, gaining entry through the ingestion of water or food or by penetrating the skin (directly or with the help of an insect vector). As they thrive under conditions of poor sanitation and malnutrition, helminth infections are common in those regions of the globe most afflicted by poverty; however, they may be found to some degree in affluent industrialized nations (an excellent example is the pinworm, a frequent unwelcome guest in schools and day-care centers). Various species of helminths reside in or migrate through almost every tissue and organ and especially common sites of infection are the gastrointestinal and hepatobiliary systems. Many helminth infections are clinically silent for years to decades, but in some hosts severe, even fatal disease may occur.

NEMATODES

Nematodes are a highly diverse group of worms that are found in almost every terrestrial and aquatic habitat. Most are free-living (i.e., nonparasitic). Although

Copyright 2004, Elsevier (USA). All rights reserved.

parasitic nematodes make up a relatively small proportion of the total, various species have evolved to exploit almost all plants and animals (vertebrate and invertebrate). Dozens of nematode species parasitize humans and many rank among the most common infections known. In addition, parasitized persons frequently harbor more than one species simultaneously. Many species of parasitic nematodes inhabit the human gastrointestinal tract and others reside in tissues (the latter are collectively known as filarial nematodes). The major species of nematodes infecting humans are listed in Table I.

Parasitic nematodes are nonsegmented roundworms of the phylum Nematoda. They possess an outer covering, or cuticle, that is chemically complex in structure and serves to protect the worm from attack (e.g., by host antibodies or digestive enzymes). The cuticle is shed each time the nematode molts during development. A muscle layer under the cuticle allows for active movement of larval and adult forms. Nematodes have a complete digestive system, with a mouth (also known as a buccal cavity), muscular esophagus, intestine, and anus. The buccal cavity of hookworms contains teeth (*Ancylostoma* species) or cutting plates (*Necator americanus*) that the worms use to attach to and lacerate intestinal mucosal tissue. Following copulation with the male, females of most intestinal nematode species produce eggs that are voided from the host in the feces. Nematode eggs hatch to release larvae following their release into the external environment (hookworms) or on ingestion by the host (pinworms, whipworms, *Ascaris lumbricoides*). An exception is *Strongyloides stercoralis*, whose eggs hatch in the host intestine immediately after being released from the female. Females of other nematode species (e.g., *Trichinella spiralis*, the tissue-dwelling filariae) are viviparous, giving birth directly to larvae. All nematode larvae undergo four molts as they develop to adults. Nematode larvae that are infectious to humans have generally undergone their first two molts in the soil (e.g., hookworms) or in an insect vector (filaria). *S. stercoralis* is unique among parasitic nematodes in that it may undergo all four molts outside the host and exist as a free-living and sexually reproducing adult. Adult parasitic nematodes vary considerably in size, from 1 cm or smaller in length (e.g., pinworms, hookworms, *S. stercoralis*) to over 1 m (*Dracunculus medinensis*). Intermediate between these extremes is the giant roundworm *A. lumbricoides*, whose size (10–30 cm) and general appearance are quite similar to those of an earthworm. Male nematodes are generally smaller than females of the same species.

Nematode infections are generally acquired through three routes: by ingestion, by skin penetration, or through the bite of an insect vector (see Table I). Several intestinal nematode species infect through the oral route, by direct ingestion of eggs (pinworms, whipworms, *A. lumbricoides*), larvae (*Ancylostoma* species hookworms), or raw/undercooked meats contaminated with larvae (*Tr. spiralis*). Intestinal nematodes (all hookworm species, *S. stercoralis*) may also infect via penetration of host skin by third-stage larvae (percutaneous infection). By contrast, the tissue-dwelling filarial nematodes are typically spread by insect vectors such as mosquitoes (lymphatic filariae) or biting flies (*Onchocerca volvulus*, *Loa loa*). A notable exception is the tissue-dwelling guinea worm (*D. medinensis*), which is spread by drinking water contaminated with the intermediate host (a copepod, or "water flea"), which may harbor infectious larvae.

The degree of pathology caused by nematodes is dependent on many variables, such as the particular species present, parasite burden, and various host factors such as immune status. Light infections with intestinal nematodes generally cause little pathology and the host is typically unaware of their presence. At higher burdens, intestinal nematodes may cause clinical disease through various mechanisms, such as irritation (pinworms, whipworms, *S. stercoralis*, and *Tr. spiralis*), physical obstruction and/or perforation of the gastrointestinal tract (*A. lumbricoides*), or laceration of the intestinal mucosa with subsequent blood-feeding (hookworms). Intestinal nematode infections may lead to nutritional deficiencies that are especially harmful to children; heavily parasitized children may suffer serious, often irreversible physical and cognitive developmental delay. The larval stages of certain intestinal nematodes (hookworms, *S. stercoralis*, and *A. lumbricoides*) may also cause pathology as they migrate through the skin and/or pulmonary system prior to taking up residence in the gastrointestinal tract. Adult filarial worms often cause pathology in the tissues they inhabit, which include the skin (*L. loa*, *O. volvulus*), the lymphatics (*Brugia malayi*, *Wuchereria bancrofti*), the eye (*L. loa*, *O. volvulus*), or subcutaneous tissues (*D. medinensis*).

TREMATODES

Trematodes, also known as flukes, are parasitic nonsegmented flatworms belonging to the class Trematoda in the phylum Platyhelminthes. The major species infecting humans are listed in Table I. There are no free-living trematodes and all species parasitic to humans require at least one intermediate host, typically a snail, to complete their life cycle. Some species also require a second intermediate host, such as a fish (*Clonorchis sinensis*) or crab (*Paragonimus westermani*). Trematodes

TABLE I Major Helminth Species Infecting Humans

Species	Intermediate host or vector	Infectious stage; Mode of infection	Major target organ(s)	Major clinical features[a]
Nematodes				
Ascaris lumbricoides (giant roundworm)	None	Egg; ingestion	Small intestine	Diarrhea, malnutrition, intestinal obstruction
			Hepatobiliary system	Biliary obstruction
Dracunculus medinensis (guinea worm)	Copepod	Larva; ingestion of infected copepod in drinking water	Subcutaneous tissues of lower extremities	Cutaneous ulcers, secondary bacterial infections
Enterobius vermicularis (pinworm)	None	Egg; ingestion/inhalation	Colon, rectum, anus	Anal itching, behavioral disturbances, appendicitis
Hookworms[b]				
Ancylostoma duodenale	None	Larva; percutaneous or oral	Small intestine	Anemia, malnutrition, growth delay of children
Ancylostoma ceylanicum	None	Larva; percutaneous or oral	Small intestine	Anemia, malnutrition, growth delay of children
Necator americanus	None	Larva; percutaneous	Small intestine	Anemia, malnutrition, growth delay of children
Loa loa	Mango fly	Larva; bite of fly	Subcutaneous tissues	Angioedema (Calabar swelling), adults in the eye
Lymphatic Filariae				
Brugia malayi	Mosquito	Larva; bite of mosquito	Lymphatic vessels	Lymphadenitis, lymphedema (elephantiasis)
Wuchereria bancrofti	Mosquito	Larva; bite of mosquito	Lymphatic vessels	Lymphadenitis, lymphedema (elephantiasis)
Onchocerca volvulus ("river blindness")	Blackfly	Larva; bite of fly	Subcutaneous tissues	Adults: subcutaneous nodules Microfilarial larvae: dermatitis, ocular lesions
Strongyloides stercoralis (threadworm)	None	Larva; percutaneous	Small intestine	Diarrhea, malnutrition, growth delay of children
Trichinella spiralis	Pigs, cattle, etc.	Larva; ingestion of raw or undercooked meat	Small intestine	Adults in intestine, diarrhea, abdominal pain
			Muscle, brain	Larvae: cysts cause damage to tissues and organs
Trichuris trichiura (whipworm)	None	Egg; ingestion	Colon, rectum	Dysentery, prolapsed rectum
Trematodes				
Clonorchis sinensis (Chinese liver fluke)	1st: snail[c]; 2nd: fish	Metacercaria; ingestion of contaminated raw fish	Liver (bile ducts)	Liver enlargement, cholangitis (bile duct inflammation), gallstones, pancreatitis
Fasciola hepatica (sheep liver fluke)	Snail	Metacercaria; ingestion of contaminated plants	Liver	Liver damage and enlargement, bile obstruction, secondary bacterial infection

continues

TABLE I Major Helminth Species Infecting Humans (*continued*)

Species	Intermediate host or vector	Infectious stage; Mode of infection	Major target organ(s)	Major clinical features[a]
Fasciolopsis buski (giant intestinal fluke)	Snail	Metacercaria; ingestion of contaminated plants	Small intestine	Ulceration, diarrhea, malnutrition, intestinal obstruction
Paragonimus westermani (lung fluke)	1st: snail[c]; 2nd: crab	Metacercaria; ingestion of contaminated raw crabs	Lung	Cough, shortness of breath, bloody sputum, lung abscesses
			Central nervous system	Seizures, visual disturbances, headache, weakness
Schistosomes (blood flukes)				
Schistosoma mansoni	Snail	Cercaria; percutaneous	Liver	Egg-induced fibrosis leads to hepatosplenomegaly
Schistosoma japonicum	Snail	Cercaria; percutaneous	Liver	Egg-induced fibrosis leads to hepatosplenomegaly
Schistosoma haematobium	Snail	Cercaria; percutaneous	Bladder	Egg-induced fibrosis leads to hematuria, dysuria
Cestodes				
Diphyllobothrium latum (broad fish tapeworm)	1st: copepod; 2nd/3rd: fish	Plerocercoid larva; ingestion of contaminated raw fish	Small intestine	Abdominal pain, diarrhea vitamin B12 deficiency leading to anemia
Echinococcus granulosus (hydatid tapeworm)	Dog[d]; sheep	Egg (from dog feces); ingestion	Liver, lungs, brain, bone marrow	Organ damage by hydatid cysts, potentially fatal anaphylaxis following rupture of cyst
Hymenolepis nana (dwarf tapeworm)	None	Egg; ingestion	Small intestine	Nausea, vomiting, diarrhea
	Flea	Cysticercoid; ingestion of flea		
Taenia saginata (beef tapeworm)	Cow	Cysticercus; ingestion of raw or undercooked beef	Small intestine	Abdominal pain, nausea, vomiting, diarrhea
Taenia solium (pork tapeworm)	Pig	Cysticercus; ingestion of raw or undercooked pork	Small intestine	Adult: abdominal pain, nausea, vomiting, diarrhea
	None	Egg (from human feces); ingestion	Brain, eye, muscle	"Cysticercosis": cysterci cause organ damage leading to seizures, hydrocephalus, visual disturbances

[a] Caused by adult worms, unless otherwise noted.

[b] The dog and cat hookworms *A. caninum, A. braziliense*, and *Uncinaria stenocephala* cannot complete their life cycles in humans but larvae may cause cutaneous larval migrans (creeping eruption of the skin). Juvenile *A. caninum* may also cause eosinophilic enteritis.

[c] The snail is a source of cercariae, which encyst in the second intermediate host.

[d] The dog is the definitive host and is infected by consuming infected sheep.

possess a cellular epithelial covering, or tegument, that protects the parasite from host immune factors and serves as an absorptive surface for the acquisition of nutrients. A muscle layer under the tegument allows for movement within the host. Trematodes have oral and ventral suckers, which they employ for attachment and migration. The digestive system originates at the oral sucker and contains a pharynx, an esophagus, and a bifurcated intestine. There is no anus; thus, waste products must be regurgitated following digestion. Aside from the schistosome species, which have separate sexes, trematodes are hermaphroditic (having both male and female reproductive organs). Reproduction of the hermaphroditic trematodes may occur through self-fertilization (*C. sinensis*, *Fasciola hepatica*, *Fasciolopsis buski*) or cross-fertilization between two worms (*P. westermani*). The largest of the major hermaphroditic trematode species is *Fasciola hepatica* (approximately 35 mm by 15 mm) and the smallest is *P. westermani* (10 mm by 5 mm). Adult schistosomes have a somewhat greater length to width ratio, with males measuring approximately 10 mm by 1 mm and females 15 mm by 0.2 mm. Schistosomes also exhibit marked sexual dimorphism. Males have a prominent longitudinal groove, or gynecophoral canal, that cradles the female as they copulate.

The hermaphroditic trematode infections are acquired orally through ingestion of the infectious metacercarial stage. The metacercariae are found encysted on water plants (*Fasciola hepatica*, *Fasciolopsis buski*), in fish (*C. sinensis*), or in crabs (*P. westermani*). Following ingestion, the metacercaria hatches in the small intestine to release a larva. The schistosome infectious stage is the cercaria, which is released from the intermediate snail host in great numbers on exposure to sunlight. Schistosome cercariae have a forked tail and are highly motile; unlike hermaphroditic metacercariae, they actively seek out the host and penetrate intact skin. Trematode eggs are typically voided from the host in the feces; they may also be passed in the urine (*Schistosoma haematobium*) or in the sputum (*P. westermani*). After exiting the host, eggs produced by trematodes hatch on contact with fresh water, releasing a ciliated free-swimming form known as a miracidium. The miracidium seeks out and penetrates the intermediate snail host, where it develops into a structure called a sporocyst that undergoes asexual reproduction, eventually giving rise to thousands of cercariae. As discussed above, schistosome cercariae are directly infectious to humans, whereas those of the hermaphroditic trematodes encyst on plants, fish, or crustaceans to become metacercariae, which are then consumed by the host.

As is the case for nematodes, trematode-induced pathology in humans is dependent on many variables, such as the infecting species, parasite burden, and host factors. The major clinical features of trematode infections are summarized in Table I. Schistosomes are unique among parasitic helminths in that their eggs are the primary cause of clinical disease, whereas adults residing in the veins cause little, if any, pathology. Many (perhaps half) of the eggs produced by *Schistosoma mansoni* and *Schistosoma japonicum* fail to penetrate the intestinal wall to be excreted and are instead carried via the portal circulation into the liver. Once lodged in hepatic tissues, the eggs induce a hypersensitive immune reaction that in chronic cases leads to fibrosis (scarring) around the portal triads of the liver (also known as "pipe stem" fibrosis). In contrast to the schistosomes, pathology associated with the hermaphroditic trematodes is primarily caused by the adult flukes at the site of infection. For example, *Fasciola hepatica* causes trauma to the liver as it burrows through tissue while migrating to the bile ducts, where in heavy infections the parasites may cause bile obstruction. Bile duct pathology may also be caused by *C. sinensis*. Other organs affected by trematodes include the small intestine (*Fasciolopsis buski*) and the lung (*P. westermani*). Furthermore, *Paragonimus* flukes may aberrantly migrate to the central nervous system, causing seizures and other serious neurological symptoms.

CESTODES

Cestodes, also known as tapeworms, are parasitic segmented flatworms belonging to the class Cestoidea in the phylum Platyhelminthes. The major species infecting humans are listed in Table I. All cestode species parasitic to humans (excepting the dwarf tapeworm, *Hymenolepis nana*) require at least one intermediate host to complete their life cycles. Like trematodes, cestodes possess a protective tegument. As cestodes do not have a gut, all nutrient acquisition occurs through the tegument using specialized submicroscopic structures known as microtriches (similar to microvilli). The basic structure of an adult tapeworm consists of a scolex (head), a neck, and the strobila, which is a chain of segments individually known as proglottids. The scolex serves to attach the tapeworm to the host intestinal tissue; in most species, it is outfitted with four suckers, and spines may also be present. The proglottids grow from the neck of the tapeworm; there may be as few as 3 (*Echinococcus granulosus*, a diminutive 5 mm in length) or as many as 4000 (*Diphyllobothrium latum*, which may reach an astounding 15 m). As the proglottids mature, they develop male and female sex organs,

with cross-fertilization generally occurring between adjacent segments. After mating, the proglottids become gravid with hundreds to thousands of eggs and detach from the tapeworm, passing out of the host in the feces. In *Taenia* species tapeworms, the eggs may then be consumed by an intermediate host (e.g., a pig or cow), where they hatch, releasing a structure called the oncosphere, which penetrates the tissues and develops into a cysticercus. Consumption of cysticerci-contaminated meat (in raw or undercooked form) by the definitive human host allows the larval tapeworm to be released in the intestine, where it develops into the adult parasite, measuring up to 10 m in length. *D. latum* are acquired by humans through the ingestion of an infected fish, which in many cases is the *third* intermediate host (the first being a copepod, which is ingested by a minnow, which in turn is ingested by a larger fish). *H. nana* tapeworms are unique in that adult infections may result from the ingestion of the intermediate host (in this case, a flea or beetle) or more commonly through direct consumption of eggs derived from human feces. Eggs of *E. granulosus* and *Taenia solium* may also be infectious to humans, leading to severe pathology (see below).

The vast majority of adult tapeworm infections cause little, if any, clinical disease, although passage of proglottids (which are often capable of caterpillar-like movement outside the host) may be quite distressing. Much more serious disease is associated with ingestion of the eggs of *E. granulosus* or *T. solium*, in effect making the human an "accidental" intermediate host for disseminated larval cysts. In the case of *E. granulosus*, ingestion of eggs shed in the feces of dogs (the definitive host) leads to large fluid-filled "hydatid" cysts in the liver, lungs, brain, and bone marrow. *T. solium* eggs are acquired from the feces of humans infected with adult tapeworms. The cysticerci (cysts) associated with *T. solium* are somewhat smaller than those of *Echinococcus* but are more likely to localize to the brain; neurocysticercosis may cause various neurological symptoms such as seizures, hydrocephalus, and visual disturbances. *T. solium* cysticerci may also cause damage to the eye and skeletal muscle.

See Also the Following Articles

Cestodes • Nematodes • Parasitic Diseases, Overview • Trematodes • *Trichinella*

Further Reading

Beaver, C. B., Jung, R. S., and Cupp, E. W. (1984). "Clinical Parasitology," 9th Ed. Lea & Febiger, Philadelphia, PA.

Despommier, D. D., Gwadz, R. W., Hotez, P. J., and Knirsch, C. A. (2000). "Parasitic Diseases," 4th Ed. Apple Trees Productions, New York.

Schmidt, G. D., and Roberts, L. S. (1989). "Foundations of Parasitology," 4th Ed. Times Mirror/Mosby College Publishing, St. Louis, MO.

Sun, T. (1999). "Parasitic Disorders," 2nd Ed. Williams & Wilkins, Baltimore, MD.

Hemobilia

SUMITA VERMA AND MICHEL MENDLER
University of Southern California, Los Angeles

endoscopic retrograde cholangiopancreaticography A procedure in which a fiber-optic endoscope is inserted into the duodenum and dye is injected via the ampulla of Vater to visualize the biliary and pancreatic ducts.

hemobilia Bleeding into the intra- or extrahepatic biliary system.

percutaneous transhepatic cholangiography Procedure during which a needle is inserted into the liver via the skin and into a bile duct. Then dye is injected, outlining the biliary system. Through the needle a catheter can be placed to drain the biliary tract, which is known as percutaneous transhepatic biliary drainage.

transarterial embolization Catheterization of the hepatic artery via the femoral artery and using various modalities to selectively embolize the bleeding vessel.

Hemobilia occurs when there is a communication between a vessel of the splanchnic circulation and the intrahepatic or extrahepatic biliary system. It is a rare entity, but nonetheless is an important differential diagnosis of gastrointestinal bleeding. It may occur in a variety of settings and its clinical course may vary from absence of symptoms to life-threatening hemorrhage.

INTRODUCTION

"I believe that if the liver is injured by a contusion, it may lead to blood leaving the body by way of vomit or the stool; for there is no doubt that the biliary duct takes unto itself (to the great good of the patient) some of the blood issuing into the liver and leads it down to the intestines; from there it is either impelled upwards through reverse peristalsis or downwards the usual way." This is the first description of hemobilia found in the medical literature and it appeared in Francis Glisson's *Anatomia Hepatis* in 1694. It was based on the clinical course of a young nobleman who had a stab injury in the right upper quadrant and died following upper gastrointestinal bleeding and on postmortem was found to have significant liver injury. In 1777, Antoine Portal recognized the first antemortem case of hemobilia. Subsequently, Quincke described the classical clinical triad of hemobilia: right upper quadrant pain, jaundice, and upper gastrointestinal tract bleeding. However, it was not until 1948 that Philip Sandblom coined the word hemobilia, while describing a few cases of traumatic hemorrhage through the biliary tract. The first successful operation for hemobilia was carried out in 1903 by Kehr when he identified and ligated an aneurysm of the right hepatic artery that had ruptured into the gallbladder neck. The first successful hepatic resection for hemobilia was reported in 1957. Finally, in 1976 Walters and colleagues used transarterial embolization (TAE) for the management of hemobilia.

ETIOLOGY AND PATHOGENESIS

Three reviews on this topic, by Sandblom, Yoshida *et al.*, and Green *et al.*, provide detailed descriptions of the various etiologies of this condition as well as the changing trends in prevalence over the past three decades (Table I). In the earliest review reported by Sandblom, the commonest cause for hemobilia was accidental trauma (mostly road traffic accidents). In contrast, the most recent series suggests that iatrogenic trauma is the most frequent cause. This is mostly due to the

TABLE I Causes of Hemobilia

Cause	Sandblom (1973) ($n = 545$)	Yoshida *et al.* (1987) ($n = 103$)	Green *et al.* (2001) ($n = 222$)
Iatrogenic trauma	15%	40%	66%
Accidental trauma	33%	19%	6%
Inflammation	28%	9.7%	6.3%
Gallbladder disease	11%	8.7%	5.8%
Vascular	7%	14.6%	9%
Tumors	5%	6.8%	5%
Others	1%	—	—

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Iatrogenic Risk of Hemobilia

Procedure	Risk of hemobilia
Percutaneous liver biopsy	0.006–1%
Laproscopic cholecystectomy	0.5%
Stenting of biliary stricture	2–10%
Percutaneous transhepatic cholangiography	4–10%
Stenting of biliary stricture	5–10%
Percutaneous transhepatic biliary drainage	3–14%

dramatic increase in the number of percutaneous liver procedures. The risk is related to the size of the needle used. Table II summarizes the incidence of hemobilia after various procedures. Although percutaneous liver biopsy (PLB) is the commonest procedure associated with hemobilia, the overall risk is extremely low (Table I). Ultrasound-guided biopsies or use of the transjugular route does not reduce the risk. In a review of 25 cases of post-PLB-related hemobilia, 10 of the 25 patients had more than one needle pass, raising the possibility that multiple passes is a significant risk factor. Percutaneous transhepatic cholangiography (PTC) also carries risk of hemobilia because of the close relationship between the bile ducts and the blood vessels. If indeed hemobilia does result after PTC, clots may form in the bile ducts and this may cause diagnostic problems as the clots can often be mistaken for stones. If percutaneous biliary drainage is performed, the risk of hemobilia can be even greater (Table II). Other iatrogenic causes include gallbladder surgery. This can result in damage to the hepatic artery due to sutures or dissection, which may result in an arteriobilary fistula or a false aneurysm that erodes into the biliary system. The vessel most often damaged is the hepatic artery or one of its branches, especially the right artery when it lies anterior to the common bile duct. The risk is increased by intraoperative explorations of the bile ducts. In 20% of ducts explored for retained stones following cholecystectomy, a clot rather than a stone was found to be the cause of the filling defect seen on the operating cholangiogram. Despite some initial concerns, there is no evidence to suggest that the incidence of hemobilia is higher after laparoscopic cholecystectomy than after conventional open cholecystectomy.

The other major cause of hemobilia remains accidental trauma (although the number of cases has declined dramatically over the years). This usually follows blunt trauma in over 80% of cases, although it can also occur after abdominal stab injuries and gunshot wounds. Approximately 2% of liver injuries result in hemobilia. The commonest extrahepatic cause is laceration or bruising of the hepatic artery at the porta hepatis. This then leads to hemobilia by one of two mechanisms: intrahepatic hematoma or arteriobiliary fistula. In the former, a cavity filled with blood and bile develops inside the liver, which enlarges, and eventually blood is evacuated into the biliary system and the gastrointestinal tract. The fact that bile retards wound healing further facilitates this mechanism. The second mechanism is by formation of a false aneurysm, which ruptures directly into a bile duct. Such aneurysms develop especially on the branches of the hepatic artery in the hilar or juxtahilar region.

Less frequent but noteworthy causes of hemobilia include gallbladder disease, cholelithiasis (erosion of the cystic artery by an impacted stone), acalculous cholecystitis, and carcinoma. Liver tumors, both primary (hepatocellular cancer, cholangiocarcinomas) and metastatic deposits, can also result in hemobilia, as can true (mycotic) aneurysms of the hepatic artery. Finally, in Asian countries such as China and Korea, hemobilia is a frequent and serious complication of ascariasis and hydatid infection.

The anatomic origin of hemobilia has been studied in detail by many investigators, including Sandblom, Yoshida *et al.*, and Merrell *et al.* Their findings are strikingly consistent. In over 50% of cases the blood originates from the liver, in approximately 20% of cases it originates from the extrahepatic ducts, in 9% of cases it originates from the gallbladder, and in approximately 2% of cases it originates from the pancreas. The bleeding is usually arterial in origin; a venobiliary communication is exceptional (the latter can occur in patients with portal hypertension especially after a transjugular portosystemic shunt placement). The fate of the blood in the bile tract varies. In cases of severe bleeding, the blood directly passes to the gastrointestinal tract and manifests as hematemesis or melena. If the hemorrhage is slow, clots can form as blood and bile do not mix due to differences in specific gravities and surface tension. Since bile is thrombolytic, if bile flow is ensured, the clots are lysed. However, in the presence of impaired bile flow, the clots persist and cause biliary obstruction, which could lead to the development of cholangitis, cholecystitis, and pancreatitis. In addition, clots that remain in the bile ducts could serve as a nidus for subsequent stone formation.

CLINICAL PRESENTATION

The clinical manifestations are protean, related to the rate of blood loss and etiology of hemobilia. Fortunately, the frequency of major hemobilia is considerably less than that of minor hemobilia. Quincke first described the classical triad of symptoms of hemobilia: right upper

quadrant pain, jaundice, and hematemesis. However, all three symptoms are present in only 22% of patients (especially if the etiology is traumatic) (Fig. 1). In approximately one-third of patients, the only symptom is gastrointestinal bleeding, and hemobilia must be remembered as an important differential diagnosis. Bleeding may be scanty or profuse, prolonged or brief, and continuous or intermittent. If bleeding is significant, it may mimic a lower gastrointestinal bleed. Slow blood loss with anemia is unusual, but three decades ago fecal occult blood testing was a diagnostic test for gallstone-induced biliary colic. In addition, patients may present with cholangitis or pancreatitis. It is important to remember that after accidental injury, there may be a delay before patients develop symptoms. The onset of bleeding may be delayed for 4 days to 5 months (average 4 weeks). Rarely the bleeding may be recurrent, with episodes of abdominal pain relieved by gastrointestinal bleeding. In contrast, after iatrogenic trauma, hemobilia manifests soon after the procedure (mean of 5 days after PLB).

Hemobilia occurs more frequently in men, except that bleeding caused by gallstones occurs more commonly in women. The age distribution is fairly uniform, from 10 to 70 years of age, with accidental trauma occurring within the first four decades and iatrogenic trauma being concentrated in the fifth to sixth decade. Laboratory tests may be normal or reveal anemia, moderate to marked elevations in serum alkaline phosphatase, and mild to moderate elevations in serum bilirubin and transaminases.

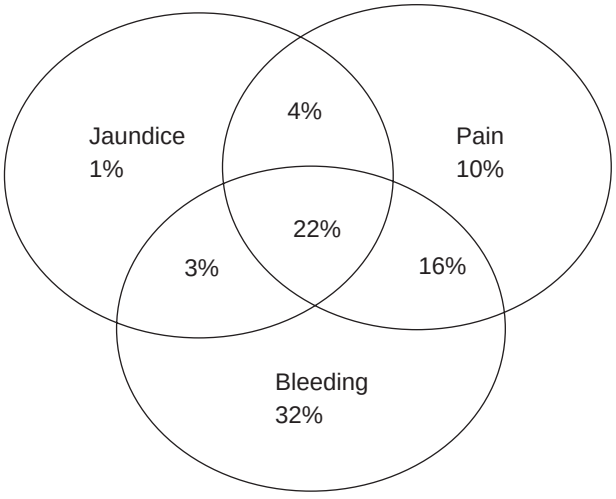


FIGURE 1 Presentation of symptoms in 121 patients with hemobilia. Reprinted from Green *et al.* (2001), with permission from Blackwell Synergy.

INVESTIGATIONS

Early diagnosis is important because the mortality is directly proportional to the delay in controlling the hemorrhage. The definitive diagnosis is established by angiography, which is the investigation of choice, though most patients usually undergo an endoscopy and/or an imaging modality beforehand. The exact sequence of investigations must be individualized depending upon the severity and etiology of the bleeding. In cases of non-life-threatening hemobilia, a diagnostic endoscopy may be reasonable followed by conservative management. Although only 12% of cases are initially diagnosed with endoscopy, it may confirm the diagnosis in an additional 30% of cases and help exclude other causes of gastrointestinal bleeding. In cases of suspected traumatic liver injury, a computerized tomography (CT) scan is useful. Although it cannot demonstrate the actual site of bleeding, and the CT grade of a liver injury alone is not a good predictor of development of hemobilia in the future, it may provide indirect evidence of hemobilia by demonstrating cavitating lesions, aneurysms, and pseudo-aneurysms. Ultrasound is probably less sensitive than a CT scan, as immediately after a bleed into the biliary tract, there may be high echogenicity similar to that of the liver, so that the bile ducts may not be seen. Cholangiography is useful if there is percutaneous access or if an ERCP has been performed. Blood may be seen oozing from the ampulla of Vater and contrast studies may show filling defects in the bile ducts. These defects may be string-like or spherical and impossible to differentiate from stones. In addition, parasitic infections of the biliary tree may be confirmed.

Angiography is the gold standard in most cases of significant bleeding. Selective arteriography of the celiac axis and the superior mesenteric artery is the ideal procedure. In addition to providing information about the exact site of bleeding, it may reveal aneurysms (both true and false) and damage to the hilar arteries. Occasionally an arteriobiliary fistula is diagnosed. In the rare case where there is a high index of suspicion for hemobilia and all investigations including angiography are normal, there may be a role for ^{99m}Tc-RBC scanning.

TREATMENT

Management depends entirely on the severity of bleeding. If the bleeding is not significant, a conservative approach is ideal. In cases of severe bleeding, the aim is to stop the bleeding and prevent/relieve biliary obstruction. Transarterial embolization is now the treatment of choice to stop bleeding in hemobilia. Embolic

occlusion is achieved by various means: balloons, microcoils, gelfoam, or cyanoacrylate. The success rate of TAE varies between 80 and 100%. It has a lower morbidity and mortality than surgery. The most serious complication of TAE is ischemia or infarction of the liver (uncommon due to dual blood supply of the liver) and inadvertent embolization of unintended organs such as the gallbladder and the pancreas. Less serious side effects include pain, fever, bacteremia, and transaminitis. As much as possible, the TAE should be superselective both to reduce the risk of ischemic liver injury and to prevent recurrent bleeding that might develop from collateral vascular channels.

If there is evidence of biliary obstruction, then drainage must be ensured either via an ERCP or through the percutaneous route. Surgery is indicated if TAE fails, for cholecystitis, for hepatic cavitating lesions, and occasionally for biliary drainage. During surgery, the aim is to ligate the offending bleeding vessel, and when this is not possible, nonselective ligation of the right or left hepatic artery and/or arterial reconstruction must be performed. If the bleeding persists, more drastic measures, such as resection of the affected liver segment or partial hepatectomy, are indicated. Other adjunctive modalities used in the management of hemobilia include pharmacological agents such as vasopressin and somatostatin.

In a recent review of 171 cases of hemobilia between 1996 and 1999, Green *et al*. reported that a conservative approach (including correcting coagulopathy and adequate drainage) successfully controlled bleeding in 43% of cases of hemobilia, TAE controlled bleeding in 36% of cases, and surgery was required in 22% of cases (including cholecystectomy in 12% and hepatic artery ligation in 2% of patients in whom TAE was unsuccessful).

PROGNOSIS

In Sandblom's initial review of hemobilia in 1972, the overall mortality was 25%, but by the late 1980s the mortality rate had been reduced by half. From 1996 to 1999, only five deaths were reported from hemobilia. This reduction in mortality rate over the past 30 years is due to multiple factors: heightened awareness, an increasing number of cases of mild hemobilia due to iatrogenic causes, and finally increasing use of and refinement in the technique of TAE.

CONCLUSIONS

Hemobilia is the phenomenon of bleeding into the biliary tree, which in two-thirds of cases is related to iatrogenic trauma, notably percutaneous liver procedures. The classical clinical presentation is with right upper abdominal pain, jaundice, and gastrointestinal bleeding. Diagnosis is established by angiography and TAE is the treatment of choice, being successful in 80–100% of the cases. Both increasing awareness of this condition and refinement in TAE techniques have contributed to the dramatic decline in the mortality resulting from this condition over the past three decades.

See Also the Following Articles

Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Percutaneous Transhepatic Cholangiography (PTC) • Trauma, Overview • Upper Gastrointestinal Bleeding

Further Reading

Bismuth, H. (1973). Current concepts: Hemobilia. *New Engl. J. Med.* **288**, 617–619.

Green, M. H. A., Duell, R. M., Joghson, C. D., and Jamieson, N. V. (2001). Hemobilia (Review). *Br. J. Surg.* **88**, 773–786.

Lichtenstein, D. R., Kim, D., and Chopra, S. (1992). Delayed massive hemobilia following percutaneous liver biopsy: Treatment by embolotherapy. *Am. J. Gastroenterol.* **87**, 1833–1837.

Merrell, S. W., and Scheider, P. D. (1991). Hemobilia: Evolution of current diagnosis and management. *West J. Med.* **155**, 621–625.

Sandblom, P. (1948). Hemorrhage into the tract following trauma: Traumatic hemobilia. *Surgery* **24**, 571–586.

Sandblom, P. (1973). Hemobilia. *Surg. Clin. North Am.* **53**, 1191–1200.

Yoshida, J., Donahue, P. E., and Nyhus, L. M. (1987). Hemobilia: Review of recent experience with a world wide problem. *Am. J. Gastroenterol.* **82**, 448–453.

Hemorrhage

DOUGLAS W. WILMORE
Brigham and Women's Hospital and Harvard University

coagulation The natural process by which blood clots to reduce or prevent bleeding.
endoscopy The process of examining the lining of the intestinal tract using a flexible device that can easily be inserted into the mouth or anus and which allows magnification and viewing of the lumen.
hemorrhage The loss of blood from blood vessels.
sclerotherapy The injection of a highly irritating substance into a blood vessel, which causes a clot to form.
sympathetic nervous system A portion of the nervous system that responds automatically to stresses on the body, such as exercise, temperature changes, and hemorrhage.
vasoconstriction The process of blood vessels closing down or constricting, to limit the amount of blood flowing through them.

Hemorrhage or bleeding is the loss of blood from blood vessels that constitute the vascular compartment. The vascular compartment is made up of arteries, veins, capillaries, and the heart, and the size of this space is approximately 7–8% of the individual's body weight (in kilograms). Thus, a 70 kg male would have approximately 5–5.5 liters of blood in the body (70 kg $\times$ 0.07 = $\sim$5.0 liters and 70 kg $\times$ 0.08 = $\sim$5.5 liters). Blood is composed of plasma, red blood cells, and white cells or leukocytes. The blood plasma (the liquid phase of blood) carries coagulation factors that facilitate blood clotting mechanisms. The red blood cells, or erythrocytes, constitute 30–45% of the blood volume and primarily carry oxygen from the lungs to tissues of the body. During acute hemorrhage, all blood components are lost in equal proportions, but during a slow occult bleed, the plasma component is rapidly replaced by the body and the primary deficit is the red blood cells.

PHYSIOLOGY

Blood Pressure

The circulating blood is propelled by the pumping action of the heart. The force of the heart and the resistance created by small arterial vessels, called arterioles, combine to cause pressure within the arterial tree (which is measured as blood pressure). This head of pressure ensures perfusion of vital organs, such as the brain and kidneys, and of the coronary arteries of the heart. As bleeding occurs, the pressure within the vascular compartment is initially maintained as the volume of blood flowing through large capacitance veins is reduced. These vessels include the superior and inferior vena cava, the pulmonary veins, and the large veins draining the extremities. With additional loss of blood, blood pressure falls, oxygen delivery to vital organs is impaired, and tissue damage or cell death may occur. This stage of bleeding is referred to as hemorrhagic shock.

Response to Bleeding

During the initial phases of acute blood loss, little initial change in blood pressure occurs. In fact, a healthy individual can generally withstand loss of 20–25% of his or her blood volume (1.0–1.5 liters) without experiencing significant symptoms (Table I). With greater degrees of acute hemorrhage, neurohormonal reflex responses are activated, in an attempt to maintain blood pressure and ensure tissue perfusion. These responses include activation of the sympathetic nervous system, which stimulates vasoconstriction of peripheral vessels and activation of the pituitary gland. This last response results in the elaboration of hormones that enhance contraction of the smooth muscles of the vascular tree, thus enhancing vasoconstriction, and stimulating salt and water retention by the kidney. The elaboration of the stress hormone epinephrine from the adrenal gland stimulates, among other things, the heart, causing tachycardia and increasing its pumping action. All of the autonomic responses reflect the body's attempt to maintain blood volume and support blood pressure, which reflects perfusion of vital tissue.

Signs Related to the Extent of Blood Loss

As acute hemorrhage progresses, the blood pressure falls and tissue perfusion diminishes. The skin becomes cool and clammy. Little urine is formed, reflecting diminished renal perfusion, and the patient may become quite confused and agitated as the oxygen supplied to the brain is reduced. Rapid restoration of blood pressure

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Signs and Symptoms Associated with Varying Degrees of Hemorrhage

Mild (20% blood loss)	Moderate (20–40% blood loss)	Severe (40% blood loss)
Pallor	Pallor	Pallor
Cool–cold extremities	Cool–cold extremities	Cool–cold extremities
Sweating	Sweating	Sweating
Collapsed subcutaneous veins	Collapsed subcutaneous veins	Collapsed subcutaneous veins
	Tachycardia	Tachycardia
	Low urine output	Low urine output
	Postural hypotension	Hypotension
	Confusion	Mental status changes

is essential at this point in order to prevent permanent organ damage or death.

Treatment

The treatment of hemorrhage is twofold: (1) stop the bleeding at its site and (2) replace the blood that has been lost. Direct pressure on a superficial vessel or a large vessel in extremity is an efficient method to employ while the patient is being transported to a medical facility for definitive treatment. A variation of this theme is followed by many emergency medical units who apply special inflatable pressure pants to the lower torso in patients with injuries to and blood loss from the lower extremities. In contrast, internal bleeding represents much more of a diagnostic and therapeutic challenge, as discussed below.

Replacement of the lost blood requires insertion of an intravenous line, usually in a vein in the extremity, followed by the immediate administration of an appropriate volume of a salt-containing intravenous solution. This solution serves as a plasma mimic and can be readily stored at room temperature in ambulances and emergency rooms. Blood pressure and urine volume are monitored to ensure adequacy of therapy. If a large volume of blood is lost, the patient's blood type is determined at the hospital and type-specific red blood cells (or occasionally whole blood or blood components, such as platelets or fresh plasma) are administered to ensure the adequacy of the oxygen-carrying capacity of the blood or to enhance coagulation mechanisms. With stabilization, further diagnostic and therapeutic methods are pursued so that a definitive treatment plan to permanently arrest the hemorrhage can be followed.

CAUSES OF HEMORRHAGE

Trauma

One obvious cause of hemorrhage is tissue injury. Falls, scrapes, and other accidents, particularly those that occur while traveling in high-speed automobiles, account for a variety of traumatic injuries associated with bleeding. The hemorrhage may be external and quite apparent or may be internal, such as that associated with a fracture of a long bone, which results in bleeding into the surrounding soft tissue. Internal bleeding may also occur in the thoracic or abdominal cavity. With blunt abdominal injury (as occurs following a fall or seat belt injury), organs such as the liver or spleen may be damaged, causing intra-abdominal hemorrhage. To aid diagnosis, a small tube is inserted into the abdomen and sterile fluid is injected and then withdrawn. The fluid is then examined for red blood cells and if they are present, bleeding has occurred. In addition to this technique of peritoneal gavage, modern scanning techniques are utilized to determine sites of tissue injury and to plan therapy.

Other causes of intracavitary hemorrhage are associated with penetrating injuries, such as those that occur when an individual falls on a pipe or rod. More common are the injuries sustained during altercations involving knives or guns. These injuries frequently require surgical exploration to stay the bleeding and repair damaged tissue.

Upper Gastrointestinal Tract Hemorrhage

Patients frequently present with bloody vomitus, which is a reflection of disease in the upper gastrointestinal tract. Fortunately, in approximately 80% of patients who present with this type of bleeding, the hemorrhage is self-limiting and diagnostic measures can proceed in a nonemergency fashion after appropriate supportive therapy has been provided. In the remaining patients, combined emergency approaches involve the gastroenterologist, who performs endoscopy, the radiologist, who performs additional diagnostic studies (and on occasion can embolize bleeding vessels), and the surgeon, who may be called upon to perform a life-saving operation.

The causes of upper gastrointestinal hemorrhage include esophageal variceal bleeding, peptic ulcer disease, tears of the lower esophagus secondary to severe

TABLE II Causes of Gastrointestinal Bleeding

Cause	% of total
Upper gastrointestinal bleeding	
Peptic ulcer	61
Mucosal erosive disease	15
Esophageal varices	6
Mallory–Weiss tear	3
Malignancy	2
Other	13
Lower gastrointestinal bleeding	
Diverticulosis	42
Malignancy	9
Ischemic colitis	9
Acute colitis	5
Hemorrhoids	5
Angiodysplasia	3
Crohn's disease	2
Other	25

vomiting (an entity called Mallory–Weiss syndrome), vascular malformations, erosive gastrointestinal mucosal lesions (caused by alcohol and drugs such as aspirin), and tumors (Table II). Many of these localized lesions can be diagnosed and treated by endoscopy and electrical coagulation or sclerotherapy, and other lesions (including tears and tumors) require operative repair.

Lower Gastrointestinal Hemorrhage

Patients frequently present with blood in their stools. Alternatively, a physician may detect a low red blood cell count and diagnose slow occult bleeding from the colon. Diagnosis of lower gastrointestinal bleeding is more difficult than the detection of upper intestinal bleeding because of the presence of fecal material in the colon, which requires removal before endoscopic and radiographic testing can satisfactorily be performed.

The causes of bleeding from the lower gastrointestinal tract are multiple and include diverticulosis, vascular anomalies, inflammatory or immune lesions of the lower tract, such as Crohn's disease or ulcerative colitis, benign or malignant tumors, or congenital anomalies, such as a Meckel's diverticulum (Table II). If bleeding is rapid, blood replacement therapy must be provided and diagnostic measures initiated. Endoscopy and arteriographic studies are the most frequent examinations performed and on occasion, in the unstable patient, emergency operative intervention is required. If no source of bleeding is discovered, a radionucleotide scan, barium studies, or combined operative and endoscopic investigations are the next usual diagnostic steps.

With slower occult bleeding, bowel preparation with the use of laxatives, enemas, and liquid diets greatly facilitates diagnostic studies of the colon and rectum. Bleeding also occurs from external hemorrhoids and anal fissures, which can be treated by local means.

Other Sites of Hemorrhage

Hemorrhage also occurs from other organs, such as the bladder, uterus, and lungs, as well as into the brain. Appropriate diagnostic studies are necessary to determine appropriate therapy for bleeding from these organs. Although hemorrhage from these or other sites is usually associated with local tissue injury or disease, patients who have abnormalities of blood coagulation mechanisms also present with episodes of bleeding from these and other organs. This is particularly true in individuals taking anticoagulant drugs (aspirin, heparin, or coumadin) or those with inherited abnormalities in coagulation.

See Also the Following Articles

Lower Gastrointestinal Bleeding and Severe Hematochezia • Occult Gastrointestinal Bleeding • Sympathetic Innervation • Trauma, Overview • Upper Gastrointestinal Bleeding

Further Reading

Lichtenstein, D. R. (2000). Management of nonvariceal upper gastrointestinal hemorrhage. *In* "Therapy of Digestive Disorders" (M. M. Wolfe, ed.), pp. 127–152. W. B. Saunders, Philadelphia, PA.

Moore, F. A., and Moore, E. E. (2002). Trauma resuscitation. *In* "ACS Surgery: Principles and Practice" (D. W. Wilmore, *et al.*, eds.), pp. 31–48. Web MD, New York.

Slivka, A., and Schoen, R. E. (2000). Lower gastrointestinal bleeding. *In* "Therapy of Digestive Disorders" (M. M. Wolfe, ed.), pp. 682–700. W. B. Saunders, Philadelphia, PA.

Hemorrhoids

SARA HUGHES JAVID AND STEPHEN JOHN FERZOCO
Brigham and Women's Hospital, Boston

external hemorrhoids Those hemorrhoids that originate below the dentate line.

hemorrhoid Symptomatic and pathologic, downward displacement of normal vascular tissue of the anal canal.

internal hemorrhoids Those hemorrhoids that originate above the dentate line. Differentiated as first through fourth degree: first degree, slide below the dentate line on straining; second degree, prolapse through the anus on straining but reduce spontaneously; third degree, prolapse through the anus on straining and require manual replacement into the anal canal; fourth degree, prolapse is not manually reducible.

Hemorrhoids constitute a condition that has been recognized and treated for over 4000 years. The word "hemorrhoid" is derived from the Greek adjective "haimorrhoides," which means bleeding . Hemorrhoids originate in the anal canal as cushions of submucosal vascular tissue that can undergo pathologic changes, producing symptoms. Hemorrhoids can be classified by their location relative to the dentate line, the junction between the rectal mucosa and the specialized epithelium of the anal canal (anoderm). Hemorrhoids that originate below the level of the dentate line are classified as external hemorrhoids, and those that originate above the dentate line are termed internal hemorrhoids. Internal hemorrhoids can be further described as stage I, II, III, or IV.

INTRODUCTION

Hemorrhoids are one of the most common anorectal complaints brought to physicians. Symptomatic hemorrhoids affect approximately 42 per 1000 persons, or 10.6 million persons per year in the United States. Of those, only 2.5 million people present to a physician for treatment. The incidence of hemorrhoids appears to peak in middle age and declines after the age of 65. An equal distribution of hemorrhoid disease is observed between the sexes.

Hemorrhoids are composed of submucosal cushions of vascular tissue found in the left lateral, right anterior, and right posterior positions of the anal wall that are devoid of muscle. Each cushion has a venous and arterial supply, as well as multiple small arteriovenous anastomoses between the two. The blood supply is derived from the superior rectal artery arising from the inferior mesenteric artery, the middle rectal arteries arising from the internal iliac arteries, and the inferior rectal arteries arising from the pudendal arteries. The venous drainage is to the portal venous system above the dentate line and to the systemic circulation below the dentate line.

The function of the hemorrhoidal cushions is to aid in anal continence. The pathologic, downward displacement of these cushions is probably caused by factors such as constipation, prolonged straining, pregnancy, aging, and increased intraabdominal pressure. This displacement is thought to result in the distension of cushion venules, which causes symptomatic hemorrhoids.

Treatment of symptomatic hemorrhoids typically begins with conservative measures, including the moderation of diet. If unsuccessful, however, more aggressive nonoperative therapies can be utilized for both external and internal hemorrhoids. Operative strategies are indicated in circumstances of nonoperative treatment failure or on presentation with acute thrombosis or prolapse of hemorrhoids.

The various clinical presentations and diagnosis of external and internal hemorrhoids are addressed in the following discussions. In addition, available therapeutic options for patients are reviewed and compared. Last, the various complications that may ensue as a result of hemorrhoid therapy are discussed.

CLINICAL PRESENTATION AND DIAGNOSIS

The most common complaint of patients suffering from hemorrhoids is painless rectal bleeding or bright red blood associated with bowel movements. Patients may also present complaining of a "bulge" from the anus, mucus discharge, or a feeling of incomplete evacuation. Pain results when hemorrhoids undergo thrombosis, incarceration, or acute prolapse with edema. As a general rule, internal hemorrhoids are typically nonpainful and present characteristically with rectal bleeding. External hemorrhoids are more apt to present as a mass protruding from the anus or irritation with constipation, but may also cause bleeding.

 Copyright 2004, Elsevier (USA). All rights reserved.

Examination of the patient using proper positioning and instruments is essential to diagnose hemorrhoids. The patient should be placed in the left lateral or prone jackknife position or in a modified left lateral (Sims') position. Inspection, palpation, digital rectal examination, and anoscopy should be performed on all patients. Internal hemorrhoids cannot be palpated, but prolapse can best be observed if the patient is asked to strain on the toilet. Anoscopy also permits clear visualization of the internal hemorrhoidal cushions. In addition, either a flexible sigmoidoscopy or a rigid procotosigmoidoscopy should be done to exclude coexisting colorectal pathology. A colonoscopy is indicated for those patients over age 40 or those patients with additional symptoms or a positive family history. Alternative etiologies for the patient's symptoms should be excluded, such as anal fissure, abscess, fistula, polyps, malignancy, inflammatory bowel disease, skin tags, anal warts, or rectal prolapse.

NONOPERATIVE MANAGEMENT

A multitude of treatment options for hemorrhoids exists, many of which have been in existence for thousands of years. These modalities include ligation, cautery, surgical excision, and possibly even the intervention of St. Fiacre, the patron saint of hemorrhoid sufferers. Treatment is reserved for symptomatic hemorrhoids only, and amelioration of these symptoms and cure are the aims to be sought. With the exception of fiber agents, hemorrhoid therapies work by fixing the anal cushions to their normal physiologic sites.

The first-line treatment of hemorrhoid disease usually begins with bulking agents (e.g., psyllium or methylcellulose) and increased water intake. These dietary modifications are thought to alleviate symptoms by softening bowel movements and reducing strain on defecation. One controlled trial has compared fiber (psyllium) with placebo in 52 patients with symptomatic hemorrhoids. These patients experienced a significant 36% reduction in bleeding when compared with placebo. Pain with defecation was also reduced. Warm sitz baths (60°C) are another effective local treatment used to alleviate symptomatic hemorrhoids. Topical creams, ointments, or foams have been shown to have minimal efficacy in randomized clinical trials (RCTs), but may reduce symptoms by exerting a local anesthetic effect.

If bleeding or other symptoms persist, other more aggressive therapies are utilized. The most commonly used nonsurgical method for treatment of first-, second-, and third-degree internal hemorrhoids is rubber band ligation (RBL). In this technique, a hemorrhoid ligator (e.g., McGown or McGivney ligator) is used to apply a rubber band to the base of the hemorrhoid cushion. This leads to strangulation, small ulcer formation, and the eventual fixation of the tissue to the underlying sphincter. For this procedure, the patient should be placed in the prone jackknife position. Typically, one band is placed on each hemorrhoid bundle, and the largest is banded first. Banding in close approximation to the dentate line or internal sphincter should be avoided. Following the procedure, patients may resume normal activities, but should be made aware of potential complications, including pain, thrombosis, delayed bleeding, or infection.

Infrared photocoagulation (IRC) is an alternative treatment for first- and second-degree hemorrhoids. Through the application of heat via infrared radiation, the hemorrhoid cushion becomes inflamed and scars, leading to its fixation. The infrared coagulator is applied three or four times near the apex of the hemorrhoid. Patients should be aware of the possible complications of pain or fissure due to inappropriate positioning of the coagulator at the dentate line.

The oldest technique in use is injection sclerotherapy, in which a sclerosing material (e.g., phenol, 5% in cottonseed oil) is injected into the submucosa above the hemorrhoid, resulting in fibrosis and fixation of the cushion. This technique is rarely used today due to high rates of stricture formation and other complications.

Although the efficacy of these techniques has never been compared to placebo in a controlled trial, there have been several small RCTs comparing them to one another. Ultimately, none has been proved superior to the others. However, one meta-analysis of five randomized trials revealed that RBL and IRC had similar efficacy rates, but sclerotherapy was inferior. Although patients with RBL had fewer symptomatic recurrences, they also experienced a fivefold greater incidence of pain compared to patients who had IRC. Therefore, the choice between RBL and IRC should be individualized based on patient and provider preference.

OPERATIVE MANAGEMENT

Surgical treatment of hemorrhoids should be reserved for those patients for whom nonoperative management fails or in cases of acute emergency. Only 5–10% of patients with symptomatic hemorrhoids require surgery. Although it has been shown in one meta-analysis comparing hemorrhoidectomy with RBL that hemorrhoidectomy was more effective, it was also associated with a significantly greater risk of complications, postprocedure pain, and cost. The most common indication for hemorrhoidectomy is the frequent, symptomatic prolapse of internal hemorrhoids leading to pain and/or anal seepage. Other indications include

difficulties in anal hygiene secondary to prolapsing internal or external hemorrhoids, or acute prolapse, thrombosis, or strangulation of hemorrhoids. Several options are available for excisional hemorrhoidectomy, including Milligan–Morgan hemorrhoidectomy, Ferguson closed hemorrhoidectomy, and Whitehead hemorrhoidectomy.

Each hemorrhoidectomy procedure is performed with the patient in the prone jackknife position with buttocks taped laterally. In the Milligan–Morgan hemorrhoidectomy, first described in 1937, the entire involved hemorrhoid complex is ligated at its base and the distal anoderm and skin are left open to heal by secondary intention. Although proved a safe and efficacious method, one obvious disadvantage to this technique is discomfort from the open wound.

In the closed Ferguson hemorrhoidectomy, an hourglass-shaped incision is made encompassing the entire hemorrhoidal complex, taking care to preserve the anal sphincters. In this technique, the entire wound is closed primarily, and the anoderm between hemorrhoid pedicles is preserved. The Whitehead hemorrhoidectomy is more technically demanding because, in addition to removing hemorrhoidal tissue, the prolapsed dentate line must be relocated to its proper position. This technique is less frequently performed because some studies have shown high complication rates of mucosal ectropion and anal stricture.

A newer technique currently used is the circular stapled hemorrhoidectomy. This procedure can be used for patients with all grades of internal hemorrhoids, but is best reserved for patients with grade III or IV hemorrhoids. It is not applicable to patients with external hemorrhoids. First, a pursestring suture is placed 4–6 cm above the dentate line, above the enlarged internal hemorrhoids. A 31-mm stapler is then used to perform a circumferential mucosectomy at this position, leading to the repositioning and fixation of the prolapsed anoderm on closure of the defect. Thus, the hemorrhoidal cushions are preserved and returned to their physiologic location. Although there are no data yet documenting the long-term results of stapled hemorrhoidectomy, preliminary studies of this approach show a decreased frequency of postprocedure pain, shorter recovery time, and similar complication rates compared to other types of hemorrhoidectomy.

Other forms of hemorrhoidectomy, such as laser hemorrhoidectomy, not only add to cost and cause increased pain, but also provide no advantage over traditional hemorrhoidectomy approaches. Anal stretch and lateral sphincterotomy methods have also been abandoned for the treatment of hemorrhoids due to the high associated risk of anal incontinence.

COMPLICATIONS OF MANAGEMENT

Pain is the most frequent complication of hemorrhoid surgery. The key to avoiding this complication lies in the usage of a combination of oral and intravenous narcotics as well as local analgesics such as bupivacaine in the perianal incision. Nonsteroidal antiinflammatory medications, including ketorolac, have also been shown effective in managing pain caused by hemorrhoidectomy. In addition, preoperative and postoperative lactulose, perioperative metronidazole, glyceryl trinitrate ointment, and warm sitz baths are all used to reduce pain.

Postoperative bleeding occurs in approximately 1% of patients within the first 24 hours of treatment, at which time it typically represents a technical failure of the procedure. Bleeding may also occur 1–2 weeks postoperatively in 0.5–4.0% of cases due to separation of the ligated pedicle before sufficient thrombosis has occurred. Bleeding at this time can be quite severe, and may require either suture ligation or tamponade by anal packing.

Urinary retention is another problem that occurs in 1–52% of patients after hemorrhoidectomy. This is best prevented by limiting fluid administration to 250 ml perioperatively, avoiding the use of spinal anesthesia or anal packing, and by the liberal use of aggressive analgesia.

See Also the Following Articles

Anal Canal • Defecation • Dietary Fiber • Rectum, Anatomy

Further Reading

Fleshman, J. (2002). Advanced technology in the management of hemorrhoids: Stapling, laser, harmonic scalpel, and ligasure. *J. Gastrointest. Surg.* **6**, 299–301.

Hulme-Moir, M., and Bartolo, D. C. (2001). Hemorrhoids. *Gastroenterol. Clin. North Am.* **30**, 183–197.

Johanson, J. F. (2002). Nonsurgical treatment of hemorrhoids. *J. Gastrointest. Surg.* **6**, 290–294.

Loder, P. B., Kamm, M. A., and Nicholls, R. J. (1994). Haemorrhoids: Pathology, pathophysiology, and aetiology. *Br. J. Surg.* **81**, 946–954.

Macleod, J. H. (1983). Rational approach to treatment of hemorrhoids based on a theory of etiology. *Arch. Surg.* **118**, 29–32.

Nivatvongs, S. (2001). Anorectal disorders. *In* "Surgery" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 3rd Ed., pp. 1158–1183. Lippincott, Williams & Wilkins, Philadelphia.

Pfenninger, J. L., and Zainea, G. G. (2001). Common anorectal conditions: Part II. *Am. Fam. Physician* **64**, 77–88.

Senagore, A. J. (2002). Surgical management of hemorrhoids. *J. Gastrointest. Surg.* **6**, 295–298.

Thomson, W. H. F. (1975). The nature of hemorrhoids. *Br. J. Surg.* **77**, 458–462.

Wexner, S. D., and Baig, K. (2001). The evaluation and physiologic assessment of hemorrhoidal disease: A review. *Tech. Coloproctol.* **5**, 165–168.

Henoch–Schönlein Purpura

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

immunoglobulin A nephropathy Immune-complex-mediated glomerulonephritis defined immunohistologically by the presence of glomerular immunoglobulin A.

palpable purpura Visible and often palpable discrete cutaneous hemorrhage, typically in dependent areas. It is occasionally confluent and extensive.

Henoch–Schönlein purpura is a small-vessel vasculitis with the classic clinical tetrad of purpura, arthritis, abdominal pain, and glomerulonephritis. Histologically, it is characterized by the deposition of immunoglobulin A in and around blood vessels, including the glomeruli of the kidneys. It is primarily a disease of children but also occurs in adolescents and adults.

EPIDEMIOLOGY

Henoch–Schönlein purpura (HSP) has a worldwide prevalence in children of 1.5/1000 and an annual incidence of 135/million. In children with familial Mediterranean fever, the prevalence is 130/1000. Upper respiratory tract infection precedes onset of HSP in children in up to 50% of cases. Girls and boys are affected equally, with a median age of onset of 4 years. In adults, the annual incidence is much lower and is estimated to be 1.2/million, but this may be an underestimation.

ETIOLOGY AND PATHOGENESIS

HSP is characterized by the deposition of immunoglobulin A (IgA)-containing immune complexes in the tissues. The glomerular lesion is identical to that of IgA nephropathy, a disease with which it likely shares a common etiology. The disease often follows an upper respiratory infection and occurs more commonly in the winter months, suggesting that infection triggers an autoimmune response. Most patients will have elevated serum levels of IgA. Of the two isotypes of IgA (IgA1 and IgA2), IgA1 is exclusively associated with IgA nephropathy, suggesting that selective increase in IgA1 production and/or clearance plays an important role in pathogenesis of this disorder. Once IgA immune complexes are deposited, immune-mediated inflammation ensues.

CLINICAL FEATURES

Cutaneous

The hallmark of HSP is the acute onset of palpable purpura of the legs. Purpura may be extensive and confluent and may involve the buttocks, trunk, and upper extremities. There may be associated edema.

Musculoskeletal

Arthritis is the second commonest feature of HSP. It is typically acute and usually involves the large joints of the legs. Arthritis may occasionally precede cutaneous disease. The arthritis is not associated with serious sequelae.

Renal

Renal involvement is very common in HSP, but is frequently asymptomatic. The incidence of renal disease increases in HSP with the patient's age. Overt disease occurs in less than 50% of children, usually within 4 weeks of the onset of systemic symptoms. When the kidneys are involved, hematuria is universal, whereas proteinuria and renal insufficiency are less common. The combination of hematuria and proteinuria is a marker for progressive renal insufficiency. The prognosis is excellent, with less than 1% of children progressing to end-stage renal disease; however, more than 10% of adults with HSP progress to end-stage renal disease.

Gastrointestinal

The majority of children with HSP have gastrointestinal symptoms and signs. These include colicky abdominal pain, nausea, vomiting, and gastrointestinal bleeding. These symptoms typically develop within 8 days of the first sign of rash. Endoscopy reveals hemorrhagic and erosive lesions, typically of the duodenum, although erosions have been documented in other regions of the gastrointestinal tract. Histology of mucosal biopsies may reveal IgA deposition. Clinically obvious gastrointestinal bleeding occurs in 25% of patients, whereas occult blood loss is present in 50%. Intussusception is a rare complication of HSP.

 Copyright 2004, Elsevier (USA). All rights reserved.

Pancreatitis and protein-losing enteropathy have been reported.

Other

Acute scrotal swelling has been reported in children. Central nervous system involvement is very rare. Asymptomatic pulmonary involvement is common, but overt pulmonary hemorrhage is rare.

DIAGNOSIS

The diagnosis of HSP in children is usually obvious, based on the classic tetrad of cutaneous, musculoskeletal, renal, and gastrointestinal symptoms. In adults, other disorders that can cause similar symptoms and signs include systemic lupus erythematosus, Wegener's granulomatosis, antiglomerular basement membrane (GBM) disease, and drug-induced cutaneous leukocytoclastic vasculitis. Skin biopsy, although often not needed for diagnosis of HSP, will show a leukocytoclastic vasculitis of the postcapillary venules. IgA deposition in vessels can be demonstrated in most cases. An elevated serum IgA is present in 50% of patients. Renal biopsy is generally reserved for patients with overt renal disease. Renal histopathology ranges from minimal change to severe crescentic glomerulonephritis. IgA deposition in the glomeruli can usually be demonstrated.

THERAPY AND PROGNOSIS

HSP is a self-limited disease in the majority of cases, especially in children, with complete recovery occurring in 90% of cases. Recurrences are common, occurring in one-third of cases. Therapy is largely supportive and consists of adequate hydration and the use of nonsteroidal antiinflammatory drugs to relieve painful edema and arthritis. Corticosteroids can hasten resolution of purpura and are also widely used to treat abdominal symptoms. Treatment for renal disease is indicated only in patients with marked proteinuria and/or renal insufficiency. There are few data regarding the efficacy of aggressive immunosuppression in advanced renal disease. Azathioprine, cyclophosphamide, and plasmapheresis have all been employed in patients with severe renal disease. Renal disease is more common and more extensive in adults. Renal transplantation may be performed in patients that progress to end-stage renal disease.

See Also the Following Articles

Intussusception • Occult Gastrointestinal Bleeding • Protein-Losing Enteropathy, Pediatric • Vasculitis

Further Reading

Donadio, J. V., and Grande, J. P. (2002). Medical progress: IgA nephropathy. *N. Engl. J. Med.* **347**, 738–748.

Esaki, M., Matsumoto, T., Nakamura, S., Kawasaki, M., Iwai, K., Hirakawa, K., Tarumi, K., Yao, T., and Iida, M. (2002). GI involvement in Henoch–Schönlein purpura. *Gastrointest. Endosc.* **56**(6), 920–923.

Kato, S., Shibuya, H., Naganuma, H., and Nakagawa, H. (1992). Gastrointestinal endoscopy in Henoch–Schönlein purpura. *Eur. J. Pediatr.* **151**(7), 482–484.

Pillebout, E., Thervet, E., Hill, G., Alberti, C., Vanhille, P., and Nochy, D. (2002). Henoch–Schönlein purpura in adults: Outcome and prognostic factors. *J. Am. Soc. Nephrol.* **13**(5), 1271–1278.

Saulsbury, F. T. (1999). Henoch–Schönlein purpura in children. Report of 100 patients and review of the literature. *Medicine (Baltimore)* **78**, 395.

Hepatic Adenomas

ANNE M. LARSON
University of Washington, Seattle

glycogen storage diseases Disorders secondary to defects in any of the proteins involved in the synthesis or degradation of glycogen, leading to storage of quantitatively or qualitative abnormal glycogen.

oral contraceptive An oral agent containing estrogen or estrogen and progesterone and used in the prevention of pregnancy.

Hepatic adenomas were virtually unknown prior to 1960. They are benign liver cell tumors resulting from the proliferation of hepatocytes that are otherwise histologically normal or near normal. In the older literature, they are often referred to as "benign hepatoma."

EPIDEMIOLOGY

Hepatic adenomas (HA) are more often seen in women (9 : 1), predominantly of childbearing age. Prior to the introduction of oral contraceptives (OCPs) in the late 1950s, HA were rarely seen. The incidence began increasing in about 1960, with the highest risk in older women (> 30 years) who had taken high-dose OCPs for 5 years or more. In the United States, the reported incidence is 1−1.3 per million in non-OCP users and 30−40 per 100,000 in long-term OCP users. There has been some decrease in incidence with the reduction of estrogen dosing in oral contraceptives.

HA have also been associated with the glycogen storage diseases (GSDs). They are seen in 52% of those with GSD type I and 25% of patients with GSD type II. In this setting, the lesions are more likely to be multiple, occur before the age of 20 years, and are more common in males (2 : 1). Malignant transformation is common at a mean age of 23 years.

Other less common associations include the use of anabolic or androgenic steroids, diabetes mellitus, pregnancy, familial adenomatous polyposis, carbamazepine, and severe combined immunodeficiency. Rarely, they can occur sporadically in patients without risk factors.

ETIOLOGY AND PATHOGENESIS

The true etiology of HA development is uncertain. Although it is postulated that they are responsive to hormonal influences, not all HA show evidence of estrogen receptors. In patients with GSD, it is postulated that there is a glucagon/insulin imbalance with glycogen overload and possible proto-oncogene activation. Diabetics may have a mechanism similar to GSD.

CLINICAL PRESENTATION

Most patients are young women (mean age 34 years) with a history of OCP use (>90%). Although patients may be asymptomatic (5−10%), many present with right upper quadrant pain or discomfort (50%) and some with an abdominal mass (25−35%). Pain may be chronic or intermittent. Approximately 25−40% will present with "classic" symptoms of severe acute abdominal pain and hemoperitoneum with subsequent hypotension followed by shock (mortality rate of 6%). The lesions are usually single, but can be multiple in 10−20% of cases. Rarely are there more than 5 in number. In the setting of GSD, multiple adenomas are more common. At the time of presentation, HA are usually > 5 cm. Hepatic laboratory studies are normal unless hemorrhage occurs. Rarely, the alkaline phosphatase and gamma-glutamyl transpeptidase (GGT) can be slightly elevated.

HISTOPATHOLOGY

Grossly, HA are large, bulging, well-circumscribed tan-yellow tumors, often 5 cm and up to 30 cm in size. There are large blood vessels traversing the surface and there is no true capsule. Central necrosis or hemorrhage may be seen and they always occur within liver parenchyma with normal architecture. Microscopically, the cells closely resemble normal hepatocytes with the hallmark feature of cells arranged in plates two to three cells thick separated by sinusoids. The cells are larger than normal hepatocytes and contain excess glycogen and fat. There is a well-defined reticulin framework, but no portal tracts or bile ducts. Bile plugs in distended canaliculi may be seen. Dilated sinusoids with thin-walled capillaries may be present and the tumor is perfused solely by the peripheral arterial feeding vessels.

 Copyright 2004, Elsevier (USA). All rights reserved.

DIAGNOSIS

The differential diagnosis includes normal liver, other liver tumors [focal nodular hyperplasia (FNH), fibrolamellar hepatocellular carcinoma, hepatocellular carcinoma], and the regenerating nodules of cirrhosis. Imaging studies all reveal a well-demarcated lesion, but are generally nonspecific. Ultrasound shows a smooth or nodular lesion with variable echogenicity. Doppler shows venous signals; however, ultrasound is not able to reliably distinguish HA from other lesions. On a noncontrast computed tomography image, the lesions are either hypo- or isodense with the surrounding liver. Irregular vascular enhancement occurs after contrast injection. The central portion of the lesion remains hypodense if hemorrhage is present and can be difficult to distinguish from FNH. On delayed images, HA is hypodense. On magnetic resonance imaging, HA shows a low signal to slightly hyperintense image on T1 with a well-defined low-intensity capsule. The tumors heterogeneously enhance on T2 images and may be difficult to distinguish from hepatocellular carcinoma.

Nuclear scintigraphy is rarely diagnostic; however, compared with normal liver, there is decreased uptake of 99m-Tc-sulcur colloid. This may be helpful differentiating HA from FNH. With angiographic evaluation, HA have a wide variety of appearances and the tumors may appear either hypovascular or hypervascular. Blood flow is from the periphery to the center with a homogenous blush during venous phase. Angiography can help distinguish hepatocellular carcinoma or focal nodular hyperplasia from HA. Definitive diagnosis, however, requires liver biopsy. This carries a small risk of bleeding and may not be diagnostic if the biopsy specimen is too small.

TREATMENT/MANAGEMENT

Management of HA is straightforward. Patients should avoid estrogens. The adenomas may decrease in size or disappear ($<20\%$). Women should be counseled to avoid pregnancy due to the increased risk of HA growth and rupture. Surgical resection, by either enucleation or resection, is the recommendation of choice, due to the risk of rupture, necrosis, and malignant transformation. Elective surgical mortality is less than 1%; however, if the surgery is performed urgently following rupture, mortality increases to 5–8%. If HA is OCP-related, an observation period is recommended to see if disappearance occurs or to "down-size" the lesion prior to surgery. No lesion >5 cm should be expectantly followed. If smaller lesions are followed, serial scans and α-fetoprotein (AFP) measurements are required. Liver transplantation should be considered if HA are found in association with type I glycogen storage disease or if the disease is multifocal or diffuse (particularly with increasing AFP).

PROGNOSIS

Prognosis is not well established. HA may improve after discontinuation of oral contraceptives (20%). Adenomas may also evolve into malignant tumors and most case reports are associated with OCP use. The risk is higher when associated with anabolic and androgenic steroids or GSD. HA may recur; therefore, postoperative surveillance with ultrasound every 2 years is recommended.

See Also the Following Articles

Glycogen Storage Disease • Hepatocellular Carcinoma (HCC)

Further Reading

Grazioli, L., Federle, M. P., Brancatelli, G., Ichikawa, T., Olivetti, L., and Blachar, A. (2001). Hepatic adenomas: Imaging and pathologic findings. *RadioGraphics* **21**, 877–894.

Molina, E. G. and Schiff, E. R. (1999). Benign solid lesions of the liver. *In* "Diseases of the Liver" (E. R. Schiff, M. F. Sorrel, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 1254–1257. Lippincott Williams & Wilkins, Philadelphia, PA.

Reddy, K. R., Kligerman, S., Levi, J., Livingston, A., Molina, E., Franceschi, D., Badalamenti, S., Jeffers, L., Tzakis, A., and Schiff, E. R. (2001). Benign and solid tumors of the liver: Relationship to sex, age, size of tumors, and outcome. *Am. Surg.* **67**, 173–178.

Trotter, J. F., and Everson, G. T. (2001). Benign focal lesions of the liver. *Clin. Liver Dis.* **5**, 17–42.

Weimann, A., Ringe, B., Klempnauer, J., *et al.* (1997). Benign liver tumors: Differential diagnosis and indications for surgery. *World J. Surg.* **21**, 983–991.

Hepatic Circulation

VIJAY SHAH
Mayo Clinic

autoregulation Process by which vascular inflow is regulated to maintain total liver blood flow at a relatively constant level.

space of Disse Virtual space located between the sinusoidal endothelial lining and the parenchymal hepatocyte population.

Under normal physiologic conditions, the liver functions as a low-resistance system, thereby enabling the accommodation of large volumes of portal blood in the absence of significant changes in pressure. Liver cells, which are predominantly responsible for function of the hepatic circulation, include endothelial cells within the sinusoids, portal venules, hepatic arteries, and hepatic venules, and contractile cells, which include hepatic stellate cells and vascular smooth muscle cells. Pathologic perturbations can impair the function of the hepatic circulation through the actions of specific vascular mediators, most notably, nitric oxide, endothelin, and perhaps carbon monoxide. In liver cirrhosis, hepatic vascular compliance is lost and portal pressure increases in response to small changes in venous inflow, due predominantly to mechanical factors as well as through imbalances in vascular factors, culminating in hepatic vascular constriction. Other clinically relevant insults such as ischemia–reperfusion, alcohol, and endotoxin impair hepatic vascular function in the absence of structural changes, probably through the production of injurious cytokines from Kupffer cells.

CELLULAR ANATOMY

Introduction

Hepatic vascular regulation occurs through the dynamic interplay of the vascular endothelium, which is uniquely situated to transduce bloodborne and mechanical stimuli onto its underlying contractile cells. However, in contradistinction to blood flow regulation in other regional vascular beds, the cells that are responsible for these hemodynamic signaling functions in liver maintain unique phenotypes (Table I).

TABLE I Major Sinusoidal Liver Cells

Cell type	Putative function
Sinusoidal endothelial cell	Vasoregulation
	Scavenging/endocytosis/transport
	Adhesion
	Antigen presentation
Hepatic stellate cell	Vitamin A storage
	Collagen deposition
	Contractility/relaxation
Kupffer cell	Phagocytosis
	Cytokine production
	Vasoregulation
	Immune functions

Liver Endothelial Cell

The liver endothelial cell encompasses distinct phenotypes determined by cell location. The hepatic macrocirculation is composed of cells that maintain a phenotype similar to that of endothelial cells throughout other regional vascular beds. However, within the liver sinusoids, which make up a microcirculatory unit that provides an anastomosis between the portal venous and hepatic venous circulation, the sinusoidal endothelial cells contain fenestrae (Fig. 1). Fenestrae in these cells are grouped into sieve plates, and these pores, in conjunction with the lack of a basement membrane, facilitate the transport of macromolecules from the vascular channels to the space of Disse, the potential space located between the abluminal portion of the endothelial cell and the adjacent hepatocytes, and within which Kupffer cells and hepatic stellate cells reside. These fenestrae are dynamic structures; their diameter is modulated by pharmacologic agents as well as through pathophysiological insults.

Hepatic Stellate Cells

Hepatic stellate cells are postulated to be a contractile cell within the sinusoid, functioning as a liver-specific pericyte, akin to the mesangial cell in the kidney. These cells serve the function of vitamin A storage under physiologic situations. However, on activation, which occurs in response to a variety of antagonistic stimuli and perhaps even in normal circumstances, these cells develop a contractile phenotype and

 Copyright 2004, Elsevier (USA). All rights reserved.

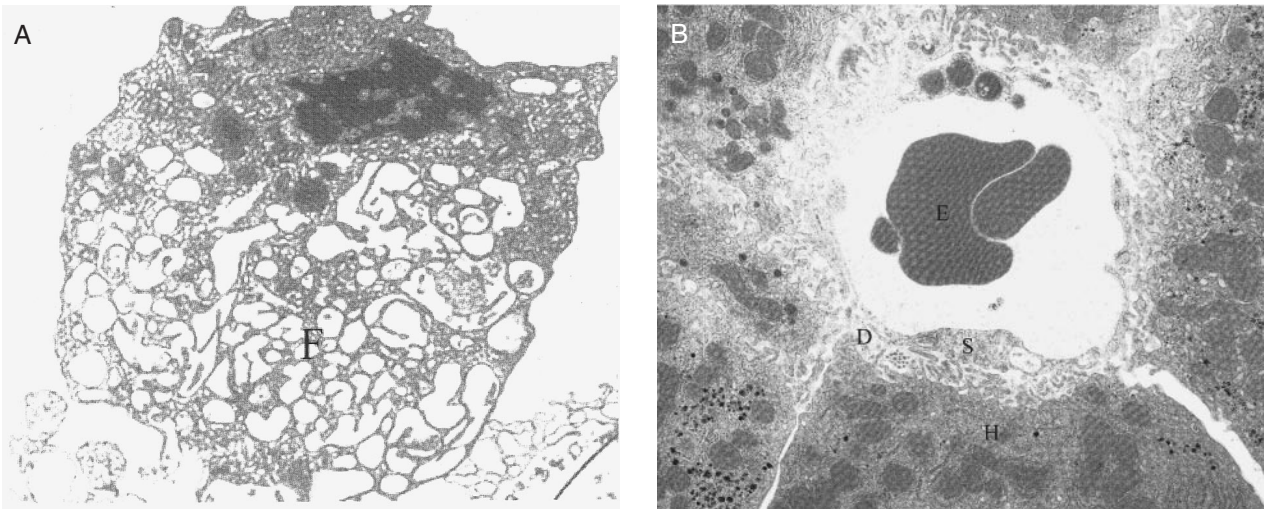


FIGURE 1 Ultrastructural analysis of the liver microcirculation. (A) Transmission electron micrograph of a hepatic sinusoid (from a study performed at the Mayo Electron Microscopy Core Facility). Note the lack of basement membrane underlying the endothelial cell. S, Sinusoidal endothelial cell; H, hepatocyte; E, erythrocyte; D, space of Disse. (B) Transmission electron micrograph of a freshly isolated sinusoidal endothelial cell. Note the presence of the spongelike fenestrae (F), which facilitate the transport of macromolecules from the sinusoidal channel to abluminal hepatocytes. Reproduced with permission from Shah *et al.*, (1997). *J. Clin. Invest.* **100**, 2923–2930.

maintain a function equivalent to a smooth muscle effector cell. In support of this concept, hepatic stellate cells have been demonstrated to possess a variety of smooth muscle cell phenotypic characteristics on activation, *in vitro* and *in vivo*, including the ability to relax in response to endothelium-derived vasodilators, most notably nitric oxide (NO).

Kupffer Cells

Kupffer cells, as members of the reticuloendothelial system, function as resident macrophages. Although their vasoregulatory function is not prominent, these cells may influence hepatic hemodynamics in certain circumstances, perhaps through the endotoxin-dependent production of potentially vasoactive cytokines such as tumor necrosis factor.

VASOREGULATORY SIGNALING SYSTEMS

Vasoregulatory signaling within the liver occurs predominantly through paracrine mechanisms, but also, in part, through neurohumoral mechanisms. Paracrine signaling occurs through the endothelium-derived production of vasoactive mediators (Table II), which act abluminally on underlying contractile cells and thereby promote alterations in vascular diameter and perfusion. Over the past few decades, several key paracrine vasoregulatory signaling systems have been identified and characterized, including nitric oxide, endothelin (ET), and carbon monoxide (CO). Relevant neurohumoral pathways include the renin–angiotensin system as well as the sympathetic adrenergic modulator norepinephrine.

Nitric Oxide

NO is produced by the heme-binding P450-like NO synthase (NOS) enzyme family, including the endothelial, neuronal, and inducible isoforms of NOS (eNOS, nNOS, and iNOS, respectively). eNOS appears to be the most relevant isoform in the maintenance of normal hepatic vascular function, and a number of experimental studies utilizing pharmacologic and molecular approaches have established the important role of this enzyme in hepatic vasodilation. eNOS-derived NO acts in a paracrine manner on adjacent contractile cells, where it produces relaxation through a cyclic

TABLE II Major Vascular Mediators in Liver

Mediator	Putative function
Nitric oxide	Vasodilation Angiogenesis/remodeling Cell injury/protection Antifibrotic
Endothelin	Vasoconstriction Vascular remodeling

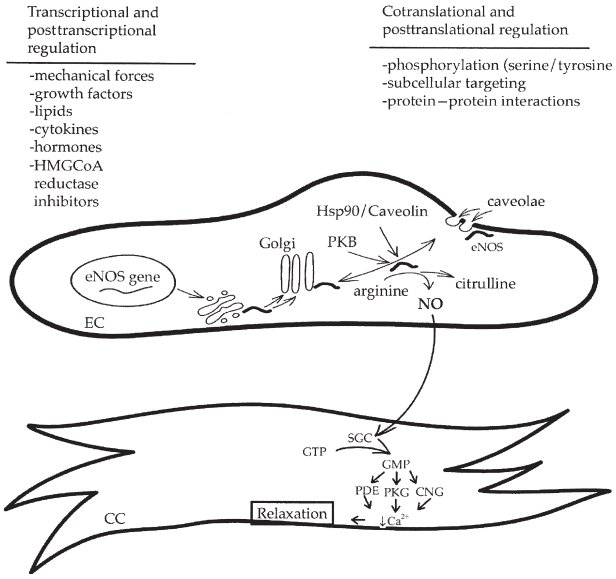


FIGURE 2 The nitric oxide (NO) signaling pathway. Endothelial nitric oxide synthase (eNOS) is regulated at multiple cellular levels, including transcription, subcellular targeting, phosphorylation and protein−protein interactions. NO activates soluble guanylate cyclase in contractile cells to induce relaxation. EC, Endothelial cell; Hsp, heat-shock protein; PKB, protein kinase B; CC, contractile cell; SGC, soluble guanylate cyclase; PDE, phosphodiesterase; PKG, protein kinase G; CNG, cyclic nucleotide gated channel. Reproduced from Shah (2001), Cellular and molecular basis of portal hypertension, *Clinics in Liver Disease* **5**, 629−644. With permission from Elsevier.

guanosine monophosphate (cGMP)-dependent signaling pathway (Fig. 2). Shear stress, caused by the frictional force of blood within the vascular channels, is one of the most potent physiologic stimuli of NO production through eNOS. The roles of nNOS and iNOS in regulation of vascular function in liver remain less defined.

Endothelin

In normal circumstances, endothelin is also synthesized and released by liver endothelial cells in response to distinct stimuli. The G protein-coupled ET-1 receptors are located on most liver cell populations. Interestingly, it appears that ligand binding with the ET-A receptor on contractile liver cells promotes hepatic constriction, whereas ligand binding on the ET-B receptor localized to liver endothelial cells can paradoxically promote vasodilation in an NO-dependent mechanism (Fig. 3). Based on studies examining portal pressure and sinusoidal diameter in response to endothelin infusion into the portal vein, the predominant action of endothelin is to constrict veins and sinusoids within the hepatic vasculature.

Carbon Monoxide

CO may also function to maintain sinusoidal tone in a relaxed state, analogous to NO, although the potency of activation of downstream guanylate cyclase by CO is very low as compared to NO. CO is synthesized by a family of enzymes termed heme oxygenases (HOs). HO-2, which is abundantly expressed in liver endothelium, appears to be responsible for basal CO production.

Neurohumoral Regulation

Both the sympathetic adrenergic agonist, norepinephrine, and the renin-dependent vasoconstrictor, angiotensin, have some vasoreactivity in the liver. The hepatic vasculature is richly innervated with sympathetic adrenergic nerves and infusion of norepinephrine results in hepatic constriction. Angiotensin also has a constrictive effect in experimental models as well as a direct contractile effect on isolated hepatic stellate cells in culture.

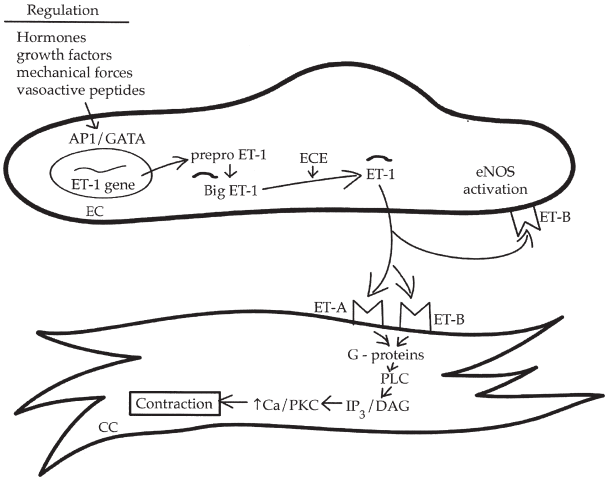


FIGURE 3 The endothelin (ET) signaling pathway. Regulation of ET-1 production occurs at the transcriptional level in response to mechanical forces, hormones, growth factors, and vasoactive peptides such as prostacyclin. Although ET-1 can paradoxically stimulate endothelial nitric oxide synthase (eNOS) activation through ligand binding on the ET-B receptor of endothelial cells (EC), the predominating action of ET-1 occurs on abluminal contractile cells through the ET-A and ET-B receptors, culminating in contraction. ECE, Endothelin-converting enzyme; PLC, phospholipase C; CC, contractile cell; DAG, diacylglycerol; IP_3, inositol trisphosphate; PKC, protein kinase C. Reproduced from Shah (2001), Cellular and molecular basis of portal hypertension, *Clinics in Liver Disease* **5**, 629−644. With permission from Elsevier.

PHYSIOLOGY OF THE HEPATIC CIRCULATION

Introduction

The hepatic vascular bed is a low-pressure system that is able to accommodate a large volume of blood; indeed, total hepatic blood flow constitutes nearly 30% of total cardiac output. The liver receives a dual blood supply (Fig. 4) from the portal vein as well as the hepatic artery, similar to the pulmonary system, which is perfused by the pulmonary artery as well as the bronchial artery. Portal venous blood comprises approximately 70–90% of total hepatic blood flow and is derived from the mesenteric venous circulation, including the digestive tract, spleen, and pancreas. Portal blood flow is determined primarily by the rate and volume of blood in the mesenteric vessels, which is, in turn, determined by dilation and constriction of the mesenteric arterioles. Hepatic arterial blood flow is derived from the celiac plexus.

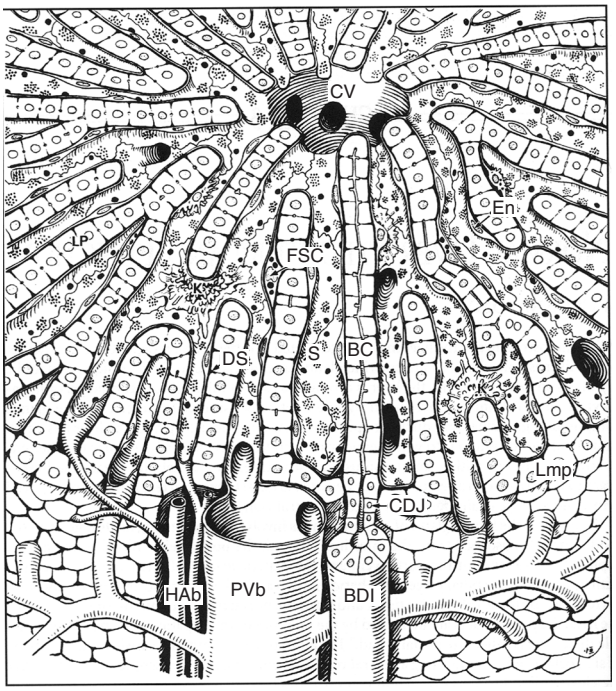


FIGURE 4 Vascular architecture of the liver. Blood flow enters the liver via the portal vein (PVb) as well as the hepatic artery (HAb). While portal blood enters directly into the sinusoids (S), hepatic arterial blood perfuses into distinct anatomic locations prior to reentering the sinusoids. Sinusoidal blood leaves the liver via the central veins (CV). Sinusoidal endothelial cells (En) form the fenestrated sinusoidal wall. Kupffer cells (K) are located within the sinusoids, whereas hepatic stellate cells, also termed fat-storing cells (FSC), lie within the space of Disse (DS), adjacent to the single layer of hepatocytes (liver plate; LP). Bile canaliculi (BC) drain bile into interlobular bile ducts (BDI) via the caniliculoductular junction (CDJ) in the opposite direction to flow in the vascular channels. Reproduced with permission from Motta, P., Muto M., and Fujita, T. (1978). "The Liver. An Atlas of Scanning Electron Microscopy," pp. 129. Igaku-Shoin Medical Publishers, Inc.

Sinusoidal Flow Regulation

Although the sinusoids are a high-compliance system, hepatic perfusion and pressure regulation are regulated, in part, through events within the hepatic sinusoids, although probably in conjunction with pre- and postsinusoidal sites, depending on the specific circumstance. Under normal physiologic circumstances, sinusoidal flow regulation occurs, in part, through the recruitment and dropout of individual sinusoidal vessels. The traditional paradigm of endothelium-dependent regulation of smooth muscle tone in sinusoids remains an area of investigation. For example, although the sinusoidal endothelial cells can produce vasoactive agents, the lack of a traditional underlying smooth muscle puts into question the precise mechanism of sinusoidal channel tone modulation. In this regard, the hepatic stellate cell may function as a sinusoidal effector cell. However, its contractile function in normal physiologic conditions remains circumspect, and the lack of contractile mechanisms in normal sinusoid has led some to believe that the site of maximal resistance within the normal liver may lie distal to the sinusoid. For example, terminal hepatic venules do possess a modest smooth muscle contractile layer and may serve as a resistance site.

Hepatic Artery and Autoregulation

Under normal physiologic conditions, the hepatic artery supplies only 10–25% of total hepatic blood flow, with the remainder of hepatic blood flow derived from the portal venous system. Although a portion of the hepatic arterial flow does eventually coalesce with portal venous blood within the hepatic sinusoids, the hepatic artery maintains a unique anatomical distribution pattern that is distinct from that of the portal vein. The major sites of primary hepatic arterial flow appear to include the peribiliary plexus, the vasa vasorum of hepatic veins, the portal tract interstitium, and Glisson's capsule, although the distribution does vary among species. Although there may be bypass tracts directly to the hepatic veins, much of the hepatic arterial flow from these sites does eventually reanastamose with portal vein derived blood within the parenchymal sinusoids, thereby providing highly oxygenated arterial blood to the hepatocytes in conjunction with the poorly

oxygenated portal venous blood. Interestingly, the high-pressure hepatic arterial flow does not contribute significantly to the overall portal pressure, although the mechanism of sinusoidal decompression of high-pressure arterial flow remains unclear. Although impairment of hepatic arterial flow has limited untoward effects on the liver under normal physiologic conditions, conditions associated with diminished portal vein flow, such as portal vein thrombosis and liver cirrhosis, result in a significant dependence on hepatic arterial flow to maintain hepatic parenchymal perfusion. Another interesting characteristic of hepatic arterial flow is its propensity to supply tumors with perfusion, a characteristic that is exploited for selective therapeutics through approaches such as hepatic artery chemoembolization of hepatoma. Another important function of the hepatic arterial blood flow is to maintain constant total hepatic blood flow through a process termed autoregulation. This requires compensatory adjustment of hepatic arterial flow in response to changes in portal blood flow. Maintenance of total hepatic blood flow at a constant level is important to maintain normal portal pressure, as well as to maintain the homeostatic balance of xenobiotics and endogenous peptides cleared within the liver. Adenosine concentration in the space of Mall is thought to mediate hepatic blood flow autoregulation through a mechanism termed the hepatic arterial buffer response.

PATHOPHYSIOLOGY OF HEPATIC CIRCULATION

Hepatic Circulation in Liver Cirrhosis

Important changes occur in the hepatic circulation in liver cirrhosis; the changes culminate in an increase in intrahepatic vascular resistance, an important component in the development of portal hypertension and its clinical sequelae. Much of the increase in intrahepatic resistance occurs through restrictive factors caused by architectural remodeling of the parenchyma and vascular channels. For example, deposition of collagen within the space of Disse and the development of regenerative nodules both likely account for a significant portion of the increase in intrahepatic resistance that occurs in cirrhosis. However, a smaller, yet important, component of the increase in intrahepatic resistance occurs through the dysfunction of hepatic vasoregulatory cells and ensuing development of hepatic vasoconstriction. Indeed, cumulative experimental evidence from independent groups has demonstrated an imbalance of vasoactive factors in liver cirrhosis, which is characterized by enhanced production of the constrictor endothelin and diminished production of the vasodilator NO. In addition to phenotypic defects in the endothelial cell production of these molecules, evidence also suggests that the effector cell mediating hepatic constriction may be the activation and enhanced contractility of the hepatic stellate cell. Elucidation of the complex regulatory mechanisms that mediate NO and ET-1 signaling in hepatic vasoregulatory cells may provide opportunities for future experimental therapeutics for treatment of cirrhosis.

Hepatic Circulation in Ischemia–Reperfusion

Clinical scenarios such as orthotopic liver transplantation are often associated with the pathophysiologic syndrome of ischemia–reperfusion, which causes parenchymal liver injury, in part, through hepatic microcirculatory dysfunction. Although the duration of ischemia plays a role in this process, the reperfusion process appears to be a particularly injurious event in the ensuing microcirculatory injury process. Reperfusion injury causes microcirculatory dysfunction through different mechanisms, including an imbalance of vasoactive agents, which promotes sinusoidal constriction and an increased production of reactive oxygen species, which alters the oxidant/antioxidant status of the vasculature in a detrimental way.

Hepatic Circulation in Response to Endotoxin

Clinical conditions such as alcoholic liver disease and sepsis syndrome are associated with hepatic vascular dysfunction. Alcoholic liver disease results in the development of increased intrahepatic resistance, often in the absence of appreciable fibrosis. Sepsis syndrome is commonly associated with liver enzyme abnormalities with changes in the hepatic microcirculation. Hepatic circulatory dysfunction in these syndromes may have a common mechanistic link, because both are associated with elevated endotoxin delivery to the portal vein. Portal venous endotoxin stimulates Kupffer cells in the sinusoids to produce tumor necrosis factor, an inflammatory cytokine, which, in turn, causes circulatory disturbances, including sinusoidal vasoconstriction, leukocyte adherence to vascular channels, and an ensuing reduction in sinusoidal perfusion. These mechanisms may be relevant to the syndromes of alcoholic liver injury and sepsis.

See Also the Following Articles

Circulation, Overview • Cirrhosis • Liver Ischemia • Portal Hypertension and Esophageal Varices

Further Reading

Clemens, M. (1999). Nitric oxide in liver injury. *Hepatology* **30**, 1–4.
Clemens, M. G., Bauer, M., Pannen, B. H., Bauer, I., and Zhang, J. X. (1997). Remodeling of hepatic microvascular responsiveness after ischemia/reperfusion. *Shock* **8**, 80–85.
Ekataksin, W., and Kaneda, K. (1999). Liver microvascular architecture: An insight into the pathophysiology of portal hypertension. *Semin. Liver Dis.* **19**, 359–382.
Groszmann, R., and Atterbury, C. (1982). The pathophysiology of portal hypertension: A basis for classification. *Semin. Liver Dis.* **2**, 177.
McCuskey, R. S., and Reilly, F. D. (1993). Hepatic microvasculature: Dynamic structure and its regulation. *Semin. Liver Dis.* **13**, 1–11.
Motta, P., Muto, M., and Fujita, T. (1978). *In* "The Liver. An Atlas of Scanning Electron Microscopy," p. 129. Igaku-Shoin Medical Publ., New York.
Rockey, D. C. (1997). The cellular pathogenesis of portal hypertension: Stellate cell contractility, endothelin, and nitric oxide. *Hepatology* **25**, 2–5.
Shah, V. (2001). Cellular and molecular basis of portal hypertension. *Clin. Liver Dis.* **5**, 629–643.
Shah, V., *et al.* (1997). Liver sinusoidal endothelial cells are responsible for nitric oxide modulation of resistance in the hepatic sinusoids. *J. Clin. Invest.* **100**, 2923–2930.
Shah, V., Garcia-Cardena, G., Sessa, W., and Groszmann, R. (1998). The hepatic circulation in health and disease: Report of a single-topic symposium. *Hepatology* **27**, 279–288.

Hepatic Encephalopathy

JAVIER VAQUERO AND ANDRES T. BLEI
Lakeside Veterans Administration Medical Center and Northwestern University Feinberg School of Medicine

asterixis Motor disturbance consisting in the failure to actively maintain posture or position; usually elicited by dorsiflexion of the patient's hand with the patient's arms outstretched and fingers separated.
confusional syndrome Disturbed orientation in regard to time, place, or person, sometimes accompanied by disordered consciousness.

Hepatic encephalopathy can be defined as a potentially reversible disturbance in central nervous system function secondary to hepatic insufficiency or portal–systemic shunting. This broad definition reflects a spectrum of neurologic manifestations, ranging from subtle alterations in neuropsychological tests to appearance of deep coma, brain edema, and intracranial hypertension. The significance of the development of hepatic encephalopathy in a patient with liver disease cannot be underemphasized. A first episode in a patient with cirrhosis carries a poor prognosis and should prompt consideration of liver transplantation. Development of hepatic encephalopathy in a patient with acute liver failure marks the transition to a life-threatening disease; in this condition, the degree of hepatic encephalopathy is a strong predictor of outcome and an important criterion for performing liver transplantation.

CLASSIFICATION

The nomenclature of hepatic encephalopathy (HE) has been a source of confusion, with some terms implying different meanings to different investigators. A new classification of HE takes into account both the type of hepatic abnormality and the duration/characteristics of HE symptoms, categorizing HE in groups and subgroups:

a. Encephalopathy associated with acute liver failure (type A, for acute).
b. Encephalopathy associated with portal–systemic bypass and no intrinsic hepatocellular disease (type B, for Br-pass).
c. Encephalopathy associated with cirrhosis and portal hypertension and/or portal–systemic shunts (type C, for cirrhosis).

We can divide type C further into:

c.1. Episodic HE. Subdivided into precipitated and spontaneous, depending on the presence of precipitating factors. "Recurrent encephalopathy" refers to the occurrence of at least two episodes of episodic HE within 1 year.

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Clemens, M. (1999). Nitric oxide in liver injury. *Hepatology* **30**, 1–4.
Clemens, M. G., Bauer, M., Pannen, B. H., Bauer, I., and Zhang, J. X. (1997). Remodeling of hepatic microvascular responsiveness after ischemia/reperfusion. *Shock* **8**, 80–85.
Ekataksin, W., and Kaneda, K. (1999). Liver microvascular architecture: An insight into the pathophysiology of portal hypertension. *Semin. Liver Dis.* **19**, 359–382.
Groszmann, R., and Atterbury, C. (1982). The pathophysiology of portal hypertension: A basis for classification. *Semin. Liver Dis.* **2**, 177.
McCuskey, R. S., and Reilly, F. D. (1993). Hepatic microvasculature: Dynamic structure and its regulation. *Semin. Liver Dis.* **13**, 1–11.
Motta, P., Muto, M., and Fujita, T. (1978). *In* "The Liver. An Atlas of Scanning Electron Microscopy," p. 129. Igaku-Shoin Medical Publ., New York.
Rockey, D. C. (1997). The cellular pathogenesis of portal hypertension: Stellate cell contractility, endothelin, and nitric oxide. *Hepatology* **25**, 2–5.
Shah, V. (2001). Cellular and molecular basis of portal hypertension. *Clin. Liver Dis.* **5**, 629–643.
Shah, V., *et al.* (1997). Liver sinusoidal endothelial cells are responsible for nitric oxide modulation of resistance in the hepatic sinusoids. *J. Clin. Invest.* **100**, 2923–2930.
Shah, V., Garcia-Cardena, G., Sessa, W., and Groszmann, R. (1998). The hepatic circulation in health and disease: Report of a single-topic symposium. *Hepatology* **27**, 279–288.

Hepatic Encephalopathy

JAVIER VAQUERO AND ANDRES T. BLEI
Lakeside Veterans Administration Medical Center and Northwestern University Feinberg School of Medicine

asterixis Motor disturbance consisting in the failure to actively maintain posture or position; usually elicited by dorsiflexion of the patient's hand with the patient's arms outstretched and fingers separated.
confusional syndrome Disturbed orientation in regard to time, place, or person, sometimes accompanied by disordered consciousness.

Hepatic encephalopathy can be defined as a potentially reversible disturbance in central nervous system function secondary to hepatic insufficiency or portal–systemic shunting. This broad definition reflects a spectrum of neurologic manifestations, ranging from subtle alterations in neuropsychological tests to appearance of deep coma, brain edema, and intracranial hypertension. The significance of the development of hepatic encephalopathy in a patient with liver disease cannot be underemphasized. A first episode in a patient with cirrhosis carries a poor prognosis and should prompt consideration of liver transplantation. Development of hepatic encephalopathy in a patient with acute liver failure marks the transition to a life-threatening disease; in this condition, the degree of hepatic encephalopathy is a strong predictor of outcome and an important criterion for performing liver transplantation.

CLASSIFICATION

The nomenclature of hepatic encephalopathy (HE) has been a source of confusion, with some terms implying different meanings to different investigators. A new classification of HE takes into account both the type of hepatic abnormality and the duration/characteristics of HE symptoms, categorizing HE in groups and subgroups:

a. Encephalopathy associated with acute liver failure (type A, for acute).
b. Encephalopathy associated with portal–systemic bypass and no intrinsic hepatocellular disease (type B, for Br-pass).
c. Encephalopathy associated with cirrhosis and portal hypertension and/or portal–systemic shunts (type C, for cirrhosis).

We can divide type C further into:

c.1. Episodic HE. Subdivided into precipitated and spontaneous, depending on the presence of precipitating factors. "Recurrent encephalopathy" refers to the occurrence of at least two episodes of episodic HE within 1 year.

Copyright 2004, Elsevier (USA). All rights reserved.

c.2. Persistent HE. Includes cognitive deficits that impact negatively on social and occupational functioning, and is subdivided into mild and severe. Treatment-dependent persistent HE is a subgroup in which overt symptoms develop promptly after discontinuing medication.

c.3. Minimal HE. Refers to abnormalities of cognition, affection/emotion, behavior, or bioregulation that are not usually detected by regular clinical examination; diagnosis requires specific neuropsychological and neurophysiological tests.

PATHOGENESIS

A common pathogenic notion is that HE is caused by substances that are efficiently metabolized by the liver. In liver disease as well as in portal-systemic shunting, these substances gain access to the systemic circulation. Increased brain exposure to these substances leads to a disturbance of normal neurotransmission and results in the clinical symptoms that are characteristic of HE.

Neuropathology

With the exception of brain edema in acute liver failure (ALF), no major anatomical alterations are found in the brains of patients in hepatic coma. However, ultrastructural studies point to the astrocyte as the main cell affected in HE. A distinctive morphologic alteration of astrocytes—Alzheimer type II astrocytosis—has been described in the HE of chronic liver disease, with cells exhibiting enlarged nuclei, peripheral margination of chromatin, prominent nucleoli, and excessive accumulation of glycogen. These morphological changes, however, are not specific to HE. In addition, selective swelling of astrocytes is a consistent finding in patients and experimental models of ALF and HE. It is noteworthy that the magnetization transfer ratio, a nuclear magnetic resonance index that reflects the water content of the tissues, is altered in patients with chronic liver disease and returns to normal after performance of liver transplantation, suggesting that low-grade brain edema may be present in the HE of chronic liver disease.

Putative Toxins

Ammonia has been viewed as the most important factor in the genesis of HE. Ammonia is detoxified in the brain through binding to glutamate, resulting in the formation of glutamine. This reaction is catalyzed by the astrocytic enzyme glutamine synthetase. An increase in brain glutamine is a consistent finding in patients and experimental models of HE. Because glutamine is an organic osmolyte, its accumulation may play a major role in astrocyte swelling. In recent years, abnormalities in the normal concentrations of brain organic osmolytes have become evident in HE, suggesting underlying disturbances in cell volume homeostasis.

Increased availability of natural benzodiazepines, which are agonist ligands of the γ-aminobutyric acid (GABA) receptor complex, has been also proposed as the cause of HE. Even though several lines of evidence point to an increased GABA-ergic tone in HE, the presence of such natural benzodiazepines has not been totally confirmed. Other compounds, such as manganese, aromatic amino acids, mercaptans, phenols, or short-chain fatty acids, have also been associated with the development of HE. Hypermanganesemia has been related to parkinsonian manifestations. The altered ratio between aromatic amino acids and branched-chain amino acids in plasma of patients with liver failure, in favor of the former, has been postulated to result in the synthesis of false neurotransmitters. This hypothesis, however, is still far from proven.

Altered Neurotransmission

Multiple neurotransmitter systems have been found to be altered in patients with HE. It is convenient to remember at this point that normal neurotransmission is highly dependent on adequate astrocyte function, and there is a linkage between astrocytic alterations and disturbed neurotransmission. Glutamate is the main excitatory neurotransmitter of the brain and the system for which most abnormalities have been shown, including decreases in total brain glutamate, increases in extracellular glutamate, and decreases in astrocytic glutamate transporters and receptors. Alterations in the GABA system, such as increased GABA-ergic tone, activation of the peripheral-type benzodiazepine receptors, and increased neurosteroids, have been also described. Increased metabolism of serotonin, decreases in dopamine (DOPA) receptors and increased degradation of DOPA, and increases in endogenous opioids and histamine receptors are alterations in other neurotransmitter systems that have been also described in HE.

CLINICAL PRESENTATION

Episode of Hepatic Encephalopathy

An acute confusional syndrome includes an impaired mental state, neuromuscular abnormalities, fetor hepaticus, and hyperventilation. The symptoms appear abruptly and develop over a period of hours

to days, with oscillation of severity over time being an important clinical feature. Initial subtle symptoms, such as changes of personality or disturbances of sleep, may progress to inappropriate behavior, confusion, stupor, and coma if no therapeutic interventions are initiated. Asterixis is very characteristic of mild and moderate stages of HE. However, asterixis is not exclusive of HE and disappears as HE progresses to coma.

Once an episode of HE has developed, relapses are common. The patients usually respond well to treatment, with reestablishment of cognitive function. Careful neurologic and neuropsychological examinations between episodes may, however, reveal abnormalities.

The clinical manifestations of an episode of HE occurring in chronic liver disease are similar, though not identical, to those seen in ALF. An episode of HE in cirrhosis is characterized by a more insidious presentation, with fluctuations of mental state; a precipitating factor is commonly identified and patients usually respond to treatment, with reversal to normal mental status within hours or few days. In contrast, HE in ALF is more progressive and less fluctuating; its evolution follows closely that of the underlying liver disease, patients respond poorly to usual treatments, and brain edema and intracranial hypertension may develop.

Persistent Hepatic Encephalopathy

Manifestations that do not reverse after adequate treatment are considered to be severe when daily activities are impaired. The most extreme manifestations include, severe parkinsonism and myelopathy, in combination with other neurologic manifestations such as ataxia, dysarthria, gait abnormalities, or tremor. This stage of the disease is rarely seen nowdays due to the availability of liver transplantation and the small number of patients that undergo surgical portal–systemic shunting.

DIAGNOSIS

Clinical history and physical examination are the most valuable tools to make the diagnosis. A history of chronic liver disease or the presence of altered liver function tests in a patient with the described symptoms and signs strongly suggests HE. However, the diagnosis of HE is a diagnosis of exclusion. Other diagnostic possibilities to exclude are intracranial occupying lesions, cerebrovascular events, infectious diseases, or other metabolic disorders. Asterixis can also be seen in patients with uremia, hypercapnia, phenytoin intoxication, or hypomagnesemia. Detailed physical examination, searching for signs of cirrhosis, and determination of plasma ammonia or performance of specific neurophysiological or neuroimaging techniques may be useful to make a positive diagnosis.

Grading of HE

Grading of HE is necessary to assess the evolution of the condition and the response to treatment. The two most useful and recommended methods of grading HE are based on clinical findings. The West Haven criteria are a semiquantitative classification that groups HE in four stages, covering the whole spectrum of HE except minimal HE. Thus, stage I includes trivial lack of awareness, euphoria, anxiety, shortened attention span, and impairment of skills such as addition or subtraction; stage II is characterized by lethargy, time disorientation, obvious personality change, and inappropriate behavior; stage III includes somnolence to semistupor but responsiveness to stimuli, confusion, gross disorientation, and bizarre behavior; and stage IV is coma and no response to noxious stimuli.

The Glasgow coma scale, initially developed for patients with neurotrauma, is less subject to observer variability, but its use has not been validated in metabolic encephalopathies. It measures the response to eye opening, verbal behavior, and motor responsiveness, and quantifies neurologic impairment in a continuous numerical scale. It is mainly useful for evaluation of advanced stages of HE.

The portal–systemic encephalopathy index, which combines assessment of mental state, arterial ammonia, electroencephalography, number connection test, and estimation of degree of asterixis, is another method that has been extensively used to grade HE. However, a consensus has been recently reached indicating that it is not adequate for clinical followup evaluation and it is not recommended for clinical trials.

Neuropsychological Tests

Neuropsychological testing plays a primary role in the diagnosis of minimal HE and in the assessment of cognitive function in patients with chronic HE. The tests are very sensitive, but may be influenced by multiple factors such as age, educational background, or repeated learning. An adequate degree of arousal is necessary for administration of these tests. The psycometric hepatic encephalopathy score (PHES) is a standardized test battery that includes number connection tests A and B and the line-tracing, the serial-dotting, and the digit-symbol tests. Its use was endorsed by a recent consensus, due to its ability to discriminate minimal HE from stage I HE.

Neurophysiological Tests

Electroencephalogram shows characteristic changes in patients with HE, including replacement of the normal background waves of 9 to 12 cycles/second by progressively slower waves and the appearance of high-voltage, low-frequency triphasic waves. The latter have been described, however, in other forms of metabolic encephalopathies. External recording of the integrated electrical response to visual, auditory, or somatosensory stimuli is the basis of the use of evoked potentials. The recordings reflect the function of the cerebral cortex and neuronal networks that are involved in each corresponding pathway. They are considered more sensitive than the conventional electroencephalogram for the diagnosis of mild forms of HE.

Neuroimaging

Computed X-ray tomography plays an important role in the differential diagnosis of the patient with acute neurological symptoms. Brain atrophy is commonly seen in the setting of HE, but no specific abnormalities of HE are observed. In patients with ALF, computed X-ray tomography is not sensitive enough for early detection of brain edema.

Magnetic resonance imaging of the brain of patients with cirrhosis characteristically shows hyperresonant globus pallidus on T1-weighted images. These images are thought to be secondary to deposition of manganese in basal ganglia. Magnetic resonance spectroscopy allows the *in vivo* measurement of common chemicals in the human brain. The characteristic pattern of patients with HE and cirrhosis in the proton spectra consists of an increase in the glutamine/glutamate peak coupled with a decrease in the myo-inositol and choline signals.

TREATMENT

Multiple therapeutic modalities have been used for the treatment of HE. However, their efficacy has been seldom assessed in well-designed clinical trials.

General Measures

The mental status of the patient with HE may change rapidly and, thus, special attention by the nursing staff and specific measures to avoid body harm are required. For patients in stage III or IV of HE, prophylactic tracheal intubation should be considered, especially in the setting of ALF. An adequate hydration and nutritional state should be maintained during the period of altered mental state. Identification and removal of precipitating factors is a mainstay of treatment in HE. Diagnosis of infection is particularly important. Gastrointestinal hemorrhage, renal and electrolyte abnormalities, use of psychoactive medications, constipation, or excessive dietary protein intake are also common precipitants. Adequate treatment for each of these factors, if present, should be promptly instituted.

Reduction of Nitrogenous Load from the Gut

Current consensus exists that prolonged periods of dietary protein restriction should be avoided. Patients should receive the maximum tolerable protein intake, preferably from vegetable and dairy sources rather than animal protein, due to their high fiber content and favorable calorie-to-nitrogen ratio, respectively.

Colonic cleansing lowers blood ammonia levels in cirrhosis. Nonabsorbable disaccharides, such as lactulose and lactitol, remain the first-line pharmacological treatment of HE. In addition to their cathartic effect, they produce acidification of the colon, which favors nonionic diffusion of ammonia into the lumen and increases fecal nitrogen excretion. Therapy is aimed at obtaining two or three soft bowel movements a day. When nonabsorbable disaccharides are taken for prolonged periods, abdominal cramping and flatulence are common, and hypernatremia can result from excessive diarrhea. Lactitol and lactulose enemas have been also shown effective, though the quality of data from all clinical trials with these agents is questionable.

Antibiotics are also a therapeutic alternative. They can reduce ammonia-producing bacteria in the colon. Neomycin, the most commonly used agent, may have actions at the level of the intestinal mucosa, in addition to its antibacterial effect. It is poorly absorbed in the intestinal tract, but oto- and nephrotoxicity may occasionally develop. Metronidazole, which affects a bacterial population different from that targeted by neomycin, can also improve HE. Dosing should be done with care. Other poorly absorbable antibiotics, such as vancomycin or rifaximin, have also shown beneficial effects.

Improvement of Extraintestinal Elimination of Ammonia

Several agents have been used to enhance the metabolism of ammonia in splanchnic and peripheral tissues, with questionable efficacy. Patients with cirrhosis are frequently zinc deficient; zinc is a cofactor in all enzymes of the urea cycle, thus administration of zinc acetate to improve ureagenesis may be specially

indicated in patients with associated malnutrition. Administration of ornithine-aspartate provides substrates for the urea cycle in the liver as well as for the synthesis of glutamine via transamination, which is especially relevant in the muscle. Preliminary experiences in acute and chronic HE have been encouraging.

Counteracting Abnormalities of Central Neurotransmission

Intravenous administration of flumazenil is effective in improving mental state in a limited group of patients with cirrhosis and HE, being particularly helpful if benzodiazepine ingestion is suspected. However, an oral preparation is not available and long-term administration has not been evaluated. Bromocriptine has been used to correct abnormalities of dopaminergic neurotransmission, but at the present time, it is indicated only for the treatment of chronic encephalopathy unresponsive to other therapy. Administration of intravenous and oral solutions of branched-chain amino acids has been evaluated in several clinical trials. However, there is no clear benefit with respect to conventional therapy, and administration of these solutions should be reserved to patients with intolerance to dietary protein.

Manipulation of Splanchnic Circulation

In HE caused by congenital portal–systemic shunts and in those patients with severe HE after placement of a transjugular intrahepatic portosystemic stent shunt, occlusion or reduction of the shunt via surgery or via interventional radiology should be considered. This may be also an option in selected patients with cirrhosis and large spontaneous portal–systemic shunts, if they present with recurrent episodes of HE without precipitating factors.

See Also the Following Articles

Cirrhosis • Fulminant Hepatic Failure • Liver Transplantation • Portal Hypertension and Esophageal Varices

Further Reading

Blei, A. T., and Cordoba, J. (2001). Hepatic encephalopathy. *Am. J. Gastroenterol.* **96**, 1968–1976.

Cordoba, J., and Blei, A. T. (2003). Hepatic encephalopathy. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 595–623. Lippincott Williams & Wilkins, Philadelphia.

Ferenci, P., Lockwood, A., Mullen, K., *et al.* (2002). Hepatic encephalopathy—Definition, nomenclature, diagnosis, and quantification: Final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* **35**, 716–721.

Hepatic Granulomas

STACEY R. VLAHAKIS
Mayo Clinic and Foundation

caseating granuloma Aggregation of granulofibrillar necrotic material, and sometimes eosinophilic granules.
fibrin-ring granuloma Tissue mass with a central cavity surrounded by a ring of fibrin and epitheloid macrophages.
lipogranuloma Tissue mass containing lipid vacuoles surrounded by macrophages, lymphocytes, and collagen fibers.
noncaseating granuloma Aggregation of epitheloid macrophages, lymphocytes, and multinucleated giant cells.

Granulomas of the liver can be found in a variety of diseases; up to 10–15% of liver biopsy specimens describe granulomas on pathological review. The presence of a granuloma signifies chronic inflammation of the liver parenchyma in reaction to a foreign agent. The inciting agent induces a cell-mediated immune response that stimulates and attracts macrophages and T cells. The macrophages transform into epitheloid cells and accumulate, forming the granuloma.

INTRODUCTION

Histologically, there are four types of granulomas; categorizing a granuloma by type can at times aid in defining the cause of the inflammation that produced the granuloma. Noncaseating granulomas, the most common type found in liver biopsies, are found in a variety of disease states, including sarcoidosis, vasculitis, primary biliary cirrhosis (PBC), and infectious diseases. Caseating granulomas are commonly found in association with mycobacterial or fungal infections. Lipogranulomas are frequently associated with fatty liver or mineral oil ingestion. Fibrin-ring granulomas result from a number of infectious diseases, such as Q fever, or hepatitis A.

CAUSES OF HEPATIC GRANULOMAS

In over half the cases of identified hepatic granulomas, the patient experiences nonspecific clinical symptoms such as fever, weight loss, anorexia, or laboratory evidence of inflammation. Hepatomegaly is reported in about half the cases. Splenomegaly is rare, as is jaundice, which occurs only when associated with bile duct injury. Radiologic imaging of the liver is usually nonspecific, but at times granulomas can coalesce into small nodules that may be detectable by computer tomography (CT) scan or ultrasound. Therefore, a liver biopsy is needed to conclusively establish the presence of hepatic granulomas.

A comprehensive list of potential causes of hepatic granulomas is given in Table I. Essentially, the causes can be separated into five categories: systemic diseases, infectious diseases, malignancies, reactions to chemicals and drugs, and idiopathic.

Sarcoidosis is the most common systemic disease to cause hepatic granulomas, accounting for up to 30% in some case series. Most patients are asymptomatic; for those who have constitutional symptoms, corticosteroid therapy may improve symptoms, but improvement in liver size or liver tests varies. Liver diseases such as PBC, primary sclerosing cholangitis (PSC), and steatosis can be associated with granulomas, but their presence has variable prognostic significance.

Infectious diseases comprise the most diverse causes of hepatic granulomas. Many types of infectious organisms, including mycobacteria, bacteria, fungi, treponemes, rickettsia, viruses, metazoa, and protozoa, can result in hepatic granulomas. Many of these infectious agents are intracellular and cause an inflammatory response in the liver that results in granuloma formation. The most common infectious cause of hepatic granulomas worldwide is *Mycobacteria tuberculosis*. This usually occurs with the miliary form of tuberculosis, and can be associated with caseating necrosis. Throughout the developing world, secondary syphilis and schistosomiasis represent other common infectious diseases causing hepatic granulomas.

Hodgkin's lymphoma is the most common malignant disease associated with hepatic granulomas, which can occur in up to 20% of affected patients. Granulomas, however, are not indicative of malignant involvement of the liver. Chemical exposures, hypersensitivity drug reactions, steatosis, cholestasis, and hepatitis can also result in hepatic granulomas. Idiopathic granulomatous hepatitis, originally described

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Causes of Hepatic Granulomas

Cause	Disease/example
Systemic diseases	Sarcoidosis, Wegner's granulomatosis, temporal arteritis, polymyalgia rheumatica, erythema nodosum, allergic granulomatosis, Crohn's disease, ulcerative colitis, idiopathic hypogammaglobulinemia, chronic granulomatous disease, systemic lupus erythematosus Liver diseases: primary biliary cirrhosis, primary sclerosing cholangitis, hypersensitivity cholangitis, steatosis
Infectious diseases	Mycobacteria: tuberculosis, *Mycobacterium avium intracellulare*, atypical mycobacteria, leprosy Bacteria: brucellosis, tularemia, yersinosis, actinomycosis, *Bartonella henslae* (cat-scratch disease), listeriosis, meliodosis, *Tropheryma whippelii* (Whipple's disease), typhoid, *Rickettsia, Coxiella burnetti* (Q fever), boutonneuse fever, *Treponema* (syphilis) Fungi: blastomycosis, histoplasmosis, coccidioidomycosis, cryptococcosis, candidiasis, aspergillosis Viruses: Epstein–Barr virus, cytomegalovirus, hepatitis C virus, hepatitis A virus, Coxsackie Metazoa: schistosomiasis, larva migrans viscerale, fascioliasis Protozoa: leishmaniasis, toxoplasmosis
Malignancy	Hodgkin's lymphoma, non-Hodgkin's lymphoma, carcinoma
Chemicals/drugs	Hypersensitivity drug reaction, sulfonamides, isoniazid, allopurinol, beryllium, throrotrast, copper sulfate
Idiopathic	Idiopathic granulomatous hepatitis

to cause noncaseating granulomas associated with constitutional symptoms and no extrahepatic involvement, may respond to steroids and has a very favorable prognosis.

DIAGNOSTIC APPROACH

Once hepatic granulomas have been confirmed by pathologic review, the diagnosis can be established using a combination of clinical and laboratory data. The remainder of the biopsy specimen should be evaluated for signs of malignancy or other liver disease, such as PBC, PSC, or steatosis. Polymerase chain reaction (PCR) can be performed to isolate infections involving mycobacteria, *Tropheryma whippelii*, *Bartonella henslae*, Epstein–Barr virus (EBV), or cytomegalovirus (CMV). In addition, gram stains, fungal stains, and cultures from the biopsy and serum can be used to determine bacterial, mycobacterial, and fungal etiologies. The patient should also be questioned about drug history and chemical exposures and should be examined for signs of extrahepatic involvement that may suggest diseases such as systemic lupus, sarcoidosis, lymphoma, or inflammatory bowel disease. Laboratory tests and serologies should be performed to screen for liver diseases and for infectious agents. If no cause is found for the hepatic granuloma, the patient should be reevaluated 3 months later. The treatment of hepatic granulomas is directed at the cause of the liver inflammation.

See Also the Following Articles

Cholangitis, Sclerosing • Liver Biopsy • Lymphomas • Mycobacterial Infection

Further Reading

Denk, H., Scheuer, P., *et al.* (1994). Guidelines for the diagnosis and interpretation of hepatic granulomas. *Histopathology* **25**, 209–218.

Lefkowitch, J. (1999). Hepatic granulomas. *J. Hepatol.* **30**, 40–45.

Valla, D., and Benhamou, J. (2000). Hepatic granulomas and hepatic sarcoidosis. *Clin. Liver Dis.* **4**, 269–285.

Hepatitis A

JOHN J. POTERUCHA
Mayo Clinic

fecal–oral transmission Acquisition of an infection by ingestion of fecally contaminated material that contains the infectious agent.
parenteral transmission Acquisition of an infection by exchange of the infectious agent via blood or blood-contaminated body fluids.
prodrome Symptoms indicating the onset of a disease.

Hepatitis A virus is an RNA virus that is a frequent cause of acute hepatitis. The virus is generally spread by person-to-person contact although transmission from contaminated food and water also occurs. The disease is characterized by liver inflammation with increases in liver enzymes and often jaundice. Children with an acute infection have less severe disease manifestations than do adults with acute hepatitis A. Supportive care results in resolution in most patients. A safe and effective vaccine has been developed and is recommended for certain individuals at high risk for the disease or its complications.

INTRODUCTION

Hepatitis A virus (HAV) is the most frequent viral cause of acute hepatitis in the United States. The U. S. Centers for Disease Control and Prevention estimates that hepatitis A causes 180,000 acute infections per year. The average cost per acute hepatitis A case ranges from $1817 to $2459. Because HAV does not cause chronic liver disease and only rarely results in death from acute liver disease, there has been less emphasis in the literature on hepatitis A than on hepatitis B or C. However, HAV causes large outbreaks due to person-to-person spread and may result in fulminant hepatic failure.

VIROLOGY

HAV is a single-stranded RNA virus that is a member of the genus Hepatovirus of the family *Picornaviridae* ("small RNA virus"). HAV is 27 nm and is a spherical, nonenveloped virus with a capsid consisting of 32 subunits. The positive-stranded RNA is approximately 7.5 kb in length and includes a 5′-untranslated segment, a single open reading frame encoding a polyprotein of 2227 amino acids, and a 3′ noncoding region with a poly(A) terminus. The polyprotein is cleaved by a 3C proteinase into 11 structural and nonstructural proteins. Although different HAV strains can be determined using molecular techniques, there are no significant antigenic differences and thus there is only one viral serotype. Infection with one strain or vaccination imparts immunity to all strains of the virus. The ability to detect different viral genotypes can be useful epidemiologically to investigate outbreaks of hepatitis A.

Humans and certain nonhuman primates are the only natural hosts of HAV, although only humans seem to develop significant hepatitis. HAV is usually transmitted by the fecal–oral route. It is postulated that the virus is absorbed in the small intestine, passes via the portal vein to the liver, and is taken up by the hepatocyte. HAV replicates in the hepatocyte cytoplasm where antigen is detected 1 week after inoculation. Virions are then secreted into the bile canaliculus or released into the systemic circulation. Virus may also replicate in the crypt cells of the small intestine, which helps account for the large amount of virus present in feces within a few weeks after exposure. Viremia occurs although HAV concentrations are much lower in blood than in feces. Viral concentrations in blood tend to be highest in the prodromal and early clinical phases of the illness. Although the amount of virus in the feces decreases dramatically soon after the onset of the clinical illness, viral antigen can be found in the stool as long as 2 weeks after the onset of symptoms. Reverse transcription–polymerase chain reaction can detect viral RNA as long as 2 months after the onset of clinical symptoms although the significance of this finding in terms of transmission is not clear.

HAV likely induces hepatocyte injury by an immune mechanism rather than by a cytopathic effect. Large quantities of virus are produced in the liver before the onset of clinical evidence of hepatitis, and HAV in cell culture is not associated with hepatocyte damage. Virus-specific cytotoxic T cells have been identified and these cells secrete cytokines that likely are involved in the inflammatory response and in the resolution of the infection.

 Copyright 2004, Elsevier (USA). All rights reserved.

EPIDEMIOLOGY AND TRANSMISSION

Hepatitis A is a robust virus and the ability to persist in a variety of unfavorable environments allows the virus to have a high attack rate, with up to 90% of those exposed becoming infected. HAV is largely transmitted by person-to-person contact although outbreaks due to ingestion of contaminated food or water also occur. Close living quarters, such as day-care centers, institutions for the disabled, and military institutions, enhance the spread of the virus. The secondary household attack rate is approximately 20%. Because of fecal–oral transmission and lack of clinical symptoms of hepatitis A, children are most likely to spread the infection and numerous outbreaks of hepatitis A in day-care centers have been reported. A short duration of the viremia and the relatively low concentration of HAV in blood help make parenteral transmission of hepatitis A unusual although there have been reported outbreaks in intravenous drug users and recipients of contaminated blood products.

Risk factors for hepatitis A acquisition in the United States are personal contact with an infected patient (12–26% of infections), day-care contact (11–16%), international travel (4–6%), and recognized food- or water-borne outbreaks (2–3%). During recent outbreaks, up to 10% of reported cases occurred in illegal drug users and homosexual men. No risk factor can be identified in 50% of patients acquiring hepatitis A infection.

The overall prevalence of prior hepatitis A infection is determined by assaying for immunoglobulin G antibody against HAV (IgG anti-HAV). The seroprevalence within a region depends on the age of the patient and on living and sanitation conditions. Countries with lower socioeconomic levels and crowded living conditions, such as Ethiopia, are characterized by high transmission rates of hepatitis A such that over 90% of the population will have IgG anti-HAV by the age of 5 years. In the United States, some of the southwestern states have the highest seroprevalence rates of hepatitis A.

Improvements in sanitation, food handling, and control of water supplies have changed the frequency of IgG anti-HAV prevalence. The most recent data show that approximately one-third of the U. S. population has serologic evidence of a prior hepatitis A infection. The highest prevalence rates, approximately 75%, are seen in those >70 years of age. The higher prevalence of infection in older individuals is likely due to a cohort effect, reflecting the fact that the chance of infection during childhood was greater years ago than it has been more recently.

Food-borne outbreaks of HAV are relatively uncommon but have been reported with frozen strawberries, fresh produce, and seafood. Consumption of raw bivalves, such as oysters and clams, is an especially important source of transmission. HAV can survive for extended periods of time in seawater, with viral nucleic acids being detected months after being experimentally seeded in salt water. Bivalves filter large volumes of seawater, thereby concentrating the amount of ingested virus. In addition, bivalves are often harvested from shallow waters near populated areas where water is more likely to be contaminated by sewage.

DIAGNOSIS

The typical clinical and serologic course of acute hepatitis A is shown in Fig. 1. Because there is only one serotype of hepatitis A, one antibody test is sufficient for diagnosis. IgM anti-HAV has a sensitivity of 100%, a specificity of 99%, and a positive predictive value of 88%. Rarely, IgM anti-HAV may not be measurable early in the disease course and repeat testing is advised if clinical suspicion remains high. IgM anti-HAV usually clears by 4 months although it persists for more than 6 months in 25% of patients. IgG anti-HAV generally appears within a few months after infection and persists for life, offering immunity against further infection.

Aminotransferases are the most sensitive markers of hepatocyte injury in acute hepatitis A, initially becoming abnormal during the prodromal phase. The peak levels of aspartate aminotransferase and alanine aminotransferase are generally 1000–2000 IU/liter and fall approximately 10% per day but can take weeks or even months to return to normal. Bilirubin is often elevated, especially in older children and adults. Bilirubin tends to fall at a rate of approximately 50% per week but can remain elevated for months in patients with a prolonged cholestatic phase. Alkaline phosphatase is normal or only modestly elevated unless the patient enters a prolonged cholestatic phase of the illness.

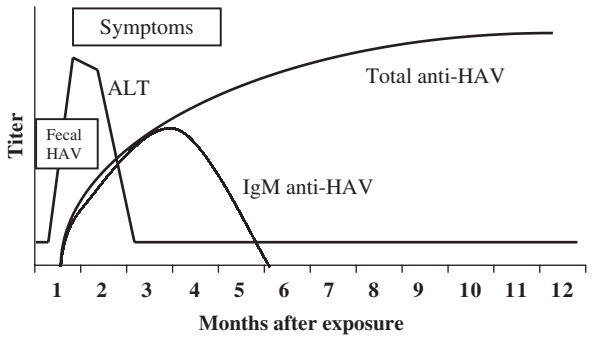


FIGURE 1 Clinical and serologic course of hepatitis A. ALT, alanine aminotransferase.

More severe episodes of hepatitis A are characterized by increases in prothrombin time and decreases in serum albumin and factor V. A mild polyclonal elevation in gamma globulin may be seen, although the elevations are less than would be seen in autoimmune hepatitis. Low-titer autoantibodies are occasionally identified in acute hepatitis A.

Liver biopsy is not necessary for diagnosis and should only rarely be performed in patients with hepatitis A. The typical features of acute hepatitis are hepatocellular degeneration with apoptotic change and variable degrees of portal and lobular inflammation. Histologic findings cannot differentiate between acute hepatitis A and acute hepatitis B. Using *in situ* hybridization techniques, HAV RNA can be detected in hepatocytes, sinusoidal cells, and inflammatory cells. The detection of viral RNA in phagocytic cells suggests that these cells are responsible for clearance of the virus, probably due to the uptake of viral–antibody complexes.

CLINICAL MANIFESTATIONS

The clinical manifestations of hepatitis A are variable, ranging from asymptomatic infection to life-threatening fulminant hepatic failure. Age at acquisition is an important factor in the severity and nature of symptoms of acute hepatitis A (see Fig. 2). Infection with hepatitis A in children less than 6 years of age leads to jaundice in fewer than 10% of cases although most patients have nonspecific symptoms such as fever, arthralgias, and diarrhea. Approximately 50% of individuals between the ages of 6 and 14 years develop jaundice with acute hepatitis A, whereas $>80\%$ of patients greater than 14 years of age develop jaundice. The nonspecific, nonicteric nature of symptoms of acute hepatitis A in children facilitates transmission of the virus because the need to address contacts is not recognized. Because of the decline in acquisition of hepatitis A during childhood, an increasing number of adults are less likely to have immunity and therefore more likely to develop severe disease at the time of infection. Pregnant women acquiring hepatitis A do not have a more severe disease course than other infected patients nor is there an increase in fetal loss.

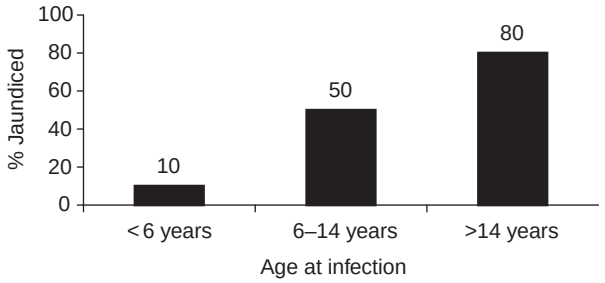


FIGURE 2 Prevalence of jaundice in acute hepatitis A varies with age at infection.

The incubation period of hepatitis A is 15–49 days with a mean of 30 days. There is a 5- to 7-day prodromal period characterized by anorexia, nausea, malaise, and occasionally fever, headache, and abdominal pain. When jaundice occurs, it happens within 2 weeks after the onset of the prodrome and is preceded by dark urine. After jaundice appears, the nonspecific symptoms often subside. The duration of jaundice is generally 5–30 days, with an average bilirubin level of 7 mg/dl. Approximately 10–15% of symptomatic persons have a prolonged illness characterized by either prolonged cholestasis or episodes of relapsing and remitting disease. Prolonged cholestasis can result in jaundice for up to 6 months. Relapsing hepatitis is associated with increases in aminotransferases and reappearance of viral particles in the stool. HAV does not cause chronic hepatitis although acute hepatitis A infection has been reported to precipitate autoimmune hepatitis.

Fulminant hepatitis A, by definition a severe hepatitis with portal systemic encephalopathy, rarely occurs. The frequency is difficult to quantify but it is estimated that the fatality rate of all hepatitis A cases is 1% and most of this mortality is due to fulminant hepatitis A. As might be expected, fatality rates vary with age. Infants and children have a fatality rate of 0.1% and those over 40 years of age have a fatality rate of 1.1%. In a recent U. S. study of fulminant hepatic failure, hepatitis A was the etiology in 7% of cases. The same investigators also suggested that the recovery from fulminant hepatitis A is 69%, significantly higher than previously thought and higher than in patients with fulminant hepatitis B.

Fulminant hepatitis A may be more common in patients who have chronic hepatitis C prior to the acquisition of hepatitis A. One study showed than 7 of 17 (41%) chronic hepatitis C patients acquiring hepatitis A developed fulminant hepatitis compared to 0 of 10 chronic hepatitis B patients developing acute HAV. It is recommended that all patients with chronic liver disease be vaccinated against hepatitis A if not already immune.

Extrahepatic manifestations of acute hepatitis A are unusual. One study reported that hemolysis and acalculous cholecystitis were the most common extrahepatic manifestations. Renal failure, pancreatitis, and manifestations of circulating immune complexes may also be seen.

MANAGEMENT

The large majority of patients with acute hepatitis A have a self-limited infection and therefore specific therapy is not needed. Attention to supportive care, especially hydration, is most important. Bed rest is not necessary although fatigue may limit certain activities. Hospitalization should be reserved for those patients who need intravenous fluids or show signs of developing encephalopathy. Corticosteroids are not beneficial and should generally be avoided. Patients with fulminant hepatic failure should be managed similar to any patient with acute liver failure and occasionally liver transplantation is necessary. The United States Fulminant Hepatitis Study Group data would suggest that most patients with acute hepatitis A resolve and therefore a slightly more conservative approach to liver transplantation than other etiologies of FHF is suggested.

PREVENTION

Because of its mode of transmission, person-to-person spread of HAV is common. Immune globulin is a preparation of concentrated antibodies made from pooled human plasma. Immune globulin provides protection against hepatitis A by passive transfer of antibody. A dose of 0.02 ml/kg provides preexposure protection for up to 3 months; 0.06 ml/kg provides protection for up to 5 months. Recently, IgG anti-HAV titers in immune globulin have decreased due to the declining prevalence in the donor population; however, there has been no definite documentation of declining efficacy. Postexposure immune globulin is useful to prevent infection if used within 2 weeks after exposure. Administration of immune globulin more than 2 weeks after exposure is ineffective. Immune serum globulin should be given to household, sexual, and day-care contacts of infected patients. If a common source food-borne outbreak is identified, immune serum globulin is also recommended for those at risk. Despite being protective, immune globulin administration does not achieve measurable levels of anti-HAV. Simultaneous vaccination with the hepatitis A vaccine is also recommended for individuals requiring immune globulin. One study suggested that the use of the hepatitis A vaccine only was effective in preventing transmission although that has not yet been accepted into clinical practice.

Hepatitis A vaccine became available in 1995 thanks to the use of cell culture techniques. HAVRIX and VAQTA, the two commercially available hepatitis A vaccines, are very effective and probably approximately equivalent in immunogenicity. Both use a formalin-inactivated, laboratory-attenuated strain of HAV and are available in two different formulations, the choice of which depends on age of the vaccinee. Both use a two-dose schedule with the second dose at least 6 months after the initial dose. The vaccines result in protective anti-HAV antibodies (IgG type) in over 90% of adults 1 month after the first dose. Approximately half of vaccinated patients are positive 14 days after the initial dose. The second dose of the vaccine ensures that nearly all patients will have protective levels of antibody. Response rates in children are even higher than in adults. Protective antibody levels seem to persist for at least 8 years after successful vaccination. The most common side effects of HAV vaccine are soreness at the infection site, headache, and malaise. Vaccination is recommended for groups at high risk of hepatitis A acquisition including intravenous drug users, day-care employees, homosexual men, and frequent recipients of blood products. Vaccine is also recommended for those in whom hepatitis A may cause a severe outcome, such as patients with chronic liver disease. Immunogenicity of the vaccine appears to be slightly less in patients with chronic liver disease than in the healthy population and, occasionally, immunity may be lost after liver transplantation. Immunosuppressed patients, such as those who have undergone organ transplantation or who have been infected with human immunodeficiency virus, also have lower immunogenicity with the HAV vaccine.

In large part due to the high response rates to the HAV vaccine, the role of prevaccination testing for anti-HAV has not yet been clarified. Some researchers have suggested that prevaccination testing for individuals at high risk of prior HAV infection is reasonable. Adults for whom prevaccination serologic testing can be considered include adults who were born or lived for a long time in areas with a high endemicity for hepatitis A, persons with a history of infection drug use, and perhaps individuals >40 years old. Because of the high rate of immunogenicity with the hepatitis vaccine, postvaccine testing is not recommended. Booster doses of the vaccine have not yet been recommended because of the long duration of protective antibody after the vaccination series.

See Also the Following Articles

AIDS, Hepatic Manifestation of • Bilirubin and Jaundice • Foodborne Diseases • Fulminant Hepatic Failure • Sexually Transmitted Diseases

Further Reading

Centers for Disease Control and Prevention (1999). Prevention of hepatitis A through active or passive immunization: Recommendations of the Advisory Committee on Immunization Practices. *Morbid. Mortal. Weekly Rep.* **48**, 1–38.

Cuthbert, J. A. (2001). Hepatitis A: Old and new. *Clin. Microbiol. Rev.* **14**, 38–58.
Hutin, Y. J., Pool, V., Cramer, E. H., *et al.* (1999). A multistate, foodborne outbreak of hepatitis A. *N. Engl. J. Med.* **340**, 595–602.
Sagliocca, L., Amorosoa, P., Stroffini, T., *et al.* (1999). Efficacy of hepatitis A vaccine in prevention of secondary hepatitis A infection: A randomized trial. *Lancet* **353**, 1136–1139.
Schiodt, F. V., Atillasoy, E., Shakil, A. O., *et al.* (1999). Etiology and outcome for 295 patients with acute liver failure in the United States. *Liver Transplant. Surg.* **5**, 29–34.
Schiodt, F. V., Davern, T. J., Shakil, A. O., *et al.* (2003). Viral hepatitis-related acute liver failure. *Am. J. Gastroenterol.* **98**, 448–453.
Sjogren, M. H. (1998). Preventing acute liver disease in patients with chronic liver disease. *Hepatology* **27**, 887–888.
Vento, S., Garofano, T., Renzini, C., *et al.* (1998). Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C. *N. Engl. J. Med.* **338**, 286–290.
Weitz, M., and Giegl, G. (1998). Hepatitis A virus: Structure and molecular virology. *In* "Viral Hepatitis" (A. Zuckerman, and H. Thomas, eds.), 2nd Ed., pp. 15–27. Churchill Livingstone, Philadelphia, PA.
Willner, I. R., Uhl, M. D., Howard, S. C., *et al.* (1998). Serious hepatitis A: An analysis of patients hospitalized during an urban epidemic in the Untied States. *Ann. Intern. Med.* **128**, 111–114.
Zachoval, R., and Deinhardt, F. (1998). Hepatitis A virus: Natural history and experimental models. *In* "Viral Hepatitis" (A. Zuckerman, and H. Thomas, eds.), 2nd Ed., pp. 43–57. Churchill Livingstone, Philadelphia, PA.

Hepatitis B

T. SHAW AND S. A. LOCARNINI
Victorian Infectious Diseases Reference Laboratory, Australia

anti-HBs, anti-HBc, anti-HBe Specific antibodies that are produced in response to their respective antigenic determinants.
hepatitis B core antigen The antigenic specificity associated with the 32 nm diameter core of hepatitis B virus.
hepatitis B e antigen The antigenic determinant that circulates as a soluble protein in serum.
hepatitis B surface antigen The antigenic determinant found on the surface of the hepatitis B virus and on the 22 nm diameter particles and filamentous forms.
hepatitis B virus The 42 nm diameter double-shelled particle, originally known as the Dane particle.
hepatitis B virus DNA The partially double-stranded circular hepatitis B virus genome.
quasispecies Heterogeneous populations of genetic variants of virus found in an infected individual.

Hepatitis B, which may be acute or chronic, is caused by infection with human hepatitis B virus (HBV), a small hepatotropic DNA virus that is distributed worldwide. Infection with HBV occurs mainly as a consequence of exposure to infected blood or body fluids. Up to 5% of cases of chronic hepatitis B progress to fatal liver disease. Effective anti-HBV drugs and vaccines have been developed, but economic and logistical obstacles prevent their use in poor countries where chronic HBV infection is endemic. Elimination of HBV is a long-term goal that will be difficult to achieve unless current paradigms of control are improved or modified.

HISTORY OF DISCOVERY

Epidemic jaundice or infectious icterus has been recognized for at least 2000 years. Outbreaks in the past were probably caused mainly by fecal–oral transmission of hepatitis A virus. Hepatitis transmitted by blood and body fluids was not recognized until the mid 19th century and it was not until the mid 20th century that infectious hepatitis (hepatitis A) and serum hepatitis (hepatitis B and "non-A, non-B" hepatitis) were shown to be caused by different viral agents. Hepatitis B surface antigen (HBsAg) was first discovered in the serum of an Australian aborigine in 1963. It was subsequently found to be associated with acute serum hepatitis in other populations worldwide and shown to be a viral

Copyright 2004, Elsevier (USA). All rights reserved.

Cuthbert, J. A. (2001). Hepatitis A: Old and new. *Clin. Microbiol. Rev.* **14**, 38–58.
Hutin, Y. J., Pool, V., Cramer, E. H., *et al.* (1999). A multistate, foodborne outbreak of hepatitis A. *N. Engl. J. Med.* **340**, 595–602.
Sagliocca, L., Amorosoa, P., Stroffini, T., *et al.* (1999). Efficacy of hepatitis A vaccine in prevention of secondary hepatitis A infection: A randomized trial. *Lancet* **353**, 1136–1139.
Schiodt, F. V., Atillasoy, E., Shakil, A. O., *et al.* (1999). Etiology and outcome for 295 patients with acute liver failure in the United States. *Liver Transplant. Surg.* **5**, 29–34.
Schiodt, F. V., Davern, T. J., Shakil, A. O., *et al.* (2003). Viral hepatitis-related acute liver failure. *Am. J. Gastroenterol.* **98**, 448–453.
Sjogren, M. H. (1998). Preventing acute liver disease in patients with chronic liver disease. *Hepatology* **27**, 887–888.
Vento, S., Garofano, T., Renzini, C., *et al.* (1998). Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C. *N. Engl. J. Med.* **338**, 286–290.
Weitz, M., and Giegl, G. (1998). Hepatitis A virus: Structure and molecular virology. *In* "Viral Hepatitis" (A. Zuckerman, and H. Thomas, eds.), 2nd Ed., pp. 15–27. Churchill Livingstone, Philadelphia, PA.
Willner, I. R., Uhl, M. D., Howard, S. C., *et al.* (1998). Serious hepatitis A: An analysis of patients hospitalized during an urban epidemic in the Untied States. *Ann. Intern. Med.* **128**, 111–114.
Zachoval, R., and Deinhardt, F. (1998). Hepatitis A virus: Natural history and experimental models. *In* "Viral Hepatitis" (A. Zuckerman, and H. Thomas, eds.), 2nd Ed., pp. 43–57. Churchill Livingstone, Philadelphia, PA.

Hepatitis B

T. SHAW AND S. A. LOCARNINI
Victorian Infectious Diseases Reference Laboratory, Australia

anti-HBs, anti-HBc, anti-HBe Specific antibodies that are produced in response to their respective antigenic determinants.
hepatitis B core antigen The antigenic specificity associated with the 32 nm diameter core of hepatitis B virus.
hepatitis B e antigen The antigenic determinant that circulates as a soluble protein in serum.
hepatitis B surface antigen The antigenic determinant found on the surface of the hepatitis B virus and on the 22 nm diameter particles and filamentous forms.
hepatitis B virus The 42 nm diameter double-shelled particle, originally known as the Dane particle.
hepatitis B virus DNA The partially double-stranded circular hepatitis B virus genome.
quasispecies Heterogeneous populations of genetic variants of virus found in an infected individual.

Hepatitis B, which may be acute or chronic, is caused by infection with human hepatitis B virus (HBV), a small hepatotropic DNA virus that is distributed worldwide. Infection with HBV occurs mainly as a consequence of exposure to infected blood or body fluids. Up to 5% of cases of chronic hepatitis B progress to fatal liver disease. Effective anti-HBV drugs and vaccines have been developed, but economic and logistical obstacles prevent their use in poor countries where chronic HBV infection is endemic. Elimination of HBV is a long-term goal that will be difficult to achieve unless current paradigms of control are improved or modified.

HISTORY OF DISCOVERY

Epidemic jaundice or infectious icterus has been recognized for at least 2000 years. Outbreaks in the past were probably caused mainly by fecal–oral transmission of hepatitis A virus. Hepatitis transmitted by blood and body fluids was not recognized until the mid 19th century and it was not until the mid 20th century that infectious hepatitis (hepatitis A) and serum hepatitis (hepatitis B and "non-A, non-B" hepatitis) were shown to be caused by different viral agents. Hepatitis B surface antigen (HBsAg) was first discovered in the serum of an Australian aborigine in 1963. It was subsequently found to be associated with acute serum hepatitis in other populations worldwide and shown to be a viral

Copyright 2004, Elsevier (USA). All rights reserved.

antigen. In 1970, Dane and colleagues identified virus-like particles that carried HBsAg on their surface in sera from hepatitis patients. Three years later, these "Dane" particles were shown to contain endogenous DNA polymerase activity, confirming their viral nature. Similar viruses were found in other animals during the next decade and have been grouped into a new virus family, the *Hepadnaviridae* (hepatotropic DNA viruses).

TAXONOMY AND CLASSIFICATION

In addition to being found in humans, rodents, and birds, hepatitis B viruses (HBVs) have recently been found in apes and primates. HBVs that infect mammals and birds are classified in separate genera, Orthohepadnavirus and Avihepadnavirus, respectively. All share common features including tissue tropism, restricted host range, virion ultrastructure, genetic organization, and replication strategy. The genome organization and replication strategy of hepadnaviruses are similar to those of retroviruses of animals and caulimoviruses of plants, suggesting a possible ancient evolutionary origin from a common ancestor. Replication of these viruses requires host-cell-mediated transcription and translation of a pregenomic viral RNA, which directs the synthesis of circular double-stranded viral DNA from which viral mRNA is transcribed.

PREVALENCE AND EPIDEMIOLOGY

Approximately 2 billion individuals—approximately one-third of the global population—have serologic evidence of past or present infection with HBV. By the end of the year 2000, nearly 400 million people were chronically infected with HBV according to World Health Organization (WHO) estimates. Chronic HBV infection is associated with a 15–25% chance of development of fatal liver disease, including cirrhosis and hepatocellular carcinoma. HBV-related disease causes approximately 1 million deaths annually. Considering these figures and the availability of safe and effective vaccines, the WHO has recommended that hepatitis B vaccination be incorporated into routine infant and childhood immunization programs in all countries. Chronically and acutely infected individuals are the only sources of HBV for new infections. In most populations, males are more likely to become chronic HBV carriers than females.

Transmission

HBV can be transmitted by sexual contact, percutaneously (for example, from contaminated needles shared by intravenous drug users), by inoculation with contaminated blood or blood products, by transplant of organs from infected donors, and perinatally from infected mothers. Infection of infants of infected mothers *in utero* is rare. Serum HBsAg and HBV DNA are considered reliable indicators of active infection, but the risk of infection from infected carriers is extremely variable because of large differences in HBV replicative activity, which results in enormous differences in viral loads in infected individuals. However, the presence of HBsAg is still regarded as a marker of infectivity irrespective of viral load.

PATHOGENESIS

HBV is not directly cytopathic to hepatocytes in cell culture systems or *in vivo*. Liver damage results mainly from the direct attack of the host's immune system on infected hepatocytes and indirect effects of the immune response on uninfected cells. Although this has been recognized for some time, concepts of the mechanisms responsible have changed radically as a result of recent studies. Noncytocidal effects of pro-oxidative cytokines, in particular, interferons α, β, and γ and tumor necrosis factor α, have been implicated as major contributors to viral clearance, whereas destruction of infected hepatocytes by cytotoxic T cells contributes to both viral clearance and the development of liver disease. Immune complex-mediated responses are believed to be responsible for extrahepatic damage when it occurs.

DIAGNOSIS

Acute and chronic HBV infection can be diagnosed and differentiated by simple serological tests that detect the presence of HBV antigens and antibodies to HBV antigens in blood. Interpretation of the results of these tests is summarized in Table I. Fluctuations in serum HBV DNA, which can be monitored by hybridization or polymerase chain reaction assays, provides a direct measurement of viral replicative activity. Leakage of liver enzymes [usually aminotransferases, in particular, alanine aminotransferase (ALT) and γ-glutamyl transferase] and bilirubin into serum, which can be assayed by biochemical tests, provides surrogate markers of liver damage. Histological examination of liver biopsies may be used to determine the stage and severity of tissue damage. More sophisticated assays may be used to monitor disease progression and response to treatment.

TABLE I Interpretation of Serological Markers in Patients with Viral Hepatitis B

Assay results			
HBsAg	**Anti-HBs**	**Anti-HBc**	**Usual interpretation**
+	−	−	Early acute HBV infection
+	+/−	+	HBV infection, either acute or chronic. Differentiate with IgM anti-HBc; determine amount of replicative activity (infectivity) from HBV DNA assay
−	+	+	Indicates previous infection and immunity to hepatitis B in most cases
−	−	+	Possibilities include: HBV infection in the remote past; "low-level" HBV carrier; "window" between disappearance of HBsAg and appearance of anti-HBs or false-positive or nonspecific reaction. Investigate with IgM anti-HBc and/or challenge with HBsAg vaccine; when present, anti-HBe helps validate the anti-HBc reactivity; assay for HBV DNA by PCR can identify "occult" infection
−	+	−	Vaccine-type response

Acute HBV Infection

In a typical case of acute hepatitis B, HBsAg is first detected 2−3 months after initial infection, which is 1−3 months before the serum ALT starts to rise and 3 to 5 weeks before the onset of symptoms such as dark urine and jaundice. The serum HBsAg reaches a peak concentration during the early acute stage of the illness and then slowly declines to undetectable levels within 4−6 months. Viremia (HBV DNA) is detected during and after the disappearance of HBsAg. Peak levels of HBV DNA exceeding 10^9 genome equivalents per milliliter are found in the late prodromal and early acute phases, reflecting the highly infectious nature of acute HBV infection. Serological markers HBeAg and anti-HBe specific IgM are typically detected after the appearance of HBsAg in blood. High levels of anti-HB-specific IgM are a useful serological test in acute hepatitis B. During early stages of convalescence and before HBsAg disappears, anti-HBe replaces HBeAg, signaling a reduction in viral replication and the beginning of resolution of the disease. Termination of acute HBV infection occurs with the disappearance of HBsAg and the development of anti-HBs.

Chronic HBV Infection

Chronic HBV infection is defined as hepatic necroinflammation due to the persistence of active HBV replication. The WHO defines chronic HBV on the basis of laboratory markers and requires the detection of HBsAg in the patient's serum on two occasions over a 6-month period. Typically, anti-HBc IgM titers are either low or undetectable, whereas anti-HBc IgG titers are high.

NATURAL HISTORY OF INFECTION

The outcome of HBV infection varies enormously. In areas of high endemicity (HBsAg prevalence >8%), HBV infection occurs early in life, usually perinatally, typically remains silent or subclinical, and frequently becomes chronic. In general, chronic infection is only established after maternal transmission of $HBeAg^+$ HBV. Neonates infected from $HBeAg^-$ mothers typically develop acute hepatitis that resolves. Under these conditions, chronic liver disease or liver cancer due to HBV infection appears much later in life and is frequently fatal. By contrast, in areas where the incidence of HBV infection is low (HBsAg prevalence <2%), most primary HBV infections occur in adults as a result of sexual contact or percutaneous infection. Such infections are more likely to be associated with clinically apparent acute hepatitis and less than 10% of cases become chronic. The incubation period from infection to onset of clinically apparent hepatitis is usually between 6 and 24 weeks.

$HBeAg^-$ Chronic Hepatitis B

$HBeAg^-$ chronic hepatitis B occurs as a result of the selection of HBV mutants that are unable to secrete the pre-core protein (HBeAg). It has become the most prevalent form of disease presentation in many parts of the world, particularly in Asian and Mediterranean countries. The most common of several mutations that can cause HBeAg negativity is a guanine to adenine transition at nucleotide position 1896 (G1896A), which creates a TAG stop codon at codon 28 of the pre-core protein. The phase of $HBeAg^-$ chronic hepatitis B

appears to be associated with rapid disease progression, resulting in cirrhosis and hepatocellular carcinoma.

VIRION ULTRASTRUCTURE

The HBV virion is a 42 nm spherical particle within a lipid envelope. The 32 nm diameter nucleocapsid contains the core antigen (HBcAg), the genomic DNA, and viral polymerase. During persistent infection, very high concentrations of incomplete viral forms, mainly aggregates of 22 nm diameter spherical particles and filamentous forms of HBsAg, are found in the blood.

GENOME ORGANIZATION

The HBV genome is a circular, partly double-stranded DNA molecule. The longer minus-sense strand is approximately 3200 nt long and is partly matched by a complementary plus strand, which varies in length from 1700 to 2800 nt. The longer strand is nicked at a point 300 nt from the 5′-end of the shorter strand. Circular conformation is maintained by base pairing between the 5′-ends of the strands, which overlap. The 5′-end of the shorter strand is capped with an oligoribonucleotide and the terminal part of the viral polymerase protein is covalently linked to the 5′-terminus of the longer strand.

Overlapping Reading Frames

HBVs encode only four genes and have evolved a remarkable genetic economy by using gene overlap: the HBV genome is only slightly longer than the longest of its four open reading frames (ORFs). The longest ORF, P, encodes a multifunctional protein that possesses RNase H and RNA- and DNA-dependent DNA polymerase activities as well as the ability to prime reverse transcription and coordinate virion assembly. The envelope (S) gene ORF lies completely within the P gene in a different reading frame and the remaining two genes, C and X, also partially overlap the P gene. Further economy is achieved by the use of alternative starting codons for transcription: the S gene encoded three co-terminal HBs proteins (Pre-S1, Pre-S2, and S). The C gene encodes both HBeAg and HBcAg and the X gene may encode more than one product. Gene overlap undoubtedly places constraints on mutation and evolution rates.

REPLICATION STRATEGY

After attachment and entry into the cell, the virus is uncoated and the viral genome is transported to the nucleus, where it is converted into a supercoiled, covalently closed circular (ccc) molecule, probably by cellular DNA repair enzymes. The supercoiled ccc DNA associates with cellular histones to form a viral minichromosome in the nucleus of infected cells. It is transcribed by cellular RNA polymerase II, producing several unspliced, capped, polyadenylated mRNA species that are translated into viral proteins. Transcription also generates a terminally redundant 3.5 kb pregenomic RNA that is packaged into viral core particles after it is translated to generate the viral polymerase. Within the capsid, the viral polymerase reverse-transcribes the pregenomic RNA, digests the RNA template, and copies the first (minus) DNA strand to regenerate the circular, partially double-stranded viral genome. Virions are enveloped and exported by budding through the endoplasmic reticulum. (See Fig. 1.)

HBV Quasispecies

RNA polymerases and reverse transcriptases are inherently error-prone, which ensures that HBV populations exist *in vivo* as heterogeneous mixtures of variants, termed quasispecies. A variety of selection pressures, including host immune surveillance, antiviral therapies, and virus–cell tropism, can affect quasispecies composition and influence the outcome of infection. A high quasispecies variability allows flexible but unpredictable evolution. Under some conditions, variation may exceed the error threshold beyond which viral survival is impossible, a situation termed “error catastrophe.” Overlap in reading frames of the polymerase, core, Pre-S/S, and X gene products probably places constraints on the number of variants that remain viable. Despite this, several viable variants of HBV that have characteristic mutations affecting the pre-core, core, polymerase, and X peptides have been described. Specific mutations that affect the polymerase may confer resistance to antivirals (see below), but the possible association between particular clinical manifestations and disease outcomes has not been firmly established for mutations that affect the other gene products.

EVOLUTION OF HBV

Until recently, it was generally believed that infection with HBV was restricted to humans, despite evidence for the presence of HBV antigens in other primates. HBVs that infect Old World apes (orangutan, gibbon, gorilla, and chimpanzee) and a New World monkey have been isolated and characterized during the past 5 years. The study of HBV evolution remains difficult because of complications introduced by

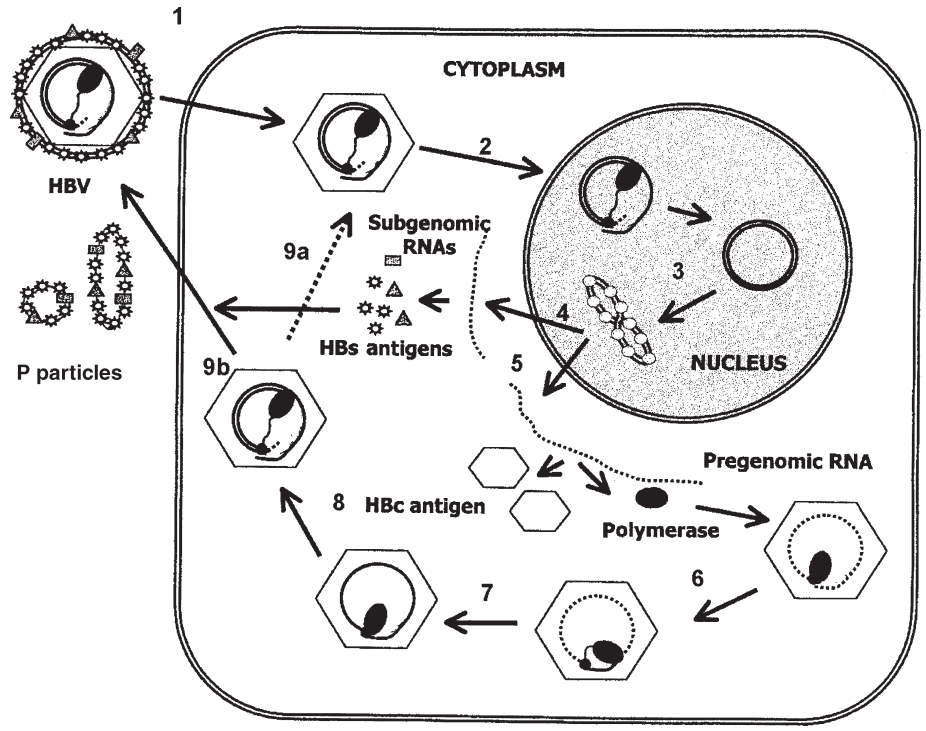


FIGURE 1 Stages in the HBV replication cycle (simplified). (1) Attachment and entry into the host cell followed by uncoating. (2) Transport to the host cell nucleus. (3) Repair of and conversion into a circular double-stranded supercoiled DNA molecule, which assembles a minichromosome. (4) Transcription of viral genes from HBV ccc DNA and export of polyadenylated subgenomic mRNAs and pregenomic RNA to the cytoplasm. (5) Translation of viral RNAs, generating the viral polymerase and structural proteins. (6) Assembly of nucleocapsid particles from core antigen, polymerase, and pregenomic RNA. (7) Covalent attachment of dNTPs to the viral polymerase to form a DNA primer and reverse transcription of the viral pregenome by extension from the DNA primer, generating the first (minus) DNA strand. (8) Degradation of the pregenomic template RNA and synthesis of a partial complementary second DNA strand. There are two alternative fates for mature intracellular nucleocapsids: delivery to the host nucleus and further replication (9a) or acquisition of a viral envelope and secretion from the cell via the endoplasmic reticulum and Golgi apparatus (9b). Aggregates of surface antigens may also be secreted. The translation and fate of nonstructural viral proteins are not shown.

intergenotype recombination, by constraints on nucleotide substitution imposed by gene overlap, and by differential selection pressures affecting different parts of the genome. Phylogenetic analyses provide evidence for a diversion of human genotypes A–E from genotype F and H and hepadnaviruses of apes. Recent analyses have also distinguished distinct patterns of recombination between genotypes A and D and between genotypes B and C and have resulted in a estimated mean nucleotide substitution rate in HBV of 4.2×10^{-5} nt/site/year, which suggests that human and ape HBVs diverged relatively recently. However, the direction, frequency, and sequence of early transmission events that generated different primate HBVs remain controversial.

HBV Genotypes

Eight HBV genotypes (named A–H) have been defined on the basis of nucleotide sequence divergences of >8%. The geographical distributions of genotypes are distinct and each is associated with different disease manifestations. Genotypes A to F occur globally, genotypes B and C predominate in Eastern Asia and Oceania, genotypes E is restricted to West Africa, and genotypes F and H are found in Central and South America, respectively. Genotype G was found in a small percentage of HBV-infected individuals in both France and Georgia (United States); it contains a unique 36 bp insertion in the core gene and two stop codons in the precore region.

Although the latter prevent the translation of HBeAg, HBV/G-infected individuals have all been shown to be $HBeAg^+$ due to co-infection with HBV/A, and HBV/A/G recombinants have also been detected *in vivo*. How co-infection occurs, the significance of recombination, and whether replication of HBV/G requires HBV/A are currently unknown. Phylogenetic trees constructed from either full-length genome sequences or individual gene sequences show that genotype G is clearly separated the other seven genotypes.

Clinical and Epidemiological Significance of Genotypes

Sequence differences between different HBV genotypes may lead to structural differences at the pregenome level, which may in turn affect translation and reverse transcription. There is accumulating evidence to suggest that clinical manifestations of infection, treatment response, and long-term prognosis of infection may all be influenced by genotype. Genotyping has also proved to be useful for tracing routes of transmission. A variety of facile and reliable methods for differentiating between different HBV genotypes have recently been developed.

PREVENTION AND CONTROL

Vaccination

Conventional Vaccines

Current HBV vaccines consist of recombinant HBsAg protein produced in yeast. Intramuscular injection of three successive doses induces protective levels of anti-HBsAg (>10 mIU/ml) in >95% of infants and children and >90% of vaccinated healthy adults, but response rates are lowered by smoking, obesity, diabetes, renal impairment, immune deficiency, and advancing age. Vaccination of all newborn infants is recommended, as is vaccination of previously unimmunized children and adolescents, as well as adults who belong to high-risk groups. High-risk groups include health care workers, injecting drug users, persons with multiple sexual contacts, partners and family members of individuals known to be infected with HBV, long-term residents of institutions, patients with chronic renal failure, recipients of clotting factor concentrates, and travelers intending to stay in regions where HBV is endemic. In high-risk groups, postvaccination testing for the presence of anti-HBs is advisable, but it is regarded as unnecessary for adolescents, children, and infants. A second three-dose course of injections is usually recommended for those who fail to respond to a primary course of vaccination. Up to 50% of nonresponders respond to a single vaccine dose and up to 75% respond to three additional doses. Responders are protected against HBV infection even after HBs antibody titers fall to undetectable levels and, except in hemodialysis patients, revaccination to boost immunity is usually not required.

Safety of Vaccines

HBV vaccines have an outstanding record of safety and efficacy. Since 1982, more than a billion doses of HBV vaccine have been administered worldwide. In countries where routine infant immunization programs have been introduced, their introduction has not been associated with increases in adverse effects and no evidence has been found to substantiate anecdotal reports of associations between HBV vaccination and a variety of syndromes and diseases in adults. Nonfatal anaphylactic reactions have been reported to occur at a frequency of approximately 1 in 600,000 vaccinations and further vaccination is obviously contraindicated in such cases.

National Vaccination Programs

Since 1991, the WHO has recommended that HBV immunization be included in national immunization programs worldwide. By March 2001, 116 countries had followed this recommendation, including most countries in Eastern and Southeast Asia, the Pacific Islands, Australia, North and South America, Western Europe, and the Middle East. Pediatric and childhood immunization programs in countries such as Taiwan, where the chronic infection frequency was previously high (up to 20% HBsAg prevalence), have reduced the childhood carrier frequency to approximately 1% within the past decade. Unfortunately, many low-income countries in sub-Saharan Africa and the Indian subcontinent and newly independent states have been unable to introduce routine HBV vaccination due mainly to cost and the lack of organizational infrastructure needed to introduce routine vaccination.

Vaccines for the Future

Vaccines that can be administered orally or intranasally are the most economically feasible for mass immunization of humans on a global scale, but very few vaccines are currently available for administration by these routes. Now that it is possible to produce genetically engineered crop plants that express biologically active proteins, efforts have been directed at producing edible plants that express HBsAg. Results of pilot experiments that show HbsAg-specific IgG production in laboratory animals and human volunteers following ingestion of transgenic plants that expressed HBsAg

are encouraging and suggest that mass immunization using dietary staples as vectors may become a reality. An alternative and equally ingenious approach involves the use of live attenuated *Salmonella typhimurium* genetically engineered to produce HBsAg as vector. Although this vaccine is only at the developmental stage, strong cellular and relatively insignificant humoral immune responses observed after oral administration of this "enterobacterial" vaccine suggest that it may be useful in some situations in which the conventional recombinant vaccines are ineffective, such as therapeutic immunization of chronic carriers.

Chemotherapy

Problems Associated with HBV Chemotherapy

Development of chemotherapy for HBV has been difficult for a variety of reasons. HBV infection has remained virtually untreatable for most of the >30 years that have passed since it was first identified as a major cause of serum hepatitis. Because the HBV genome does not encode virus-specific enzymes for (deoxy)-nucleoside salvage, its replication depends on host cell enzymes for its supply of DNA precursors. Furthermore, its unusual strategy of replication via an RNA intermediate means that transcription and translation of the RNA intermediate can continue even when reverse transcription and duplication of the resulting DNA transcript are completely blocked. Nucleoside analogues that inhibit the HBV polymerase effectively have no direct effect on the viral minichromosome, which can be removed only by the destruction of infected hepatocytes. Chemotherapeutic strategies that rely solely on inhibitors of the viral polymerase must inevitably be virustatic rather than virucidal, and, not surprisingly, HBV replication invariably resumes following withdrawal of treatment. Prolonged treatment carries disadvantages besides cost and the risk of cumulative toxicity. Antiviral monotherapy (continued exposure to a single drug) selects for mutations that cause drug resistance, typically within 12 months of initiation of treatment.

Currently Approved Drugs

Interferon-α Until very recently, interferon-α was the only drug licensed for use against chronic HBV infection. Although a subgroup of patients with active liver disease (indicated by elevated serum transaminase levels) and low-level viremia respond favorably to interferon treatment, overall response rates are poor. Individuals who belong to certain ethnic groups as well as those who are co-infected with hepatitis delta virus (HDV) or human immunodeficiency virus (HIV) or who are otherwise immunocompromised usually fail to respond. Furthermore, interferon is generally administered by subcutaneous or intramuscular injection, which is inconvenient, and its use is associated with several undesirable side effects. Thus, the need for more efficacious and better-tolerated orally available anti-HBV agents has been evident for some time. Two safe and efficacious new anti-HBV drugs are now available and several more are at advanced stages of clinical development.

Lamivudine Lamivudine, the β-L-[−]enantiomer of 2′,3′-dideoxy-3′-thiacytidine, is an analogue of deoxycytidine. It is activated intracellularly by phosphorylations catalyzed by cellular enzymes. The triphosphate inhibits HBV polymerase activity by terminating newly synthesized viral DNA chains. Originally developed for treatment of HIV infection, lamivudine shows good oral bioavailability and is well tolerated, even when given for long periods. Treatment with lamivudine at the usual dose of 100 mg/day produces a rapid decrease in serum HBV DNA, clearance of HBeAg, improvement in liver histology, and normalization of serum transaminases in many cases. Unfortunately, drug-resistant HBV mutants appear almost invariably during long-term (>12 months) lamivudine treatment (see below) and their appearance is associated with further disease progression.

Adefovir dipivoxil Adefovir dipivoxil is a prodrug for the broad-spectrum antiviral deoxyadenosine monophosphate analogue adefovir (9-[2-phosphonylmethoxyethyl]-adenine). Large-scale phase III clinical trials have shown that treatment with adefovir dipivoxil at a dose of 10 mg/day produces antiviral effects comparable to those of lamivudine. Resistance to adefovir has not been observed during 48-week clinical trials and it is active against lamivudine-resistant HBV mutants. Adefovir dipivoxil has been approved for use as an anti-HBV agent by the Food and Drug Administration in the United States and is expected to gain approval from regulatory authorities in other countries in the near future.

Drug-Resistant HBV and Combination Chemotherapy

As noted above, HBV exhibits a high mutation rate because of the inherently poor copying fidelity of both host RNA polymerase II and the HBVl polymerase, both of which lack proofreading 3′-5′ exonuclease activity. Populations of HBV quasispecies in chronically infected individuals include preexisting variants that are potentially drug-resistant. Initiation of treatment may select for drug-resistant variants ("mutants") that may become the predominant population as treatment

progresses. Lamivudine resistance is usually due to mutations (analogous to those that confer lamivudine resistance in HIV) that cause the methionine in the active site YMDD (tyrosine–methionine–aspartate–aspartate) motif of the viral polymerase to be replaced by isoleucine, valine, or serine. Lamivudine resistance appears to confer cross-resistance to other L-nucleoside analogues, but does not confer cross-resistance to adefovir dipivoxil. Lamivudine-resistant HBV mutants are eclipsed by wild-type virus after drug withdrawal, but can still be detected for at least 12 months, and their transmission can cause acute hepatitis B. Current consensus of opinion predicts that multidrug combination regimes analogous to highly active antiretroviral therapy that have been used to treat HIV infection will be required to avoid development of drug resistance and will most likely be used routinely in the future to control chronic HBV infection. Although a high viral load ensures the preexistence of variants that are potentially resistant to one or two antiviral drugs, it has been shown that the preexistence of variants with the potential for resistance to three or more drugs is unlikely.

FUTURE PROSPECTS

Prophylactic vaccination may ultimately be able to control and perhaps eliminate HBV infection at some time in the distant future. Universal hepatitis B vaccination will require concerted efforts and cooperation between governments, public health authorities, and pharmaceutical companies in many countries. Meanwhile, control of HBV infection in the millions of individuals already infected with HBV will depend on the development of new chemotherapeutic strategies, which will similarly require high-level organization and planning for optimal effect. Continuing advances in the understanding of HBV biology, in drug design, in monitoring disease progression, and in vaccine development should, in principle, make the task of elimination of HBV increasingly easy, provided that economic and political obstacles can be overcome.

See Also the Following Articles

AIDS, Hepatic Manifestations of • Hepatitis A • Hepatitis D • Liver Transplantation • Sexually Transmitted Diseases

Further Reading

Beutels, P. (2001). Economic evaluations of hepatitis B immunization: A global review of recent studies (1994–2000). *Health Econ.* **10**, 751–774.

Guidotti, L. G., and Chisari, F. V. (1999). Cytokine-induced viral purging—Role in viral pathogenesis. *Curr. Opin. Microbiol.* **2**, 388–391.

Hepatitis Foundation International home page. Available at http://www.hepfi.org/.

Hollinger, F. B. (2001). Hepatitis B virus. *In* "Fields' Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, *et al.,* eds.), 4th Ed., Chap. 86, pp. 2739–2807. Lippincott-Raven, Philadelphia, PA.

Jung, M. C., and Pape, G. R. (2002). Immunology of hepatitis B infection. *Lancet Infect. Dis.* **2**, 43–50.

Millinship, S. "Hepatitis B, Version 4.1." Downloadable document from Health on the Net web site. Available at http://www.hon.ch/Library/Theme/Hepatitis.

Pawlotsky, J. M. (2002). Molecular diagnosis of viral hepatitis. *Gastroenterology* **122**, 1554–1568.

Rabe, C., Cheng, B., and Caselmann, W. H. (2001). Molecular mechanisms of hepatitis B virus-associated liver cancer. *Digest. Dis.* **19**, 279–287.

Seeger, C., and Mason, W. S. (2000). Hepatitis B virus biology. *Microbiol. Mol. Biol. Rev.* **64**, 51–68.

Staschke, K. A., and Colacino, J. M. (2001). Drug discovery and development of antiviral agents for the treatment of chronic hepatitis B virus infection. *Prog. Drug Res. Special Issue*, 111–183.

World Health Organization Hepatitis B Fact Sheet. Available at http://www.who.int/inf-fs/en/fact204.html.

Hepatitis C

ANDREA D. BRANCH
Mount Sinai School of Medicine, New York

acute infection First 6 months of hepatitis C virus infection; patients are typically asymptomatic. Serum alanine aminotransferase levels often reach 10 times the upper limit of normal; the course may be self-limiting or may progress to chronic infection.

alanine aminotransferase Enzyme that is released from the liver during times of stress and injury; can be measured in serum and plasma.

chronic infection Hepatitis C viremia that persists for longer than 6 months. The clinical course is highly variable, and the disease spectrum ranges from minimal disease to liver failure and hepatocellular carcinoma.

cryoglobulinemia Most common extrahepatic manifestation of chronic hepatitis C. Cryoglobulins are detected in lab tests based on their precipitation at cold temperatures; symptomatic disease is uncommon (less than 1% of patients) and results from local deposition of immune complexes.

hepatitis C virus genotypes Broad families of hepatitis C viruses that circulate worldwide; some genotypes are more responsive to interferon treatment than others. Depending on the genotyping system, the global population of viruses is divided into 6 or 11 genotypes. Full-length RNA sequences of the members of a genotype are identical to each other at 30% or more of the nucleic acid positions.

interferons Family of naturally occurring peptides with both antiviral and antiproliferative effects; several α interferons (2a, 2b, and a consensus molecule, for example) have been developed as pharmaceutical agents for HCV.

internal ribosome entry site Intricate RNA structure that is located near the beginning of hepatitis C virus RNA; promotes the initiation of viral protein synthesis.

pegylated α interferons Pharmaceutical agents in which the interferon is linked to polyethylene glycol to improve the pharmokinetic profile (half-life) and efficacy.

polymerase chain reaction Process that is used to copy nucleic acids multiple times so that there is sufficient material for analytical detection.

quasispecies Population of hepatitis C viruses present in a particular individual at a particular time. The full-length sequences comprising the quasispecies are identical to each other at a minimum of 90% of the 9600 positions in the viral RNA molecules.

ribavirin Guanosine analogue that has antiviral and immunomodulatory effects; increases the sustained viral response rate of interferon when the combination is used to treat patients with chronic hepatitis C virus.

Hepatitis C virus is primarily transmitted through contaminated blood and is an important cause of liver failure and hepatocellular carcinoma. Acute infection is usually asymptomatic, although clinically significant and protracted disease can occur during this phase, which lasts up to 6 months. Hepatitis C virus is difficult to study. It has no known hosts other than humans and chimpanzees, does not propagate well in culture, and can barely be detected in the liver during chronic infection. Despite these limitations, an effective therapy has been developed, and about 50% of patients who can tolerate treatment are cured of chronic infection.

INTRODUCTION

Hepatitis C virus (HCV) infects about 3% of the world's population. In the United States, it accounts for about 15% of acute viral hepatitis, 60–70% of chronic hepatitis, and up to 50% of cirrhosis and end-stage liver disease. Chronic infection accounts for virtually all of the serious hepatitis C virus-related liver disease and causes an estimated 8000 to 10,000 deaths in the United States annually.

Routes of Transmission and Epidemiology

The routes of HCV transmission are central to HCV epidemiology, liver disease, and virology. Transmission usually occurs as a result of percutaneous or permucosal exposure to contaminated blood or blood products. Because blood exchanges are relatively rare events, HCV has few opportunities to move from one person to another. It has evolved strategies that allow it to elude the immune system and to establish chronic infections that maintain high levels of infectious virus in the blood for years, while causing minimal liver damage, at least in the short term.

HCV does not have an insect vector for transport from one person's bloodstream into another's. It relies on humans. Unsafe medical practices cause HCV epidemics in many parts of the world. Intravenous drug use is largely responsible for the epidemic in the United States. According to Miriam Alter and colleagues, the

 Copyright 2004, Elsevier (USA). All rights reserved.

Centers for Disease Control and Prevention (CDC) estimate that the recreational use of intravenous drugs has transferred HCV to 60–70% of the 3.9 million living Americans (1.8% of the population) who have been infected with HCV. Transfusions account for only about 7% of the transmissions, and sexual activity for about 15%. Infections resulting from sexual activity comprise a relatively small fraction of the total because transmission of HCV between sexual partners in monogamous relationships is very inefficient compared to that of other blood-borne viruses, such as human immunodeficiency virus (HIV) and hepatitis B virus (HBV). The transmission rate is so low that the CDC recommends that couples use their own discretion when deciding whether to use condoms to further reduce the already minimal risk. Vertical transmission from viremic mothers to their babies also occurs at a relatively low rate, about 4–7% per pregnancy. Coinfection with HIV makes transmission more likely. The actual time and mode of vertical transmission are not known. Babies of viremic mothers should be tested for HCV between the ages of 2 and 6 months and again at 18 to 24 months. CDC guidelines for HCV screening of members of high-risk groups are available on their web site.

The number of new cases of HCV in the United States has declined dramatically in recent years, according to CDC estimates. During the period from 1985 to 1989, there were 242,000 acute cases each year; in 1998, there were 40,000 new infections. The change is almost entirely due to a reduction in new HCV infections among recreational intravenous drug users. The introduction of specific screening tests to remove HCV from the blood supply was an important step for clinical medicine, but it had limited impact on the epidemic. Transfusions never accounted for more than 20% of HCV transmissions, at most. Screening for hepatitis B markers, begun in the 1970s, had already eliminated most posttransfusion HCV in the United States before specific tests for this virus became available. The factors leading to the decline in new cases of HCV that took place during the 1990s need to be elucidated so that this information can be used to avoid a repeat of the previous outbreak. Unfortunately, despite the drop in new infections, the burden of HCV disease is projected to climb as individuals exposed in previous decades advance into the later stages of pathology.

Natural History and Antiviral Therapy

The natural history of HCV infection is highly variable. Acute infection is usually asymptomatic and leads to self-limited disease in about 10–50% of patients. Patients who fail to clear the virus during the first 6 months progress to chronic infection. Chronic infection causes only mild inflammation and minimal fibrosis in many patients, but has serious consequences for others. It is estimated that about 20% of patients develop cirrhosis, usually after decades of chronic viremia. Once it arises, cirrhosis can be followed by liver failure and/or hepatocellular carcinoma in a few years. Disease progression is hastened by excessive alcohol consumption. It is more rapid in men than in women, and is faster in both sexes when infection is acquired after the age of 40. Liver damage during chronic infection is thought to be largely immune mediated, although viral cytotoxicity cannot be ruled out with existing data.

It is not clear how HCV eludes the immune system and establishes chronic infection. Viral mutability may play a role. HCV has a high mutation rate and circulates in each patient as a population of closely related sequences called a quasispecies. Sequence diversity and mutation may allow resistant viral variants to emerge faster than the adaptive immune system can cope with them. Hepatotropism may also be a factor. The liver can be chronically infected by several pathogens, suggesting that it is inherently vulnerable to this type of infection.

Antiviral treatment is available for patients with either acute or chronic infection. HCV can be eliminated in a significant percentage of patients with chronic infection by treatment for 6 to 12 months with interferon α (IFNα) and ribavirin. The duration and efficacy of combination therapy are influenced by the genotype of the virus. Genotype 1, the most common genotype in the United States, is relatively resistant; however, in selected populations receiving optimized treatment, almost 50% of patients with genotype 1 virus achieve a sustained viral response. More efficacious and economical treatments with fewer side effects are needed. Several groups are actively seeking a vaccine.

ACUTE INFECTION

Clinical Course and Serology

Primary HCV infection is asymptomatic in the vast majority of cases. Symptoms, when they occur, include right upper quadrant pain, nausea, vomiting, malaise, and fever. Jaundice develops in less than 20% of patients. Some evidence suggests that patients who become jaundiced are more likely to clear the infection, but this observation has not been confirmed in all studies. The innocuous presentation leaves most individuals unaware that they are infected. Because most patients with acute infection do not seek medical

attention, the natural history of this phase of HCV infection remains in question. Much of the available data were obtained from subjects who had a risk factor for exposure to the virus, rather than from the general population.

The first detectable serum marker of acute HCV infection, HCV RNA, initially can be found 7 to 21 days after exposure. HCV antibodies usually become detectable at about 50 days after infection, but may appear as early as 20 days, as late as 6 months, or even later. During the first month, HCV RNA levels often rise rapidly and reach high titers prior to the onset of alanine transaminase (ALT) elevations or symptoms. During the first 6 months, HCV RNA serum levels often fluctuate and may range over several orders of magnitude at various times (see Fig. 1). HCV RNA may become temporarily undetectable in patients who later develop chronic disease. Because of this variability, patients presumed to have cleared the infection should be retested 6 to 12 months later. True clearance of viral RNA is usually followed by a normalization of ALT levels within a few weeks or months and a complete recovery.

The incubation period—the interval between exposure and the elevation of aminotransferases to twice the upper limit of normal—ranges from 2 to 26 weeks, with an average of about 6–7 weeks. In 90% of adult cases, the peak ALT level exceeds 10 times the upper limit of normal, but this peak may be missed in patients who are not monitored at frequent intervals. ALT values may not reach their peak or plateau level until 8–12 weeks after inoculation, or later. The lack of correlation between the timing of the RNA and ALT elevations has been taken as a sign that HCV is not directly cytopathic. This conclusion may be an oversimplification. The processes leading to damage in the HCV-infected liver are not well understood.

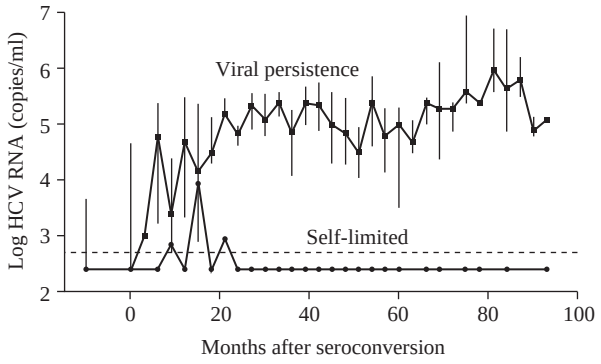


FIGURE 1 Hepatitis C virus (HCV) RNA levels in self-limited versus persistent infection. The median HCV RNA levels for each 3-month interval among 28 individuals with viral persistence and 6 individuals with self-limiting infection are indicated. All members of a group were included if data were available. Vertical lines represent the 25th to the 75th percentile of values for viremic patients. The dashed line indicates the lower limit of the assay (500 copies/ml). Reproduced with permission from Villano *et al.* (1999).

Histology

Acute HCV has histological features that are typical of acute viral hepatitis in general. These features include evidence of liver cell injury and lobular necrosis. During acute infection, inflammation in the portal areas is in proportion to the inflammation throughout the specimen. Inflammatory cells are often present in areas of liver cell necrosis. Thus, during acute infection, inflammation is less concentrated in portal areas than during chronic infection. Studies of acute HCV (non-A, non-B hepatitis) often report eosinophilic degeneration of hepatocytes, intense sinusoidal mononuclear cell infiltration, Kupffer cell activation, hepatocellular fat accumulation (steatosis), and bile duct injury. Because acute HCV is generally asymptomatic and is rarely assessed histologically, these findings may reflect liver pathology that is unusually severe.

Diagnosis of Acute Infection in the Clinical Setting

There are no currently validated methods that allow acute and chronic infection to be distinguished with certainty except in the rare patient who is observed to convert from seronegativity for HCV antibodies and RNA to seropositivity. High levels of serum aminotransferases in a patient with HCV RNA and antibodies may indicate acute infection, but may also indicate exposure to a second hepatitis virus, a toxic drug, or alcohol. Thus far, immunoglobulin M (IgM) responses have not proved to be helpful; similar levels may be present in both acutely and chronically infected patients. Some investigators are exploring anticore IgM as a possible marker of acute infection.

Treatment of Acute Infection

No guidelines are currently available in the United States for the antiviral treatment of patients with acute HCV infection. Several trials of IFNα monotherapy have been conducted. In some series, nearly 100% of the patients eliminated the virus. Further trials are needed to determine the optimal timing, duration, and dose of interferon and to explore the potential benefit of combining interferon and ribavirin for the treatment of acute HCV infection. If there is a window of opportunity during primary infection, combination therapy may

extend its boundaries and reduce the pressure to initiate treatment in patients who would experience spontaneous clearance if given the opportunity.

The successful outcome of interferon monotherapy for the treatment of acute HCV suggests that virtually all genotypes of HCV are inherently sensitive to interferon. Because treatment has oftentimes been delayed for 3 to 4 months, the high rate of viral clearance also indicates that HCV strains remain interferon sensitive and vulnerable to elimination even after they have been given the opportunity to diversify.

As illustrated in Fig. 2, the successful outcome of early interferon treatment indicates that a large percentage of the patients who would otherwise progress to chronic infection can mount an effective defense if they are aided by interferon during the acute phase. Evidently, if the virus is not expunged during this critical period, the patient and the virus adapt to each other and chronicity becomes fully established. The mechanisms of this adaptation are only partially understood. Once chronicity is established, only a minority of patients can clear the virus when treated with interferon monotherapy.

Fulminant HCV

In rare cases, acute HCV has a fulminant presentation. In fulminant hepatitis C infection, HCV RNA titers are very high until late in the clinical course and liver failure rapidly ensues. Because of the compressed time frame, HCV antibodies may not be present at any point.

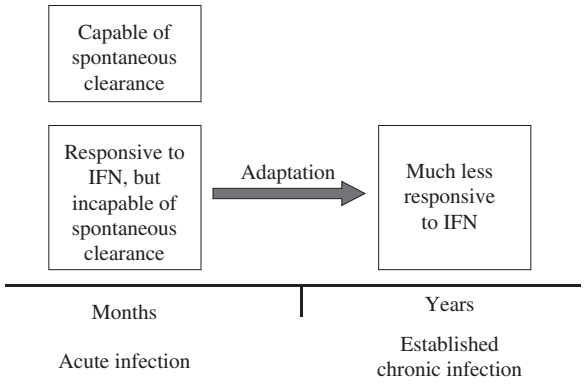


FIGURE 2 The transition to chronic infection. During the acute phase, most individuals have the capacity to clear the virus spontaneously, or in response to interferon (IFN) monotherapy. Later, most viremic patients can no longer clear the infection when treated with interferon alone. The adaptations responsible for this conversion are likely to include modulation of the host responses and changes in the viral quasispecies.

The detection of serum HCV RNA is the earliest and most valuable marker for the diagnosis of fulminant hepatitis C.

THE TRANSITION TO CHRONICITY

ALT Values Fall, but High-Level Viremia Persists

Traditionally, HCV infection has been strictly divided into acute and chronic phases, and a diagnosis of chronic infection has been made when aminotransferases remain elevated for 6 months or more. The persistence of HCV RNA is now the hallmark feature, and the line dividing acute and chronic infection has become blurred by the recognition that spontaneous viral clearance can occur more than 6 months after inoculation (see Fig. 1). The timing of viral clearance is of little clinical significance provided that it occurs before extensive scarring takes place.

In patients who fail to clear the virus, fluctuations of viral load may parallel ALT fluctuations for a time. Later, this temporal association is lost. During established chronic infection, ALT levels are usually much lower than during the peak of acute infection, and may be persistently normal, elevated, or variable. Approximately 30% of patients with chronic hepatitis C have normal ALT levels, and another 40% have ALT levels that are less than twice the upper limit of normal. In contrast to the decline in ALT levels, HCV RNA titers may remain at approximately the same level as during acute infection (see Fig. 1). The large number of persistently viremic patients who have minimal liver damage indicates that HCV quasispecies and the individual patient can become adapted to each other and coexist for decades.

Demographic Factors Associated with Higher and Lower Chronicity Rates

The percentage of acute infections that progress to chronicity ranges from about 50 to 85%, depending on the population. The value most commonly cited for the overall chronicity rate is closer to 85%. The low incidence of symptomatic acute HCV infection in the general population makes it difficult to obtain the prospective data needed to measure this important value directly. In the absence of longitudinal data, the chronicity rate is often calculated from cross-sectional data as follows: the fraction of the subjects positive for both HCV antibodies and HCV RNA (those with chronic infection) is divided by the fraction positive for HCV antibodies (those with chronic infection plus those who

cleared HCV). This approach misses individuals who lack HCV antibodies, including people with self-limited infections who lost antibodies after clearing the infection. Loss of antibodies has been reported in up to 50% of patients with self-limited infections. Thus, the chronicity rate of 75–80% may be an overestimate for the population at large, and is surely an overestimate for certain demographic groups. It is important to delineate the viral and host factors that determine whether acute infection spontaneously clears or progresses to chronicity. This knowledge may provide useful insights for pharmaceutical and vaccine development.

Self-limited infection appears to be more common among the young, females, and Caucasians; however, prospective studies have not yet identified any clinical, serologic, or virologic features that predict the outcome in an individual patient. The survey of the United States population in the National Health and Nutrition Examination Survey (NHANES III) revealed that African-Americans had a chronicity rate of 86%, whereas Caucasians had a chronicity rate of 68%. This disparity is unexplained, but is confirmed by other studies. Its cause may underlie the lower success rate of interferon treatment in African-American patients compared to Caucasians. NHANES III also demonstrated that African-American men had a higher chronicity rate than African-American women (98 versus 70%). Although this survey showed no difference between the men and women of other racial and ethnic groups, several cohort studies reveal that women, especially young women, clear HCV at a relatively high rate. At least 45% of the young healthy women exposed to HCV through contaminated immune globulin had self-limited infection. Youthfulness also favors HCV clearance. Only about 50% of children exposed to HCV while undergoing cardiac surgery had chronic infection when tested as adults, indicating that at least half cleared the infection. Infants born to HCV-infected mothers may clear HCV at a similarly high rate, although more studies are needed to firmly establish the infection frequency in newborns.

Because the incidence of acute HCV infection in the general population is low, the populations enrolled in most cohort studies are composed of individuals at high risk of infection, such as health care workers exposed through needle sticks or other accidents, recipients of blood products, babies of HCV-infected mothers, and recreational users of intravenous drugs. In some cases, the distinctive features of the study population may have an impact on the outcome of HCV infection. Transfusion recipients are often chronically ill, elderly, and immunocompromised; the 80% chronicity rate of these individuals may not be representative of the population as a whole. Similarly, individuals who use intravenous drugs for recreation often have repeated exposure to blood-borne pathogens, including HIV and HBV. Their immune responses to HCV may be distinctive. The special characteristics of the study populations must be considered when drawing general conclusions from the results.

Features of Efficacious Immune Responses

Intense, broad-based, rapid, and enduring cellular immune responses are associated with self-limited infection, whereas, weak, narrow, delayed, and fleeting T cell responses are associated with viral persistence. Humoral immune responses may also play a role in viral clearance; however, this point is controversial. Moreover, the precise characteristics of effective T cell responses are not yet known.

A detailed longitudinal study of individuals exposed through needle stick accidents revealed that HCV infection was controlled only in the single patient who had an early and vigorous CD4+ and CD8+ T cell response that evolved during the course of the infection. In this study, Francis Chisari and colleagues first detected HCV-specific CD8+ T cells in the peripheral blood 7 weeks after infection. Cells were CD38+ (activated) and their appearance coincided with the onset of liver disease. From week 12 onward, the CD8+ T cells became CD38− and began to produce IFNγ for the first time. This change in the CD8+ T cell phenotype coincided with a 5 log drop in the titer of HCV RNA and a precipitous decline in the ALT level. If CD8+ CD38− T cells that produce IFNγ play an important role in viral eradication and if they achieve viral clearance through noncytopathic mechanisms that do not destroy liver cells, as Chisari and colleagues suggest, greater knowledge of these cells may open the door to improved antiviral therapies. This knowledge may also shed light on the immunological mechanisms that allow some chronically infected patients to maintain virus replication at a stable level while experiencing little or no liver damage.

The patient who cleared the virus had an unusually high HCV RNA titer. This high titer was thought to enhance immune responses by providing a robust signal; however, in other studies, low viral titers were a favorable prognostic sign for viral clearance. The predictive value of RNA titer is not clear. Further research is also needed to relate immune responses detected in peripheral blood cells to those occurring in the liver, and to determine whether a single series of events, including the emergence of CD8+ CD38− T cells that produce IFNγ, is required for spontaneous clearance, or if a variety of pathways can lead to resolution.

Viral Strategies to Evade the Immune System and Establish Chronic Infection

Many features of HCV are likely to help it avoid host antiviral defenses. These include (1) a tropism for the liver, which is a large and resilient organ that supports a variety of chronic viruses, (2) the ability to infect lymphocytes, potentially establishing an extrahepatic reservoir that may be resistant to eradication, (3) the production of protein domains, such as the interferon sensitivity-determining region (ISDR), which may dampen innate cellular antiviral defenses and alter cell physiology, and (4) rapid expansion and diversification that outpaces antiviral defenses. Of these escape strategies, rapid expansion and diversification have received the most attention.

A sudden burst of replication during the first weeks of infection may allow the virus to invade a large number of cells and to diversify before the adaptive immune system can bring the infecting strain under control. Diversification creates a mixed population, which may contain escape mutants. Patrizia Farci and colleagues found that diversification and evolution of the HCV population typified patients who progressed to chronicity, whereas relative evolutionary stasis of the viral population occurred in patients who later cleared the infection. The evolutionary stasis observed during self-limited infection may have reflected the immune system's ability to eliminate individual components of the quasispecies before they could generate resistant progeny, or it may have resulted from the absence of selective pressure exerted by the adaptive immune system. The latter circumstance would indicate that broad-spectrum antiviral factors, such as cytokines, e.g., IFNγ, played a major role.

In addition to diversification, other strategies are also important, as indicated by an experiment in which chimpanzees were inoculated with a single pure HCV sequence. Chronic infection developed, but without the accumulation of mutations in the hypervariable region—the domain most strongly associated with escape mutants in people. Although mutations in other regions of the genome did occur, and may have generated escape mutants, it is likely that host factors also contributed to the outcome.

The liver can be chronically infected by a number of pathogens, including HBV, HCV, the delta agent, and the flavivirus, GB virus C (GBV-C)/hepatitis G virus (HGV). Even the hepatotropic viruses that do not establish chronic infections, such as hepatitis A and E viruses (HAV and HEV), persist for weeks or months. Specific features of liver physiology and immunology may contribute to the protracted course of viral infections. The liver contributes to oral tolerance, and it may play a role in other processes involving active down-modulation of immune responses. If mechanisms similar to those leading to oral tolerance became activated during viral infection, the "healthy carrier" state that characterizes many chronically infected HCV patients might result.

CHRONIC INFECTION

The Spectrum of Disease

HCV is a slowly progressive disease. In many patients, liver pathology is limited to mild inflammation and minimal fibrosis. Only about 20% of patients develop cirrhosis, usually after decades of chronic viremia. The major life-threatening consequences of chronic hepatitis C are cirrhosis, complications of end-stage liver disease (portal hypertension and hemorrhage), and hepatocellular carcinoma (HCC). Cirrhosis may develop silently in patients who are unaware of their infection and may not be discovered until overt hepatic decompensation occurs. The likelihood of the development of HCC is estimated to be 1−3% after 30 years of chronic infection. Once cirrhosis is established, liver cancer develops at an annual rate of about 1−4%. Many patients with chronic infection suffer from fatigue and report a seriously diminished quality of life. This problem is most pronounced in patients with cirrhosis. Symptomatic cryoglobulinemia (which is associated with local deposition of immune complexes) and other extrahepatic manifestations occur in a small percentage of patients. Fortunately, many patients escape the serious adverse effects of chronic HCV infection for at least 20 years. It is possible that most will retain adequate liver function indefinitely, but the dearth of long-term natural viral history studies leaves the final outcome in question.

Liver Damage

Much of the serious liver dysfunction in patients with chronic HCV infection is due to the deposition of scar tissue, or fibrosis. Scarring is a generalized response to insults. It is one of the downstream consequences of inflammation and cell injury. Inflammation contributes to liver damage by promoting fibrosis and by causing multiple rounds of cell death, mutation, and cell division—thereby setting the stage for transformation and carcinogenesis. Some investigators believe that all HCV-associated liver injury is indirect and immune mediated, but this point is not settled.

Whatever the driving forces, key variables in the outcome of chronic HCV infection are the extent of

scarring and the rate of scar tissue accumulation. The extent of fibrosis can be accurately determined only by biopsy. Many scoring systems are used to measure fibrosis, including the histology activity index (HAI; Knodell score), the Ishak modification of the HAI score, and the Metavir score. Both the stage of fibrosis and the grade of necroinflammation have prognostic significance; however, unlike scar formation, necroinflammatory activity is not cumulative.

Because liver biopsy is uncomfortable for patients and has a significant morbidity and mortality, efforts are underway to identify a panel of serum markers that might be used in combination with a second modality, such as ultrasound, to measure fibrosis. No noninvasive method is currently available, and liver biopsy remains the gold standard. However, longitudinal studies show that higher ALT levels are associated with a high rate of fibrosis progression, and, conversely, that persistently normal ALT levels are associated with a low rate. Such associations raise hope that in the future noninvasive fibrosis assessment will become possible.

Host Factors Affecting the Fibrosis Progression Rate

All of the important variables that are known to impact the rate of fibrosis progression are characteristics of the patient. HCV genotype, viral load, and the diversity of the quasispecies have not consistently emerged as significant factors (although genotype strongly affects the response to antiviral therapy). Three independent human factors are associated with an increased rate of fibrosis progression: age, alcohol consumption, and male sex. The median time between infection and the development of cirrhosis is estimated to range from a high of 42 years (for abstinent women who were infected before the age of 40) to a low of only 13 years (for men infected after the age of 40 who consume large quantities of alcohol). It is likely that the fibrosis progression rate varies over time in the same individual. Immune responses are generally less vigorous in the elderly, raising concern that the fibrosis progression rate may increase as an individual ages.

It is interesting that two of the factors associated with higher fibrosis progression rates—greater age and male sex—are also associated with increased chronicity rates. This similarity suggests that the inability to clear the virus and the inability to avoid HCV-associated liver damage stem from the same root cause. The cause is likely to be a subtle perturbation, or lesion, in immune responses.

The fibrosis progression rate is accelerated by a number of comorbid conditions. HIV infection, genetic immunodeficiencies, and immunosuppressive drugs have a significant impact. Almost 24% of HCV-infected patients with agammaglobulinemia develop end-stage liver failure within 5 years. Among immunosuppressed liver transplant recipients, the estimated mean time to the development of cirrhosis is only 12 years. Coinfection with hepatitis B virus and other liver pathogens, and diseases such as nonalcoholic steatohepatitis (NASH), may also accelerate fibrosis. To avoid the deleterious effects of coinfection, all patients with chronic HCV should receive vaccines for HAV and HBV unless they are already immune.

Diagnostic Tests for HCV: Detection of Antibodies and HCV RNA

An enzyme immunoassay (EIA) for HCV antibodies is usually the first diagnostic test performed on a patient suspected of having HCV. Commercial assays include epitopes of the nucleocapsid protein (core) and three additional viral proteins (NS3, NS4, and NS5). The specificity of the current EIAs for HCV antibodies is greater than 99%. A second type of antibody test, an immunoblot assay, can be used to confirm EIA results. Recent improvements in EIAs have reduced the need for immunoblot assays in most clinical settings. However, these tests remain useful when used in combination with EIAs during the screening of individuals who have a low risk of HCV infection, such as blood donors. Assays that detect the HCV core antigen have recently been developed and their usefulness is under investigation.

Three types of assays are used to characterize HCV RNA. Qualitative tests establish the presence or absence of HCV RNA in serum and are the most sensitive methods for HCV RNA detection. They are especially valuable when viremia is low. Quantitative tests measure the titer of HCV RNA in serum and are often used to monitor the response to antiviral therapy. When quantitative testing is used for this purpose, all of the measurements should be performed in the same laboratory with the same method. Fortunately, a common system for reporting HCV RNA titers has recently been adopted. Formulas are available for converting the units measured in previous eras to the currently agreed-upon international units (IUs) of HCV RNA.

Genotyping is an additional test that is applied to HCV RNA. This test is performed to identify the type of HCV that infects a particular patient. This information is needed to guide treatment decisions because some genotypes are more sensitive to interferon than others. Genotype is usually determined by amplifying a portion of the HCV RNA with the polymerase chain reaction (PCR) and then analyzing the resulting products.

Available methods reliably distinguish the six major genotypes and many subgenotypes.

Antiviral Treatment for Chronic HCV Infection

Antiviral treatments for HCV have improved dramatically in recent years and now include longer half-life pegylated interferons (peginterferons) that are used in combination with ribavirin. Treatment guidelines continue to evolve. The results of clinical trials should be consulted frequently for updates. The primary goal of antiviral treatment is viral eradication. Several studies show that viral clearance is followed by an improvement in liver histology. Cirrhosis is not completely irreversible, as was once thought. It is hoped that the incidence of HCC will also be reduced by viral clearance, but longer followup is needed to measure this end point and to obtain information about the impact of viral clearance on conditions such as portal hypertension. Some studies suggest that IFN treatment may reduce the rate of disease progression even among patients who remain viremic.

Two forms of pegylated IFNα are currently approved by the Food and Drug Administration. Most protocols use weekly injections of between 0.5 and 1.5 μg/kg of body weight. More research is needed to identify the optimal dose of these new pharmaceuticals. In recent trials, patients with genotype 2 or 3 had a sustained virological response rate of 70–80% when treated with peginterferon and 800 mg/day of ribavirin for 24 weeks. Patients with genotype 1 had a 40–50% sustained response rate after treatment for 48 weeks with weekly peginterferon and daily ribavirin at a dose of 1000 to 1200 mg.

Because of the lower response rate of patients with genotype 1 virus, it is useful to obtain a baseline measurement of the HCV RNA levels in these individuals prior to initiating treatment, and to repeat this test 12 weeks later. If HCV RNA is not detectable, or if the titer has dropped 100-fold or more, treatment should be continued. In the absence of such a drop, a sustained viral response is very improbable. The wisest choice may be to discontinue treatment, or continue with the understanding that a slowing of disease progression is the best result that can be realistically considered as a therapeutic goal.

At the end of treatment, HCV RNA should be measured with the most sensitive test available. A positive HCV RNA test strongly predicts relapse. Negative results should be repeated 24 weeks later, to ensure that the virus has been eradicated. Because of the relatively low success rate (in the case of patients with genotype 1) and the paucity of information about success rates (in the case of patients with genotypes 4, 5, and 6), some investigators feel that it may be advisable for patients with genotypes 1, 4, 5, and 6 who have little evidence of liver injury to be monitored at regular intervals, but to delay therapy until more efficacious treatments become available. Conversely, patients with extensive fibrosis and/or necroinflammatory activity may wish to initiate treatment using current protocols, unless counterindications are involved.

Many patients cannot tolerate current antiviral treatments. The expense and side effects of combination therapy are an impediment to treatment. Moreover, because a wide variety of medical and psychiatric conditions have been exclusion criteria in many of the trials, it is an open question as to what response rates can be achieved in the broad population of chronically infected HCV patients. More than half of all HCV-infected patients would not satisfy the inclusion and exclusion criteria for enrollment in the major peginterferon and ribavirin trials. Fortunately, new pharmaceutical agents may be available soon. These new agents include small-molecule inhibitors of HCV enzymes that are analogous to HIV protease inhibitors. Safety trials of an HCV protease inhibitor are underway.

Liver Transplantation

Liver transplantation replaces a failed or tumorous liver with a graft from a cadaveric or live donor. HCV is the leading indication for liver transplantation in the United States. Reinfection of the graft is virtually universal and immediate. Although some recipients have a favorable course, reinfection often leads to recurrent hepatitis and graft damage. Cirrhosis of the graft occurs in up to 30% of patients within 5 years of transplantation. In recent years, the incidence and severity of recurrent hepatitis C has increased, raising concern that long-term survival rates may be compromised. The factors that may be contributing to this increase in recurrent hepatitis, such as an increase in the average age of liver donors and changes in immunosuppressive drug protocols, are under investigation.

VIROLOGY

HCV RNA and Proteins

Michael Houghton and colleagues first positively identified HCV in 1989. The availability of specific tests for this pathogen soon revealed that HCV is the cause of over 90% of the disease previously known as non-A, non-B (NANB) hepatitis. The HCV-specific tests opened the door to detailed studies of HCV virology and revealed that HCV is in the family Flaviviridae. Yellow fever virus

and GBV-C/HGV are also in this family, although they are rather distant relatives of HCV and are members of another genus. HCV is a positive-sense RNA virus, meaning that both the genomic and the antigenomic strands are composed of RNA, and the RNA that is released from the infectious particle is capable of serving as a messenger RNA once it is reaches a compatible cytoplasm.

The RNA genome of HCV is about 9600 nucleotides in length and has nontranslated regions (NTRs) at both ends (see Fig. 3). Portions of the NTRs form RNA structures that are required for replication, a process that takes place in association with membranous structures in the cytoplasm. An internal ribosome entry site (IRES) mediates the initiation of protein synthesis. This highly conserved region is a target in many HCV RNA detection and quantitation tests, such as the polymerase chain reaction (PCR) assay. The conventional HCV proteins are encoded in the main open reading frame (ORF) of the genomic RNA and are released from a polyprotein precursor while the nascent polyprotein is undergoing synthesis.

The conventional proteins include three structural proteins that are thought to be included in the viral particle: core, and two glycoproteins, E1 and E2. The remaining proteins encoded by the main ORF include two molecules that are not required for HCV RNA replication, p7 and nonstructural protein 2 (NS2); p7 is a small hydrophobic transmembrane protein reported to self-assemble into a pore, and NS2 promotes a metal-dependent self-cleavage reaction that severs the bond

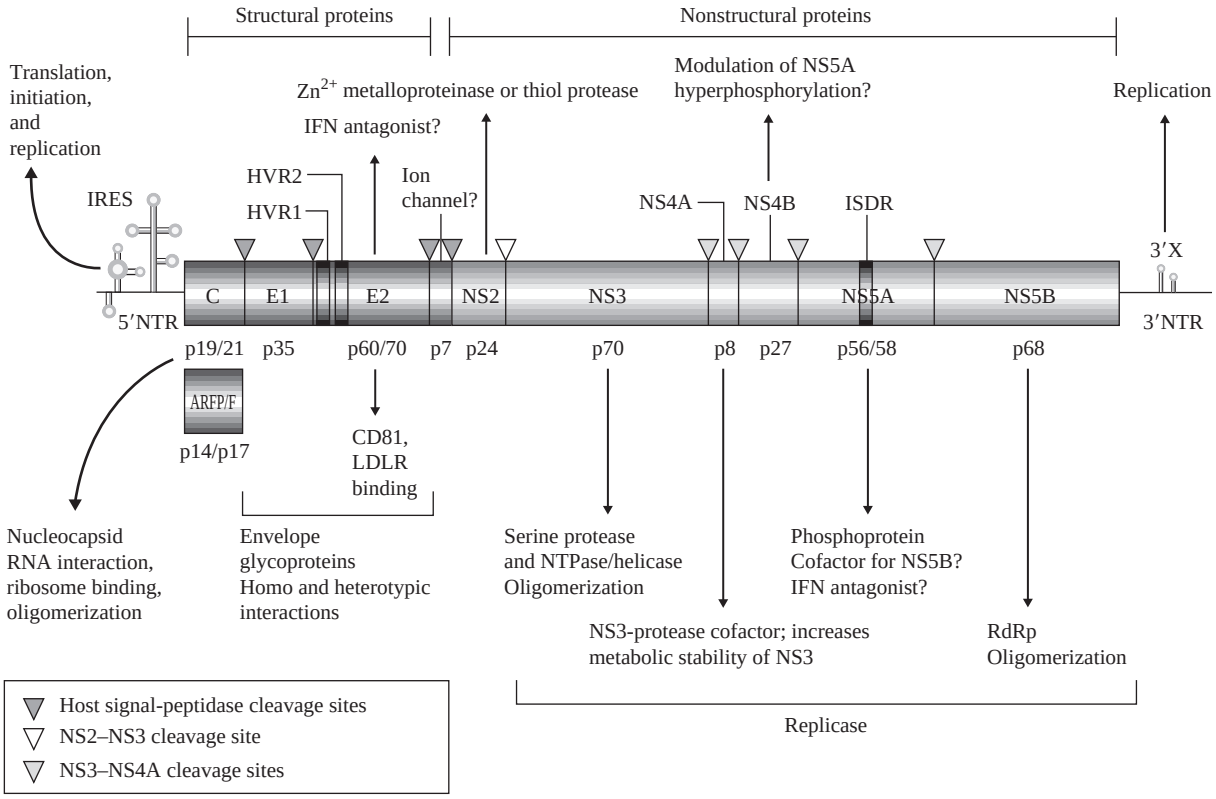


FIGURE 3 Map of the hepatitis C virus (HCV) genome. The boxed area corresponds to the single open reading frame of the HCV genome. The stem–loop structures represent the 5′ and 3′ nontranslated (NTR) regions, including the internal ribosome-entry site (IRES) and 3′X regions. The functions and molecular masses (in kilodaltons) of the gene products after polyprotein processing are shown. Core (C)–E1, E1–E2, E2–p7, and p7–nonstructural protein 2 (NS2) junctions are cleaved by cellular signal peptidases to yield structural proteins. The NS2–NS3 metalloproteinase undergoes autocatalytic cleavage, which releases the mature NS3 serine protease. NS3 cleaves the remainder of the NS polypeptide. The two regions that have extreme sequence variability in E2, known as hypervariable regions 1 and 2 (HVR1 and HVR2), are indicated. A region in NS5A, known as the interferon (IFN) sensitivity-determining region (ISDR), has been linked to the response to IFNα therapy in some strains of HCV. Both NS5A and E2 have been implicated as antagonists of IFN. ARFP/F, Alternative reading-frame protein/frameshift protein; LDLR, low-density lipoprotein receptor; RdRp, RNA-dependent RNA polymerase. Reproduced by permission from Nature Reviews Drug Discovery, Tan *et al.* (2002) Hepatitis C therapeutics: Current status and emerging strategies. **11**, 867–881. Copyright 2002 Macmillan Publishers Ltd.

between itself and NS3. NS3 has a serine proteinase domain at one end, and a helicase domain at the other. It combines with a 54-amino-acid-long membrane protein called NS4a. The NS3–NS4a heterodimeric complex processes the polyprotein, releasing the downstream viral proteins. The function of NS4b is not known, but its membrane associations may aid in the assembly of a multiprotein replication complex. The function of NS5a in viral replication is also unknown; however, this phosphoprotein contains the interferon sensitivity-determining region and confers resistance to IFNα. NS5b is the RNA-dependent RNA polymerase. In addition to the 9100-base-long ORF, genomic HCV RNA has an alternate reading frame (ARF) overlapping the core gene. The ARF gives rise to antigenically active products whose functions are under investigation.

Open Questions

What Is the Form of the Infectious Particle and the Identity of the HCV Receptor?

It is generally accepted that the HCV virion is a flavivirus-like particle with an outer envelope composed of lipids and two viral proteins, E1 and E2. Inside this protective envelope, the HCV core protein forms an inner shell and encases a single full-length copy of HCV genomic RNA. A few electron microscope images of such a particle have been obtained from serum samples of infected patients, although visualization has proved to be unexpectedly difficult. Sedimentation studies indicate that HCV infectivity peaks in serum fractions that are less dense than those that contain traditional flavivirus particles. This unusual sedimentation pattern could be a consequence of particle degradation caused by the centrifugation process, or it could indicate that a portion of the HCV RNA is associated with components of lipoprotein particles, as has been proposed. In keeping with this possibility, the low-density lipoprotein (LDL) receptor mediates the uptake of certain HCV RNA-containing particles—although it is not clear that this pathway can initiate an infectious cycle. CD81 has also been identified as a possible cellular receptor for HCV, and its significance is also under investigation.

Do the Enzymatic Viral Proteins Have Additional Functions?

The proteins encoded by the main ORF are derived from a single polyprotein precursor. Thus, of necessity, they are generated in equimolar amounts. If HCV has a flavivirus-like infectious particle, each virion contains about 500 molecules of the core protein (and one genomic RNA). Every time a virion is produced, 500 molecules of each of the other proteins are also produced. Whereas the virion constituents (core, E1, and E2) are released into the circulation, the other proteins have no export pathway. This design would seem to generate viral enzymes, such as NS5b, in excess of the amount needed for enzymatic activity. If the viral enzymes have additional functions, one may be to provide scaffolding within the replication complex. If these proteins are not all needed for enzymatic activity and scaffolding, their relatively high concentration may position them to interact with cellular constituents.

Does the Core/ARF Region of the HCV Genome Express an Oncogenic Protein?

Epidemiology and natural history studies demonstrate an association between HCV and HCC, but the nature of the association between HCV and HCC remains unexplained at the molecular level. Although HCV may induce HCC primarily by causing chronic inflammation, rather than by producing a frankly oncogenic protein, it appears that chronic inflammation is not solely responsible for HCV-associated HCC. Certain strains of HCV transgenic mice expressing the core/ARF gene develop cancer in the absence of inflammation. Many cell culture studies implicate this segment of the HCV genome in cellular transformation and tumorigenesis. The oncogenic potential of other gene regions is also under investigation. The discoveries that the HCV IRES functions more efficiently during mitosis and that full-length HCV replicons reach higher levels in dividing cells suggest that HCV may replicate more efficiently in cells that are actively dividing and that the virus is under selective pressure to express a protein that promotes cell cycling.

Which Liver Cells Support HCV Replication during Chronic Infection?

Although chronically infected patients often have more than 1 million copies of HCV RNA/ml of blood, detecting HCV proteins and RNA in liver cells is difficult, and the nature of the infected cell is in doubt. As noted previously, experimental systems indicate that dividing cells favor HCV replication; however, this awaits confirmation in human liver specimens. Most data point to a widespread infection of hepatocytes, with each infected cell expressing a very low level of HCV proteins. Based on this model, it is estimated that each cell produces only about 10 to 100 virions per day—such attenuated expression is thought to promote chronicity by confounding the immune system's efforts to identify and destroy infected cells. Ironically, both high-level expression in a very small number of cells and very low-level expression in a large number of cells

would lead to a similar result: a liver in which no signal can be detected in the vast majority of cells, and technical questions about the rare cells emitting a signal. New experimental approaches are needed to resolve this matter and to understand the complex interplay between the virus and the patient that takes place in the HCV-infected liver.

CONCLUSIONS

HCV is a highly unusual pathogen. Its novel characteristics confer advantages and disadvantages from the standpoint of containment and intervention. For example, compared to other viruses that establish chronic infection, HCV has an unprecedented sensitivity to antiviral therapy and can be eradicated in 50% of patients. This sensitivity is clearly an advantage to patients with HCV.

A second novel feature of HCV, the complexity of the core/ARF gene region, is more ambiguous in its clinical implications. The position of this region at the beginning of the main ORF, and its high content of RNA signals, suggest that this domain has regulatory functions. Regulation of viral gene expression could be an advantage to patients if it results in reduced cell damage, but it could be a disadvantage if it promotes chronicity by attenuating viral protein production to the point that the immune system has a difficult time recognizing infected cells.

The nearly total reliance of HCV on direct exposure to contaminated blood for transmission differs from that of other blood-borne viruses, such as HIV and HBV, and has minimized its dissemination into the general United States population. The epidemic is settling into a selected segment of the population: The CDC studies estimate that 70% of new HCV infections occur in individuals who are using intravenous drugs for recreation. This pattern has the potential to allow targeted interventions to be deployed efficiently, but it also has the potential to harm HCV-infected patients by increasing their isolation.

One feature of HCV that is not unusual is its tendency to rise and fall rapidly in the population. The recent decline in new HCV infections provides a window of opportunity: if the reduced incidence can be sustained, the benefits will be great. Pathogens, such as HCV, that cause outbreaks need to be monitored closely. The timely NHANES III study has been extremely valuable; however, it was limited to individuals in households, and missed homeless people and those in institutions. HCV surveillance should continue to have a high priority and should be expanded to include all sectors of society, especially those at high risk of infection.

Many aspects of HCV molecular virology, pathogenesis, and epidemiology have been characterized, but much remains to be learned about this highly unusual virus. The unanswered questions, the expanding burden of HCV-associated liver disease, and the lack of a vaccine combine to make HCV investigation one of the most active research areas in gastroenterology.

Acknowledgments

Helpful insights were provided by Joseph Odin and Efsevia Albanis (Mount Sinai) and Kerry Willis (National Kidney Foundation). Support was provided by NIDDK R01DK52071.

See Also the Following Articles

Cirrhosis • Fibrogenesis • Hepatocellular Carcinoma (HCC) • Liver Transplantation • Portal Hypertension and Esophageal Varices • Sexually Transmitted Diseases

Further Reading

Alter, M. J. (1997). Epidemiology of hepatitis C. *Hepatology* **26**, 62S–65S.

Alter, M. J., Kruszon-Moran, D., Nainan, O. V., McQuillan, G. M., Gao, F., Moyer, L. A., Kaslow, R. A., and Margolis, H. S. (1999). The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N. Engl. J. Med.* **341**, 556–562.

Alter, M. J., Margolis, H. S., Krawczynski, K., Judson, F. N., Mares, A., Alexander, W. J., Hu, P. Y., Miller, J. K., Gerber, M. A., Sampliner, R. E., *et al.* (1992). The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. *N. Engl. J. Med.* **327**, 1899–1905.

Armstrong, G. L., Alter, M. J., McQuillan, G. M., and Margolis, H. S. (2000). The past incidence of hepatitis C virus infection: Implications for the future burden of chronic liver disease in the United States. *Hepatology* **31**, 777–782.

Arthur, M. J. (2002). Reversibility of liver fibrosis and cirrhosis following treatment for hepatitis C. *Gastroenterology* **122**, 1525–1528.

Bonham, M. A., Brown, S., Boyd, A. L., Brown, P. H., Bruckenstein, D. A., Hanvey, J. C., Thomson, S. A., Pipe, A., Hassman, F., Bisi, J. E., *et al.* (1995). An assessment of the antisense properties of RNase H-competent and steric-blocking oligomers. *Nucleic Acids Res.* **23**, 1197–1203.

Branch, A. D., and Seeff, L. B. (eds.). (2000). Hepatitis C: State of the art at the millennium (parts 1 and 2). *Semin. Liver Dis*, **20**.

Centers for Disease Control and Prevention (2003). Web site: http://www.cdc.gov.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244** 359–362.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J. C., Strazzera, A., Chien, D. Y., Munoz, S. J., Balestrieri, A., Purcell, R. H., and Alter, H. J. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344.

Gerlach, J. T., Diepolder, H. M., Jung, M. C., Gruener, N. H., Schraut, W. W., Zachoval, R., Hoffmann, R., Schirren, C. A., Santantonio, T., and Pape, G. R. (1999). Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. *Gastroenterology* **117**, 933–941.

Jaeckel, E., Cornberg, M., Wedemeyer, H., Santantonio, T., Mayer, J., Zankel, M., Pastore, G., Dietrich, M., Trautwein, C., and Manns, M. P. (2001). Treatment of acute hepatitis C with interferon alpha-2b. *N. Engl. J. Med.* **345**, 1452–1457.

National Institutes of Health (2002). Consensus Development Conference: Management of hepatitis C. *Hepatology* **36** (Suppl. 5). W. B. Saunders, Philadelphia.

Orland, J. R., Wright, T. L., and Cooper, S. (2001). Acute hepatitis C. *Hepatology* **33**, 321–327.

Poynard, T., Bedossa, P., and Opolon, P. (1997). Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* **349**, 825–832.

Poynard, T., McHutchison, J., Manns, M., Trepo, C., Lindsay, K., Goodman, Z., Ling, M. H., and Albrecht, J. (2002). Impact of pegylated interferon alpha-2b and ribavirin on liver fibrosis in patients with chronic hepatitis C. *Gastroenterology* **122**, 1303–1313.

Tan, S. L., Pause, A., Shi, Y., and Sonenberg, N. (2002). Hepatitis C therapeutics: Current status and emerging strategies. *Nat. Rev. Drug Discov.* **11**, 867–881.

Thimme, R., Oldach, D., Chang, K. M., Steiger, C., Ray, S. C., and Chisari, F. V. (2001). Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J. Exp. Med.* **194**, 1395–1406.

Villano, S. A., Vlahov, D., Nelson, K. E., Cohn, S., and Thomas, D. L. (1999). Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. *Hepatology* **29**, 908–914.

Walewski, J. L., Keller, T. R., Stump, D. D., and Branch, A. D. (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**, 710–721.

Hepatitis D

MARIO RIZZETTO
University of Turin, Italy

hepatitis D virus Subviral human pathogen requiring helper functions of the hepatitis B virus to replicate and induce disease.

hepatitis D virus RNA/hepatitis D antigen Genome/genome product of hepatitis D virus, identification of which in serum and/or liver establishes the diagnosis of ongoing hepatitis D virus infection.

immunoglobulin M/total immunoglobulin antihepatitis D Antibodies to hepatitis D virus, identification of which in serum establishes the diagnosis of exposure to hepatitis D virus.

Clinical and infectivity studies carried out in Italy and the United States in the late 1970s and early 1980s indicated the existence of a new and unique defective pathogenic agent, initially called "delta" and then named hepatitis D virus. This virus behaves like a viral parasite of hepatitis B virus and is capable only of infecting carriers of the hepatitis B surface antigen. Chronic hepatitis D virus infection occurs worldwide and is a serious medical problem; about 5% of the hepatitis B surface antigen carriers (about 15 million patients) have been estimated to harbor the infection.

INTRODUCTION

The hepatitis delta virus (HDV) is a defective RNA virus requiring helper functions provided by the hepatitis B virus (HBV). The most important factor influencing efficiency of transmission is an underlying HBV infection, and carriers of the hepatitis B surface antigen (HBsAg) are therefore the major risk group; they become infected through superinfection because their HBV condition rescues and rapidly activates HDV and supports its indefinite synthesis. HDV can also be acquired by coinfection with HBV, but in this setting the synthesis of helper HBV is transient and expression of HDV is short lived. HDV infection is present worldwide, but

Copyright 2004, Elsevier (USA). All rights reserved.

Gerlach, J. T., Diepolder, H. M., Jung, M. C., Gruener, N. H., Schraut, W. W., Zachoval, R., Hoffmann, R., Schirren, C. A., Santantonio, T., and Pape, G. R. (1999). Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. *Gastroenterology* **117**, 933–941.

Jaeckel, E., Cornberg, M., Wedemeyer, H., Santantonio, T., Mayer, J., Zankel, M., Pastore, G., Dietrich, M., Trautwein, C., and Manns, M. P. (2001). Treatment of acute hepatitis C with interferon alpha-2b. *N. Engl. J. Med.* **345**, 1452–1457.

National Institutes of Health (2002). Consensus Development Conference: Management of hepatitis C. *Hepatology* **36** (Suppl. 5). W. B. Saunders, Philadelphia.

Orland, J. R., Wright, T. L., and Cooper, S. (2001). Acute hepatitis C. *Hepatology* **33**, 321–327.

Poynard, T., Bedossa, P., and Opolon, P. (1997). Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* **349**, 825–832.

Poynard, T., McHutchison, J., Manns, M., Trepo, C., Lindsay, K., Goodman, Z., Ling, M. H., and Albrecht, J. (2002). Impact of pegylated interferon alpha-2b and ribavirin on liver fibrosis in patients with chronic hepatitis C. *Gastroenterology* **122**, 1303–1313.

Tan, S. L., Pause, A., Shi, Y., and Sonenberg, N. (2002). Hepatitis C therapeutics: Current status and emerging strategies. *Nat. Rev. Drug Discov.* **11**, 867–881.

Thimme, R., Oldach, D., Chang, K. M., Steiger, C., Ray, S. C., and Chisari, F. V. (2001). Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J. Exp. Med.* **194**, 1395–1406.

Villano, S. A., Vlahov, D., Nelson, K. E., Cohn, S., and Thomas, D. L. (1999). Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. *Hepatology* **29**, 908–914.

Walewski, J. L., Keller, T. R., Stump, D. D., and Branch, A. D. (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**, 710–721.

Hepatitis D

MARIO RIZZETTO
University of Turin, Italy

hepatitis D virus Subviral human pathogen requiring helper functions of the hepatitis B virus to replicate and induce disease.

hepatitis D virus RNA/hepatitis D antigen Genome/genome product of hepatitis D virus, identification of which in serum and/or liver establishes the diagnosis of ongoing hepatitis D virus infection.

immunoglobulin M/total immunoglobulin antihepatitis D Antibodies to hepatitis D virus, identification of which in serum establishes the diagnosis of exposure to hepatitis D virus.

Clinical and infectivity studies carried out in Italy and the United States in the late 1970s and early 1980s indicated the existence of a new and unique defective pathogenic agent, initially called "delta" and then named hepatitis D virus. This virus behaves like a viral parasite of hepatitis B virus and is capable only of infecting carriers of the hepatitis B surface antigen. Chronic hepatitis D virus infection occurs worldwide and is a serious medical problem; about 5% of the hepatitis B surface antigen carriers (about 15 million patients) have been estimated to harbor the infection.

INTRODUCTION

The hepatitis delta virus (HDV) is a defective RNA virus requiring helper functions provided by the hepatitis B virus (HBV). The most important factor influencing efficiency of transmission is an underlying HBV infection, and carriers of the hepatitis B surface antigen (HBsAg) are therefore the major risk group; they become infected through superinfection because their HBV condition rescues and rapidly activates HDV and supports its indefinite synthesis. HDV can also be acquired by coinfection with HBV, but in this setting the synthesis of helper HBV is transient and expression of HDV is short lived. HDV infection is present worldwide, but

Copyright 2004, Elsevier (USA). All rights reserved.

predominates in tropical and subtropical areas. With the increased control of HBV achieved in recent years in developed countries, the prevalence of HDV infection has greatly declined in the Western world.

The clinical expression of HDV infections varies from asymptomatic infections to severe forms of acute and chronic liver disorders. Most HDV-infected individuals who develop chronic hepatitis D suffer from a serious and progressive liver disorder. First-line diagnosis is based on the finding of immunoglobulin (IgM and IgG) antibodies to HDV. The ultimate diagnostic gold standard is the detection of HDV RNA in serum or hepatitis D antigen (HDAg) in liver.

The only approved therapy for chronic hepatitis D is interferon α. This, however, cures no more than 15–20% of noncirrhotic forms of the disease. Transplantation offers a safe, therapeutic option for end-stage HDV disease. Provided that prophylaxis with lamivudine and standard hepatitis B immunoglobulins is given, reinfection occurs very rarely.

VIROLOGY

Hepatitis D virus is the only member of the Deltaviridae family. It consists of a negative-stranded RNA virus that depends on HBV for propagation but not for RNA replication. The helper HBV does not share sequence homologies or functional similarities with HDV; the latter shares functional and structural properties with small plant viruses such as viroids and virusoids. There are three major genotypes of HDV.

The virion is a 35- to 37-nm particle consisting of the circular ribonucleic acid (HDV RNA) and delta antigen coated by the HBV surface antigen. The genome is a 1.7-kb single-stranded circular RNA that can fold itself into an unbranched rodlike structure; this folding is caused by base pairing of over 70% of its nucleotides. It has two unique features: transcription by a host RNA polymerase and an autocatalytic capacity (self-cleaving and self-ligation). The autocatalytic RNA segments are functionally similar to the "hammerhead" ribozymes described in plant viruses, but differ structurally, suggesting that the catalytic domain of HDV RNA represents a new and different ribozyme motif.

The HDV encodes a single protein, the delta antigen, from the transcription of a 0.8-kb mRNA. The HDAg consists of two isoforms, the small (S-HDAg, p24) HDAg and the large (L-HDAg, p27) HDAg, derived from a common open reading frame; the isoforms differ by 19 to 20 amino acids at the C terminus. The addition of these amino acids is the consequence of RNA editing at the *amber/W* site. The mRNA contains the open reading frame for the S-HDAg, a 195-amino-acid species that is essential for HDV genome replication. A host cell RNA polymerase that is normally DNA dependent is redirected to act on HDV RNA as template. The S-HDAg p24 promotes HDV replication and is produced by infectious RNA. The L-HDAg p27 inhibits replication and is required for virion assembly. Replication of HDV occurs via the rolling circle model. Circular genomic RNA is transcribed by poly II to yield multimeric linear transcripts of antigenomic sense that undergo autocatalytic cleavage and ligation to produce circular monomeric antigenomic RNA; then, the antigenomic RNA serves as a template for replication of circular genomic RNA by similar transcription and processing.

The host range of HDV infections includes humans, chimpanzees, woodchucks, and ducks carrying the HBV-related hepadnavirus. Though spread of the virus from cell to cell requires the presence of the HBsAg, HDV infection can be initiated by the direct injection of cDNA clones into the liver or by cell transfection.

NATURAL HISTORY

Hepatitis D virus can be acquired by coinfection with HBV or by superinfection of a carrier of HBV. Individuals with antibody to the HBsAg (anti-HBs) in their serum are also protected from HDV. In healthy persons simultaneously coinfected by the two viruses, activation of HDV is dependent on the previous activation of HBV. Usually the underlying HBV infection is self-limited and HDV coinfection resolves. In superinfected HBsAg carriers, the biological effect necessary to support HDV is provided by the HBV colonizing the infected host. Superinfected carriers stand a high risk of becoming chronic HDV carriers; progression to chronicity occurs in over 90% of cases.

The natural history of chronic HDV disease varies. In active double HBV and HDV infections, disease may run a rapidly progressive course, leading to liver failure in less than 2 years. Within the general population, the disease runs a benign nonprogressive course in a minority of the patients (about 15%); in the majority it runs a progressive course, leading to cirrhosis within a few years. Over a median of 6.6 years of follow up, HDV cirrhosis increases twofold the risk for mortality and threefold the risk for hepatocellular carcinoma compared to ordinary HBV cirrhosis, and over a period of 12 years liver cancer has developed in 40% of HDV cirrhotic patients. Patients with chronic hepatitis D may exhibit liver/kidney antibody microsome type 3 (LKM3), which is distinct from LKM1 of autoimmune type 2 hepatitis. LKM3 reacts with uridyl diphosphate (UDP) glucuronyltransferase.

Most HDV infections in endemic areas are contracted in adolescence or early adulthood and present clinically with signs of decompensated cirrhosis in the fourth or fifth decade of life, i.e., one to two decades earlier than patients presenting with HBV or HCV cirrhosis. Of note, the current clinical scenario of HDV disease in the developed world reflects the recent profound decrease in the prevalence of HDV and is different from the clinical spectrum of the disease described in the past. With the control of HBV infection in recent years the number of new HDV infections has markedly declined and the clinical scenario of hepatitis has distinctly changed in the developed world; fresh and florid forms of the disease have become rare and a majority of contemporary HDV cases represent cohorts of patients with long-standing infection, and clinical expression is usually an advanced cirrhosis.

TRANSMISSION AND EPIDEMIOLOGY

Transmission of HDV occurs by the parenteral route, either overtly or covertly, and the same as for HBV transmission. Significant rates of HDV transmission were reported in the past in specific epidemiologic groups, including medical personnel, hemodialysis patients, institutionalized patients, and prisoners, with prevalence rates depending on local hygienic standards and effectiveness of HBV prophylaxis; the control of HBV achieved in recent years in developed countries has also greatly diminished HDV infection in these settings.

Sexual contacts provide HDV transmission, as attested by the prevalence of anti-HDV in prostitutes and in sexual partners of HDV-infected carriers. Transmission of HDV is promoted by household overcrowding; in Italy, cohabitation with a carrier of HDV has been identified as a critical risk factor, and the reduction of family size has been important in determining the decrease of the infection in recent years.

Infection occurs worldwide, but distribution varies. Prevalence variations are in keeping with local prevalence rates of HBV but not dependent on them. The prevalence of HDV infection is very low in North America and Northern Europe. Hepatitis D predominates in several tropical and subtropical countries, where HBV is still hyperendemic. New foci have been identified in the Mato Grosso in Brazil, in Samara, Russia, and among indigenous people of the Amazon basin of Ecuador.

Of the three well-characterized genotypes of HDV, genotype I is predominant and is found in many parts of the world. Genotype II predominates in Taiwan and Japan. Genotype III has so far been found only in the Amazon Basin.

DIAGNOSTIC AND CLINICAL FEATURES

Active HDV infection is diagnosed by the finding in serum or liver of the HDV RNA or of the HDAg. Viremia can be detected with genetic probes that hybridize with complementary HDV RNA; transcription polymerase chain reaction (PCR) can detect in serum 10 to 100 copies of the viral genome.

Attempts to measure HDAg in serum are unrewarding. In immunocompetent patients with chronic hepatitis D, serum HDAg is undetectable by enzyme immunoassay (EIA) or radioimmunoassay (RIA) because these patients have invariably high titers of the homologous antibody, which blocks the antigen in immunocomplexes that are unavailable to the assays. The HDAg can be detected in the liver with immunohistological techniques. Diagnostic screening is performed with indirect antibody markers—the IgM antibody to HDAg (IgM anti-HD), which is measured with μ capture immunoassays, and total antibody to the HDAg (anti-HD), which detects predominantly the IgG antibody and is measured with competitive radioimmunoassays. The IgM anti-HD is the first reaction in primary infection; it persists, progressing to chronicity, and as a rule is detectable in high titers in patients with chronic hepatitis D and represents a surrogate marker of HDV-induced liver damage. The IgG antibodies develop some weeks after primary infection, reaching high titers with progression of infection, and persist in immunocompetent patients with chronic hepatitis D. They do not confer protection and may also be present in HBsAg-negative individuals, representing a serological "scar" of a double HBV/HDV infection that has resolved.

The clinical course of acute HBV/HDV coinfections varies according to the degree of HBV and HDV expression and the interplay between the two viruses, ranging from icteric hepatitis, with full but transient expression of both viruses, to milder anicteric forms associated with diminished HDV expression. Early repression of HBV can inhibit HBsAg synthesis to the extent that this marker is not detectable in serum.

Acute hepatitis D is clinically and histologically indistinguishable from ordinary hepatitis B; distinction requires specific serologic testing. The markers of HBV infection are the HBsAg and IgM anti-HBc; finding the latter is required for diagnosing HBV/HDV coinfection as opposed to superinfection. Coinfection hepatitis is diagnosed from the increase of IgM anti-HD; the IgG antibody also increases but is usually elevated after a delay of several weeks, compared to the IgM antibody.

Severe forms of the illness are accompanied by the full battery of HDV markers, including early HD antigenemia. HDV RNA can be detected in serum very soon during the course of infection and disappears with resolution of disease.

Superinfection results in hepatitis in previously healthy carriers of the HBsAg; it may lead to liver failure in carriers with preexisting chronic hepatitis B. In about 15% of the cases, the chronic disease is a mild portal and periportal hepatitis with no or minimal histological detection of fibrosis. In the other 85% of cases, histology shows extensive inflammation and necrosis with widespread lobular involvement. Disease can be present in HDV carriers with normal liver enzymes; it is often severe in HDV-positive children. Concomitant HIV infection does not appear to modify the course of chronic hepatitis D; these patients may have blunted or absent IgM and IgG anti-HD responses.

PREVENTION AND THERAPY

Vaccination against HBV is safe and efficacious prophylaxis against hepatitis D, because it prevents the helper infection on which HDV thrives. No specific form of prophylaxis is available for the HBsAg carrier at risk of HDV. The only licensed therapy of chronic hepatitis D is interferon α (IFNα), which appears efficacious only in a minority of patients; although 40–70% of treated noncirrhotic patients normalize liver enzymes during therapy, 60–97% of the responders relapse after stopping therapy. Permanent response is shown by the progressive decline of IgM anti-HD and the disappearance of HDV RNA during therapy and by the subsequent seroconversion from HBsAg to anti-HBs.

Doses of 5–9 MU thrice weekly or of 5 MU daily are required and therapy should be prolonged for at least for 12 additional months after normalization of aminotransferases has been achieved. This regimen may cause important side effects, in particular psychiatric alterations. With the recent epidemiologic decline of HBV and HDV infection in Southern Europe and the striking diminution of new cases of both infections, the ratio of cases of long-standing advanced disease over fresh forms of the disease has increased; these forms are more resistant to IFN, and the results of therapy are even less encouraging than reported in clinical trials in the 1980s. Treatment with lamivudine in order to further inhibit and possibly eradicate the underlying HBV infection has been unsuccessful. Combination therapy of IFN with lamivudine is currently being explored; initial results are not promising.

Liver transplantation provides a valid therapeutic option for end-stage HDV disease; dual HBV/HDV infection protects from graft reinfection because in most cases HBV synthesis is repressed by the concomitant HDV infection, so that HDV transplants cannot transmit the HBV necessary to rescue HDV in the liver graft. Combined prophylaxis with lamivudine (pre- and posttransplantation) and standard hepatitis B immunoglobulins posttransplantation have virtually abolished the risk of HDV reinfection in the graft.

See Also the Following Articles

AIDS, Hepatic Manifestations of • Hepatitis B • Liver Transplantation

Further Reading

Casey, J. L., Polson, A. G., Bass, B. L., and Gerin, J. L. (1997). Molecular biology of HDV: Analysis of RNA editing and genotype variations. *In* "Viral Hepatitis and Liver Disease" (M. Rizzetto, R. H. Purcell, J. L. Gerin, and G. Verme, eds.), pp. 290–294. Minerva Medica, Torino.

Gaeta, G. B., Stroffolini, T., Chiaramonte, M., *et al.* (2000). Chronic hepatitis D: A vanishing disease? An Italian multicenter study. *Hepatology* **32**, 824–827.

Gudima, S., Wu, S. Y., Chiang, C. M., Moraleda, G., and Taylor, J. (2000). Origin of hepatitis delta virus mRNA. *J. Virol.* **74**, 7204–7210.

Rizzetto, M., and Rosina, F. (1993). Hepatitis D virus: Treatment. *In* "Viral Hepatitis" (A. J. Zuckerman and H. C. Thomas, eds.), pp. 363–369. Churchill Livingstone, Edinburgh.

Rizzetto, M., Ponzetto, A., and Forzani, I. (1991). Epidemiology of hepatitis delta virus: Overview. *Prog. Clin. Biol. Res.* **364**, 1–20.

Rosina, F., Conoscitore, P, Cuppone, R, *et al.* (1999). Changing pattern of chronic hepatitis D in Southern Europe. *Gastroenterology* **117**, 161–166.

Smedile, A., Niro, M. G., and Rizzetto, M. (1997). Hepatitis D virus. *In* "Clinical Virology" (D. D. Richman, R. J. Whitley, and F. G. Hayden, eds.), pp. 1237–1284. Churchill Livingstone, New York.

Smedile, A., Rizzetto, M., and Gerin, J. L. (1994). Advances in hepatitis D virus biology and disease. *In* "Progress in Liver Disease" (J. L. Boyer and R. K. Okner, eds.), Vol. 12, pp. 157–175. W. B. Saunders, Philadelphia.

Taylor, J. M. (1990). Structure and replication of hepatitis delta virus. *Semin. Virol.* **1**, 135–141.

Hepatitis E

PIET C. DE GROEN
Mayo Clinic and Foundation

bilirubin A bile pigment formed during the catabolism of heme-containing compounds, primarily hemoglobin.

jaundice Excessive accumulation of bilirubin, as a result of enhanced production or impaired elimination, resulting in yellow discoloration of the skin, sclera, and mucous membranes.

The hepatitis E virus is responsible for major epidemics of enterically transmitted non-A, non-B viral hepatitis in sub-tropical and tropical countries. Similar to hepatitis A, clinical manifestations associated with symptomatic disease are mostly those of mild cholestatic hepatitis, treatment is supportive, and medical emphasis is placed on prevention and development of effective vaccines.

VIRAL AGENT AND DIAGNOSTIC TESTS

The hepatitis E virus, responsible for major epidemics of enterically transmitted non-A, non-B viral hepatitis in subtropical and tropical countries, was cloned in 1990. The virus is currently classified in a separate genus—*Hepatitis E-like viruses*—and consists of a single, plus-strand RNA genome of approximately 7.2 kb without an envelope (Fig. 1). The virus contains at least three open reading frames encoding viral proteins against which antibodies are made on exposure. These antibodies, especially those against the capsid protein derived from the second open reading frame and a small protein of unknown function derived from the third open reading frame, are detected by the current serologic assays. As for hepatitis A, immunoglobulin M (IgM) antibodies and the ratio of IgM/IgG antibodies are markers of acute infection (Fig. 2). IgM antibodies are present in most patients with acute infection as defined by the presence of hepatitis E DNA in serum. Detection of DNA via polymerase chain reaction is used mainly for research purposes. Hepatitis E virus isolates segregate into four major genotypes and at least nine different groups based on full-length sequence comparisons.

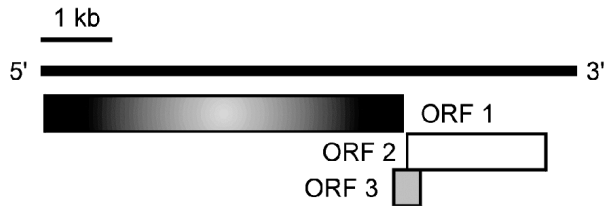


FIGURE 1 Schematic representation of the hepatitis E virus. 5′, 5′-end; 3′, poly(A) tail and 3′-end; ORF, open reading frame.

EPIDEMIOLOGY

Epidemics of symptomatic disease due to hepatitis E virus occur only in tropical or subtropical climates. Endemic areas include Africa (Genotype I), Mexico (Genotype II), and the Middle East and Asia (Genotypes I and IV). Retrospective studies on stored sera from past epidemics of viral, enterically transmitted non-A, non-B hepatitis in Mexico, Africa, Afghanistan, Pakistan, India, Bangladesh, Burma, Nepal, and Borneo have revealed that all were caused by strains of hepatitis E. In addition, hepatitis E was found to be responsible for the hepatitis epidemic of the southern area of Xinjiang, China, in which 120,000 individuals

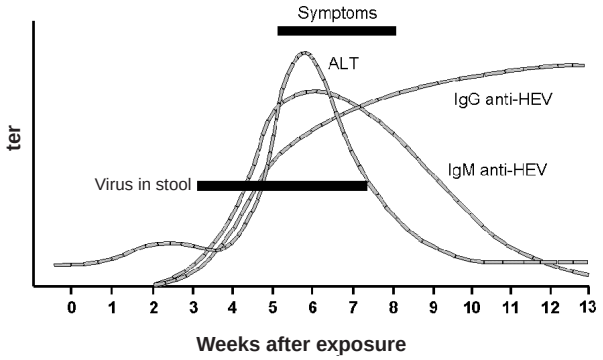


FIGURE 2 Typical serologic course. HEV, hepatitis E virus; ALT, alanine aminotransferase. Adapted from the Centers for Disease Control and Prevention Internet Presentation on HEV (available at http://www.cdc.gov/), with permission.

 Copyright 2004, Elsevier (USA). All rights reserved.

became infected between September 1986 and April 1988.

Most reports of symptomatic hepatitis E in the United States and Europe describe cases of travelers who acquired the disease in endemic areas, such as Mexico, Africa, or the Far East, and became symptomatic after returning home. However, a small number of cases, mostly asymptomatic, likely are acquired in the United States or Europe. The different genotype (Genotype III) isolated from these cases supports infection from a local source. Sporadic cases have been reported from numerous nonendemic countries, including the United States, Japan, and the Netherlands.

TRANSMISSION

Hepatitis E is transmitted via the fecal–oral route, predominantly via fecally contaminated drinking water supplies. However, preliminary reports also suggest transmission of the hepatitis E virus via blood transfusions. Volunteer studies confirmed the presence of the virus in serum and feces before and during clinical disease. The virus is shed into feces approximately 1 week before symptoms develop. The incubation period varies from 2 to 9 weeks with a mean of approximately 45 days. Swine constitute the main animal reservoir for hepatitis E. Indeed, analyses of hepatitis E genotypes in swine show a close homology for each geographic area with the human genotypes, suggesting that hepatitis E is a zoonosis, with swine as one of its hosts. Other animal sources in Asia and the United States are rodents, in particular, rats.

SYMPTOMS AND SURVIVAL

Hepatitis E predominantly affects young adults (15–40 years old). The symptoms of hepatitis E are similar to those of hepatitis A. Frequently, a prodrome consisting of anorexia, nausea, low-grade fever, and right upper abdominal pain is present 3–7 days before jaundice develops. Aminotransferase levels peak (mostly between 1000 and 2000 U/liter) near the onset of symptoms; bilirubin levels (10–20 mg/dl) peak later. Liver biopsy specimens may show prominent canalicular bile stasis. Jaundice usually resolves after 1–2 weeks. In approximately 10% of cases, the disease is fulminant, especially in pregnant women, in whom mortality rates as high as 20% due to hemorrhagic and thrombotic complications have been reported. The disease may also be fulminant in elderly patients and cause decompensation in patients with chronic liver disease. There is no evidence suggesting that hepatitis E can cause chronic infection.

TREATMENT AND PREVENTION

No specific treatment for symptomatic hepatitis E exists; as for hepatitis A, the disease is self-limited and supportive care is provided as necessary. Prevention is achieved by improving hygienic conditions, e.g., clean drinking water, thereby preventing fecal–oral transmission. Postexposure administration of immune serum globulins from endemic areas has not decreased infection rates during epidemics. However, several experimental vaccines—employing recombinant protein, DNA, or virus-like particles—have been shown to induce protective immunity in animals as well as in human volunteers. A recombinant hepatitis E vaccine is currently in phase III clinical trials.

See Also the Following Articles

Bilirubin and Jaundice • Hepatitis A

Further Reading

de Groen, P. C. (1997). Hepatitis E in the United States: A case of "hog fever"? *Mayo Clin. Proc.* **72**, 1197–1198.

Reyes, G. R., *et al.* (1990). Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* **247**, 1335–1339.

Worm, H. C., van der Poel, W. H., and Brandstatter G. (2002). Hepatitis E: An overview. *Microbes Infect.* **4**, 657–666.

Hepatocellular Carcinoma (HCC)

CHRISTOPHER D. COSTANZA AND KHALED SELIM
University of California, San Diego

aflatoxin B1 Toxic metabolite produced by the fungi *Aspergillus flavus* and *Aspergillus parasiticus,* which contaminate improperly stored peanuts, grain, and rice in hot and humid parts of the world.

ascites Accumulation of fluid that has leaked out of the liver and collected in the abdominal cavity.

Child–Pugh score System that makes use of simple laboratory values (total bilirubin, albumin, and prothrombin time) and clinical examination (ascites and hepatic encephalopathy) to predict which patients with end-stage liver disease will survive a major abdominal surgery; also used to assess the severity of cirrhosis.

chronic liver disease Ongoing injury to the liver lasting at least 6 months.

chronic viral hepatitis Infection with hepatitis B or hepatitis C virus lasting more than 6 months.

cirrhosis Condition involving the liver parenchyma; microscopic assessment demonstrates replacement of normal hepatocytes with fibrosis and regenerative nodules.

cryosurgery Ablative procedure in which subzero liquid nitrogen is injected into the hepatocellular cancer, with a goal of killing the tumor cells.

end-stage liver disease Signs and symptoms of a patient who has decompensated cirrhosis, e.g., ascites, encephalopathy, and esophageal variceal bleeding.

α-fetoprotein Normally produced to a greater degree in fetal life, but also made by malignant hepatocytes in hepatocellular carcinoma; can be elevated in the serum of patients with liver cancer and can be used to screen in patients at risk to develop hepatocellular carcinoma.

fibrolamellar hepatocellular carcinoma Tumor that is considered a histological variant of hepatocellular carcinoma, but has distinctive histological and clinical features. Tumors usually develop in younger, noncirrhotic patients have a more favorable prognosis.

hemochromatosis Inherited disease in which iron overload can affect the liver, heart, pancreas, and skin. Complications of liver involvement include cirrhosis and hepatocellular carcinoma.

hepatic artery chemoembolization Ablative therapy in which the blood supply of the tumor is selectively decreased by embolization to cause tumor cell death. In addition to decreasing the blood supply, a chemotherapeutic agent can be applied directly to the tumor.

hepatic encephalopathy Altered mental status in a patient with end-stage liver disease, ranging from mild confusion to coma; stems from the fact that in a cirrhotic liver, the cells have been replaced with scar tissue and are no longer able to remove toxins from the blood.

hepatitis B DNA virus that infects humans and is transmitted by blood, sex, or childbirth. Some chronically infected patients can progress to cirrhosis and hepatocellular carcinoma.

hepatitis C RNA virus that infects humans and is transmitted primarily through blood. Most patients become chronically infected and some will develop cirrhosis and hepatocellular carcinoma.

hepatocellular carcinoma (hepatoma) Liver cancer.

hepatocyte Normal liver cell.

liver biopsy Procedure in which a needle is placed into the liver and a small portion of the liver is removed for histological examination; most commonly done through the skin as an outpatient procedure.

percutaneous ethanol injection Ablative therapy in which a needle is inserted into the mass and ethanol is injected, with the goal of killing all the tumor cells.

radiofrequency ablation Ablative therapy in which a probe is inserted through the skin into the tumor, with the goal of killing the tumor cells with heat.

surgical resection Treatment that offers best chance of a cure, but is rarely possible because most patients present with symptomatic tumors; the surgery is technically challenging and recurrence rate is high.

transplantation Offers excellent chance for cure if patients are chosen carefully.

Hepatocellular carcinoma (HCC) is a serious health problem and is responsible for as many as 1 million deaths per year worldwide. Major risk factors for HCC are hepatitis B, hepatitis C, cirrhosis of any kind, and exposure to aflatoxin B1. In Africa and Asia, where both the incidence of vertically transmitted hepatitis B and aflatoxin B1 exposure are increased, rates of HCC are nearly 15 times higher than in the United States. If detected early in its course, HCC can be cured with surgical resection, transplantation, and possibly by nonsurgical ablative therapies. Unfortunately, many patients do not have any symptoms from HCC and are diagnosed after the cancer has become advanced and incurable. Most patients who suffer from HCC will die from the tumor. Screening and surveillance strategies to detect HCC at an earlier and more curable stage are being developed and implemented in countries around the world. The effects of such strategies on the natural history of HCC are unknown.

 Copyright 2004, Elsevier (USA). All rights reserved.

Decreasing rates of vertical transmission of hepatitis B, increasing the vaccination rate of hepatitis B, and improving the food storage conditions in areas where aflatoxin B1 flourishes seem to represent the best strategies to decrease the morbidity and mortality associated with HCC.

INTRODUCTION

Hepatocellular carcinoma is the most common primary malignant tumor of the liver, it represents up to 90% of all hepatic malignancies. Worldwide, the incidence of HCC is generally correlated with the prevalence of hepatitis B virus (HBV) infection. The incidence of HCC has risen sharply in Japan and other developed nations, including the United States, possibly due to the "silent" hepatitis C virus (HCV) epidemic. Patients infected with HCV can go several decades without symptoms, but after 20–30 years they can develop symptoms from complications of chronic liver failure, including HCC. HCC occurs predominantly in the setting of chronic liver disease and is most often associated with cirrhosis of the liver. Microscopically, cirrhosis is defined by the destruction and replacement of the normal liver cells by scar or fibrosis. Cirrhosis is caused by a variety of chronic insults to the liver, including HBV, HCV, aflatoxin B1, and alcohol. Once a patient develops cirrhosis, the rate of developing a subsequent HCC is approximately 1–3% per year. HCC is more common in men and the incidence increases with age. HCC is one of the few human cancers for which an underlying etiology, such as HBV or HCV, can be identified in most cases. Surveillance for HCC in high-risk individuals with serum α-fetoprotein (AFP) and ultrasound is being recommended in the United States and other countries. The goal of a surveillance program is to detect small, asymptomatic masses in a group of high-risk patients before the tumor is deemed unresectable. Early diagnosis of HCC is a critical prognostic factor because smaller tumors are generally associated with better outcomes. Several different systems have been used to stage HCC clinically, but, in general, vascular invasion, poor histological differentiation, and large tumors portend a poor prognosis. Surgical resection, transplantation, and ablative therapies are the mainstay of treatment for HCC.

EPIDEMIOLOGY

HCC is the sixth most common cancer in men and eleventh most common cancer in women worldwide. However, because almost every individual who develops liver cancer dies of the disease, HCC is the third most common cause of death in men and the seventh most common in women. HCC accounts for approximately 12,000 deaths per year in the United States and more than 1 million deaths per year worldwide. The epidemiology of HCC differs among geographic regions, ethnic groups, and gender. The prevalence rate of chronic HBV correlates with the incidence of HCC in most geographic regions. Vertical transmission of HBV at birth (from the mother to child) is the major risk in areas with increased incidence of HBV and HCC, such as Southeast Asia, Japan, Korea and sub-Saharan Africa (see Table I for regional incidence rates of HCC). In some parts of the world, HCC is the most common form of internal malignancy and the most common cause of death from cancer. In a given geographical region, different ethnic groups can have markedly varied rates of HCC. For example, in the United States, the incidence of HCC is higher among individuals of East Asian origin, in part because of the increased prevalence of HBV in these patients.

HCC is less common in the majority of developed, Western countries, but appears to be increasing in incidence in these areas. It has been suggested that this increase is due, in part, to the silent epidemic of hepatitis C virus (HCV). Incidence rates are expected to continue to rise for the next 20 years because of HCV. In the US, 50% of all patients who develop HCC will be infected with HCV.

RISK FACTORS

HCC is one of the few human cancers for which at least one risk factor can be identified in most cases. Table II shows the most common risk factors for development of HCC and the relative amount that each factor accounts for in the United States. The most important etiologic factors for HCC are HBV, HCV, cirrhosis, and exposure to aflatoxin B1. Regardless of the etiology, patients with cirrhosis have a fourfold increase in incidence of HCC as compared to patients without cirrhosis. The estimated annual incidence of HCC in patients with cirrhosis is estimated between 3 and 6% per year.

About 85% of the world's cases of HCC occur in the Southeast Asia, eastern Asia, and sub-Saharan Africa.

TABLE I Regional Incidence Rates of Hepatocellular Carcinoma

Region	Age-adjusted incidence
China	60–130/100,000
Taiwan	50–120/100,000
Korea	40–100/100,000
Japan	20–50/100,000
Southern Europe	10–40/100,000
Northern Europe	<5/100,000
United States	<5/100,000

TABLE II Risk Factors for Hepatocellular Carcinoma

Etiology	Percentage of HCC cases caused by, in the United States
Cirrhosis of any kind[a]	
Idiopathic	15–40%
Chronic viral hepatitis	
Hepatitis B	5–10%
Hepatitis D	<1%
Hepatitis C	30–55%
Metabolic disorders	
Hemochromatosis	<1%
α1-Antitrypsin deficiency	<1%
Toxins	
Aflatoxin	<1%

[a] Possible etiologies: anabolic steroids, estrogens, oral contraceptives, and smoking.

HBV is endemic in these regions, with up to 15% of the population being chronically infected. Persistent HBV infection increases the risk of HCC by 100-fold. Men infected at birth with HBV have an estimated lifetime risk equal to 50% for developing HCC.

In HBV infection, random integration into the host genome by a portion of the viral genome contributes to genetic instability by microdeletion or secondary chromosomal rearrangement, increasing the chance for a mutational event. Unlike HBV, HCV is not integrated into the host's genome and the precise mechanism of how HCV leads to HCC is not known, but most likely it is secondary to cirrhosis.

In the United States and Europe, HBV and HCV have equivalent contributions to the incidence rate of HCC. Alcohol-related cirrhosis, alone or in combination with HCV, causes a fair amount of HCC in the United States and Europe. Alcohol acts synergistically with HCV and thus all patients with HCV must be instructed to abstain from alcohol.

Inherited metabolic disorders such as hemochromatosis, tyrosinemia, α1-antitrypsin deficiency, Wilson's disease, and glycogen storage diseases are all associated with a significantly increased risk for development of HCC. Hemochromatosis is a familial disease that is the result of iron overload in various organs, including the liver. In the liver, iron is accumulated in the hepatocytes, causing cellular degeneration and inflammation that can lead to cirrhosis and significant increased risk of developing HCC. Men, for unknown reasons, are at an increased risk for HCC from all conditions in which cirrhosis develops, including chronic HBV infections. In addition to gender, increasing age correlates with incidence of HCC in all geographical and ethnic groups.

Exposure to mycotoxin, aflatoxin B1 (a by-product of the fungi *Aspergillus flavus* and *Aspergillus parasiticus*, which infects grains, rice, and peanuts in hot and humid parts of the world), Thorotrast (a radiographic contrast agent used until the 1950s); C_{17}-alkylated androgenic steroids, and vinyl chloride is another risk factor for the development of HCC.

DIAGNOSIS

Diagnosing HCC at an early stage is critical in determining prognosis. An early diagnosis leaves open the possibility of a surgical cure, via resection or transplantation. Sadly, most patients are diagnosed with HCC after symptoms have developed and the tumor is advanced and incurable.

Clinical Features

Up to 40% of patients with HCC will be asymptomatic at time of diagnosis. When symptoms from HCC are present, it is often difficult to distinguish from symptoms associated with cirrhosis or end-stage liver disease. Patients with both chronic liver disease and HCC often complain of fatigue, jaundice, abdominal pain, and weight loss. On physical exam, many patients will have stigmata of end-stage liver disease such as ascites, peripheral edema, muscle wasting, and jaundice. Other common signs on physical exam are hepatomegaly, splenomegaly, and hepatic bruit, and in some cases a liver mass may be obvious. HCC should be included in the differential diagnosis in any patient with chronic liver disease who experiences sudden worsening of their general condition.

Laboratory Features

Routine laboratory tests for the evaluation of liver function (alanine aminotransferase, aspartate aminotransferase, albumin, prothrombin time, bilirubin, and alkaline phosphate), markers of specific diseases (hepatitis B serologies, hepatitis C antibodies, hemochromatosis, ceruloplasmin, and α1-antitrypsin), and serum α-fetoprotein can be measured in the evaluation for HCC. AFP is an oncofetal antigen; a protein normally expressed in fetal life, but can be reproduced by neoplastic cells, including HCC cells. AFP can be used as a tumor marker and has been shown to be elevated (above 20 mg/ml) in up to 70% of patients diagnosed with HCC. However, AFP utility is limited, because elevations as high as 1000 ng/ml may be seen in patients with very active chronic viral hepatitis without HCC. Although AFP can be produced from other tumors, including undifferentiated teratocarcinoma and embryonal cell

carcinoma, a rising AFP in a patient with cirrhosis is highly suspicious for HCC. AFP can also be used to monitor a response to therapy for HCC and in detecting recurrence of HCC in that patient if AFP was elevated prior to treatment. Some patients with HCC will present with polylcythemia (elevated hemoglobin), hypercalcemia, and hypoglycemia, elevations that are most likely due to tumor-generated hormones.

Liver Biopsy

Preliminary evidence suggests that when performing liver biopsy in a patient with HCC, the needle tracks can become seeded with HCC. The true frequency with which this occurs is unknown, but has been reported as high as 1%. Because of possible malignant seeding as well as other more common complications (bleeding or pneumothorax), some physicians advocate to not perform a needle-biopsy before possible curative surgical therapy; resection, or transplantation. Some examples of indications supporting a liver biopsy include cases in which the diagnosis is not clearly established, when multifocal disease is suspected, and when nonoperative ablative therapy is being considered.

The histological appearance of HCC is of a liver cell cord-like structure, composed of hepatocytes of varying cell differentiation. In some well-differentiated tumors, the cancer cells are so well-differentiated that it is difficult to distinguish tumor cells from normal hepatocytes. On the other extreme, some tumors have cells with severe anaplastic features and only minimal evidence of hepatocellular differentiation. Fibrolamellar hepatocellular carcinoma, usually considered a histological variant of HCC, occurs in a different population and has a better prognosis and deserves special mention. Fibrolamellar HCC usually develops in noncirrhotic livers and does not produce AFP. Fibrolamellar HCC is more common in Caucasian patients and is rarely seen in Asians. The tumors usually grow to a large size without metastasizing and therefore are more amenable to surgical cure at time of presentation.

Radiographic Features

Radiological studies have played an important role in diagnosis and management of HCC for over 20 years. Radiological evaluation is used to identify and characterize the tumor, to direct liver biopsy, and to assess the extent of the disease. Radiological technology has evolved from radioisotope scans of the liver to highly specialized computed tomography (CT) and magnetic resonance imaging (MRI). Ultrasound examination also plays a key role in HCC and has been studied extensively in screening and surveillance populations. Ultrasound has been found to be highly sensitive, 66–84% for lesions less than 3 cm, specific, and less expensive than CT or MRI. Additional benefits of ultrasound include its wide availability, noninvasive nature, and ability to estimate the severity of the underlying liver disease. However, ultrasound is operator dependent and is not as sensitive, specific, or reliable as dual-contrast CT scans and MRI in detecting lesions smaller than 1 cm. Because of this, multiphase contrast-enhancing CT scans and MRI have largely replaced ultrasound for characterizing HCC. The term "triple-phase" CT scan has been coined to describe the very rapid scanning of the liver before contrast injection, during the filling of the hepatic artery, and during filling of the portal vein with contrast. CT scans or MRI scans are particularly useful for patients in whom ultrasound has not been able to accurately define a suspected lesion. Additionally, CT scans can provide volumetric data that can be used to determine the extent of resection in suitable candidates. Using specialized contrast agents can also increase the sensitivity and specificity of the CT and MRI scans. Lipiodol CT scans use lipiodol oil, which is iodized poppy seed oil that selectively enhances the vascular HCC liver nodules. For lesions less than 1 cm, a lipiodol CT scan has a sensitivity of about 70%. Even with the latest and most advanced technology, radiological imaging will not detect a substantial number of small tumor nodules and thus periodic surveillance in high-risk cirrhotic patients is recommended.

SCREENING AND SURVEILLANCE

HCC screening programs for the average-risk patient are not currently recommended in the United States. Surveillance programs for patients at high risk for the development of HCC, such as cirrhotics, are being done in the United States and other nations. Table III shows the suggested guidelines for surveillance for HCC. Serum AFP and radiological imaging, including ultrasound, CT scan, or MRI, are used to screen high-risk individuals. Using AFP alone to screen patients in populations with an increased incidence results in a sensitivity of 80–90%. In areas with a low incidence of HCC, such as the United States, the sensitivity of serum AFP drops to between 50 and 70%. Specificity is constant at 80–90% in both high- and low-incidence populations. In the United States and Western nations, serum AFP and an ultrasound examination are performed every 6 months for high-risk patients. Patients with persistently elevated AFP, or ill-defined lesions on previous imaging, represent a higher risk group of patients, and should be followed more closely with AFP every 3 months and ultrasound, CT scan, or MRI every 3–6 months.

TABLE III Guidelines for Hepatocellular Carcinoma Surveillance

Surveillance is recommended for the following patients
Cirrhosis from any cause
Chronic hepatitis B
Family history of HCC
Previously resected HCC
Chronic liver disease with documented fibrosis
Surveillance is not recommended for the following patients
Healthy patients with no liver disease
Chronic hepatitis C without cirrhosis
Patients who are not candidates for possible treatment options (surgical resection, transplantation, or ablative therapies)

Up to 50% of patients who are diagnosed with HCC are not aware that they had underlying liver disease and therefore would not benefit from a surveillance program. Additionally, for a surveillance program to have an effect and to achieve the modest survival benefit reported from Western tertiary centers, a population must have adequate financial and surgical support. Therefore, it is unlikely that surveillance programs will have a major impact on HCC mortality rates throughout the world. It is more likely that strategies such as universal vaccination against HBV, improved eradication of HBV and HCV with medical treatment, and removal of aflatoxins from the food chain will have a more significant impact on prevention.

STAGING

Once a patient has been diagnosed with HCC, the next step is to determine the stage of the cancer. As in other types of malignancies, the goal of staging in HCC is to stratify the patient based on historical outcomes, to predict mortality, and to then determine the most appropriate treatment modality. There are several different staging systems available, including the pathologic tumor–node–metastasis (pTMN), International Union Against Cancer (UICC), Cancer of the Liver Italian Program (CLIP), Okuda, and Barcelona Clinic Liver Cancer (BCLC) systems. HCC differs from other cancers in that underlying hepatic function and not the biology of the tumor best correlates with survival. Because the pTMN system does not assess the underlying hepatic function, its prognostic value is limited. The Okuda system was the first system to incorporate the biology of the tumor and the status of underlying liver disease. All the clinical staging systems have shortcomings, thus it is necessary to consider many aspects, including performance status and severity of hepatic dysfunction, tumor extent, and comorbid conditions, prior to recommending a treatment strategy.

TREATMENT

Surgical Treatment

Surgical removal of the tumor, via resection or liver transplantation, is the treatment of choice and offers the best chance at a cure. However, less than 20% of patients with HCC will present with a resectable lesion. Patients who present with unresectable HCC have a mean survival time of about 6 months. If a patient has symptoms at the time of diagnosis, the mean survival/doubling time of the tumor is about 3 months. Relative contradictions to surgical resection include extrahepatic tumor, jaundice without obstruction, hepatic encephalopathy, refractory ascites, renal insufficiency, poor hepatic function, multifocal disease, portal vein involvement, and/or portal hypertension. The most common sites for distant metastasis are lungs and bone.

Surgical Resection

For patients with HCC in an otherwise normal liver, i.e., without underlying hepatic dysfunction, resection can be done with low morbidity and mortality ($< 5\%$). Long-term results are also good, with a 5-year survival ranging between 35 and 50%. Patients with a single, small tumor (less than 5 cm), negative resection margins, and an absence of nodal or vasculature involvement have better survival. Partial hepatectomy is the preferred surgical technique for noncirrhotics or patients with mild (Child's A) cirrhosis. Whereas individuals with well-compensated cirrhosis tolerate a partial hepatectomy, patients with Child's B or C cirrhosis do poorly with such a large segment of their liver removed. Therefore, when possible, it is best to preserve as much residual liver mass by using a smaller operation, such as a segmental or nonanatomical resection, in patients with severe liver dysfunction. Hepatic resection is most often done for curative purposes, but other indications for resection include ruptured HCC, mass effect, paraneoplastic symptom relief, and for diagnostic purposes. The most common reported complications following resection include bile leaks, bleeding, and liver failure. Tumor recurrence in the residual liver occurs in approximately 50–70% of patients, either in a *de novo* or a metachronous lesion.

Liver Transplantation

Liver transplantation is an appealing surgical approach to HCC because it removes the cancer and

results in the widest possible resection margin, removing the remaining liver, while simultaneously restoring hepatic function. Transplantation is the treatment of choice for patients with small HCC and advanced cirrhosis (Child–Pugh score B or C). Since 1983, the National Institutes of Health has endorsed liver transplant for selected patients with HCC. Transplantation is contraindicated if there is extrahepatic spread of the tumor, if a single tumor is greater than 5 cm, or if there are more than three tumors and one of them is greater than 3 cm in diameter. Several series have found similar survival rates in patients with HCC compared with those without HCC as long as the aforementioned guidelines were followed. Historically, patients with HCC have done poorly while waiting on the transplant list, because the rate of tumor progression is about 70% per year. Recently, to allocate organs and give a higher priority to patients with HCC, the Model for End-Stage Liver Disease (MELD) system has been implemented in the United States. Whether this improves the outcomes of patients with HCC on the waiting list remains to be seen. Other solutions to this problem may include living donor transplantation, split-liver techniques, and accepting marginal donor organs for patients with HCC. Living donation eliminates the waiting period and may be the preferred treatment strategy for patients with HCC. However, it is a relatively new strategy and the long-term survival analysis is not yet available. Although there is little evidence to suggest that preoperative or neoadjuvant ablative therapy improves survival, it is generally recommended to retard tumor growth and metastatic spread.

Ablative Therapy

Tumors that are not amenable to surgical resection, either because the tumors are multiple or inaccessible or there is severe hepatic dysfunction, can be treated with one of several ablative techniques. The first nonsurgical ablative technique was ethanol injection, which is relatively safe and effective and is still widely used. Cryosurgery, radiofrequency ablation, hepatic arterial embolization, and chemoembolization are additional ablative therapies that can be used on nonresectable tumors.

Percutaneous Ethanol Injection

In percutaneous ethanol injection (PEI), using ultrasound for guidance, the tumor is injected with ethanol, with a goal of achieving complete tumor necrosis. Most tumors require multiple injections on separate days. PEI is used to treat tumors less than 5 cm in diameter in patients with three nodules, each less than 3 cm. Contraindications to PEI include the presence of ascites, coagulopathy, and/or obstructive jaundice. Results vary with the size of the tumor, with an 80% response rate for tumors less than 3 cm. The advantages of PEI are that it is minimally invasive, simple, safe, low cost, and quite efficacious.

Cryosurgery

Cryosurgery is an operative procedure in which subzero liquid nitrogen is administered directly into the tumor. It can be done in conjunction with hepatic resection, especially when the tumor is discovered in the contralateral lobe. Survival rates are comparable to resection but randomized clinical trials are lacking.

Radiofrequency Ablation

A relatively new technique, radiofrequency ablation (RFA), uses a heat-generating probe to destroy the tumor cells. The probe can be introduced through the skin, laparoscopically, or via an open operation. The long-term results for RFA patient survival has not been assessed. Short-term results appear to be similar to those for PEI, but fewer applications are needed. Complications include tumor seeding of the tract in 3–12% of the cases. Most physicians reserve this technique for nonresectable lesions.

Hepatic Artery Chemoembolization

Most HCC tumors receive all their blood supply from the hepatic artery, whereas the normal liver receives most of its blood supply (60–70%) from the portal vein. The goal of hepatic artery chemoembolization (HAC) is to reduce blood supply selectively to the tumor, causing induced hypoxic death of tumor cells. The procedure consists of doing an angiogram and identifying the specific artery that is supplying the tumor with blood. Next, an emulsion of lipiodol, gelatin sponges, and a chemotherapeutic agent is infused into the tumor. Randomized trials have failed to show any survival benefit of HAC versus no treatment. HAC is a palliative approach when surgery and other percutaneous ablative therapy are contraindicated. Complications include infection, abscess, liver failure, and death.

CONCLUSION

HCC is a major worldwide health issue and appears to be increasing in developed Western countries. Hepatitis B, hepatitis C, cirrhosis, and exposure to aflatoxin B1 are the major risk factors for development of HCC. If diagnosed earlier, smaller tumors are more amenable to

potentially curative management, including surgical resection, liver transplantation, and possibly ablation. Unfortunately, despite screening directed at high-risk populations, 80% of patients present with advanced HCC that cannot be resected or cured. Prevention of HCC with HBV vaccination programs and medical treatment of HCV and HBV may have an impact on HCC outcomes in the future.

See Also the Following Articles

Cancer, Overview • Cirrhosis • Hepatic Adenomas • Hepatitis B • Hepatitis C • Hereditary Hemochromatosis • Liver Biopsy • Liver Transplantation

Further Reading

Befler, A. S., and Di Biscegie, A. M. (2002). Hepatocellular carcinoma: Diagnosis and treatment. *Gastroenterology* **122**, 1609–1619.

Bisthmuth, H., Chiche, L., and Adam, R. (1993). Liver resection vs. transplantation for hepatocellular carcinoma: A multivariate analysis of survival. *Semin. Oncol.* **218**, 145–151.

Collier, J., and Sherman, M. (1998). Screening for hepatocellular carcinoma. *Hepatology* **27**, 273–278.

Di Bisceglie, A. M. (1999). Malignant neoplasms of the liver. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 1281–1304. Lippincott-Raven Publ., Philadelphia.

El-Serag, H. B. (2001). Epidemiology of hepatocellular carcinoma. *Clin. Liver Dis.* **5**, 87–107.

El-Serag, H. B., and Mason, A. C. (1999). Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.* **340**, 745–750.

Hussain, S. A., Ferry, D. R., El-Guzzaz, G., Mirza, D. F., James, N. D., McMaster, P., and Kerr, D. J. (2001). Hepatocellular carcinoma. *Ann. Oncol.* **12**, 161–172.

Jagannath, S., and Thuluvath, P. J. (2002). Hepatocellular carcinoma. *Pract. Gastroenterol.* **26**, 33–54.

Johnson, P. J. (2001). The role of alpha fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. *Clin. Liver Dis.* **5**, 145–149.

Kew, M. C. (2002). Hepatic tumors and cysts. *In* "Gastroenterology and Liver Diseases" (M. Feldman, B. Scharschmidt, and M. H. Sleisenger, eds.), 7th Ed., Vol. 1, pp. 1577–1587. W. B. Saunders, Philadelphia.

Hepatocytes

SANJEEV GUPTA AND VINAY KUMARAN
Albert Einstein College of Medicine, New York

allograft Transplantation of cells from the same species with differences in major histocompatibility complex antigens.

mature hepatocyte Parenchymal epithelial liver cell expressing liver genes expected in fully differentiated cells.

progenitor cell Cell with the capacity to replicate extensively, exhibit patterns of gene expression found in immature cells, and differentiate into mature cells.

stem cell Cell with the capacity to produce differentiated cells of more than one lineage or multiple lineages without losing the capacity for self-renewal.

xenograft Transplantation of cells from a different species with incompatible major histocompatibility complex antigens.

The term "hepatocyte" refers to parenchymal liver cells, which exhibit a variety of synthetic, metabolic, and detoxification functions necessary for sustaining life. Loss of an adequate functioning mass of hepatocytes leads to liver failure, which may occur suddenly (acute liver failure) or gradually over months or years following the onset of progressive liver disease (chronic liver failure). Liver failure is characterized by global abnormalities in hepatic function leading to or associated with perturbation of brain function (hepatic encephalopathy), kidney function (hepatorenal syndrome), sepsis, multiorgan failure, and eventually death. Deficiency or loss of specific gene function may lead to other types of liver diseases or affliction of distant organs without hepatic injury. Consideration of cell therapy is appropriate for liver failure, as well as numerous genetic and metabolic diseases. Moreover, hepatocytes possess the machinery to express many therapeutic genes, which makes the liver an attractive

Copyright 2004, Elsevier (USA). All rights reserved.

potentially curative management, including surgical resection, liver transplantation, and possibly ablation. Unfortunately, despite screening directed at high-risk populations, 80% of patients present with advanced HCC that cannot be resected or cured. Prevention of HCC with HBV vaccination programs and medical treatment of HCV and HBV may have an impact on HCC outcomes in the future.

See Also the Following Articles

Cancer, Overview • Cirrhosis • Hepatic Adenomas • Hepatitis B • Hepatitis C • Hereditary Hemochromatosis • Liver Biopsy • Liver Transplantation

Further Reading

Befler, A. S., and Di Biscegie, A. M. (2002). Hepatocellular carcinoma: Diagnosis and treatment. *Gastroenterology* **122**, 1609–1619.

Bisthmuth, H., Chiche, L., and Adam, R. (1993). Liver resection vs. transplantation for hepatocellular carcinoma: A multivariate analysis of survival. *Semin. Oncol.* **218**, 145–151.

Collier, J., and Sherman, M. (1998). Screening for hepatocellular carcinoma. *Hepatology* **27**, 273–278.

Di Bisceglie, A. M. (1999). Malignant neoplasms of the liver. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 1281–1304. Lippincott-Raven Publ., Philadelphia.

El-Serag, H. B. (2001). Epidemiology of hepatocellular carcinoma. *Clin. Liver Dis.* **5**, 87–107.

El-Serag, H. B., and Mason, A. C. (1999). Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.* **340**, 745–750.

Hussain, S. A., Ferry, D. R., El-Guzzaz, G., Mirza, D. F., James, N. D., McMaster, P., and Kerr, D. J. (2001). Hepatocellular carcinoma. *Ann. Oncol.* **12**, 161–172.

Jagannath, S., and Thuluvath, P. J. (2002). Hepatocellular carcinoma. *Pract. Gastroenterol.* **26**, 33–54.

Johnson, P. J. (2001). The role of alpha fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. *Clin. Liver Dis.* **5**, 145–149.

Kew, M. C. (2002). Hepatic tumors and cysts. *In* "Gastroenterology and Liver Diseases" (M. Feldman, B. Scharschmidt, and M. H. Sleisenger, eds.), 7th Ed., Vol. 1, pp. 1577–1587. W. B. Saunders, Philadelphia.

Hepatocytes

SANJEEV GUPTA AND VINAY KUMARAN
Albert Einstein College of Medicine, New York

allograft Transplantation of cells from the same species with differences in major histocompatibility complex antigens.

mature hepatocyte Parenchymal epithelial liver cell expressing liver genes expected in fully differentiated cells.

progenitor cell Cell with the capacity to replicate extensively, exhibit patterns of gene expression found in immature cells, and differentiate into mature cells.

stem cell Cell with the capacity to produce differentiated cells of more than one lineage or multiple lineages without losing the capacity for self-renewal.

xenograft Transplantation of cells from a different species with incompatible major histocompatibility complex antigens.

The term "hepatocyte" refers to parenchymal liver cells, which exhibit a variety of synthetic, metabolic, and detoxification functions necessary for sustaining life. Loss of an adequate functioning mass of hepatocytes leads to liver failure, which may occur suddenly (acute liver failure) or gradually over months or years following the onset of progressive liver disease (chronic liver failure). Liver failure is characterized by global abnormalities in hepatic function leading to or associated with perturbation of brain function (hepatic encephalopathy), kidney function (hepatorenal syndrome), sepsis, multiorgan failure, and eventually death. Deficiency or loss of specific gene function may lead to other types of liver diseases or affliction of distant organs without hepatic injury. Consideration of cell therapy is appropriate for liver failure, as well as numerous genetic and metabolic diseases. Moreover, hepatocytes possess the machinery to express many therapeutic genes, which makes the liver an attractive

Copyright 2004, Elsevier (USA). All rights reserved.

target for gene therapy. Also, hepatocytes can be used to produce bioartificial liver-assist devices to support patients with liver failure. Hepatocytes constitute the most abundant type of liver cell (~60%). Other types of liver cells include bile duct cells; hepatic stellate cells, which are present in the space of Disse, juxtaposed between hepatocytes and liver sinusoids; endothelial cells, which line the liver sinusoids; fibroblasts; pit cells; and Kupffer cells. The function of the liver as a cohesive unit depends on cross talk among various cell types through release of cytokines, chemicals, and signaling molecules capable of modulating cell function. Each of these cell types may impact liver function in health and disease.

ONTOGENIC CONSIDERATIONS

The human liver originates from the foregut endoderm after 4 weeks of pregnancy and develops rapidly. The structure of the liver assumes an adult form during the subsequent several weeks with production of bile by 14 weeks, indicating establishment of hepatobiliary function. The development of hepatocytes requires "specification" of embryonic cells along the hepatic lineage, which occurs with the activation of master transcription factor switches regulating gene expression, e.g., hepatocyte nuclear factor-3, as well as "differentiation" of specified cells into mature hepatocytes under regulation by additional sets of genes. The liver stroma originates from primitive cardiac mesoderm and the development of endothelial cells is critical for hepatocytes at this stage. During fetal development, hepatocytes express some genes, e.g., α-fetoprotein, which is replaced after birth by abundant expression of the albumin gene. Also, fetal hepatoblasts express both hepatic and biliary genes and may differentiate along either of these cell lineages. In contrast, the prevalence of progenitor cells in the adult liver declines markedly, although in certain circumstances, progenitor cells can reappear in the adult liver.

RELATIONSHIP WITH STEM/PROGENITOR CELLS

To sustain an adult organ, cell losses arising from normal "wear and tear" must be balanced by equivalent replacement of healthy cells. The normal adult liver is characterized by extremely limited rates of cell losses. However, in response to trauma or surgical resection, hepatocytes replicate and restore the liver mass. This property places the liver among the few organs in the body with the capacity to regenerate itself. Studies established that in appropriate circumstances hepatocytes divide indefinitely and repopulate the liver multiple times. On the other hand, the life span of individual hepatocytes may be limited by the ability of cells to maintain normal chromosomal lengths. In animals lacking telomerase, an essential protein required for reconstituting the ends of chromosomes during cell division, hepatocytes become more susceptible to injury and liver disease. In situations where hepatocytes are unable to proliferate, additional types of liver cells, designated "oval cells," may appear. Oval cells originally referred to poorly differentiated epithelial cells arising after exposure to cancer-inducing chemicals. However, oval cells also arise during liver regeneration following hepatotoxicity with chemicals, hepatitis virus infection, and acute liver failure. Cells similar to oval cells have been isolated from the pancreas, which in common with the liver arises from the foregut endoderm. Oval cells may produce hepatocytes, bile duct cells, and other cell types. Similarly, hepatocytes may be derived from hematopoietic stem cells, occasionally in the normal liver but much more frequently in the diseased liver. It is noteworthy that the fetal liver is a major site for the production of hematopoietic cells. This function is no longer necessary in adults, although hematopoiesis may be encountered in the adult liver during some diseases. Further examples of stem cell plasticity have been encountered in cells from other sources, e.g., multipotent adult progenitor cells of mesenchymal origin, with the capacity to differentiate into hepatocytes. Also, differentiation of pluripotential embryonic stem cells into mature hepatocytes is under active investigation. These considerations of liver renewal are relevant for understanding pathophysiological mechanisms and for identifying suitable types of cells for accomplishing cell therapy since the supply of donor livers for transplantation is very limited.

PRINCIPLES OF CELL ISOLATION AND CULTURE

The ability to isolate primary hepatocytes offers a variety of cell culture-based methods to elucidate biological mechanisms, to develop novel drugs, to establish drug metabolism and toxicity, and to advance cell therapy and other applications. The standard procedures employed to dissociate liver cells require the use of enzymes, e.g., collagenase, to perfuse the entire liver or portions of the liver, followed by mechanical dispersion of cells and separation of various liver cell types through gravity, by inert gradients using cell density, and by the use of fluorescence-activated cell sorting or magnetic cell sorting with specific antibodies. These methods

create discrete populations of hepatocytes and other liver cells for establishing cell cultures or for immediate use. Unlike some cell types, however, hepatocytes are easily injured during cell isolation and survive for only short periods (hours) while suspended in culture medium. On the other hand, when hepatocytes are provided anchorage to tissue-culture plastic or other substrates, cell survival improves greatly and cells can survive for several days or weeks. The survival and function of cultured hepatocytes can be improved by incorporation of specific extracellular matrix (ECM) components, e.g., collagen, fibronectin, and laminin. Hepatocytes interact with the ECM via specific cell surface integrins and such interactions modulate the proliferation of the cells as well as the expression of differentiated genes. Nonetheless, it is generally difficult to induce proliferation in cultured hepatocytes. A variety of growth factors affect DNA synthesis in cultured hepatocytes. Hepatocyte growth factor (HGF) and transforming growth factor (TGF)-α are among the most potent inducers, whereas TGF-β is a most potent inhibitor of hepatocellular DNA synthesis. Incorporation of hepatic growth factors permits the survival of hepatocytes in culture over several months. It should be noted that subpopulations of hepatocytes exhibit differences in proliferation capacity under culture conditions as well as in intact animals. Hepatocytes characterized by a "small size" exhibit greater proliferation. On the other hand, hepatocytes with evidence of DNA damage show a significantly lower level of proliferation capacity. Also, the behavior of cultured hepatocytes differs depending on the condition of the donor liver. Cells isolated from a donor liver with excessive amounts of fat, fibrosis, or inflammation show poor viability and survival under culture conditions. The survival and function of cultured hepatocytes improve in the presence of additional liver cell types, especially hepatic stellate cells, which express a variety of hepatotrophic growth factors and cytokines, including HGF. Hepatocytes isolated from the adult liver do not survive in serial culture. However, hepatocytes isolated from the liver of young subjects or from the fetal liver can be cultured serially by repeated subpassaging over several months, with hepatocytes from fetal liver performing better than those from young subjects (Fig. 1). Manipulations aimed at increasing the replication potential of hepatocytes involve the introduction of oncogenes, e.g., simian virus 40T antigen, or reconstitution of telomerase activity, which could represent a more natural approach. These methods have been successful in obtaining hepatocytes that can undergo hundreds of population doublings. Such manipulations offer the potential for development of novel substrates for *in vitro* studies, development of bioartificial liver devices to support a failing liver, and use of these cells for transplantation, although whether prolonged cell culture will introduce deleterious genetic aberrations needs further analysis.

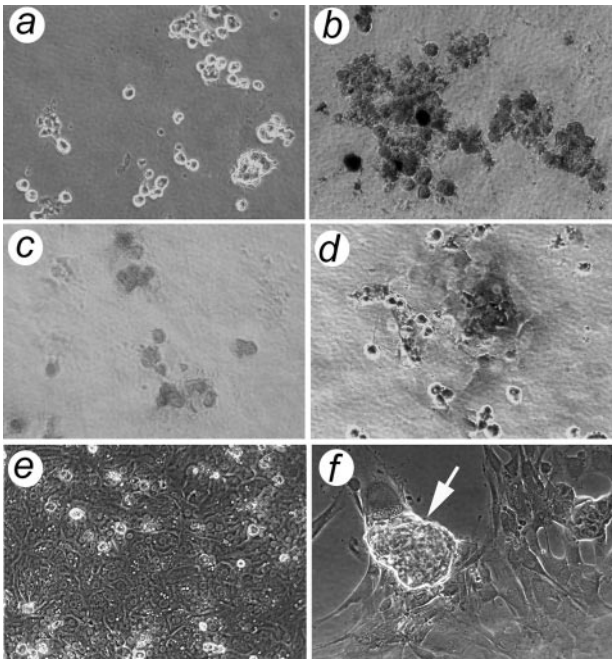


FIGURE 1 Representative examples of primary hepatocytes. (a–d) Primary hepatocytes isolated from an adult human liver. The cells were cultured in the short term on tissue-culture plastic and are shown under phase-contrast microscopy. (b, c, and d) Histochemical demonstration of glucose-6-phosphatase activity, dipeptidyl peptidase IV activity, and glycogen activity, respectively. (e) Primary hepatocytes isolated from an adult rat liver with typical confluent cell cultures on tissue-culture plastic. (f) Hepatocytes isolated from a rat pup within a few days after birth. These cells had been passaged 16 times in culture and showed significant proliferation, as shown by the area of amassed cells (arrow). Original magnification, ×200.

MANIPULATION OF CELLS FOR *IN VITRO* AND OTHER APPLICATIONS

The liver represents a complex unit of multiple cell types with a unique morphological arrangement, including dual blood supply with oxygenated blood arriving from the hepatic artery and blood enriched in nutrients and hormones arriving from the portal vein. Hepatocytes show position-dependent differences in their biological behavior. For instance, hepatocytes in the periportal areas of the liver lobule undergo earlier and greater DNA synthesis during liver regeneration, whereas hepatocytes in the perivenous areas of the liver lobule show greater expression of cytochrome

TABLE I Selected Conditions Considered Amenable to Liver-Directed Cell Therapy

Liver suffers from disease	Liver spared from disease with diseases in other target organs
Metabolic diseases	**Metabolic deficiency states**
Wilson's disease	Congenital hyperbilirubinemia, e.g., Crigler–Najjar syndrome
α-1 Antitrypsin deficiency	Familial hypercholesterolemia
Erythropoietic protoporphyria	Hyperammonemia syndromes
Lipidoses, e.g., Niemann-Pick disease	Defects of carbohydrate metabolism
Tyrosinemia type 1	Oxalosis
Acquired disorders	**Hereditary coagulation disorders**
Chronic viral hepatitis	Hemophilia
Liver cirrhosis	Factor IX deficiency
Chronic liver failure	**Inherited immune disorder**
Fulminant liver failure due to viral hepatitis, drug toxicity, etc.	Hereditary angioedema

P450 genes that metabolize or detoxify injurious by-products, chemicals, and drugs. Reproducing the behavior of hepatocytes isolated from this complex microenvironment under culture conditions is obviously difficult. A variety of manipulations have been tested for augmenting specific properties of cells, e.g., culture of cells with ECM, in gels composed of various biomaterials, with feeder cells. However, maintenance of the most differentiated function of cultured primary hepatocytes is an unresolved challenge at present. Similarly, expansion of hepatocytes under culture conditions, such that a bioartificial liver device could be seeded with relatively few cells, followed by cell proliferation has not been possible. Gene transfer with retroviral vectors, which integrate in the host genome and confer permanent gene transfer, has been hampered by limited DNA synthesis in cultured hepatocytes. Another area concerns cryopreservation of hepatocytes, which would provide the ability to store cells for use at short notice. Although progress has been made in developing suitable conditions for freezing primary hepatocytes, further work is needed to develop superior methods.

MECHANISMS IN CELL THERAPY

A major use of isolated hepatocytes involves transplantation to correct or ameliorate a variety of disorders (Table I). Numerous animal studies have been conducted to establish fundamental mechanisms in cell engraftment and liver repopulation with transplanted cells. A major goal of these studies is to establish what mass of transplanted hepatocytes will be necessary for correcting specific diseases. Animal models were established initially to localize transplanted cells containing unique genetic markers for distinction between native and transplanted hepatocytes. The use of such rodents established that transplanted hepatocytes engraft, survive, and function in a most optimal manner within the liver, rather than in ectopic sites, of recipients. Transplanted hepatocytes survive in the liver of animals during the entire life span of mice and rats (approximately 2 years), when rejection of mismatched cells is avoided. Of course, transplantation of incompatible hepatocytes with mismatched major histocompatibility complex antigens leads to rejection within several days. Analysis of the biological properties and fate of human hepatocytes requires the availability of immunodeficient mice or

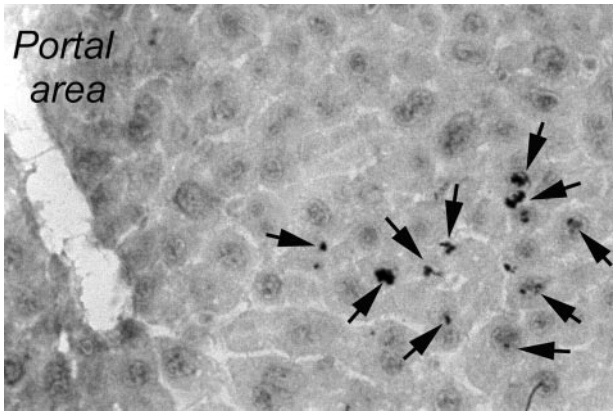


FIGURE 2 Transplanted hepatocytes in the mouse liver. An example is shown of how transplanted hepatocytes become integrated in the liver parenchyma. Primary human hepatocytes were transplanted into the liver of a NOD–SCID mouse by intrasplenic injection. Transplanted cells were identified by a human pan-centromeric DNA probe labeled with digoxigenin. This probe produced multiple dark-reaction products in transplanted cell nuclei (arrows) after detection with peroxidase-conjugated anti-digoxigenin antibody and color development with diaminobenzidine. Original magnification, ×400.

other suitable animals capable of tolerating xenografts. Fortunately, mice with severe combined immunodeficiency (SCID) represent such a host with a deficiency of T and B lymphocytes that minimizes immune responses against transplanted tissues. Mice that exhibit natural onset diabetes (NOD) in the background of SCID show additional depletion of natural killer lymphocytes and are superior animals for the analysis of transplanted human cells (Fig. 2). Other types of immunodeficient mice are being tested, e.g., NOD−SCID mice lacking in β-2-microglobulin expression. Moreover, the use of fetal sheep has been effective for analyzing the fate of human cells.

The general approach is to transplant "normal" hepatocytes into humans to correct genetically determined metabolic disorders, although the use of stem cells capable of differentiating into hepatocytes with amenability to immunological manipulations for avoiding rejection is gaining interest. Treatment of patients with liver failure could potentially utilize healthy hepatocytes from any source. Studies in authentic animal models of disease, e.g., the Gunn rat, which models Crigler−Najjar syndrome (congenital jaundice due to impaired bilirubin UDP-glucoronosyl transferase activity), the Long-Evans Cinnamon (LEC) rat, which models Wilson's disease (copper toxicosis due to ATP7B gene mutations), Nagase analbuminemic rat (NAR; hypoalbuminemia due to albumin gene mutation), Watanabe heritable hyperlipidemic rabbits (WHHL), which model familial hypercholesterolemia (low-density lipoprotein receptor mutations), and a variety of induced mutations in mice, including models for hereditary tyrosinemia, progressive familial intrahepatic cholestasis, and apolipoprotein deficiency, have been most valuable. Similarly, studies are being conducted in animals with acute liver failure or chronic liver disease and cirrhosis. In some animal models, e.g., NAR, Gunn rats, LEC rats, and WHHL rabbits, outcomes of cell therapy are monitored by demonstrating changes in serum proteins (albumin, bilirubin, ceruloplasmin, and total cholesterol, respectively). Assessment of improved outcomes by analyzing manifestations of liver disease, e.g., encephalopathy, coagulopathy, and jaundice, requires other types of physiological measurements or analysis of mortality. Insights from animal studies have begun to be translated into people. To date, initial studies of hepatocyte transplantation in approximately 50 patients worldwide indicate that cells can be transplanted safely. In selected situations, transplantation of allogeneic human hepatocytes has been effective in ameliorating metabolic diseases. These encouraging initial studies establish the framework for further development of liver cell therapies and pave the way for eventual applications of stem cells for a similar purpose.

Acknowledgments

This work was supported in part by National Institutes of Health Grants RO1 DK46952 and P30 DK41296.

See Also the Following Articles

Biliary Tract, Development • Growth Factors • Hepatotoxins • Liver, Development • Liver Failure, Pediatric • Transforming Growth Factor-β (TGF-β)

Further Reading

Allen, J. W., Hassanein, T., and Bhatia, S. N. (2001). Advances in bioartificial liver devices. *Hepatology* **34**, 447−455.

Berry, M. N., Grivell, A. R., Grivell, M. B., and Phillips, J. W. (1997). Isolated hepatocytes—Past, present and future. *Cell Biol. Toxicol.* **13**, 223−233.

Forbes, S., Vig, P., Poulsom, R., Thomas, H., and Alison, M. (2002). Hepatic stem cells. *J. Pathol.* **197**, 510−518.

Gupta, S., and Chowdhury, J. R. (2002). Therapeutic potential of hepatocyte transplantation. *Semin. Cell Dev. Biol.* **13**, 439−446.

Hengstler, J. G., Utesch, D., Steinberg, P., Platt, K. L., Diener, B., Ringel, M., Swales, N., Fischer, T., Biefang, K., Gerl, M., Bottger, T., and Oesch, F. (2000). Cryopreserved primary hepatocytes as a constantly available *in vitro* model for the evaluation of human and animal drug metabolism and enzyme induction. *Drug Metab. Rev.* **32**, 81−118.

Jiang, Y., Jahagirdar, B. N., Reinhardt, R. L., Schwartz, R. E., Keene, C. D., Ortiz-Gonzalez, X. R., Reyes, M., Lenvik, T., Lund, T., Blackstad, M., Du, J., Aldrich, S., Lisberg, A., Low, W. C., Largaespada, D. A., and Verfaillie, C. M. (2002). Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418**, 41−49.

Kmiec, Z. (2001). Cooperation of liver cells in health and disease. *Adv. Anat. Embryol. Cell Biol.* **161**, 1−151.

Runge, D., Michalopoulos, G. K., Strom, S. C., and Runge, D. M. (2000). Recent advances in human hepatocyte culture systems. *Biochem. Biophys. Res. Commun.* **274**, 1−3.

Wege, H., Le, H. T., Chui, M. S., Liu, L., Wu, J., Giri, R. K., Malhi, H., Sappal, B. S., Kumaran, V., Gupta, S., and Zern, M. A. (2003). Telomerase reconstitution immortalizes human fetal hepatocytes without disrupting their differentiation potential. *Gastroenterology* **124**, 432−444.

Zaret, K. S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nat. Rev. Genet.* **3**, 499−512.

Hepatorenal Syndrome

JOSH LEVITSKY*, TARA N. TROY*, AND THOMAS W. FAUST†
*University of Chicago Hospitals, Illinois and †University of Pennsylvania Health System, Philadelphia

acute tubular necrosis Abrupt and sustained decline in glomerular filtration rate in response to an acute ischemic or nephrotoxic insult, often characterized by brown tubular casts in the urine.

fractional excretion of sodium Percent of filtered sodium excreted in the urine, calculated based on concentrations of urinary (U) and plasma (P) sodium (Na) and creatinine (Cr): $([Na_U][Cr_P]/[Na_P][Cr_U])100$.

glomerulonephritis Inflammation of the renal glomerulus often due to infectious or immune disorders; red blood cells, and red and white blood cell casts commonly seen on urinalysis.

hepatorenal syndrome Development of renal failure in patients with acute or chronic liver failure in the absence of any other known cause of renal disease.

splanchnic Mesenteric or intestinal in origin.

transjugular intrahepatic portosystemic shunt Placement of a self-expanding metal mesh stent between the hepatic vein and a branch of the portal vein via angiographic catheters inserted through a jugular vein.

Hepatorenal syndrome is characterized by progressive renal insufficiency in patients with acute or chronic liver failure. Two types of hepatorenal syndrome exist and are differentiated by the rapidity and severity of renal dysfunction. Treatment includes the avoidance of or discontinuation of nephrotoxic medications or procedures, volume support, and therapy directed at reversing splanchnic vasodilation and improving renal perfusion.

INTRODUCTION

Renal failure is a frequent complication in patients with fulminant hepatic failure and end-stage liver disease, occurring in 40–80% of all patients. If a known cause of renal dysfunction is not determined and renal function continues to deteriorate, the patient is typically diagnosed with hepatorenal syndrome (HRS). Two types of HRS exist according to the International Ascites Club (1996). In the past, various treatments for HRS have been utilized with limited success. However, recent studies have shown improvement in HRS with the use of splanchnic vasoconstrictors or transjugular intrahepatic portosystemic shunts (TIPS). The epidemiology, pathophysiology, diagnosis, treatment, and prevention of HRS are reviewed in the following discussions.

EPIDEMIOLOGY

Because the formal diagnostic criteria for HRS were created by the International Ascites Club only recently (1996), the epidemiology of HRS has not yet been well defined. Retrospective studies have demonstrated that HRS is present in ~17% of patients hospitalized with ascites, in 30–50% of patients with spontaneous bacterial peritonitis (SBP), and in >50% of patients dying from end-stage liver disease. In a prospective study of 229 nonazotemic patients with cirrhosis and ascites, HRS developed in 18% at 1 year and in 39% at 5 years. Unfortunately, recent studies have shown that physicians often diagnose HRS when other causes of renal failure are present.

PATHOPHYSIOLOGY

Since 1956, when HRS was first recognized, various pathophysiological mechanisms have been proposed to explain the pathophysiology of the syndrome. Initially, it was believed that, despite the increased plasma volume in cirrhosis, the effective circulating plasma volume reaching vital organs, such as the kidney, is actually decreased due to pooling in the peripheral circulation. Volume expanders and surgical shunts have been used to deliver more volume to the systemic circulation and thus increase renal perfusion. Over time, it has become clear that splanchnic arterial vasodilation due to portal hypertension is more important than is reduced effective circulating volume in cirrhotic circulatory dysfunction. Endogenous splanchnic vasodilators, such as nitric oxide, are thought to be major contributors to cirrhotic circulatory dysfunction.

Three abnormalities affect renal function in patients with cirrhosis. The earliest abnormality to manifest is the lack of sodium excretion by the kidneys. At this stage, the mechanism of salt retention is unknown and ascites typically develops. Later in the course of the disease, the renin–angiotensin and sympathetic nervous systems become activated in response to splanchnic arterial vasodilation. Glomerular filtration rate (GFR) is maintained, likely due to a balance of renal vasodilation (prostaglandins) and vasoconstriction. The last

 Copyright 2004, Elsevier (USA). All rights reserved.

abnormality to manifest, in which either rapid (type 1 HRS) or slow (type 2 HRS) reductions in GFR occur, is mainly caused by systemic vasodilation and intrarenal vasoconstriction overwhelming the protective effects of intrarenal vasodilators. Most patients develop hyponatremia, extremely low urinary sodium excretion, and ascites unresponsive to diuretic therapy.

CLINICAL PRESENTATION

Hepatorenal syndrome should be suspected in any patient with advanced liver disease when there is an increase in serum creatinine level to greater than 1.5 mg/dl. Two forms of hepatorenal syndrome have been described; classification is mainly based on the severity of onset. Type 1 HRS is a rapidly progressive form associated with an 80% mortality at 2 weeks. It is often precipitated by an acute insult such as infection, nephrotoxic agents, or volume loss from gastrointestinal bleeding or overly aggressive diuresis. Deterioration of renal function is defined as a doubling of the serum creatinine level to greater than 2.5 mg/dl or a 50% reduction of the 24-hour creatinine clearance (< 20 ml/minute) within a 2-week period. Patients are usually oliguric or anuric. Type 2 HRS is the less severe, more insidious form. Deterioration of renal function occurs over months rather than weeks. Patients often have ascites resistant to diuretic therapy or are highly sensitive to the renal effects of diuretics. Although survival time is longer than type 1 HRS, the prognosis is still poor.

DIAGNOSIS

The International Ascites Club has identified specific criteria for the diagnosis of HRS (Table I). To diagnose HRS, there must be documented lack of improvement in renal function following a trial of adequate fluid repletion and the discontinuation of potential nephrotoxic agents. The major criteria must be present to diagnose HRS, whereas minor criteria serve to provide support for the diagnosis. It is important to remember that HRS is a diagnosis of exclusion; it is imperative to rule out other potential causes of acute renal failure.

DIFFERENTIAL DIAGNOSIS

Because treatment and prognosis differ greatly, it is vital to differentiate HRS from other potentially reversible causes of renal insufficiency. Patients should have a careful physical examination assessing volume status, urinalysis with microscopic examination, appropriate cultures for infection, and a renal ultrasound to evaluate for renal parenchymal disease and exclude obstruction. The three most common, reversible causes of acute renal failure in patients with advanced liver disease are prerenal azotemia, acute tubular necrosis (ATN), and glomerulonephritis.

Prerenal disease can be induced by gastrointestinal losses (e.g., vomiting and diarrhea), gastrointestinal bleeding, medications that reduce renal perfusion (e.g., diuretics and nonsteroidal antiinflammatory drugs), and cardiac dysfunction. The prognosis is excellent if the cause is identified and treated early. Volume support and achieving hemostasis are the mainstays of therapy.

ATN typically occurs after the administration of a nephrotoxic agent (nonsteroidal antiinflammatory-drugs, intravenous contrast, or aminoglycosides) or after an episode of prolonged, untreated prerenal azotemia or hypotension. ATN is usually associated with a fractional excretion of sodium (FENa) >2% and "muddy" brown casts in the urine sediment. However, in cirrhotics, the FENa may be < 1% and urinary casts may be due to hyperbilirubinemia. ATN carries a good prognosis

TABLE I International Ascites Club (1996) Major and Minor Criteria for Diagnosis of Hepatorenal Syndrome[a]

Major Criteria	Minor Criteria
Must be present for diagnosis	Provide support for the diagnosis
1. Chronic or acute liver disease with advanced hepatic failure and portal hypertension	1. Urine volume <500 ml/day
2. Low GFR, indicated by serum creatinine value >1.5 mg/dl or creatinine clearance <40 ml/minute	2. Urine Na <10 mEq/liter
3. Absence of treatment with nephrotoxic drugs, shock, infection, or significant recent fluid losses	3. Urine osmolality greater than plasma osmolality
4. No sustained improvement in renal function after diuretic withdrawal and volume expansion with 1.5 liters of isotonic saline	4. Urine red blood cells <50 per high-powered field
5. Proteinuria <500 mg/dl and no ultrasonographic evidence of obstruction or parenchymal renal disease	5. Serum Na concentration <130 mEq/liter

[a] Adapted with permission from W.B. Saunders Company, from Arroyo, V., Gines, P., Gerbes, A.L., *et al.* (1996). Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. *Hepatology* **23**, 164.

and usually reverses with supportive care (avoidance of nephrotoxic agents, correction of hypovolemia, treatment of sepsis, and, occasionally, hemodialysis).

Glomerulonephritis (GN) or vasculitis should be suspected in patients with a urine sediment containing red cells or red cell casts. Hepatitis B and C can be associated with cryoglobulinemia, polyarteritis nodosa, membranous GN, and membranoproliferative GN. A renal ultrasound and postvoid residual should be performed to exclude urinary obstruction as the cause of renal insufficiency.

TREATMENT

Because liver failure and portal hypertension are responsible for HRS, liver transplantation (LT) is the treatment of choice. Kidneys in patients with HRS generally function normally when transplanted into patients without liver disease. However, patients with HRS who undergo LT have more complications, have a higher in-hospital mortality, and more often require hemodialysis than do patients without HRS undergoing LT. Fortunately, the 3-year survival after LT of patients with HRS (60–65%) is only slightly lower than that of patients without HRS (70–80%); therefore, LT is an important option and carries a good prognosis.

If LT is not immediately available or recovery of liver function remains a possibility (e.g., alcoholic hepatitis or fulminant liver failure), other therapies should be attempted to improve renal function until liver function recovers. Because type 2 HRS is slowly progressive, vasoactive medications are uncommonly used. Avoidance and treatment of any process known to affect renal function (nephrotoxic drugs, excessive diuresis, gastrointestinal bleeding, infection, etc.) are the mainstays of management for type 2 HRS. Large-volume paracentesis should be performed for tense ascites, mainly to reduce intraabdominal pressure and thus improve cardiac and renal function; however, overzealous paracentesis should be avoided.

Because the greatest degree of splanchnic vasodilation and renal vasoconstriction occurs in type 1 HRS, most of the recent treatments have focused on reversing this complication. Prior studies have failed to show that surgical shunts and nonspecific vasoactive drugs (dopamine, fenoldopam, phentolamine, and metaraminol) improve renal function in HRS. Although most of the studies have been small, uncontrolled trials, recent data have demonstrated that vasoconstrictors specific to the splanchnic circulation can rapidly reverse HRS. Ornipressin, a vasopressin analogue, was found to mprove circulatory function and HRS in a small number of patients with HRS, although the studies were complicated by ischemic complications. Terlipressin with albumin given to seven patients with HRS over 5–15 days resulted in reversal of HRS and was not complicated by ischemia. A recent pilot study showed that norepinephrine combined with albumin and furosemide was safe and effective for type 1 HRS. Octreotide (a long-acting somatostatin analogue) in combination with midodrine (an oral α-adrenergic agonist) was shown to reverse type 1 HRS in five patients. Other medications, such as *N*-acetylcysteine and misoprostol, have shown limited success in some patients.

Other therapies have been attempted in the treatment of HRS. Reversal of portal hypertension after TIPS has been shown to cause a significant reduction in renin–angiotensin system activity and improvement in the GFR in both type 1 and type 2 HRS. However, this procedure is often complicated by hepatic encephalopathy and should be considered only in patients unresponsive to medical therapy. Hemodialysis and arteriovenous or venovenous hemofiltration may temporarily support patients with progressive HRS, but their efficacy is unknown and treatment may be complicated by electrolyte abnormalities and hypotension. The molecular adsorbent recirculating system (MARS), a modified dialysis method that removes albumin-bound substances from the blood, has shown promise in reversing type 1 HRS and improving short-term survival.

PREVENTION

Specific measures can significantly reduce the development or progression of HRS in patients with cirrhosis. Avoidance of nephrotoxic medications and intravenous contrast media in patients with cirrhosis and ascites is absolutely mandatory. Judicious use of diuretics is also recommended. In patients undergoing large-volume paracentesis, albumin should probably be given as fluid replacement, although studies have not shown that this treatment actually prevents HRS. However, in patients with SBP, the addition of albumin [1.5 g/kg, intravenously (iv) on admission, and 1 g/kg iv 48 hours later] to cefotaxime significantly reduces the occurrence of type 1 HRS and lowers 3-month mortality rates. Another study showed that the tumor necrosis factor inhibitor pentoxifylline (400 mg, three times daily) significantly lowers the incidence of type 1 HRS compared to placebo in patients with acute alcoholic hepatitis.

CONCLUSION

HRS represents progressive renal insufficiency in patients with liver failure characterized by splanchnic vasodilation, stimulation of the renin–angiotensin and

sympathetic systems, followed by impaired renal perfusion. Other renal diseases must be excluded before a diagnosis of HRS can be made. Type 2 HRS is characterized by a slow reduction in the GFR, usually over months, and is managed by avoiding known causes of renal dysfunction and careful use of diuretics, paracentesis, and TIPS. Type 1 HRS is a rapidly progressive disease with a high mortality if left untreated. Treatment of type 1 HRS with specific splanchnic vasoconstrictors appears to be evolving as effective therapy in liver transplant candidates and in patients with potential for hepatic recovery. TIPS and MARS are potential options for selected patients who either are not candidates for or have failed medical therapy.

See Also the Following Articles

Ascites • Cirrhosis • Fulminant Hepatic Failure • Liver Transplantation • Portal Hypertension and Esophageal Varices

Further Reading

Arroyo, V., Gines, P., Gerbes, A. L., *et al.* (1996). Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. *Hepatology* **23**, 164.

Arroyo, V, Guevara, M, and Gines, P. (2002). Hepatorenal syndrome in cirrhosis: Pathogenesis and treatment. *Gastroenterology* **122**, 1658–1676.

Brensing, K. A., Textor, J., Perz, J., *et al.* (2000). Long term outcome after transjugular intrahepatic portosystemic stent-shunt in nontransplant cirrhotics with hepatorenal syndrome: A phase II study. *Gut* **47**(2), 288–295.

Dagher, L., and Moore, K. (2001). The hepatorenal syndrome [review]. *Gut* **49**(5), 729–737.

Dagher, L., Patch, D., Marley, R., *et al.* (2000). Review article: Pharmacological treatment of the hepatorenal syndrome in cirrhotic patients. *Aliment. Pharmacol. Ther.* **14**, 515–521.

Ortega, R., Gines, P., Uriz, J., *et al.* (2002). Terlipressin therapy with and without albumin for patients with hepatorenal syndrome: Results of a prospective, nonrandomized study. *Hepatology* **36**(4), 941–948.

Hepatotoxicity, Drug-Induced

DOMINIQUE LARREY AND GEORGES-PHILIPPE PAGEAUX
Hôpital Saint-Eloi, Montpellier School of Medicine, France

apoptosis Biological process in which cell death is progressive and programmed.

causality assessment Analytical method allowing evaluation of the relationship between drug administration and the occurrence of a liver injury.

cholangitis Inflammation of biliary cells. The process may be acute or chronic and may affect either the small bile duct, visible only with optic microscopy, or the large bile ducts, which can be assessed radiologically.

cholestatic liver injury Clinical and biological patterns reflecting impaired bile flow. The syndrome can be isolated (pure cholestasis) or associated with inflammation (cholestatic hepatitis).

cytolytic liver injury Clinical and biological patterns reflecting liver cell necrosis.

hepatitis Inflammation of the liver. The process may be acute or chronic according to the level of exposure to the damaging agent.

hepatotoxicity State of toxic damage to the liver.

mixed hepatitis Combination of cytolytic and cholestatic hepatitis.

pharmacogenetics Study of the interrelation of hereditary constitution and response to drugs.

pharmacovigilance General term including all actions evaluating drug safety.

reactive metabolites Products normally resulting from the transformation of the parent compound by enzymatic reactions, principally in the liver. A majority of produced metabolites are nontoxic and facilitate the elimination of the drugs. Occasionally, metabolites are unstable and react with cellular structures, thereby causing cell

Copyright 2004, Elsevier (USA). All rights reserved.

sympathetic systems, followed by impaired renal perfusion. Other renal diseases must be excluded before a diagnosis of HRS can be made. Type 2 HRS is characterized by a slow reduction in the GFR, usually over months, and is managed by avoiding known causes of renal dysfunction and careful use of diuretics, paracentesis, and TIPS. Type 1 HRS is a rapidly progressive disease with a high mortality if left untreated. Treatment of type 1 HRS with specific splanchnic vasoconstrictors appears to be evolving as effective therapy in liver transplant candidates and in patients with potential for hepatic recovery. TIPS and MARS are potential options for selected patients who either are not candidates for or have failed medical therapy.

See Also the Following Articles

Ascites • Cirrhosis • Fulminant Hepatic Failure • Liver Transplantation • Portal Hypertension and Esophageal Varices

Further Reading

Arroyo, V., Gines, P., Gerbes, A. L., *et al.* (1996). Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. *Hepatology* **23**, 164.

Arroyo, V, Guevara, M, and Gines, P. (2002). Hepatorenal syndrome in cirrhosis: Pathogenesis and treatment. *Gastroenterology* **122**, 1658–1676.

Brensing, K. A., Textor, J., Perz, J., *et al.* (2000). Long term outcome after transjugular intrahepatic portosystemic stent-shunt in nontransplant cirrhotics with hepatorenal syndrome: A phase II study. *Gut* **47**(2), 288–295.

Dagher, L., and Moore, K. (2001). The hepatorenal syndrome [review]. *Gut* **49**(5), 729–737.

Dagher, L., Patch, D., Marley, R., *et al.* (2000). Review article: Pharmacological treatment of the hepatorenal syndrome in cirrhotic patients. *Aliment. Pharmacol. Ther.* **14**, 515–521.

Ortega, R., Gines, P., Uriz, J., *et al.* (2002). Terlipressin therapy with and without albumin for patients with hepatorenal syndrome: Results of a prospective, nonrandomized study. *Hepatology* **36**(4), 941–948.

Hepatotoxicity, Drug-Induced

DOMINIQUE LARREY AND GEORGES-PHILIPPE PAGEAUX
Hôpital Saint-Eloi, Montpellier School of Medicine, France

apoptosis Biological process in which cell death is progressive and programmed.

causality assessment Analytical method allowing evaluation of the relationship between drug administration and the occurrence of a liver injury.

cholangitis Inflammation of biliary cells. The process may be acute or chronic and may affect either the small bile duct, visible only with optic microscopy, or the large bile ducts, which can be assessed radiologically.

cholestatic liver injury Clinical and biological patterns reflecting impaired bile flow. The syndrome can be isolated (pure cholestasis) or associated with inflammation (cholestatic hepatitis).

cytolytic liver injury Clinical and biological patterns reflecting liver cell necrosis.

hepatitis Inflammation of the liver. The process may be acute or chronic according to the level of exposure to the damaging agent.

hepatotoxicity State of toxic damage to the liver.

mixed hepatitis Combination of cytolytic and cholestatic hepatitis.

pharmacogenetics Study of the interrelation of hereditary constitution and response to drugs.

pharmacovigilance General term including all actions evaluating drug safety.

reactive metabolites Products normally resulting from the transformation of the parent compound by enzymatic reactions, principally in the liver. A majority of produced metabolites are nontoxic and facilitate the elimination of the drugs. Occasionally, metabolites are unstable and react with cellular structures, thereby causing cell

Copyright 2004, Elsevier (USA). All rights reserved.

damage that represents a major mechanism leading to drug-induced hepatotoxicity.

Hepatotoxicity is the primary cause of drug-related deaths and the principal reason that pharmaceuticals are withdrawn from the market. Despite improvements in toxicological studies and in safety analyses in clinical trial protocols, the frequency of drug-induced liver injury has not decreased over the past 10 years. All liver cells can be affected by drugs. The types of lesions vary according to the mechanism of drug action, the role of the parent drug or one of its metabolites, the route of drug administration, the drug dosage, and the susceptibility of the patient. Consequently, adverse reactions to drugs affecting the liver can reproduce the entire spectrum of noniatrogenic liver diseases (Table I). More than 1100 classic drugs are known to provoke hepatotoxic reactions. Furthermore, some drug-associated excipients, herbal medicines, and recreational or illegal compounds (such as amphetamines and cocaine) are also hepatotoxic. Drug-induced morbidity and mortality of are of concern to physicians, health authorities, and pharmaceutical companies, and it is important to understand and manage the most critical aspects of potential hepatotoxicity.

EPIDEMIOLOGICAL ASPECTS

Current knowledge of drug hepatotoxicity remains limited despite many effort to obtain reliable information. Data have been collected for the past 20 years, through the establishment of drug safety agencies in many countries. A limiting factor is the difficulty in determining the causality of a drug in relation to a hepatic injury, even with the development of sophisticated diagnostic methods. A second limiting factor is the method employed to collect data. Presently, a two-step strategy—cost/benefit toxicity evaluation screens and trials and postmarketing epidemiological surveillance—is being used.

The first step, the evaluation of the safety of a new drug in order to determine the benefit/risk ratio, is necessary to obtain marketing authorization. An initial screening is performed through animal toxicological experiments and, increasingly, by using cellular models. Toxicity is then assessed prospectively in healthy volunteers and patients. This allows analysis of the prevalence and type of liver test abnormalities, and triggers investigations to determine the potential mechanisms, the susceptibility factors, if any, and the ability to predict the liver injury and its severity. The number of participants enrolled in clinical trials generally ranges from 2000 to 5000. Only frequent events (>1%) can be detected. In the majority of cases, such an event leads to a halt in the development of the new drug. However, in some exceptional situations, the importance of the compound is such that liver toxicity may be acceptable. An example is tacrine, the first drug shown to have some efficacy in Alzheimer's disease.

The risk of hepatotoxicity varies widely from one drug to another. The high risk for marketed drugs is around 1%, the major examples being isoniazid, chlorpromazine, and tacrine. In contrast, for other drugs (for instance, antihistaminic compounds, minocycline, or penicillins), the risk of hepatotoxicity is exceedingly low, affecting less than 1/100,000 or even 1/1,000,000 persons. This makes it difficult to prove hepatotoxicity, especially when liver injury exhibits no specificity, and it is practically impossible to detect such an event at an early stage. For most marketed drugs, the risk of toxicity is moderate, ranging from 1/10,000 to 1/100,000 affected persons. Such is the case with troglitazone, for which hepatotoxicity has been estimated to be around 1/50,000 persons exposed (this is probably why toxicity was not clearly detected among the 5000 patients

TABLE I Cell Types and Characteristics of Lesions Comprising Drug-Induced Hepatotoxicity

Hepatocyte	Cholangiocyte (bile duct epithelial cell)	Endothelial cell	Stellate cell
Acute cytolytic hepatitis	Acute and chronic cholangitis	Venoocclusive disease	Perisinusoidal fibrosis
Pure cholestasis	Sclerosing cholangitis	Sinusoidal dilation	
Cholestatic hepatitis		Peliosis hepatis	
Chronic hepatitis		Budd–Chiari syndrome	
Cirrhosis		Regenerative nodular hyperplasia	
Macrovacuolar steatosis			
Microvesicular steatosis			
Phospholipidosis			
Steatohepatitis			
Granulomatous hepatitis			
Adenoma			

included in the therapeutic trials). These risk and detection factors explain why the first cases of drug-related hepatotoxicity are generally described only after a drug has been marketed for 1 to 2 years, when the drug has been administered to a statistically sufficient number of patients.

The second step in managing drug hepatotoxicity is postmarketing surveillance, which mainly relies on the spontaneous reporting of liver side effects. The process is clearly imperfect and many events are overlooked. There are very few data for estimating the proportion of underreporting and the true incidence of hepatotoxicity. The rare epidemiological studies of drug hepatotoxicity have been performed mainly on a retrospective basis and have been aimed at determining the frequencies and types of liver injuries caused by drugs, compared to other causes, over a given period. The importance of underreporting has been recently stressed by a unique prospective population-based study performed over 3 years. The analysis showed a global crude hepatotoxicity incidence of 13.9 per 100,000 persons. This emphasizes that spontaneously reported cases represent less than 10% of real events and that drug hepatotoxicity is largely underestimated as a cause of liver disease.

Although imperfect, retrospective studies provide some interesting information. A study made over 10 years in the liver unit of Hôpital Beaujon, Paris, France, revealed that among all adult patients admitted with hepatitis, 10% of the cases were related to drug toxicity. The prevalence of drug hepatotoxicity exceeded 40% for patients over 50 years. Another French national survey conducted in 1983 collected data from 980 cases of hepatitis due to drugs; 63% of the patients were women and most were over 50 years old. Similarly, the Danish Board of Adverse Reactions to Drugs recorded 572 cases of drug-induced hepatitis during 1968–1978 (6% of all drug-related side effects); again, most cases occurred in women over 50 years old. Drugs have been estimated to cause around 15–20% of all cases of fulminant and subfulminant hepatitis in Western countries, and 10% of all cases in Japan. The risk of a fulminant course in patients with drug-induced hepatitis with jaundice is around 20%, which is much higher than the risk in patients suffering from acute viral hepatitis with jaundice (1%). Further, in 70% of patients with drug-induced hepatitis with encephalopathy, the disease runs a subfulminant course.

MECHANISMS OF HEPATOTOXICITY

Despite many advances in the past 20 years, the mechanisms of hepatotoxicity of most drugs remain unknown. A single drug may have several toxic effects on the liver, or may produce either toxic or allergic hepatitis in different patients. The formation of reactive metabolites is relatively frequent. Despite several protective mechanisms, a drug may cause covalent binding of electrophilic metabolites to proteins, stimulation of lipid peroxidation by free radicals, and depletion and/or oxidation of glutathione. There follow a number of structural and functional lesions, including a sustained increase in cytosolic calcium, eventually leading to cell death. The formation of reactive metabolites may lead to two major types of hepatitis in humans: (1) toxic hepatitis, which occurs predictably after massive overdosage, e.g., with paracetamol, and (2) immunoallergic hepatitis, in which the drug triggers an adverse immune response directed against the liver. The main clinical features are the dose-independent response; the association with hypersensitivity reactions such as fever, chills, skin rash, hypereosinophilia, and immunoallergic thrombopenia; the shortened delay in response time on rechallenge; and the occasional presence of serum autoantibodies. These autoantibodies have been observed in cases of hepatitis caused by halothane, tienilic acid, dihydralazine, anticonvulsants, papaverine, and nitrofurantoin. The following mechanism has been postulated: (1) the drug is first metabolized into a reactive metabolite, which binds to the enzyme that generates it. (2) This produces a neoantigen, which, once presented to the immune system, may trigger an immune response characterized by the production of antibodies recognizing both the natural and/or the modified protein. (3) Rechallenge leads to increased neoantigen production, a situation in which the presence of antibodies may induce cell necrosis.

Several lines of evidence have recently suggested that apoptosis and mitochondrial dysfunction can also markedly contribute to acute hepatotoxicity. Other mechanisms of hepatotoxicity have been demonstrated to vary with the type of liver injury, as summarized in the Table II. For instance, microvesicular steatosis, which is induced by several drugs, including valproic acid, nonsteroidal antiinflammatory drugs, some tricyclic antidepressants, tetracycline, and acetylsalicylic acid, is caused by an inhibition at different stages of mitochondrial β-oxidation of fatty acids and/or alteration of the respiratory chain.

FACTORS MODULATING HEPATOTOXICITY

The risk of developing drug hepatotoxicity is influenced by various physiological and genetic factors, as summarized in Tables III and IV.

TABLE II Primary Mechanisms of Hepatotoxicity

Liver injuries	Mechanisms
Acute hepatitis	Metabolite-mediated toxicity
	Metabolite-mediated immunoallergy and/or autoimmunity
	Apoptosis
	Mitochondrial dysfunction
Acute cholestasis	Inhibition of biliary secretion
Macrovacuolar steatosis	Decreased secretion of lipoproteins
Microvacuolar steatosis	Inhibition of fatty acid mitochondrial β-oxidation
Nonalcoholic steatohepatitis	Mitochondrial dysfunction
	Cytokine release
Phospholipidosis	Inhibition of lysosomal phospholipases
Chronic hepatitis	Metabolite-mediated immunoallergic reaction
Vanishing bile duct syndrome	Autoimmune destruction of small bile ducts
	Abnormal multidrug resistance protein system?
Sclerosing cholangitis	Biliary ischemia caused by arterial lesions
Venoocclusive disease	Metabolite-mediated endothelial lesions
Perisinusoidal fibrosis	Activation of Ito cells

Physiological Factors

Age above 60 years is a promoting factor for hepatitis induced by isoniazid, halothane, and nitrofurantoin; children are more subject to the toxicity of valproic acid and salicylates. In particular, aspirin more frequently induces microvesicular steatosis and Reye's syndrome. Gender is also involved, women being more frequently exposed to hepatotoxic methyldopa, diclofenac, and nitrofurantoin, whereas men are more prone to azathioprine-induced injury. Pregnancy appears to be an influencing factor. For instance, most of the tetracycline-induced severe hepatitis has been observed in pregnant women receiving high doses of the drug by the intravenous route. Experimentally, pregnant mice are particularly susceptible to acetaminophen hepatotoxicity, probably because of the glutathione requirement of the fetus and placenta, which depletes hepatocyte glutathione stores.

Nutrition status can affect hepatotoxicity in different ways. For instance, obesity promotes halothane and

TABLE III Acquired Factors Contributing to Drug Hepatotoxicity

Factor	Drug
Age	
>60 years old, children	Isoniazid, nitrofurantoin, halothane, valproic acid, salicylates
Gender	
Women	Nitrofurantoin, diclofenac, isoniazid
Men	Azathioprine
Nutrition	
Obesity	Halothane, methotrexate
Fasting/malnutrition	Acetaminophen
Pregnancy	Acetaminophen, tetracycline
Chronic alcohol abuse	Acetaminophen, methotrexate
Combined drug effects	
Enzyme induction	Rifampicin–isoniazid; other anticonvulsant drugs and valproic acid
Enzyme inhibition	Troleandomycin–estrogens, phenobarbital–antidepressants
Toxicity combination	Isoniazid–pyrazinamide, isoniazid–acetaminophen
Disease	
HIV infection	Cotrimoxazole/sulfonamides, isoniazid
Hyperthyroidism	Halothane
Hepatitis B and C virus infection with liver fibrosis	Antituberculosis, antiretroviral agents, nonsteroidal antiinflammatory drugs?
Nonalcoholic steatohepatitis	Methotrexate

TABLE IV Genetic Factors Contributing to Drug Hepatotoxicity

Factor	Drug
Deficiency in CYP2D6	Perhexiline
Deficiency in CYP2C19	Atrium, troglitazone
Deficiency in NAT2[a]	Sulfonamides, dihydralazine
Deficiency in sulfoxidation	Chlorpromazine?
Deficiency in glutathione synthetase	Acetaminophen
Deficiency in GSTase type T[b]	Tacrine?
Deficiency in detoxication of reactive metabolites (unknown mechanisms)	Halothane, phenytoin, carbamazepine, amineptine, sulfonamides
HLA system	
A11	Halothane
DR6 and DR2	Nitrofurantoin
A8	Clometacin
A11	Tricyclic antidepressants
A11	Diclofenac
DR6	Chlorpromazine
DRB1 1501, DRQBI 0602	Amoxicillin–clavulanic acid

[a] NAT2, *N*-Acetyltransferase 2.
[b] GSTase, Glutathione *S*-transferase.

methotrexate hepatotoxicity whereas fasting and malnutrition facilitate acetaminophen (paracetamol) hepatitis, again, probably by depleting glutathione stores in hepatocytes. Chronic alcohol abuse facilitates acetaminophen hepatotoxicity, probably by complex mechanisms involving the induction of critical enzymes such as cytochrome P450 2E1, forming toxic metabolites (*N*-acetyl-*p*-quinone imine), and by decreasing the defense against these metabolites because of glutathione depletion. Chronic alcohol abuse also increases methotrexate toxicity.

Drug interactions can also contribute to drug hepatotoxicity. For instance, enzyme induction leading to increased formation of toxic metabolites from one drug has been demonstrated with the rifampicin–isoniazid combination; rifampicin facilitates the transformation of isoniazid into toxic metabolites. Similarly, phenobarbital increases the hepatotoxicity of antidepressants. In contrast, enzyme inhibition can also be implicated. This is demonstrated by the troleandomycin–estrogen interaction. Troleandomycin inhibits estrogen metabolism by blocking cytochrome P450 3A, thereby creating an estrogen overdose with cholestatic effects. Some drug combinations lead to a higher risk of toxicity by additional production of toxic metabolites (Table III).

A preexisting liver disease may contribute to the hepatotoxicity of some drugs, such as antiretroviral agents and methotrexate. Conversely, antiretroviral agents may exacerbate latent hepatitis C or B by restoring an efficient immune system and by modifying immunotolerance. Chronic viral hepatitis also may increase the hepatotoxicity of antituberculosis compounds and nonsteroidal antiinflammatory drugs. Some extrahepatic diseases can also contribute to drug hepatotoxicity. For instance, hyperthyroidism promotes halothane hepatitis and HIV infection is a contributing factor to the hepatotoxicity of cotrimoxazole.

Genetic Factors Contributing to Drug Hepatotoxicity

A deficiency in cytochrome P450 2D6 (CYP2D6) appears to be a major determinant of perhexilline hepatotoxicity. Indeed, more than 75% of patients with perhexilline hepatotoxicity are deficient for CYP2D6. This enzyme deficiency is present in 6–8% of Caucasian populations and is transmitted as an autosomal recessive trait in relationship with more than 10 mutations of the gene for CYP2D6.

A deficiency in cytochrome P450 2C19 (CYP2C19) may be involved in Atrium hepatotoxicity. Atrium is a complex drug combining febarbamate, difebarbamate, and phenobarbital. A recent study comprising a small number of patients with a previous history of Atrium hepatitis showed that all of them had a partial or a complete deficiency in CYP2C19. The prevalence of this deficiency in Caucasians is around 3–5% and around 20% in Asians. This small study, however, deserves to be confirmed by a study of a larger number of patients. Troglitazone hepatotoxicity also appears to be associated with CYP2C19

deficiency, with the proportion of poor metabolizers being twice as high in affected patients compared to a control group.

A deficiency in acetylation capacity related to an inactive *N*-acetyltransferase 2 (NAT2) has been shown to be involved in the hepatotoxicity of sulfonamides and hydralazine. This defect is transmitted as an autosomal recessive trait. More than 25 mutant alleles have been associated with the slow acetylation phenotype. The high frequency of the slow acetylation phenotype in most populations suggests that the defect may contribute, but is not sufficient alone, to the toxicity of sulfonamides and hydralazine.

A deficiency in sulfoxidation has been noted in one study assessing patients with chlorpromazine hepatitis. However, the lack of reproducibility of the method used for analyzing sulfoxidation has made this conclusion questionable. Further studies using reproducible tests for sulfoxidation polymorphism are required.

A deficiency in glutathione synthetase is an uncommon condition (1/10,000), responsible for a syndrome comprising oxoprolinuria and hemolytic anemia. By using an *in vitro* lymphotoxicity assay on specimens from patients with this deficiency, it has been shown that the patients may be more susceptible to acetaminophen hepatotoxicity.

Other metabolic deficiencies and liver susceptibility to drugs have been linked. By using the *in vitro* lymphotoxicity assay developed by Spielberg, several groups have shown deficiencies in detoxification capacity for reactive metabolites in patients with drug-induced hepatitis. This deficiency was also observed in some family members, suggesting a genetic defect. Such observations have been made for halothane, phenytoin, carbamazepine, amineptine, and sulfonamides. The precise defect involved in these susceptibilities has still not been identified. In another study, it has been proposed that tacrine hepatotoxicity may be promoted by a deficiency in glutathione *S*-transferase, type T. However, this has not been confirmed by any other study.

Finally, genetic variations in the immune system may be involved in drug hepatotoxicity. Indeed, an association has been observed between several human leukocyte antigen (HLA) haplotypes and some drugs (Table IV). The association between amoxicillin–clavulanate and HLA DRB1 1501 is particularly strong.

DEFINITIONS AND CLASSIFICATIONS

An international consensus meeting under the auspices of the Council for International Organizations of Medical Sciences (CIOMS) and gathering a panel of experts have come together to define a common language between the different partners involved in the assessment of drug-induced liver injury.

Definitions

The following designations for drug-induced liver disorders have been proposed, based mainly on liver test abnormalities. When a liver biopsy has been performed, the lesion should be named according to the histological findings, e.g., cirrhosis, acute hepatitis, chronic hepatitis, or hepatic necrosis. In the absence of histological data (and this corresponds to the majority of cases), the preferred term is "liver injury." Liver injury is defined by an increase of over two times the upper limit of normal range (2*N*) in serum alanine aminotransferase (ALT) or conjugated bilirubin, or a combined increase of aspartate aminotransferase (AST), alkaline phosphatase (AP), and total bilirubin, provided one of them is above 2*N*.

Clinicopathological Classifications

Albeit drug hepatotoxicity can reproduce practically the whole spectrum of liver disease, acute liver injuries (characterized by abnormalities lasting less than 3 months) predominate (about 90% of cases). Acute liver injuries can be classified in three groups using biochemical criteria based on ALT and AP, and their ratio (R). This classification has the advantage of separating types of hepatitis with different courses and prognostic features.

Acute Hepatocellular Hepatitis

Acute hepatocellular hepatitis is defined by an ALT above 2*N* or an ALT/AP ratio of ≥ 5. Acute hepatocellular hepatitis generally has no specific features and mimics acute viral hepatitis. Liver injury may remain asymptomatic, revealed only by an increase in ALT or by nonspecific symptoms such as asthenia or anorexia. Jaundice is inconstant. The main biological feature is the marked increase in transaminases. The major pathological finding is liver cell necrosis, generally associated with inflammatory infiltration. The presence of eosinophils in the infiltrate and the centrilobular predominance of the lesions argue for drug hepatotoxicity.

Hepatitis may be associated with hypersensitivity manifestations such as fever, chills, skin rash, hypereosinophilia, and immunoallergic thrombopenia, which suggest an immunoallergic mechanism. Several hundred of drugs can induce this type of hepatitis, which may also be produced by herbal medicines, illegal

TABLE V Primary Drugs Responsible for Acute Hepatocellular Hepatitis

Primary conventional drugs	
	Without hypersensitivity: acetaminophen, isoniazid, pyrazinamide, ketoconazole, valproic acid, troglitazone, tacrine, paroxetine, tolcapone, antiretroviral drugs (didanosine, zidovudine, zalcitabine, stavudine, ritonavir, indinavir, saquinavir, efevirenz, nelfinavir), anti-mycotics (itraconzole, ketoconazole, terbinafine), cytokines and growth factors (interleukin-2, interleukin-12, interleukin-2F,interleukin-3, interleukin-6, granulocyte colony-stimulating factor), leflunomide, gemcitabine
	With hypersensitivity: nonsteroidal inflammatory drugs (almost all drugs), sulfonamides, almost all antidepressants (tricyclic, iproniazid, fluoxetine), halothane and derivatives, abacavir, halothane and its derivatives
Herbal medicines	
	Pyrrolizidine alkaloids (crotaloria, senecio), *Atractylis gummifera* L., germander, chinese herbal preparations, chapparal leaf, senna, skullcap, valerian, mentha-containing pennyroyal oil, kava-kava, great celandine
Illegal compounds	
	Cocaine, "Ecstasy" (*N*-methyl-3,4-methylene dioxyamphetamine)
Excipients	
	Sodium saccharinate (dihydroergocristine; chlordemethyldiazepam; prednisolone, 20 mg), polysorbate (E-Ferol, intravenous amiodarone), propylene glycol (multivitamin intravenous solutions)

compounds, and even some excipients. The main causes are indicated in the Table V. In most instances, the discontinuation of treatment is followed by a quick improvement of symptoms and a complete recovery within 1 to 3 months. Sometimes, however, acute hepatitis may be followed by fulminant or subfulminant hepatitis. The course is then very severe, with a spontaneous mortality around 90%. The only treatment is emergency liver transplantation. The risk to develop fulminant or subfulminant hepatitis is particularly high when drug administration is continued despite the occurrence of jaundice. The course of the disease may be more insidious in the progressive development of chronic hepatitis or even cirrhosis.

Acute Cholestatic Liver Injury

Acute cholestatic liver injury is characterized by an isolated increase of serum alkaline phosphatase above 2*N*, or by an ALT/AP ratio of <2. There are two subtypes, pure (bland) cholestasis and cholestatic hepatitis.

Pure cholestasis Pure cholestasis is mainly characterized by jaundice, pruritus, and dark urine. Biologically, there is an increase in alkaline phosphatase, (-glutamyl transpeptidase, and conjugated bilirubin. Transaminases are normal or only slightly increased; liver biopsy mainly shows bilirubin deposits in hepatocytes and more or less dilated biliary canaliculi containing biliary pigments. Cholestasis predominates in centrilobular area. Pure cholestasis is observed with a few drugs, mainly hormonal derivatives (Table VI). Discontinuation of the causative drug is followed by a complete recovery.

Acute cholestatic hepatitis In addition to manifestations of pure cholestasis, acute cholestatic hepatitis liver injury may be associated with abdominal pain, fever, and chills, which can mimic acute biliary obstruction. Hypersensitivity manifestations are frequent. Liver biopsy shows cholestasis associated with inflammatory infiltration in portal tracts. The prognosis is much better than that for hepatocellular hepatitis. After drug withdrawal, symptoms rapidly disappear and recovery occurs within a few weeks. Rarely, however, chronic cholestasis simulating primary biliary cirrhosis may develop. This particular course is observed in cases with associated cholangitis. Several hundreds of drugs

TABLE VI Primary Drugs Responsible for Acute Cholestatic or Mixed Hepatitis and Cholangitis

Pure cholestasis	
	Oral contraceptives, estrogens, estrogens plus troleandomycin or erythromycin, androgens, tamoxifen, azathioprine, cytarabine
Acute cholestatic/mixed hepatitis	
	Phenothiazines, tricyclic antidepressants, nonsteroidal antiinflammatory drugs, carbamazepine, macrolide antibiotics, amoxicillin/clavulanic acid, sulfonamides, gold salts, β-lactam antibiotics, propoxyphene, antiretrovirals (didanosine, zidovudine, stavudine, ritonavir), interleukins (IL-2, IL-6, IL-2)
Acute cholangitis	
	Phenothiazines, carbamazepine, tricyclic antidepressants, macrolide antibodies, amoxicillin−clavulanic acid
Chronic cholangitis	
	Phenothiazines, arsenical derivatives, tricyclic antidepressants, macrolide antibodies, thiabendazole, tetracycline, fenofibrate

are able to cause cholestatic hepatitis; the main ones are indicated in the Table VI.

Mixed-Pattern Acute Liver Injury

Mixed-pattern acute liver injury is characterized by an ALT/AP ratio of between 2 and 5. The clinicopathological manifestations are a mixture of those observed with hepatocellular and cholestatic hepatitis, but also include granulomatous reactions. A mixed-pattern liver injury is frequently associated with immunoallergic manifestations. The prognosis is generally very good. The main causative drugs are indicated in Table VI.

Other Types of Liver Injury

Other types of liver injury were not well characterized by the international consensus meeting. It was, however, mentioned that the term "chronic" applies to liver diseases with a course longer than 3 months, without any reference to the underlying lesion. This is to distinguish from drug-induced chronic hepatitis, well characterized by histological examination. The lack of precision of this system, based on biological tests, has been recently stressed by several clinicians. The definitions of chronic and other drug-induced liver disorders rely principally on histological findings. Their frequency is globally low. Drugs are the cause of less than 1% of cases of chronic hepatitis and cirrhosis. However, for some rare lesions, drugs are an important cause. Such is the case for estrogens and hepatic adenoma, and the contribution of thiopurines and antineoplastic agents to hepatic vascular disorders such as nodular regenerative hyperplasia and peliosis hepatitis. Such drug-induced injuries and the primary drugs involved are indicated in Table VII.

TABLE VII Other Types of Drug-Induced Liver Injury and Causative Drugs

Chronic hepatitis and/or cirrhosis	
	Valproic acid, amiodarone, aspirin, benzarone, halothane, iproniazid, isoniazid, methotrexate, methyldopa, nitrofurantoin, vitamin A, papaverine, tamoxifen
Granulomatous hepatitis	
	Allopurinol, carbamazepine, quinidine, sulfonamides, gold salts, phenylbutazone, penicillamine, dapsone
Macrovacuolar steatotis	
	Corticosteroids, methotrexate, asparaginase
Microvesicular steatosis	
	Nonsteroidal antiinflammatory drugs, valproic acid, tetracycline, amineptine, tianeptine, aspirin antiretroviral drugs (dideoxynucleotides)
Steatohepatitis with phospholipidosis	
	Amiodarone, perhexilline, diethylaminoethoxyhexestrol
Steatohepatitis without phospholipidosis	
	Nifedipine, diltiazem, tamoxifen
Vascular diseases of the liver	
	Perisinusoidal fibrosis: vitamin A, azathioprine, 6-mercaptopurine, methotrexate
	Sinusoidal dilatation and peliosis hepatitis: oral contraceptives, androgens, estrogens, azathioprine
	Venoocclusive disease: pyrrolyzidine alkaloids, azathioprine, anticancerous agents
	Budd–Chiari syndrome: oral contraceptives, dacarbazine
Tumors	
	Adenoma and hepatocellular carcinoma: androgens, oral contraceptives, estrogens

DIAGNOSIS

There are generally no specific markers or tests for the diagnosis of drug-induced liver injury. Therefore, the diagnosis entirely relies on circumstantial evidence. To address the problem, various analytical methods have been developed to assess the causality of a given drug in the occurrence of liver injury. The principles of all these methods are similar. Causality assessment relies on chronological and clinical criteria to allow elimination of other causes and to demonstrate the role of the offending drug (Table VIII).

Chronological Criteria

The first criterion is the time interval between the beginning of treatment with the suspected agent and the onset of liver injury; this "latent period to onset" varies widely. It is considered suggestive when the interval is between 1 and 12 weeks. A shorter duration (1 or 2 days) may be observed in patients who have been previously exposed to the compound and have become sensitized. A delay of between 3 and 12 months remains compatible but is less common. A delay above 1 year is very uncommon; it generally makes the role of the suspected drug very unlikely in the case of acute hepatitis, but note that with unusual forms of chronic liver disease (such as hepatic adenoma and some types of vascular injury mentioned earlier), the latent period to onset is often several years.

The second criterion is the resolution of liver test abnormalities after withdrawal of the treatment. This is very suggestive when clinical features disappear within a few days, and when aminotransferases decrease by more than 50% in a week. Usually, complete recovery is obtained within a few weeks.

The third criterion is a relapse of liver test abnormalities after an accidental readministration of the offending drug. This is a very good diagnostic criterion.

TABLE VIII Diagnostic Criteria for Drug-Induced Liver Diseases

Chronological criteria
Interval between the beginning of the treatment and the onset of liver injury: 1–12 weeks
Regression of liver abnormalities after withdrawal of the treatment
Relapse of liver abnormalities after accidental readministration (rechallenge) of the offending drug; deliberate rechallenge must be avoided
Clinical criteria
Elimination of other causes
Previous hepatic or biliary disease
Alcohol abuse
Viral hepatitis (HAV, HBV, HCV, HDV, HEV, cytomegalovirus, Epstein–Barr virus, herpes
Biliary obstruction (ultrasonography, magnetic resonance imaging, etc.)
Autoimmune hepatitis/cholangitis
Ischemia and congestion of the liver
Wilson's disease
Bacterial infection (*Listeria*, *Campylobacter*, *Salmonella*, Lyme disease)
Positive clinical criteria
Age >50 years
Intake of many drugs
Intake of a known hepatotoxic drug
Specific serum autoantibodies: anti-M6, anti-LKM2, anti-CYP1A2, anti-CYP2E1
Positive blood analysis for acetaminophen, vitamin A
Liver biopsy showing drug deposits (vitamin A), microvesicular steatosis, eosinophil infiltration, centrilobular zonal necrosis, associated bile duct injury

However, such reexposure should not be done on purpose because it can be very dangerous, particularly in cases of immunoallergic hepatitis. In this situation, the readministration of a single tablet can occasionally provoke fulminant hepatic failure.

Clinical Criteria

Clinical criteria are based on the exclusion of other causes that might explain the liver injury, and on the presence of features tending to favor a drug etiology.

Eliminating or Negative Criteria

Analytical features vary according to the type of liver injury. For acute hepatitis, it is important to seek a history of hepatitis or biliary disease, alcohol abuse, or epidemiological circumstances that are compatible with viral infection (intravenous drug use, blood transfusion, recent surgery, travel in an endemic area). Appropriate serological tests should be performed to exclude viral hepatitis (hepatitis A, B, C, D, and E viruses), and in some circumstances, cytomegalovirus, Epstein–Barr virus, and herpesviruses. There does not appear to be a useful role of testing for hepatitis G virus (HGV) or transfusion-transmitted virus (TTV). However, the potential for hepatic ischemia and congestion related to cardiorespiratory failure should be excluded, particularly in the elderly and after cardiac surgery. Biliary obstruction should be eliminated by ultrasonography or other appropriate examinations (magnetic resonance imaging, endoscopic ultrasonography). Autoimmune hepatitis or cholangitis should also be ruled out, as should specific bacterial infections that can produce acute hepatitis, such as infection by *Campylobacter*, *Salmonella*, *Listeria*, and Lyme disease. Finally, Wilson's disease should be considered in children and young adults.

Positive Criteria

The presence in serum of specific autoantibodies—for instance, antimitochondrial type 6, anti-LKM2, anti-CYP1A2 and anti-CYP2E1—indicates potentially important diagnostic markers, although the diagnostic accuracy of these tests has not been studied and they are not widely available. Drug analyses of blood and liver tissue may be useful in detecting acetaminophen and vitamin A overdoses.

The presence of hypersensitivity manifestations, albeit not completely specific, is a positive argument for involvement of a drug and for an immunoallergic mechanism, as typified by the reactive metabolite syndrome for some anticonvulsants, sulfonamides, and protease inhibitors. Finally, liver biopsy may also contribute to the diagnosis by showing the presence of drug deposits (vitamin A) or lesions suggestive of drug reactions—for instance, microvesicular steatosis, eosinophilic infiltrates, centrilobular necrosis, or associated bile duct lesions.

Causality Assessment by Scoring Systems

The analytical approach is insufficient in many situations; complementary methods have been developed to provide a better diagnostic assessment, by ascribing scores to groups of items and then by calculating a global score. The main scoring system, the Roussel Uclaf Causality Assessment Method (RUCAM), was presented in the original publication of the CIOMS (Table IX). RUCAM (CIOMS) has been independently validated in a large number of patients. The global score for acute liver injury theoretically ranges from −5 to 14. On this basis, the diagnosis is classified as follows: ≤0, relationship with the drug excluded;

TABLE IX Method for Causality Assessment of Adverse Drug Reactions[a]

Criterion	Interpretation	Score
1. Time to onset of the reaction	*Highly suggestive*	+3
	Suggestive	+2
	Compatible	+1
	Inconclusive	0
If incompatible, then case "unrelated"		
If information not available, then case "insufficiently documented"		
2. Course of the reaction	Highly suggestive	+3
	Suggestive	+2
	Compatible	+1
	Against the role of the drug	−2
	Inconclusive or not available	0
3. Risk factors(s) for drug reaction[b]	Presence	+1 to 2
	Absence	0
4. Concomitant drug(s)[c]	Time to onset incompatible	0
	Time to onset compatible but unknown reaction	−1
	Time to onset compatible and unknown reaction	−2
	Role proved in this case	−3
	None or information not available	0
5. Non-drug-related causes[d]	Ruled out	+2
	Possible or not investigated [d]	+1 to 2
	Probable	−3
6. Previous information on the drug	Reaction unknown	0
	Reaction published but unlabeled	+1
	Reaction labeled in the product's characteristics	+2
7. Response to readministration	Positive	+3
	Compatible	+1
	Negative	−2
	Not available or not interpretable	0
(or plasma concentration of the drug known as toxic or validated laboratory test with high specificity, sensitivity, and predictive values)	Positive	+3
	Negative	−3
	Not interpretable or not available	0

[a] Using the Roussel Uclaf Causality Assessment Method. After Benichou *et al*. (1993).
[b] One additional point for every validated risk factor (maximal value +2).
[c] Sum of negative values of criteria 4 and 5 cannot be lower than −4.
[d] Depending on the nature of the reaction.

1−2, unlikely; 3−5, possible; 6−8, probable; >8, highly probable.

A simplified clinical diagnostic scale (CDS) for the diagnosis of drug-induced liver injury has been recently proposed by a Spanish group. The criteria are very similar to those used in the RUCAM (CIOMS). They lead to a score separating adverse reactions: score >17, definite; 14−17, probable; 10−13, possible; 6−9, unlikely; <6, drug hepatotoxicity excluded.

The agreement between the RUCAM (CIOMS) and CDS scoring systems has been tested in two recent studies. Discrepancies were observed in cases of cholestasis, metabolic idiosyncratic, fulminant hepatitis, or death. The best correlation was found in cases of liver injury with immunoallergic reactions. Presently, there are no data showing whether the CDS method has additional advantages over the RUCAM (CIOMS) method.

Performance of Causality Assessment

In rare circumstances, a link between drug and hepatotoxicity may be probable or highly probable (RUCAM score >6 and CDS score >17): examples

include clear drug overdose (acetaminophen), relapse after accidental readministration, and presence of specific features for drug hepatitis (dihydralazine, iproniazid). More often, the diagnosis is possible: liver injury exhibits no specific features, the history is chronologically suggestive, and other causes of liver injury have been reasonably excluded (RUCAM score 3–5 and CDS score 10–13). Frequently, however, the diagnosis remains doubtful or unlikely: there are no specific features to the liver injury and critical pieces of information regarding chronological or clinical data are missing (RUCAM score 1–2 and CDS score 6–9). It is noteworthy that fulminant hepatitis must always be classified in this way because it is usually impossible to assess complete recovery after drug withdrawal due to the poor outcome or treatment by liver transplantation.

The diagnosis is excluded when another cause has been demonstrated (viral infection), when chronology is not compatible (e.g., when treatment had been started at a time when symptoms were already present), or when there is a delay exceeding 15 days between the end of the treatment and the onset of liver injury (RUCAM score (0 and CDS score <6). However, again there are some exceptions. With hepatitis caused by halothane and related haloalkane anesthetics, reactions often occur 3 weeks after the first exposure (the latent period is shorter with repeat exposures). Similarly, for a clavulanic acid–amoxicillin combination, acute hepatitis frequently occurs 3–4 weeks after the discontinuation of the treatment.

TREATMENT AND PREVENTION

The main therapeutic measure to prevent the progression to a more severe liver injury consists of discontinuing the administration of the offending drug. In some circumstances, this not sufficient and liver injury continues to develop, as seen for fulminant and subfulminant hepatitis, some cases of chronic liver diseases with storage of the drug in the liver or fat tissue, or protracted immunologic reactions. The main active treatment for acetaminophen intoxication is administration of *N*-acetylcysteine as antidote. Thus, prevention of further damage represents a determining point.

Many improvements are required in managing hepatotoxicity. First, it is important to have a better evaluation of drug hepatotoxicity epidemiology. Postmarketing surveillance on the basis of spontaneous reports is clearly insufficient. Prospective studies are required, in particular to identify the most severe reactions, fulminant and subfulminant hepatitis. These represent a major consideration for the benefit/risk ratio of a drug. An unfavorable ratio will lead to withdrawal of the compound from the market. Second, improvements in the definition, classification, and diagnostic methods of drug-induced liver disease provide another challenge. The new scoring systems already represent an improvement, but they also have some limitations. They are not easy to handle in routine clinical practice and a substantial number of cases remain indeterminate or doubtful in nature because key data are missing. Therefore, it is important to find more specific markers of hepatotoxicity; some of these already exist for rare drugs. Third, it is also very important to determine which patients are most at risk. Several physiological and genetic factors can promote hepatotoxicity. The development and the wide availability of genotyping tests for drug-metabolizing enzyme polymorphisms should help to better understand the influence of genetics on drug hepatotoxicity and to prevent side effects. Finally, the importance of monitoring susceptible patients in followup settings should be reinforced.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Cholangitis, Sclerosing • Cholestatic Diseases, Chronic • Cytochrome P450 • Hepatotoxins • Hyperthyroidism • Sinusoidal Obstruction Syndrome (Hepatic Venoocclusive Disease)

Further Reading

Benichou, C., Danan, G., and Flahault, A. (1993). Causality assessment of adverse reactions to drug—II. An original model for validation of drug causality assessment methods: Case reports with positive rechallenge. *J. Clin. Epidemiol.* **46**, 1331–1336.

Bissell, D. M., Gores, G. J., Laskin, D. L., and Hoofnagle, J. H. (2001). Drug-induced liver injury: Mechanism and test systems. *Hepatology* **33**, 1009–1013.

Farrell, G. C. (1994). Drug-induced liver disease. Churchill Livingstone, London.

International Consensus Meeting. (1990). Report. Criteria of drug-induced liver disorders. *J. Hepatol.* **11**, 272–276.

Kaplowitz, N. (2001). Causality assessment versus guilt-by-association in drug hepatotoxicity. *Hepatology* **33**, 308–309.

Larrey, D. (1997). Hepatotoxicity of herbal medicines. *J. Hepatol.* **26**(Suppl. 1), 47–51.

Larrey, D. (2000). Drug-induced liver diseases. *J. Hepatol.* **32**, 77–88.

Larrey, D. (2002). Epidemiology and individual susceptibility to adverse drug reactions affecting the liver. *Semin. Liver Dis.* **22**, 145–167.

Lee, W. M. (2000). Assessing causality in drug-induced liver injury. *J. Hepatol.* **33**, 1003–1006.

Lucena, I., Camargo, R., Andrade, J., Perez-Sanchez, C. J., and Sanchez De La Questa, F. (2001). Comparison of two clinical scales for causality assessment in hepatotoxicity. *Hepatology* **33**, 123–130.

Maria, V. A. J., and Victorino, R. M. M. (1997). Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* **26**, 664–669.

Pessayre, D., Larrey, D., and Biour, M. (1999). Drug-induced liver injury. *In* "Oxford Textbook of Clinical Hepatology" (J. Bircher, J. P. Benhamou, N. McIntyre, M. Rizzetto, and J. Rodés, eds.), 2nd Ed., Vol. 2, pp. 1261–1315. Oxford University Press, Oxford.

Schuppan, D., Jia, J. D., Brinkhaus, B., and Hahn, E. G. (1999). Herbal products for liver diseases: A therapeutic challenge for the new millennium. *Hepatology* **30**, 1099–1104.

Zimmerman, H. J. (1999). "The Adverse Effects of Drugs and Other Chemicals on the Liver," 2nd Ed. Lippincott Williams & Wilkins, Baltimore, MD.

Hepatotoxins

MARK J. CZAJA
Albert Einstein College of Medicine, New York

cytochrome P450s A superfamily of enzymes that mediate the oxidative metabolism of many endogenous and foreign compounds.

lipopolysaccharide A lipid component of the cell wall of gram-negative bacteria.

tumor necrosis factor α A cytokine released from macrophages that induces biological effects, including cytotoxicity, on other cells.

Hepatotoxins are synthetic or naturally occurring compounds that cause a variety of forms of liver injury through their direct or indirect damage to hepatocytes. Depending on the toxin, its likelihood to induce liver injury may be predictable or unpredictable (idiosyncratic). The type of liver disease that may develop depends on the toxin and can range from only asymptomatic elevations in serum liver function tests to acute or chronic liver failure. The mechanisms by which hepatotoxins injure the liver are complex and include both direct forms of biochemical cellular injury and indirect injury resulting from products of the accompanying inflammatory response.

INTRODUCTION

Many chemical compounds are able to cause hepatic injury and the vast majority of them cause injury limited only to the liver. The liver is particularly susceptible to toxic injury for two reasons. The first is that the portal blood supply from the intestine drains directly to the liver. Hepatocytes are therefore unique in their direct exposure to high levels of ingested toxins. The second is that one of the prime functions of the liver is to metabolize endogenous and foreign compounds. To perform this function, hepatocytes express high levels of a number of metabolic enzymes, such as those in the cytochrome P450 family. Although the biotransformation of these compounds is protective to the body as a whole, it may result in the generation of a metabolite that becomes toxic when large amounts accumulate in the hepatocyte. Hepatotoxins exist in a number of forms. The vast majority of clinically important toxin-induced human liver disease results from chronic alcohol ingestion. A number of medicinal drugs can also cause hepatotoxicity even when used in the usual therapeutic doses. Certain varieties of mushrooms and an increasing number of herbs usually used as alternative medicines are known hepatotoxins. Industrial or environmental chemicals are currently rare causes of hepatotoxicity.

FORMS OF HEPATOTOXIN-INDUCED LIVER INJURY

Hepatotoxins can have either predictable or idiosyncratic liver toxicities. For example, the pain reliever acetaminophen has a predictable hepatotoxicity that increases as the amount ingested exceeds the known therapeutic range. In contrast, other agents, such as inhaled anesthetics, rarely cause a sporadic, idiosyncratic hepatotoxicity even when they are used in the

Copyright 2004, Elsevier (USA). All rights reserved.

Maria, V. A. J., and Victorino, R. M. M. (1997). Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* **26**, 664–669.

Pessayre, D., Larrey, D., and Biour, M. (1999). Drug-induced liver injury. *In* "Oxford Textbook of Clinical Hepatology" (J. Bircher, J. P. Benhamou, N. McIntyre, M. Rizzetto, and J. Rodés, eds.), 2nd Ed., Vol. 2, pp. 1261–1315. Oxford University Press, Oxford.

Schuppan, D., Jia, J. D., Brinkhaus, B., and Hahn, E. G. (1999). Herbal products for liver diseases: A therapeutic challenge for the new millennium. *Hepatology* **30**, 1099–1104.

Zimmerman, H. J. (1999). "The Adverse Effects of Drugs and Other Chemicals on the Liver," 2nd Ed. Lippincott Williams & Wilkins, Baltimore, MD.

Hepatotoxins

MARK J. CZAJA
Albert Einstein College of Medicine, New York

cytochrome P450s A superfamily of enzymes that mediate the oxidative metabolism of many endogenous and foreign compounds.

lipopolysaccharide A lipid component of the cell wall of gram-negative bacteria.

tumor necrosis factor α A cytokine released from macrophages that induces biological effects, including cytotoxicity, on other cells.

Hepatotoxins are synthetic or naturally occurring compounds that cause a variety of forms of liver injury through their direct or indirect damage to hepatocytes. Depending on the toxin, its likelihood to induce liver injury may be predictable or unpredictable (idiosyncratic). The type of liver disease that may develop depends on the toxin and can range from only asymptomatic elevations in serum liver function tests to acute or chronic liver failure. The mechanisms by which hepatotoxins injure the liver are complex and include both direct forms of biochemical cellular injury and indirect injury resulting from products of the accompanying inflammatory response.

INTRODUCTION

Many chemical compounds are able to cause hepatic injury and the vast majority of them cause injury limited only to the liver. The liver is particularly susceptible to toxic injury for two reasons. The first is that the portal blood supply from the intestine drains directly to the liver. Hepatocytes are therefore unique in their direct exposure to high levels of ingested toxins. The second is that one of the prime functions of the liver is to metabolize endogenous and foreign compounds. To perform this function, hepatocytes express high levels of a number of metabolic enzymes, such as those in the cytochrome P450 family. Although the biotransformation of these compounds is protective to the body as a whole, it may result in the generation of a metabolite that becomes toxic when large amounts accumulate in the hepatocyte. Hepatotoxins exist in a number of forms. The vast majority of clinically important toxin-induced human liver disease results from chronic alcohol ingestion. A number of medicinal drugs can also cause hepatotoxicity even when used in the usual therapeutic doses. Certain varieties of mushrooms and an increasing number of herbs usually used as alternative medicines are known hepatotoxins. Industrial or environmental chemicals are currently rare causes of hepatotoxicity.

FORMS OF HEPATOTOXIN-INDUCED LIVER INJURY

Hepatotoxins can have either predictable or idiosyncratic liver toxicities. For example, the pain reliever acetaminophen has a predictable hepatotoxicity that increases as the amount ingested exceeds the known therapeutic range. In contrast, other agents, such as inhaled anesthetics, rarely cause a sporadic, idiosyncratic hepatotoxicity even when they are used in the

Copyright 2004, Elsevier (USA). All rights reserved.

prescribed therapeutic amounts. Host factors clearly play a role in determining whether hepatotoxicity occurs, even in the cases of predictable hepatotoxicity. Important host factors include gender, age, racial background, and genetic polymorphisms, particularly those related to hepatotoxin metabolism. Complex interactions may determine the toxicity of a compound, such as with the increased toxicity of acetaminophen found in chronic alcoholics due to the ability of alcohol to induce expression of enzymes that metabolize acetaminophen. Depending on the toxin, a variety of forms of liver injury can result from hepatotoxin exposure. Frequently the injury is mild and manifests itself only as asymptomatic elevations in serum liver transaminases or cholestasis. These injuries usually resolve when hepatotoxin exposure ceases. Occasionally, acute liver failure develops, with the majority of these cases being caused by the ingestion of toxic amounts of acetaminophen. Continuous hepatotoxin exposure, such as that which occurs with chronic alcohol abuse, can result in chronic liver disease and cirrhosis. Although the likelihood of developing cirrhosis from alcohol abuse clearly increases with the amount and duration of ingestion, host factors presumably explain the widely variable development of chronic liver disease among individuals.

DIRECT MECHANISMS OF HEPATOTOXIC INJURY

Hepatotoxins are taken up by hepatocytes and can injure the cell by damaging any of a number of cellular constituents. As discussed previously, typically the parent compound is harmless until it is metabolized into a toxic intermediate. For example, acetaminophen is metabolized by the cytochrome P450 enzyme isoform 2E1 to the toxin *N*-acetyl-*p*-benzoquinone amine. This potent oxidant is normally neutralized by glucuronide and sulfate conjugation. With large doses of acetaminophen, these protective factors become depleted and oxidative damage ensues. The creation of such an oxidative stress is the final common pathway of cellular injury for a number of hepatotoxins in addition to acetaminophen. Metabolized hepatotoxins may act as oxidants themselves or cause the generation of reactive oxygen intermediates, such as superoxide and hydroxyl free radicals. These highly reactive compounds oxidize and alter cellular lipids, proteins, or DNA, leading to cell injury and,\,death. The oxidative destruction of these cellular components has classically been thought to result in a form of cell death, termed necrosis. In necrotic death, the cell swells, the membranes rupture, and the cell disintegrates. However, oxygen radicals also induce hepatocyte death by apoptosis—an active process in which gene expression triggers a cell death marked by cell shrinkage and chromatin compaction. In both apoptosis and necrosis, death is regulated by cellular signaling pathways that are activated by toxin-induced changes in the redox homeostasis of the cell. Particularly important is activation of the mitogen-activated protein kinase pathway and the resultant activation of transcription factors, such as activating protein-1 and nuclear factor κB (NF-κB). These transcription factors presumably regulate the expression of genes that in turn promote or block cell death, although the identity of these genes remains unknown.

ROLE OF KUPFFER CELLS AND CYTOKINES IN TOXIN-INDUCED LIVER INJURY

One of the unique aspects of the cellular environment of the liver is that it contains 80−90% of the fixed macrophages in the body. During toxin-induced liver injury, additional macrophages are recruited from the circulation into the liver and these cells, along with the resident macrophages or Kupffer cells, undergo a morphological and physiological change, termed activation. Activated cells produce a number of products, including reactive oxygen intermediates, proteolytic enzymes, nitric oxide, eicosanoids, and cytokines (see Fig. 1). Cytokines are a large family of small secreted proteins that exert biological effects on other cells. Activated macrophages promote liver damage from hepatotoxins because toxin-induced liver injury is markedly decreased when liver macrophages are depleted or functionally inactivated. The actions of gut-derived lipopolysaccharide (LPS) are critical in this process because LPS neutralization also prevents much of the liver injury induced by toxins. LPS presumably promotes injury through its ability to activate macrophages. Although all of the previously mentioned products of activated macrophages may promote hepatocyte injury, experimental studies suggest that toxin-induced liver injury depends in large part on the effects of one cytokine, tumor necrosis factor α (TNFα). In animal models of liver injury, blocking the activity of TNFα dramatically reduces the amount of liver injury from a number of toxins, including alcohol and acetaminophen. Taken together, these findings suggest that during hepatotoxic injury LPS causes macrophage activation, leading to the production of TNFα, which then causes liver cell injury. Surprisingly, hepatotoxic liver injury in large part results therefore not from the direct effects of the toxin,

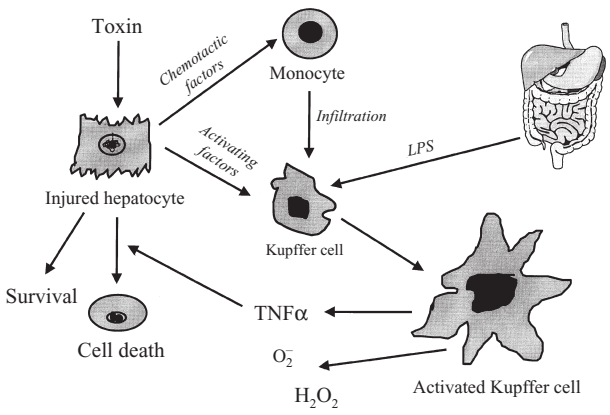


FIGURE 1 Critical events during hepatotoxic liver injury. A toxin is taken up by the hepatocyte, resulting in direct cellular injury. Injured cells release products (chemotactic factors) that attract blood monocytes to infiltrate the injured portion of the liver. These cells, along with preexistent Kupffer cells, respond to factors from injured hepatocytes and to gut-derived lipopolysaccharide (LPS) by undergoing a process termed activation. These activated macrophages release reactive oxygen intermediates, such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2), and a number of cytokines, including tumor necrosis factor α (TNFα). Depending on the degree of injury, hepatotoxin-injured hepatocytes may survive or undergo cell death that is mediated in large part by their sensitization to death from macrophage-produced TNFα.

but from the actions of TNFα produced by macrophages. The mechanism by which TNFα causes hepatocyte injury remains unclear. TNFα is normally toxic to tumor cells but not to nontransformed cells, including hepatocytes. Inhibition of hepatocyte RNA or protein synthesis sensitizes hepatocytes to injury and death from TNFα. Therefore, hepatocyte resistance to the toxic effects of TNFα depends on the induction by TNFα of a protective gene(s). Since hepatotoxins almost invariably interfere with macromolecular synthesis, they presumably block expression of the protective gene, thereby sensitizing hepatocytes to death from TNFα. Although the identity of this protective gene in hepatocytes is currently unknown, it has been established that a critical factor in hepatocyte resistance to TNFα toxicity is the activation of the transcription factor NF-κB.

THERAPIES FOR HEPATOTOXIC INJURY

For many clinical episodes of hepatotoxicity, particularly those related to medicinal drugs, halting toxin exposure is sufficient treatment by itself. For more severe cases of hepatotoxicity that result in acute liver failure, treatment initially involves only supportive care and then liver transplantation if hepatic function fails to improve. Devices that are mechanical, or more recently devices that contain living hepatocytes, have been used experimentally as a temporary liver substitute until the patient's liver recovers. With increasing knowledge of the mechanisms by which hepatotoxins damage the liver, novel therapies are being developed to treat these forms of liver injury. These measures include the administration of antioxidant compounds, especially compounds targeted specifically for the liver or at particular organelles within the liver cell that are vulnerable to toxic injury, such as the mitochondria. Interventions to neutralize TNFα and its toxicity may also become effective therapies against toxin-induced liver injury. Finally, an identification of the common signaling pathways that ultimately lead to the death of the hepatocyte may eventually result in the development of even more specific therapies to prevent hepatotoxic injury.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Cytochrome P450 • Hepatocytes • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Jones, B. E., and Czaja, M. J. (1998). Mechanisms of hepatic toxicity. III. Intracellular signaling in response to toxic liver injury. *Am. J. Physiol.* **275**, G874–G878.

Kaplowitz, N. (2000). Mechanisms of liver cell injury. *J. Hepatol.* **32**(Suppl. 1), 39–47.

Larrey, D. (1997). Hepatotoxicity of herbal remedies. *J. Hepatol.* **26**(Suppl. 1), 47–51.

Nelson, S. D. (1990). Molecular mechanisms of the hepatotoxicity caused by acetaminophen. *Semin. Liver Dis.* **10**, 267–278.

Thurman, R. G. (1998). Alcoholic liver injury involves activation of Kupffer cells by endotoxin. *Am. J. Physiol.* **275**, G605–G611.

Hereditary Fructose Intolerance

DAVID A. RUDNICK
Washington University, St. Louis, Missouri

Fanconi's syndrome A disorder of the proximal kidney tubules characterized by urinary excretion of large amounts of amino acids, glucose, and phosphate despite normal blood levels of these molecules.
gluconeogenesis The biochemical process by which glucose is synthesized from noncarbohydrate substrate.
glycolysis The sequential enzymatic conversion of glucose to lactic acid.

Hereditary fructose intolerance is associated with mutations in the aldolase B gene. Dietary exposure to fructose leads to the clinical syndrome, with disease severity related to degree of nutritional exposure. Traditionally, the diagnosis has depended on fructose tolerance testing or enzymatic assay of biopsy material; however, the recognition of specific aldolase B mutations has prompted recent development of genetic testing strategies. Treatment relies on strict adherence to a fructose-free diet, which leads to complete symptomatic resolution.

INTRODUCTION

Hereditary fructose intolerance (HFI) is a recessively inherited inborn error of metabolism in which ingestion of fructose leads to toxic effects on the liver, intestines, and kidney. Acute ingestion can cause hypoglycemia and gastrointestinal symptoms and chronic ingestion can lead to hepatic or renal injury and growth disturbance. Deficiency of fructose-1,6-bisphosphate aldolase enzyme activity (aldolase B; EC 4.1.2.13), an enzyme of fructose-1-phosphate metabolism, causes HFI. Several mutations in the human aldolase B gene have been associated with HFI. Severity of clinical disease is related to the degree of nutritional exposure and the age of the affected individual. This article examines fructose metabolism and the molecular basis, clinical presentation, pathogenesis, diagnosis, and treatment of HFI.

FRUCTOSE METABOLISM

Fructose exists as a monosaccharide in fruits and honey and is present in the disaccharide sucrose. Sucrose is hydrolyzed to glucose and fructose by the enzyme sucrase, present in the brush border membrane of intestinal epithelial cells. Fructose is transported across the plasma membranes of enterocytes and hepatocytes and metabolized in the liver, kidney, and small intestine. After cellular uptake, fructose is phosphorylated to fructose-1-phosphate, which aldolase B then metabolizes to dihydroxyacetone phosphate and D-glyceraldehyde. D-Glyceraldehyde is phosphorylated to D-glyceraldehyde-3-phosphate and further metabolized through glycolysis or gluconeogenesis (Fig. 1).

ALDOLASES

Three eukaryotic aldolases have been identified, each with distinct tissue-specific and developmental patterns of expression. Aldolase B is found in liver, kidney, and small intestine. Aldolase A exists in most tissues but predominates in muscle. Aldolase C is present in the brain. Aldolases A and C have greater activity against fructose-1,6-diphosphate than does aldolase B; thus, gluconeogenesis can proceed even in the absence of aldolase B (Fig. 1). Aldolase A is the predominant fetal isoform. During embryonic development, aldolase A is repressed and aldolase B is induced in the liver,

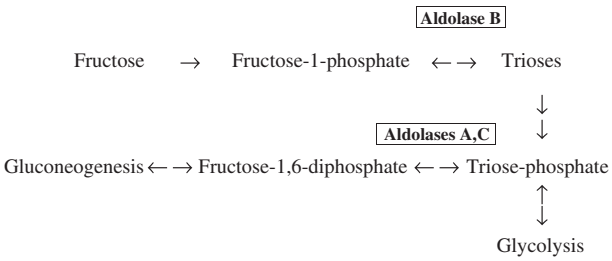


FIGURE 1 Summary of fructose metabolism. Aldolase B catalyzes the reversible cleavage of fructose-1-phosphate to the trioses, D-glyceraldehyde and dihydroxyacetone phosphate. Aldolases A and C have greater activity against fructose-1,6-diphosphate than does aldolase B.

 Copyright 2004, Elsevier (USA). All rights reserved.

kidney, and intestine. After birth, expression of aldolase B is regulated by dietary exposure to fructose, whereas aldolases A and C are constitutively expressed.

HEREDITARY FRUCTOSE INTOLERANCE

In HFI, aldolase B deficiency leads to impaired fructose metabolism. The first clinical description, in 1956, was of a 24-year-old woman with faintness, abdominal pain, and nausea after ingestion of fructose but not glucose. Investigators suspected that symptoms were secondary to fructose-induced hypoglycemia. In 1957, the same syndrome was reported in several members of a family. Aldolase deficiency was shown to be responsible based on enzymatic analyses of liver biopsies. Subsequent studies have shown that the phenotype is particularly severe in individuals unable to avoid fructose-containing sugars, particularly infants at the time of weaning from breast milk.

Molecular Basis and Epidemiology

The gene encoding aldolase B (GenBank Accession No. X01098) maps to chromosome 9q22.3, contains 9 exons, and encodes a 364-amino-acid polypeptide. HFI occurs when mutant alleles of the aldolase B gene whose protein products are deficient in enzyme activity are inherited in an autosomal recessive fashion. The disease is manifest only in homozygotes, with an estimated frequency of 1 in 20,000 individuals. In affected individuals, the enzymatic activities of aldolases A and C are normal.

Genetic Mutations of Aldolase B

The first identified mutation of the aldolase B gene associated with HFI, described in 1988, was a homozygous single base-pair change (G → C) resulting in mutation of Ala-149 to proline (A149P). A149P remains the most common mutation identified to date, accounting for more than half of mutant North American and northern European alleles. Two other mutations (A174D and A174E) have been frequently identified in HFI patients from Italy, Switzerland, and Yugoslavia but not in those from North America or northern Europe. Testing for these three mutations, using polymerase chain reaction (PCR) to amplify aldolase B DNA, can identify more than 95% of patients with HFI. A total of 22 genetic lesions in aldolase B, including several null mutations, have now been associated with HFI disease. Patients with null mutations are healthy as long as they avoid fructose. They have no morbidity associated with fasting, proving that aldolase B, though required for the metabolism of fructose and fructose-containing sugars, is not required for gluconeogenesis or glycolysis in the absence of fructose.

Structure–Function Relationships

Most of the characterized mutations in aldolase B associated with HFI abolish enzyme activity. The molecular effects of the two most common mutations, A149P and A174D, have been examined by analyses of the recombinant proteins *in vitro*. Both enzymatic activity and structural stability are profoundly disturbed in the mutant enzymes. Correlations between specific genetic mutations and the severity of clinical phenotype have not been recognized. Rather, disease severity appears to be related primarily to dietary exposure to fructose.

Clinical Presentation and Pathogenesis

Signs and Symptoms

Individuals with HFI are asymptomatic and healthy until they ingest foods containing fructose (or complex fructose-containing molecules such as sucrose and sorbitol). The signs and symptoms of acute exposure to fructose include nausea, vomiting, tremor, dizziness, and lethargy. Large doses may induce seizures or coma. Chronic exposure can be associated with failure to thrive, jaundice, cirrhosis, vomiting, diarrhea, and feeding difficulties. Symptoms are the result of fructose-1-phosphate accumulation in liver, kidney, and small intestine. Accumulation in the small intestine leads to abdominal distension, pain, colic, vomiting, and diarrhea. In the liver and kidney, buildup of fructose-1-phosphate sequesters phosphate, leading to hypophosphatemia, hypoglycemia, hyperlactic acidemia, and hyperuricemia. Hepatotoxicity may manifest as hepatosplenomegaly, edema, ascites, cholestasis, coagulopathy, and liver failure. Renal toxicity may present as a Fanconi's syndrome with generalized amino aciduria, phosphaturia, and bicarbonate wasting. With chronic exposure, the injury may be irreversible. In some children, the only manifestation of HFI may be nutritional deficiency states, such as rickets or failure to thrive, and hepatomegaly.

The severity of clinical presentation depends on age at the time of exposure and the amount of the exposure. Younger patients and those exposed to greater amounts of fructose exhibit more severe reactions. Infants with HFI are asymptomatic while breast-feeding or on sucrose-free formulas. Symptoms first appear at the time of introduction of fructose- or sucrose-containing foods in the diet, e.g., fruits, juices, vegetables, or

sucrose-containing formulas. Older children, adolescents, or adults with HFI, who are not exposed to fructose or diagnosed with HFI in infancy, learn by experience to avoid foods containing even small amounts of fructose. Such individuals may come to medical attention because of unusual feeding behaviors involving self-restriction from sweet-tasting foods. They also have low incidence of dental caries related to low sucrose intake.

Laboratory Abnormalities

Laboratory abnormalities are nonspecific and related to target end-organ dysfunction. Common findings include hypertransaminasemia, hyperbilirubinemia, prolonged prothrombin time, hypoproteinemia, hypokalemia, and hypophosphatemia. Renal Fanconi's syndrome can occur with increased urinary reducing substances (particularly fructose), protein, amino acids, and organic acids. Hypoglycemia, anemia, and thrombocytopenia also occur.

Pathogenesis and Biochemical Abnormalities

The metabolic derangements that occur after acute fructose ingestion in HFI have been examined by analyses *in vitro* and in animals, healthy controls, and affected patients. Aldolase B deficiency results in hepatic, renal, and intestinal intracellular accumulation of fructose-1-phosphate. After exposure to fructose, sequestration of intracellular phosphate as fructose-1-phosphate leads to the reduction of free intracellular phosphate pools and ATP levels. The accumulation of fructose-1-phosphate impairs gluconeogenesis and glycogenolysis, resulting in hypoglycemia. Excess fructose-1-phosphate also inhibits the activity of enzymes involved in glycoprotein synthesis and processing, leading to derangement in normal protein glycosylation patterns, i.e., a secondary carbohydrate-deficient glycoprotein syndrome. Fructose exposure also causes hyperuricemia and lactic acidosis. Excessive uric acid is generated from AMP as a result of activation of adenosine deaminase and associated purine degradation in response to hypophosphatemia and decreased ATP levels. Hypermagnesemia occurs secondary to depletion of intracellular ATP, leading to the release of magnesium ions from intracellular stores. Fructosemia and fructosuria occur as a result of fructose-1-phosphate-mediated inhibition of fructokinase, preventing further accumulation of intracellular fructose. Depletion of intracellular inorganic phosphate and ATP ultimately disrupts protein synthesis and other functions leading to cellular and organ-system toxicity. Additional biochemical disturbances occur as a result of such toxicity including hypokalemia, coagulopathy, and elevated serum amino acids.

Diagnosis

The clinical presentation of HFI is variable and nonspecific. Young infants often present with difficulty feeding, poor growth, vomiting, and hepatomegaly. Older infants and children may present with similar symptoms or for evaluation of abnormal behaviors. The differential diagnosis includes intrauterine infection, sepsis, hepatitis, hemolytic–uremic syndrome, and metabolic liver diseases including galactosemia, tyrosinemia, Wilson's disease, and others. The key to diagnosis is a careful nutritional history correlating fructose ingestion with onset of symptoms. Once HFI is suspected, all sources of fructose, sucrose, and sorbitol must be removed from the diet. Confirmation of the diagnosis is important because strict compliance with dietary exclusion of fructose prevents tissue injury and growth retardation.

Traditional Diagnostic Tests

Historically, the specific diagnosis of HFI could be made only by enzymatic analysis for aldolase B activity in hepatic and intestinal biopsies or by a controlled intravenous fructose tolerance test and simultaneous analysis for characteristic clinical and biochemical changes. Fructose-1-phosphate aldolase activity from hepatic or small intestinal biopsy material is decreased in HFI. Other tissues, such as serum, blood cells, cultured skin fibroblasts, and placenta, are not useful for diagnosis because aldolase A is the predominant isoform present and enzyme activity is not decreased. Biopsy of the liver offers advantages compared to the intestine in that histological assessment of hepatic tissue damage can also be performed. In HFI, the liver may exhibit steatosis, hepatocellular necrosis, and intralobular or periportal fibrosis or cirrhosis. Many investigators have increasingly discouraged the fructose tolerance test because of severe toxic effects observed during such studies and because of the availability of definitive genetic testing.

Genetic Testing

The development of PCR-based methodologies for the detection of mutations in the aldolase B gene offers noninvasive screening to confirm clinical diagnoses of HFI. At present, more than 95% of patients with HFI can be diagnosed by PCR-based DNA amplification performed on material obtained from blood using a limited number of allele-specific oligonucleotide pairs.

Treatment

Immediate and complete removal of *all* sources of fructose from the diet is the only effective strategy for

management of HFI and prevention of the acute and long-term consequences associated with fructose exposure. Particular care should be taken to avoid foods and medicines with fructose-containing food additives. For example, sucrose and sorbitol are often used as supplements for delivery of medications in a palatable form. Monitoring growth and nutritional status to ensure that strict adherence to the restricted diet does not lead to specific deficiencies is important.

Outcome

The natural history of HFI depends on the age at which the diagnosis is made and the degree of dietary exposure to fructose. Infants with HFI, who are unable to control their own diet, are at the greatest risk of morbidity and mortality associated with fructose exposure. Treatment of acute metabolic crises involves supportive care to correct associated metabolic derangements and complications. This may include intravenous fluids and glucose to restore normal fluid and electrolyte status and correct hypoglycemia, blood products to treat coagulopathy or bleeding, and vitamin and nutritional support to restore growth. Complete and long-term abstinence from fructose exposure often leads to resolution of hepatomegaly, reversal of organ dysfunction, normalization of intellectual function, and catch-up growth. Persistent hepatomegaly or growth failure may indicate ongoing unrecognized ingestion of fructose. Adults with HFI who adhere to an appropriate diet are usually asymptomatic and healthy. Morbidity and mortality have been reported in people with undiagnosed or unrecognized HFI who are exposed to fructose-containing parenteral nutrition. Such feeding regimens, once thought to be of benefit for patients with diabetes mellitus, have been greatly restricted because of the description of fatal and near-fatal cases resulting from their use in subjects with undiagnosed HFI.

OTHER DISORDERS OF FRUCTOSE METABOLISM

HFI is one of three known disorders of fructose metabolism. The others are essential fructosuria and fructose-1,6-diphosphate deficiency. Essential fructosuria, also known as hepatic fructokinase deficiency, is a benign condition characterized by hyperfructosemia and fructosuria in association with dietary fructose intake. It is typically diagnosed after detection of urinary reducing substances in otherwise healthy individuals. Fructose-1,6-diphosphatase deficiency is a severe inborn error of metabolism associated with deficiency of fructose-1,6-diphosphatase, an enzyme required for gluconeogenesis. It is characterized by fructose- and fasting-induced hypoglycemia, lactic acidosis, and hepatomegaly. Symptoms usually occur within the first 4 days of life and can be life-threatening. A milder variant has been described in which affected patients are resistant to the development of lactic acidosis, presumably because of retention of partial enzyme activity.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Food Intolerance • Galactosemia • Glycogen Storage Disease • Tyrosinemia

Further Reading

Ali, M., Rellos, P., and Cox, T. M. (1998). Hereditary fructose intolerance. *J. Med. Genet.* **35**, 353–365.

Gishan, F. K. (2003). Inborn errors of metabolism that lead to permanent hepatic injury. *In* "Hepatology: A Textbook of Liver Disease" (D. Zakim and T. Boyer, eds.), 4th Ed., pp. 1399–1404. W. B. Saunders, Philadelphia, PA.

Gishan, F. K., and Ballew, M. P. (2001). Inborn errors of carbohydrate metabolism: Disorders of fructose metabolism. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Ballistreri, eds.), 2nd Ed., pp. 573–577. Lippincott Williams & Wilkins, Philadelphia, PA.

Stanley, C. A. (2000). Genetic and metabolic hepatic disorders. 1. Disorders of carbohydrate metabolism. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (W. A. Walker, P. R. Durie, J. R. Hamilton, J. A. Walker-Smith, and J. B. Watkins, eds.), 3rd Ed., p. 1071. B. C. Decker, Lewiston, NY.

Teckman, J. H., and Perlmutter, D. H. (1999). Metabolic disorders of the liver. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (R. Wyllie, and J. S. Hyams, eds.), 2nd Ed., pp. 590–591. W. B. Saunders, Philadelphia, PA.

Hereditary Hemochromatosis

DAVID J. BRANDHAGEN
Mayo Clinic and Foundation

C282Y The main hemochromatosis gene mutation, present in approximately 85% of Caucasians with iron overload in the United States; tyrosine for cysteine substitution at amino acid position 282.

compound heterozygote One copy of two different mutations. In patients with iron overload, describes the presence of one copy of the C282Y mutation and one copy of the H63D mutation. The genotype is present in 1–2% of the Caucasian population; most do not develop iron overload but a few do (usually mild).

divalent metal transporter 1 Protein located in the apical villous surface of enterocytes; imports non-transferrin-bound iron from the small bowel lumen to the enterocyte cytoplasm.

duodenal cytochrome b Ferric reductase located on the apical villous surface of enterocytes; converts ferric (Fe^{3+}) iron in the small bowel lumen to ferrous (Fe^{2+}) iron, thereby facilitating its uptake into the enterocyte.

ferritin Molecule involved in cytosolic iron storage.

ferroportin 1 Protein located on the basolateral surface of villous enterocytes; functions as an exporter of iron from the enterocyte to the plasma; a mutation in ferroportin 1 is associated with an autosomal dominant form of iron overload in hereditary hemochromatosis type 4.

H63D Hemochromatosis gene (*HFE*) mutation that involves substituting an aspartate for a histidine at amino acid position 63. Approximately 25% of the general population is heterozygous for H63D and one copy of this mutation is not associated with iron overload. Homozygosity for H63D is a rare cause of iron overload (usually mild).

hepcidin Small protein produced by hepatocytes; may be the soluble iron stores regulator that controls iron absorption in intestinal crypt cells.

hephaestin Protein located on the basolateral surface of villous enterocytes; functions as a ferroxidase, converting ferrous (Fe^{2+}) to ferric (Fe^{3+}) iron, facilitating its transfer across the basolateral membrane of the enterocyte to the plasma.

***HFE* gene** Hemochromatosis gene located on chromosome 6p. Two main mutations (C282Y and H63D) were initially described.

iron regulatory elements Stem–loop structures present on the 3′ or 5′ untranslated regions of mRNA of several molecules involved in iron regulation.

iron regulatory proteins Molecules that increase in response to iron deficiency and bind to iron regulatory elements; such binding at the 3′ untranslated region of mRNA of the transferrin receptor and the divalent metal transporter 1 will stabilize the transcript and increase expression of these molecules. Conversely, binding on the 5′ untranslated region of ferritin will block translation and decrease ferritin levels.

transferrin Circulating iron carrier; each transferrin molecule binds two atoms of diferric iron.

transferrin receptor Protein involved in endocytosis of diferric transferrin; binds the *HFE* molecule.

transferrin receptor 2 Recently discovered protein expressed predominantly on hepatocytes; also involved in endocytosis of diferric transferrin., but it is uncertain if this receptor binds *HFE*. An autosomal recessive form of hemochromatosis (hereditary hemochromatosis type 3) associated with several mutations in the transferrin receptor 2 gene on chromosome 7 has recently been described.

Hereditary hemochromatosis (HH) is an autosomal recessive disorder characterized by increased intestinal absorption of iron; excess iron is deposited in the liver, pancreas, and other organs. It is the most common genetically inherited disorder in the United States Caucasian population, with a prevalence of 0.5% and a carrier frequency of 10%. Clinical manifestations may include fatigue, diabetes, arthritis, impotence, increased skin pigmentation, cirrhosis, hepatocellular carcinoma, heart failure, and cardiac arrhythmias. Diagnosis is based on a combination of clinical, laboratory, and pathologic criteria, including an elevated transferrin saturation assay. Treatment by phlebotomy, if initiated before the development of cirrhosis or diabetes, is associated with a normal life expectancy. Approximately 85% of patients with HH are homozygous for the *C282Y HFE* gene mutation. The *HFE* gene test is useful in confirming a diagnosis of HH, screening adult blood relatives of probands, and resolving ambiguous cases of iron overload.

CLINICAL FEATURES

Iron accumulation in hereditary hemochromatosis is a slow, insidious process; only a few milligrams of excess iron may be absorbed from the duodenum each day. Clinical manifestations frequently do not occur until

 Copyright 2004, Elsevier (USA). All rights reserved.

at least the fifth decade, when 15–40 g of iron have accumulated (normal body iron stores are approximately 3–4 g). Disease expression may occur earlier in some persons and may not occur at all in others. Clinical expression is influenced by age, sex, iron content of the diet, blood loss as occurs in menstruation and pregnancy, and other unknown factors. Although the gene frequency is similar in males and females, the disease is expressed less frequently in women than in men. Factors such as alcohol and hepatitis C may also influence disease expression. Several recent studies have concluded that clinically significant disease expression may not occur in a majority of individuals who are homozygous for the C282Y mutation, even if undiagnosed and untreated.

The classic description of HH is cutaneous hyperpigmentation, diabetes mellitus, and cirrhosis of the liver (bronze diabetes). Other clinical manifestations include fatigue, abdominal pain, hepatomegaly, abnormal liver tests, hepatocellular carcinoma, cardiomyopathy, cardiac conduction disorders, hypothyroidism, hypogonadism, impotence, and arthropathy.

In the past, HH was usually diagnosed at an advanced stage; currently, most patients with newly diagnosed HH are asymptomatic. This shift toward earlier diagnosis may be due in part to increased physician awareness. Of HH patients who are symptomatic, fatigue, arthralgias, and impotence are most common. Most if not all clinical manifestations are preventable if HH is diagnosed early and treated appropriately. Some of the disease manifestations, such as skin bronzing, cardiomyopathy, cardiac conduction disorders, hepatomegaly and abnormal liver tests, frequently are reversible once excess iron stores are removed. Most of the other clinical manifestations are not reversible, however.

DIAGNOSIS

Serum Iron Studies

A diagnosis of HH is based on a combination of clinical, laboratory, and pathologic criteria, including an elevated serum transferrin saturation [100 (serum iron concentration ÷ total iron binding capacity)] result and an elevated serum ferritin concentration. There is diurnal variation in serum iron values, and measurements may be affected by the ingestion of food; therefore, an elevated transferrin saturation assay should be repeated as a fasting early morning determination. A transferrin saturation above 50% is the earliest phenotypic abnormality in HH.

Although serum transferrin saturation is the best initial screening test, the results may be normal early in the course of HH. In addition, the serum ferritin concentration and transferrin saturation may be elevated in up to 50% of patients with viral hepatitis, nonalcoholic steatohepatitis, and alcoholic liver disease, and in end-stage liver disease of various etiologies. Serum ferritin usually provides a reasonable estimate of total body iron stores, but it is also an acute-phase reactant and is elevated in a variety of infectious and inflammatory conditions in the absence of iron overload. For this reason, it should not be used as the initial screening test to detect HH.

HH should be distinguished from iron overload that is due to other causes. Secondary iron overload should be suspected in patients with chronic anemias with ineffective erythropoiesis, multiple blood transfusions, prolonged iron supplementation, and chronic liver disease. Disorders associated with iron overload are summarized in Table I.

TABLE I Disorders Associated with Iron Overload

- **Hereditary hemochromatosis**
  - HFE related
    - C282Y homozygous
    - C282Y and H63D heterozygous (compound heterozygote)
    - Other
  - Non-HFE related
    - HH type 2 (juvenile hemochromatosis)
    - HH type 3 (TfR2 mutation)
    - HH type 4 (ferroportin mutation)
- **Secondary Iron Overload**
  - **Chronic anemias**
    - Thalassemia major
    - Sideroblastic anemia
    - Congenital dyserythropoietic anemia
    - Congenital atransferrinemia
  - **Exogenous iron overload**
    - Chronic iron supplementation (in the absence of blood loss)
      - Transfusion
      - Iron dextran
      - Oral supplements (rare)
  - **Chronic liver disease**
    - Cirrhosis
    - Viral hepatitis
    - Alcoholic liver disease
    - Nonalcoholic steatohepatitis
    - Porphyria cutanea tarda
    - Portacaval shunt
  - **Miscellaneous**
    - Iron overload in sub-Saharan Africa
    - African-American iron overload
    - Neonatal iron overload
    - Aceruloplasminemia
    - Congenital atransferrinemia

Hepatic Iron

Prior to the availability of the *HFE* gene test, a liver biopsy was often necessary to confirm a diagnosis of HH. Hepatic iron may be assessed with an iron stain. In HH, iron initially accumulates in periportal hepatocytes but is eventually distributed throughout the liver. In secondary iron overload, iron is often present predominately in Kupffer cells, which may help distinguish it from HH. Histologic distinction between HH and secondary iron overload is often not possible once severe iron overload has developed. Hepatic histological features of HH and secondary iron overload are demonstrated in Fig. 1.

In HH, there is a progressive, age-related increase in hepatic iron stores. This knowledge led to the development of the hepatic iron index (HII), which is the hepatic iron concentration (in micromoles/gram dry weight liver) divided by the patient's age in years. The original intent of the HII was to distinguish HH homozygotes from heterozygotes, and those with alcoholic liver disease. In the initial study, all HH homozygotes had an HII >1.9, whereas all of the HH heterozygotes or patients with alcoholic liver disease had an HII <1.9. An HII >1.9 is not diagnostic of HH because patients with severe iron overload of any cause may have an HII >1.9. In addition, because HH is increasingly diagnosed at an earlier stage, not all

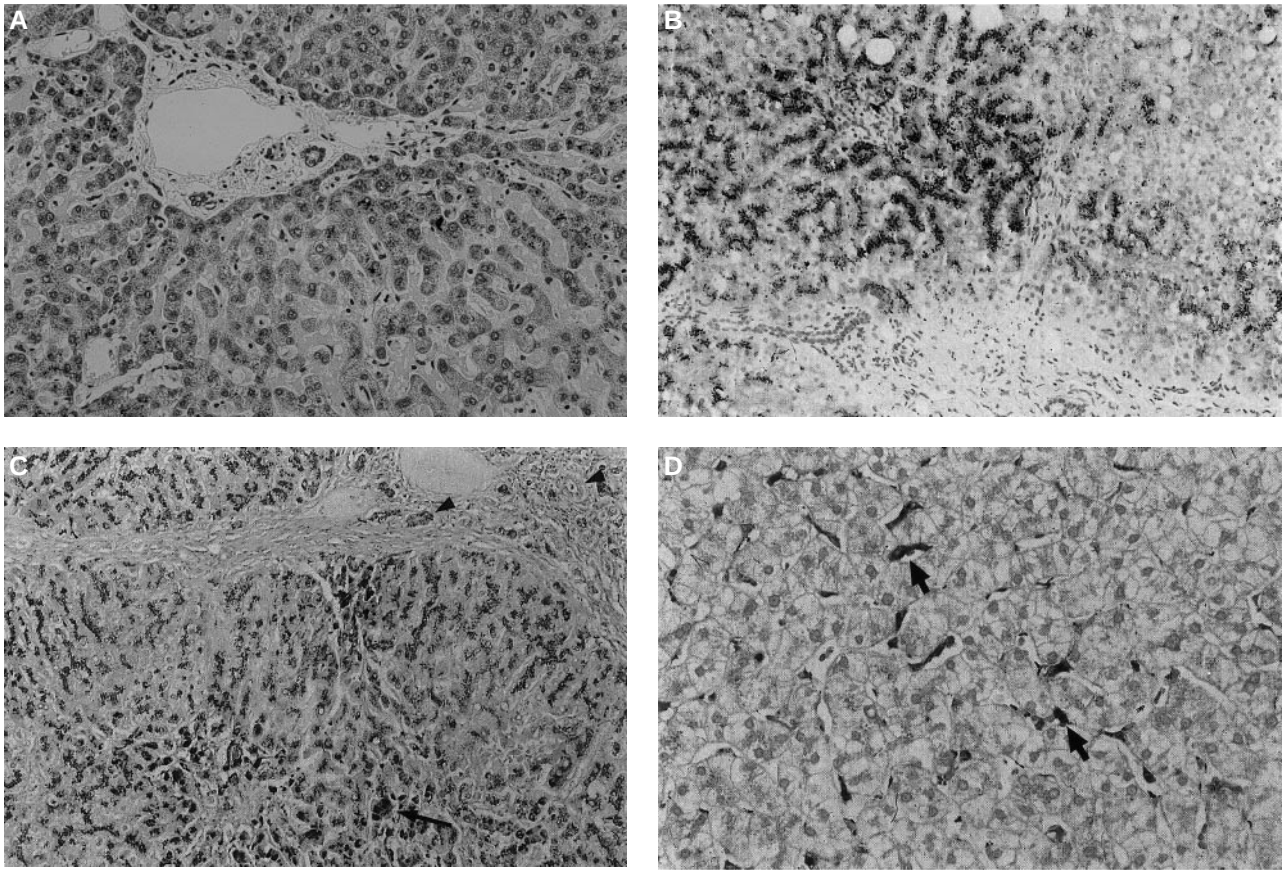


FIGURE 1 Histological hepatic iron deposition. (A) Mild (grade 1 of 4) iron deposition in hepatocytes. Reproduced with permission from *Liver Transplantation* 7(8), 663–672 (2001). (B) Moderate hemosiderin deposition in precirrhotic homozygous hemochromatosis. Zone 1 hepatocytes are predominantly involved, biliary hemosiderin is not evident, and fibrosis has not yet occurred, all indicating relatively early precirrhotic disease (hepatic iron concentration, 10,307 μg Fe/g dry weight; iron index, 3.2). (Perls' Prussian blue staining; original magnification, ×133.) (C) Marked hemosiderosis and cirrhosis in homozygous hemochromatosis. Although most iron is in the hepatocytes, some Kupffer cell (arrow) and biliary iron (arrowheads) is also present (Perls' Prussian blue staining; original magnification, ×133.) (D) Kupffer cell hemosiderosis. The presence of hemosiderin in Kupffer cells alone (arrows) is typical of mild transfusion hemosiderosis, is nonspecific, and should not prompt further consideration of hemochromatosis (Perls' Prussian blue staining; original modification, ×240). Reproduced with permission from *Hemochromatosis: Genetics, Pathophysiology, Diagnosis, and Treatment* (J. C. Barton and C. Q. Edwards, eds.), pp. 192–194 (2000).

HH homozygotes will have an HII >1.9. Since HFE gene testing has become the standard test for confirming the diagnosis of HH, the HII has assumed a less important role.

THE *HFE* GENE

The gene associated with HH, *HFE*, is located on the short arm of chromosome 6. The *HFE* gene encodes a 343-amino-acid protein that resembles a human leukocyte antigen (HLA) class I molecule. Two point mutations, designated C282Y and H63D, were initially described. Other mutations have since been discovered, including S65C, G93R, I105T, Q127H, and R330M, in addition to a couple of frameshift mutations, nonsense mutations, and one splice site mutation. Of these other mutations, S65C is most common, occurring in 2–3% of patients with iron overload. All of these mutations are rare and are not likely to be of major clinical importance. In nearly all cases, patients with these other mutations who develop iron overload are C282Y compound heterozygotes.

The greatest risk for iron overload exists in those who are homozygous for the C282Y mutation. Several studies from the United States and Europe have found that 60–93% of patients with iron overload are homozygous for C282Y. The wide range in the prevalence of C282Y homozygotes may be due in part to different diagnostic criteria for HH as well as to geographic differences in the prevalence of *HFE* mutations. Iron overload also occurs in a minority of those with other *HFE* mutations, but it is usually of lesser severity. Approximately 1–2% of Caucasians are compound heterozygotes (one copy of C282Y and one copy of H63D) or H63D homozygotes, but only a few percent will develop problems with iron overload. In addition, approximately 25% of the United States Caucasian population is heterozygous for H63D. A single copy of H63D is not associated with an increased risk for developing iron overload. Nearly all studies have demonstrated that severe iron overload can occur in the absence of *HFE* gene mutations. A negative *HFE* gene test, therefore, does not exclude iron overload. The role of *HFE* mutation analysis in the diagnosis of iron overload disorders is summarized in Fig. 2.

The gene test is most useful in screening adult blood relatives of an identified proband. Screening blood relatives is crucial, because 25% of siblings and 5% of children of a proband will have HH. *HFE* gene testing should replace more expensive HLA typing previously used to screen siblings. In addition, *HFE* gene testing often is useful in helping to resolve ambiguous cases,

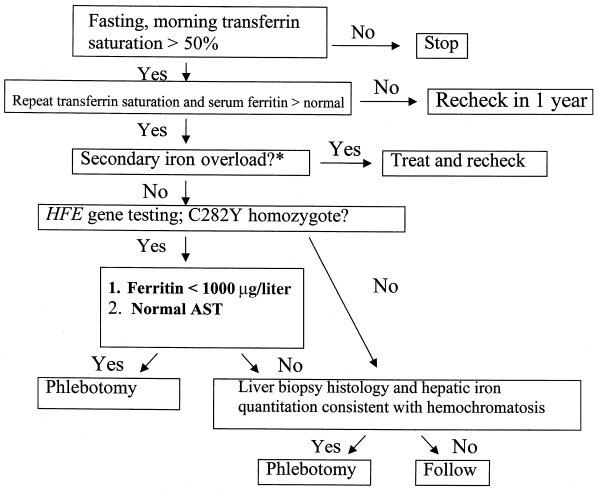


FIGURE 2 Diagnostic algorithm for hereditary hemochromatosis. AST, Aspartate transaminase. Modified with permission from Brandhagen *et al., Mayo Clinic Proceedings* 74, 917–921 (1999).

such as iron overload associated with hepatitis C, alcoholic liver disease, or other causes of end-stage liver disease. Prior to obtaining the *HFE* gene test, an individual should be counseled about the risks, benefits, and alternatives of genetic testing by a qualified professional. There is concern about the possibility of insurance, employment, or other discrimination based on *HFE* test results. For this reason, *HFE* gene testing usually is not recommended for anyone younger than 18 years old.

PATHOPHYSIOLOGY

The discovery of the *HFE* gene has furthered our understanding of the pathophysiology of iron overload. The HFE protein is expressed throughout the intestine as well as in many other tissues. It is expressed on epithelial cells throughout the gastrointestinal tract except in the small intestine, where it is localized in crypt cells. HFE protein is not expressed on hepatocytes and is, therefore, probably not directly involved in hepatic iron loading. The normal (wild-type) HFE protein binds to β2-microglobulin and is expressed on the cell surface. The HFE β2-microglobulin complex binds the transferrin receptor (TfR) and may lower its affinity for diferric transferrin. The C282Y-mutated protein does not bind β2-microglobulin and is not expressed on the cell surface. Failure of the mutated HFE protein to bind to β2-microglobulin may be important in the development of iron overload because β2-microglobulin knockout mice develop hepatic iron overload in a pattern

similar to that observed in humans with HH. The H63D protein associates with β2-microglobulin but does not lower the affinity of the TfR for diferric transferrin.

Several other iron regulatory proteins have recently been discovered. Divalent metal transporter 1 (DMT1; also known as DCT1 and Nramp2), a protein located on the apical surface of villus enterocytes, functions to import non-transferrin-bound iron from the small bowel lumen to the enterocyte cytoplasm. This is facilitated by duodenal cytochrome b (Dcytb), a ferric reductase that converts iron in the small bowel lumen from the ferric to the ferrous form, thus facilitating its uptake into the enterocyte. Once in the enterocyte, iron can be stored as ferritin or transferred to the plasma to bind to transferrin. Ferroportin 1 (FP1; also known as Ireg1), a recently discovered protein located in the basolateral surface of villus enterocytes, functions as the basolateral transporter of iron from the enterocyte to the plasma. This is facilitated by a protein closely related to ceruloplasmin, hephaestin, which functions as a ferroxidase, converting ferrous iron to ferric iron, thereby facilitating its transfer across the basolateral membrane of the enterocyte. An autosomal dominant form of HH (type 4) has recently been described in patients with a mutation in the ferroportin gene.

The duodenal crypt cell is likely involved in sensing the body's iron requirements. Dietary iron intake, iron from storage sites such as the liver, and iron needed for erythropoiesis may all influence intestinal iron absorption. Hepcidin is a recently discovered small protein produced by hepatocytes. Hepcidin mRNA levels are increased in mouse models of dietary iron loading and genetic iron overload, and hepcidin may be the soluble iron stores regulator that controls iron absorption in intestinal crypt cells.

Many of the iron regulatory proteins contain iron regulatory elements (IREs) on the 3′ or 5′ untranslated region (UTR) of their mRNA. IREs bind iron regulatory proteins (IRPs) produced in response to iron deficiency sensed at the level of the crypt enterocyte. The binding of IRPs to an IRE at the 5′ UTR of mRNA from molecules such as ferritin will block translation and decrease ferritin levels. Conversely, the binding of IRPs to an IRE at the 3′ UTR of mRNA from the TfR and DMT1 will stabilize the transcript and increase levels of TfR and DMT1. Although much has been learned about molecular aspects of iron absorption in the past few years, a great deal is still unknown, including the specifics of how the mutated HFE protein leads to the development of iron overload. Intestinal iron overload is summarized in Fig. 3.

TREATMENT

Therapeutic Phlebotomy

The treatment of patients with HH is usually reserved for those with iron overload as evidenced by

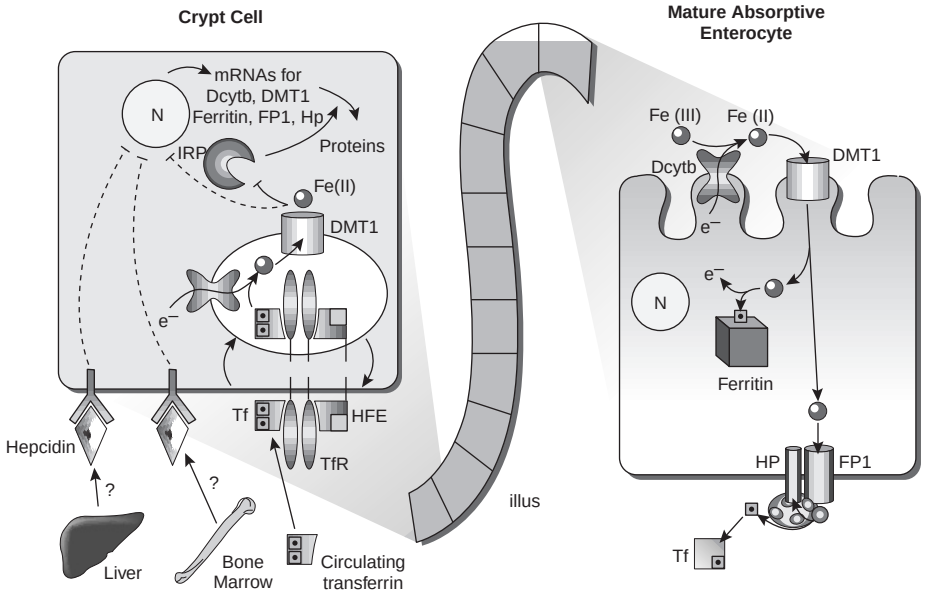


FIGURE 3 Intestinal iron absorption. Dcytb, Duodenal cytochrome b; DMT1, divalent metal transporter 1; FP1, ferroportin 1; Hp, hephaestin; IRP, iron regulatory protein; Tf, transferrin; TfR, transferrin receptor. Reproduced with permission from Philpott (2002), *Herpatology* **35**, 993–1001.

an elevated serum ferritin level. Therapeutic phlebotomy is the preferred treatment because it is simple, inexpensive, and effective. During the initial phase of therapeutic phlebotomy, 500 ml of blood is removed on a weekly basis. A hemoglobin should be checked prior to each phlebotomy. Weekly phlebotomy should continue as long as the hemoglobin is above a preselected value (usually 12–13 g/dl). If the hemoglobin is below the selected cutoff value, phlebotomy should not be performed. Once the hemoglobin remains below the cutoff for 3 consecutive weeks without phlebotomy, the serum ferritin and transferrin saturation should be rechecked. Iron depletion is confirmed if the ferritin level is less than 50 μg/liter with a transferrin saturation in the low normal range. Once iron depletion is accomplished, most patients require four to eight "maintenance" phlebotomies per year to keep the ferritin level lower than 50 μg/liter.

Iron Chelators

Iron chelators such as desferrioxamine are rarely utilized to treat iron overload in HH. Iron chelators are expensive and must be given by subcutaneous infusion. They also are much less effective at removing excess iron compared to phlebotomy. At present, an effective oral iron chelator is not routinely available in the United States.

Dietary

Patients with HH should refrain from using iron supplements, including multivitamins with iron as well as high-dose vitamin C supplements. A "low-iron diet" is not necessary, but red meat should be consumed in moderation. Patients with HH should avoid consuming raw shellfish due to an increased risk of infections with *Vibrio vulnificus*. Persons with HH should also avoid or minimize alcohol use because iron and alcohol are synergistic hepatotoxins.

Liver Transplantation

Liver transplantation may be an option for HH patients with cirrhosis and complications of end-stage liver disease. Despite being common, HH is an uncommon indication for orthotopic liver transplantation (OLT), accounting for less than 1% of all liver transplants performed in the United States. Survival in HH patients undergoing liver transplantation is poor. The 1-year patient survival rate after OLT is approximately 50–60% for patients with HH compared to over 90% for most other indications. Deaths in HH OLT recipients are usually due to cardiac or infectious complications. There are preliminary reports of improved survival in patients with HH who are depleted of excess iron stores prior to OLT.

PROGNOSIS

Patients with HH who are diagnosed and treated before the development of diabetes and cirrhosis have normal age-and sex-adjusted survival rates. Once cirrhosis or diabetes develops, the rate of survival is markedly reduced. Deaths in cirrhotic patients with HH are often due to hepatocellular carcinoma. Patients with cirrhosis due to HH usually do not develop decompensated liver disease. Up to one-third of patients with HH and cirrhosis will develop hepatocellular carcinoma. This represents a 200-fold increased risk for the development of liver cancer. HH patients without bridging fibrosis or cirrhosis do not have an increased risk for developing hepatocellular carcinoma.

ROLE OF LIVER BIOPSY

HFE gene testing may eliminate the need for a liver biopsy in many cases. Traditionally, a liver biopsy has been performed in patients with iron overload to confirm the diagnosis of HH and to exclude cirrhosis. Patients who are homozygous for the C282Y mutation with an elevated serum iron and transferrin saturation without secondary iron overload do not need a liver biopsy to confirm the diagnosis of HH. Liver biopsy still remains the "gold standard" for assessing the degree of fibrosis. Definitively excluding cirrhosis is important because of the increased risk of developing hepatocellular carcinoma. The risk for cancer persists even after patients are depleted of excess iron stores. In such patients, screening with an ultrasound scan and α-fetoprotein every 6 months may be appropriate.

There may be a subset of HH patients whose risk of cirrhosis is minimal, and a liver biopsy would be unnecessary. Several recent studies have confirmed that certain predictive noninvasive assessments are accurate in excluding cirrhosis in C282Y homozygotes. In these studies, cirrhosis was extremely uncommon in C282Y homozygotes who had serum ferritin levels lower than 1000 μg/liter and normal aspartate aminotransferase values. A serum ferritin of <1000 μg/liter seems to be the best predictor of the absence of cirrhosis in C282Y homozygotes. The positive predictive value of a serum ferritin of >1000 μg/liter is poor, however, because only about 50% of those with serum ferritin values >1000 μg/liter had cirrhosis. A recent study found that cirrhosis was present in approximately

80% of C282Y homozygotes with serum ferritin of >1000 μg/liter and a platelet count <200 k and an elevated AST. Until these findings are confirmed, a liver biopsy is advisable in C282Y homozygotes with serum ferritin values >1000 μg/liter to definitely assess for the presence of cirrhosis. There are limited data on noninvasive predictors of cirrhosis for non-C282Y homozygotes. A liver biopsy may be necessary in this group of patients to confirm the diagnosis of HH and exclude cirrhosis.

SCREENING FOR HEMOCHROMATOSIS

There is currently disagreement among experts regarding the utility of screening for HH in the general population. HH fulfills most if not all of the World Health Organization (WHO) criteria of a condition amenable to population screening. HH is common and has a long presymptomatic phase. A simple, noninvasive, inexpensive screening test exists, and effective treatment that improves survival is available. Several studies have used reasonable estimates of the number of patients who will develop life-threatening disease and concluded that population screening for hemochromatosis would be cost effective.

Despite the fact HH fulfills many of the criteria of a condition appropriate for population screening, many public health experts have not recommended screening for HH. They cite a lack of information about burden of disease and disease expression in those with *HFE* mutations as reasons why they do not endorse population screening for HH. In support of this, a recent population screening study in the United States of over 40,000 individuals did not find a difference in symptoms reported by C282Y homozygotes compared to non-C282Y homozygotes.

Unfortunately, the natural history HH in an asymptomatic patient identified by population screening may never be known because many would consider it unethical to withhold treatment once a patient develops iron overload. At present, several studies have performed both *HFE* genotyping and serum iron studies in subjects selected from the general population. Between 25 and 81% of C282Y homozygotes in these studies had a normal serum ferritin with no clinical evidence of iron overload. More data from a greater number of individuals, including those from diverse populations, are necessary before reaching definitive conclusions. Without this information, true estimates on the cost effectiveness of screening for hemochromatosis in the general population will be difficult to obtain. With the information currently available, it seems reasonable to screen for iron overload in persons with a chronic liver disease, symptoms or signs suggestive of HH, or a family history of HH.

See Also the Following Articles

Iron Absorption • Liver Biopsy • Liver Transplantation • Neonatal Hemochromatosis

Further Reading

Adams, P. C. (1999). Population screening for hemochromatosis. *Hepatology* **29**, 1324–1327.

Bacon, B. R., Powell, L. W., Adams, P. C., Kresina, T. F., and Hoofnagle, J. H. (1999). Molecular medicine and hemochromatosis: At the crossroads. *Gastroenterology* **116**, 193–207.

McCullen, M. A., Crawford, D. H. G., and Hickman, P. E. (2002). Screening for hemochromatosis. *Clin. Chim. Acta.* **315**, 169–186.

Philpott, C. C. (2002). Molecular aspects of iron absorption: Insights into the role of HFE in hemochromatosis. *Hepatology* **35**, 993–1001.

Tavill, A. S. (2001). Diagnosis and management of hemochromatosis. *Hepatology* **33**, 1321–1328.

World Health Organization, the National Institute of Health, and the Centers for Disease Control and Prevention. (2000). EASL International Consensus Conference on Haemochromatosis. *Hepatol.* **33**, 485–504.

Hernias

ZARA COOPER AND STEPHEN JOHN FERZOCO
Brigham and Women's Hospital, Boston

Ehlers–Danlos syndrome Heterogeneous group of heritable collagen disorders characterized by joint hypermobility and increased skin elasticity and tissue fragility.

Marfan syndrome Heritable disorder of the connective tissue (caused by a mutation in the fibrillin gene) that affects many organ systems.

A hernia is the abnormal protrusion of a structure through the tissues that normally contain it. A vast majority of abdominal wall hernias are located in the groin. A hernia is reducible if the protruding contents can be put back in their original location. Herniated structures that cannot be restored to their normal location are called irreducible or incarcerated, and are at risk for obstruction or strangulation. Strangulation, which implies compromise to the hernia blood supply, can eventually lead to tissue death and serious infection.

INTRODUCTION

Hernias are very common, and hernia repair is the most common reason for primary care physicians to make surgical referral. Over 700,000 groin hernias are repaired in the United States and Europe annually. Males are seven times more likely than females to develop a hernia and have a 5% lifetime risk of developing a groin hernia. Risk factors include obesity, pulmonary disease with chronic cough, steroid use, straining with exercise, urinary retention, constipation, and other conditions that chronically increase intraabdominal pressure. Connective tissue disorders such as Ehlers–Danlos and Marfan syndromes have also been associated with increased risk of hernia.

All abdominal wall hernias involve a peritoneal sac protruding through a weakness in the abdominal wall. Each side of the abdominal wall includes an inguinal canal, a canal-shaped opening to the groin. In women, this canal contains the round ligament, which supports the uterus. In men, the canal contains the spermatic cord, which includes the blood supply to the testicles and the vas deferens, through which semen flows from the testes to the penis. In the anteromedial portion of the cord, there is a potential space through which abdominal contents can herniate.

ANATOMY

The anatomy of the lower abdominal wall is complex, and even after centuries of study, there is still debate about its structure. The lower wall is composed of three musculoaponeurotic layers: the external oblique, the internal oblique, and the transversus abdominus. Inferiorly, the external oblique aponeurosis forms the superficial inguinal ring through which the round ligament or spermatic cord exits the inguinal canal. This triangular opening is just superior to the medial portion of the inguinal ligament. The internal oblique is the middle layer of the abdominal wall. In 5–10% of the population, the internal oblique aponeurosis will merge with the transversus abdominus aponeurosis to form the conjoint tendon. The conjoint tendon is in the medial aspect of the canal and can be used in reconstruction of the abdominal floor during hernia repair. When the conjoint tendon is not present, the internal oblique contributes little to the integrity of the groin. The deepest layer of the abdominal wall is the transversus abdominus, which is encased in a fascial sheath called the transversalis fascia. The deep inguinal ring, through which the round ligament or spermatic cord enters the inguinal canal, is a defect in the transversalis fascia. This is located at the midpoint of the anterior superior iliac spine and the pubic tubercle. The transversalis fascia forms the floor of the inguinal canal. The roof of the canal is composed of the external oblique aponeurosis and fibers of the internal oblique muscle. Fibers of the internal oblique and transversus abdominus form the superior wall of the canal, and the inguinal and lacunar ligaments form the inferior wall. The illiohypogastric, illioinguinal, and genitofemoral nerves all traverse the region and provide key motor and sensory functions. Nerve entrapment is a risk of herniorraphy and can cause postoperative disability.

The femoral canal is a potential space underneath the inguinal ligament, through which visceral herniation may occur. Its borders are the inguinal ligament

 Copyright 2004, Elsevier (USA). All rights reserved.

superiorly, the external iliac vein medially, Cooper's ligament posteriorly, and fibers of the illiopubic tract as they converge onto Cooper's ligament laterally. The entrance of the canal is the fossa ovalis, a defect in the fascia lata. Here, the femoral and greater saphenous veins merge.

CLASSIFICATION

Hernias are classified by their location. There are a number of types of groin hernias: indirect, direct, pantaloon, and femoral. Indirect inguinal hernias are due to congenital patency of the processus vaginalis and weakening of the transversalis fascia around the internal ring. The processus vaginalis is the peritoneal tunnel through which the testes migrate from the retroperitoneum toward the scrotum during embryological development. Here, there will be herniation of abdominal contents through the internal inguinal ring. Direct inguinal hernias are secondary to weakness in the transversalis fascia inferior to the inferior epigastric artery. The region is known as Hesselbach's triangle. When both direct and indirect hernias are present, it is called a pantaloon hernia. Femoral hernias occur when intraabdominal contents pass beneath the inguinal ligament and through the femoral canal medial to the femoral vein. This type of hernia is most likely to occur in elderly women and precipitant weakness in the abdominal floor may be related to pregnancy.

Anterior abdominal wall hernias are umbilical, incisional, and epigastric. Umbilical hernias may be congenital and appear as a protruding belly button. Umbilical and incisional hernias may also be acquired as defects through previous fascial closures. Epigastric hernias are secondary to congenital defects in the linea alba.

A lump heralds most groin hernias. Patients often describe a sudden bulge that occurred while straining while lifting or moving their bowels. Pain can be associated with an enlarging hernia and may be referred to the scrotum. With the exception of dull pain and a dragging sensation, most reducible hernias are asymptomatic. Patients must be examined while standing and supine, and while coughing and straining to fully evaluate the hernia. A finger should be inserted through the upper scrotum and into the internal ring. If a hernia is present, a sudden impulse or lump will be felt on the fingertip with cough or strain. Bowel sounds may be heard on auscultation of the hernia. It is usually difficult to discern direct from indirect hernias on physical exam, and in most cases the type of hernia is not accurately identified before surgery. However, if contents descend into the scrotum, an indirect, rather than direct, hernia is present. Most hernias in childhood and early adulthood are indirect inguinal hernias.

If a hernia is tender, discolored, or incarcerated, an experienced hand can attempt manual reduction first. Any difficulty with this maneuver signifies that strangulation of the hernia is likely, and emergent operation is indicated. Patients with strangulated hernias may be toxic, dehydrated, and febrile. In this case, reduction should not be attempted. The patients should be adequately resuscitated and taken to the operating room as soon as possible. The risks of incarceration, strangulation, and obstruction are greater compared to risks of hernia repair. Consequently, expeditious elective operative repair is standard unless there are specific contraindications or operative risk is too high. This applies to groin and ventral hernias. When all goes well, elective herniorraphy is often an outpatient procedure done with local anesthesia.

SURGERY

The surgical principles of hernia repair include division of the external oblique and transversalis fascia, isolation of the spermatic cord, high ligation of the hernia sac, and reconstruction of the inguinal canal. In general, there are three approaches to surgical repair of groin hernias: (1) open repair, (2) tension-free repair with mesh or Lichtenstein repair, and (3) laparoscopic repair. Open repair involves freeing the spermatic cord and opening the transversalis fascia to access the inguinal canal. The hernia sac is identified and ligated at its base. Hernia contents are restored to their correct position, and the floor of the inguinal canal is reconstructed with permanent sutures and minimal tension. In tension-free repair with mesh, prosthetic mesh instead of sutures is used to construct the canal. Although there is higher risk of infection with this technique, the subsequent fibrotic reaction secures the repair, resulting in a recurrence rate of less than 1%. Laparoscopic hernia repair is done via smaller incisions, which may result in less pain and disability. One advantage is that bilateral hernias can be repaired with one large piece of abdominal mesh placed posteriorly. Another advantage is that recurrent hernias can be repaired without interfering with previous surgical sites. Drawbacks to laparoscopic repair include requirement of general anesthesia and significantly more expense than open techniques. Because laparoscopic herniorraphy is a relatively new technique, long-term results are unavailable. Thus far, complication and recurrence rates are similar to those of open techniques.

Complications following hernia repair are usually minor; they include wound infections, hematoma, and

nerve entrapment. Injury to the illioinguinal nerve can cause groin and scrotal pain. Entrapment of the genitofemoral nerve is associated with ejaculatory dysfunction and hypersensitivity of the groin, scrotum, and upper thigh. Recurrence may not occur for up to 20 years and further recurrence is more likely after two or more repairs. Recurrence is more likely in the setting of related comorbidities.

Umbilical hernias are present in 10% of White infants and in greater than 40% of Black infants. A hernia in infancy is a congenital defect that usually regresses within 2 years. It is for this reason that repair is postponed until after 2 years of age. In adults, the defect does not regress. Like groin hernias, umbilical hernias tend to expand over time and are aggravated by conditions that increase intraabdominal pressure. The risk of incarceration or strangulation is greater than the risk of elective repair, and surgical treatment is preferred in the absence of contraindications. In umbilical hernias, the sac usually contains omentum. However, strangulated omentum can pose a serious risk to the patient and should not be underestimated in considering urgent repair. Epigastric hernias are often multiple, and until they become symptomatic, repair is usually postponed.

About 10% of incisional hernias occur within the first postoperative year, almost 65% occur within the first 5 years, and the remainder occur after that. In addition to risk factors for other types of hernias, age, poor nutrition, and jaundice can also predispose patients to incisional hernias. Of course, poor surgical technique and wound infection can also predispose hernia formation. Surgery is indicated for repair to prevent incarceration and strangulation. Small defects can be closed primarily, but most incisional hernias require mesh for adequate closure. Drains are often indicated to prevent fluid collections, which may contribute to infection.

In summary, abdominal wall hernias are common and usually cause minimal stress or discomfort. They should be treated with early surgical repair to prevent untoward complications, which can cause significant morbidity and mortality. There are a number of approaches to hernia repair, and the type of repair will vary by patient need and expertise of the surgeon.

See Also the Following Articles

Ehlers–Danlos Syndrome • Hiatal Hernia • Laparoscopy • Minimally Invasive Surgery

Further Reading

Deveney, K. E. (1994). Hernias and other lesions of the abdominal wall. *In* "Current Surgical Diagnosis and Treatment" (L. W. Way, ed.), 10th Ed., p. 712. Appleton & Lange, East Norwalk.

Liem, M. S. *et al.* (1997). Comparison of conventional anterior surgery and laparoscopic surgery for inguinal-hernia repair. *N. Engl. J. Med.* **336**(22), 1541.

Skandalakis, J. E., *et al.* (2001). Anatomy of the hernia rings. *In* "Mastery of Surgery" (R. J. Baker and J. E. Fischer, eds.), 4th Ed., p. 1881. J. B. Lippincott, Philadelphia.

Skandalakis, J. E., Gray, S. W., Mansberger, A. R., *et al.* (1989). "Hernia: Surgical Anatomy and Technique." McGraw Hill, New York.

Wantz, G. E. (2002). Open repair of hernias of the abdominal wall. *In* "ACS Surgery Principles and Practice" (D. W. Wilmore, ed.), p. 857. WEB MD Reference, New York.

Hiatal Hernia

SUBROTO PAUL*RAPHAEL BUENO†
*Brigham and Women's Hospital and †Harvard University

esophagogastroduodenoscopy Endoscopic examination of the esophagus, stomach, and duodenum.
fundoplication Procedure in which the stomach fundus is wrapped around the esophagus.

The esophageal hiatus is formed by the muscle fibers of the right crus of the diaphragm with limited contribution from the left crus. The esophageal hiatal crural fibers form a tunnel through which the esophagus travels from the thorax to the abdomen. The distal esophagus is further attached to the hiatus by the phreno-esophageal ligament, which arises from the fusion of endothoracic and endoabdominal fascia. A hiatal hernia results from the herniation of abdominal contents through the esophageal hiatus into the thoracic cavity or the mediastinum. Small hiatal hernias are common radiographic findings and may or may not be associated with clinical significance. This article focuses on the diagnosis, classification, pathophysiology, and treatment of hiatal hernias.

CLASSIFICATION AND PATHOPHYSIOLOGY

There are multiple ways to classify hiatal hernias. The two basic types are small sliding hernias that are associated with reflux and much larger hernias that progressively contain the stomach and other intra-abdominal contents. This article will use a four-type hernia classification scheme based on the extent of the defects.

Type I

The type I hiatal hernia or sliding hernia is the most common. Type I paraesophageal hernia is defined as one in which the gastroesophageal junction has migrated from the intraabdominal position through the esophageal hiatus and subsequently occupies an intrathoracic position. Type I hernia occurs chiefly as a result of a weakening or thinning of the phreno-esophageal membrane. This type of hernia is rather common and is reportedly found in 10% of Northern American adults examined by barium swallow. This defect can be temporarily induced in many normal individuals by any maneuver that increases intraabdominal pressure, such as vomiting or the Valsalva maneuver. The significance of this defect is dependent on the degree of associated symptoms of esophageal reflux as discussed below.

There are multiple potential etiological causes for type I hiatal hernias. Some individuals may have a congenitally inferior insertion of the phreno-esophageal ligament into the esophagus or a congenitally weakened ligament leading to an intra-thoracic gastro-esophageal junction and resulting in an intra-thoracic location of the gastric cardia. Some of these patients may have had symptoms of reflux or regurgitation since childhood. However, in most cases, it is difficult to know with certainty whether the defect was acquired or congenital in nature. The development of a type I hernia is also thought to result from an increase in intra-abdominal pressure that occurs in conditions such as pregnancy, obesity, protracted vomiting, or direct trauma from blunt force to the abdomen. Protracted vomiting may also cause vigorous esophageal contraction leading to the elevation of the gastro-esophageal junction into the chest.

Type I sliding hiatal hernia rarely causes symptoms. Symptoms are present in a fraction of patients with this defect and result from reflux attributed to a lower esophageal sphincter rendered dysfunctional by its intrathoracic location. Although most patients with sliding hernia do not have reflux, between 50 and 90% of patients with reflux have an associated sliding hiatal hernia. Gastro-esophageal reflux symptoms commonly present as a sensation of "heartburn" that is worsened by positional changes such as lying down. Heartburn is a nonspecific symptom complex, as it may mean many different things to different patients. In cases of hiatal hernia, heartburn is a burning sensation in the epigastrium that in severe cases can radiate to the left shoulder and be mistaken for myocardial infarction. This symptom complex is usually present after meals and may often be relieved by the ingestion of food or antacid medications.

Regurgitation of gastric contents is also an associated symptom of reflux. Patients usually describe the taste of sour liquid in their mouth. Regurgitation

 Copyright 2004, Elsevier (USA). All rights reserved.

of food is not typically reported as a symptom of reflux associated with type I hiatal hernias and the presence of such a symptom should lead to a search for another etiology such as a type II hiatal hernia or esophageal diverticulum. Dysphagia is often also associated with reflux seen with type I hiatal hernias and is thought to result from esophageal spasm, irritation, or stricture of the esophagus brought about by acid exposure in the distal esophagus. However, esophageal carcinoma can also present with dysphagia and hence these symptoms should be evaluated aggressively usually with a barium swallow and most likely with an esophagogastroduodenoscopy (EGD).

Patients with gastro-esophageal reflux may also present with respiratory symptoms resulting from aspiration of gastric acid. Typically, such patients present with a nocturnal cough that awakens the patient from sleep. Hoarseness may or may not be noted in the morning. Asthma may be a manifestation of this disease, but it is currently unclear to what extent and how frequently asthma results from aspiration. Severe aspiration episodes can ultimately lead to pneumonia, lung abscess, or bronchiectasis. There is some evidence that patients with severe uncorrected reflux after lung transplantation have a higher incidence of rejection.

Other less common symptoms associated with type I hiatal hernias include hemorrhage from diffuse erosive gastritis as well as ulcerative esophagitis with referred pain to the back and neck. Patients with the latter condition often complain of a "lump in their throat" that is unrelieved by swallowing. Frequently, these patients are misdiagnosed or dismissed as having psychological problems.

Type II

The type II hiatal hernia or para-esophageal hernia is an uncommon disorder that is anatomically distinct from the type I sliding hernia. The problem is a defect in the phreno-esophageal membrane and the underlying endothoracic and endoabdominal fasciae rather than a circumferential weakening as in type I defects (Fig. 1). The defect present in type II hernia is usually located in the left ventral aspect of the esophageal hiatus (Fig. 1), but occasionally defects can be present in other locations. Through these defects, a peritoneal sac containing a portion of the stomach protrudes along with the peritoneum into the thoracic cavity. Without the constraints of the fascia to prevent migration and the pressure gradient from the abdominal to thoracic cavity, the hiatal defect progressively enlarges and the stomach progressively migrates into the chest with total herniation of the stomach into the chest in advanced cases.

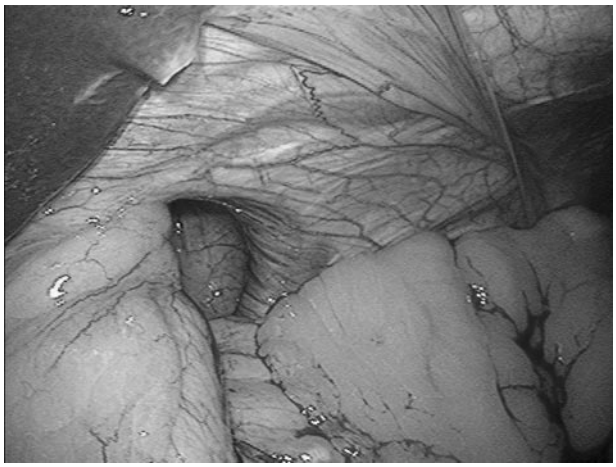


FIGURE 1 A laparoscopic view of the hernia sac of a reduced type II hiatal hernia. The stomach is reduced into the abdomen and is below and to the left of the diaphragmatic defect that is bordered on the right by the left crus.

However, since the distal esophagus and cardia of the stomach are usually still fixed by the phreno-esophageal membrane to the right crus and remain in the abdominal cavity even in advanced cases, the pylorus can come to lie next to the distal esophagus and cardia. Fixation at this single point allows for the development of gastric volvulus and its often fatal consequences. Hence, type II or para-esophageal hernias often require prompt attention.

Typically, patients with type II hernias are asymptomatic until the hernia becomes large. Most patients present in their fourth to fifth decades of life and a fair number present when very old. Symptoms of type II hernias include retching and postprandial nausea. Reflux symptoms may be present as well if a combined type I defect is also present to form a type III defect, which will be described below. In type II hernias with fixation at a single point, the intra-thoracic stomach can twist, leading to the development of gastric volvulus, gastric obstruction, gastric dilation, and gastric strangulation. With intermittent gastric distention and obstruction, patients can experience a sense of substernal pain often exacerbated after eating, as well as vomiting from obstructive pathology. Other symptoms include bleeding from gastric ulceration in the intrathoracic portion of the stomach and aspiration of gastric contents, leading to pulmonary complication as discussed for type I hernia.

The most feared complication of type II hiatal hernia is gastric volvulus and strangulation. Patients with this complication experience substernal pain and progressive dysphagia followed by protracted nausea and vomiting. As the stomach becomes progressively ischemic,

pain increases and is followed by symptoms of shortness of breath, shock, and sepsis intervening as the ischemia progresses and perforation occurs, leading to mediastinitis. Without prompt surgical management, patients rapidly deteriorate and expire. Due to the dire consequences of type II hernias, prompt surgical correction is recommended even in asymptomatic patients with significant type II paraesophageal hernias. Patients with known type II hernias who present with unremitting pain should be resuscitated and have the stomach gently decompressed with the aid of a carefully placed nasogastric tube under fluoroscopic guidance. Those who improve dramatically can be electively repaired surgically. Those who do not may require an emergent exploration usually transabdominally.

Type III

Type III, or mixed, hiatal hernia is a progression and combination of type I and II hiatal hernias. If a type I hiatal hernia is present, the phreno-esophageal membrane may weaken over time, leading to a type II defect. Conversely, the presence of a type II defect may over time weaken the phreno-esophageal ligament circumferentially, leading to a type I defect. The rate of conversion from a type I or type II hiatal hernia to a type II hernia is not known.

The symptoms of a type III hiatal hernia manifest as a combination of those seen in type I and II hiatal hernia. Typically, the symptoms of the larger defect predominate. However, once diagnosed, type III hiatal hernia must promptly be fixed as the stomach may strangulate.

Type IV

In type IV, hiatal hernia is defined as one where the stomach and other intra-abdominal organs have herniated into the chest through the esophageal hiatus. These herniate in order with the omentum followed by the transverse colon. The spleen or small bowel may be involved as well. Type IV hernias are rarely asymptomatic and symptoms are chiefly those of obstruction and strangulation. Hence, these hernias require prompt surgical treatment once diagnosed.

DIAGNOSIS AND MANAGEMENT

Most hiatal hernias are first incidentally noted on chest radiographs. The most frequent radiographic finding is a retro-cardiac air bubble. Patients are subsequently triaged according to the type of hernia and degree of symptoms. Those patients with dysphagia should undergo an aggressive workup to rule out esophageal carcinoma with EGD and biopsies if indicated. Subsequent evaluation with a barium swallow, chest and/or abdominal computed tomography scan, or EGD confirms the diagnosis in most patients.

Type II, III, and IV hiatal hernias often require prompt surgical patients even for asymptomatic treatment in order to avoid the mortality and morbidity of gastric and bowel strangulation. Operative repair can be accomplished through an abdominal or thoracic approach and by laparoscopy. There have been reports of a combined thoracoscopic or laparoscopic approach as well. Although strong proponents exist for each approach, there are no data to support the superiority of one surgical approach over another. There are several elements to the surgical repair. First, the stomach and any additional viscera must be returned to the abdominal cavity and the hernia sac excised (Fig. 2). Then, the hernia defect is repaired usually primarily and occasionally with the use of a prosthetic mesh. The esophagus is occasionally lengthened to ensure that the gastroesophageal junction remains in the abdomen. Patients who clearly have documentation of gastroesophageal reflux will also have an anti-reflux procedure (see below) performed at the same time. Whether all patients with paraesophageal hernias who are undergoing surgical repair ought to have an anti-reflux procedure is a controversial matter. The authors' group routinely performs an anti-reflux procedure as part of all advanced hernia repairs. Finally, patients with type III or IV hernias may also benefit from fixation of the stomach to the abdominal wall with a temporary gastrostomy in order

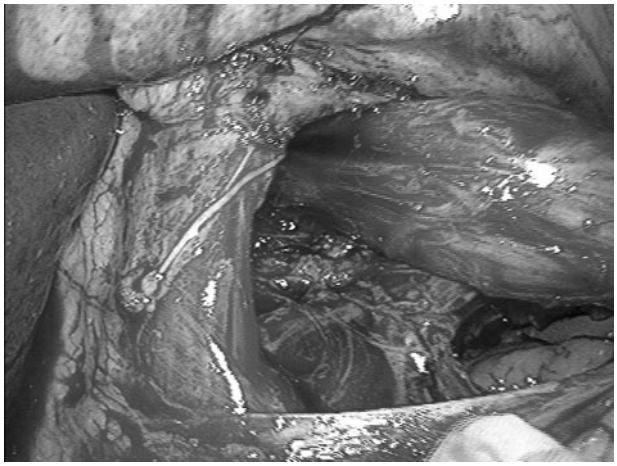


FIGURE 2 A laparoscopic view of the entire dissected hiatus in the same patient. The right crus is prominently displayed on the left side of the picture. The esophagus is coming out toward the right side of the picture and covering a portion of the left crus. Note the wide defect between the two crurae that will need to be surgically closed.

to reduce the chance of recurrent herniation and provide a simple portal to an often sluggish stomach.

Therapy for type I hiatal hernia is more controversial. The only clear recommendation is that for asymptomatic patients with small defects, expectant therapy is indicated. For symptomatic patients with gastro-esophageal reflux, therapeutic options include the following: (1) medical therapy; (2) endoscopic therapy; and (3) surgical therapy.

Medical Therapy

With the advent of proton pump inhibitors, most patients can be adequately treated for their reflux symptoms with medical therapy. However, controversy exists concerning the long-term effects of prolonged proton pump therapy. It is unclear whether acid suppression alone can prevent the metaplastic changes of Barrett's esophagus and progression to dysplasia. Also, it is unclear what constitutes maximal medical therapy and its failure. Furthermore, many patients do not wish to be on prolonged medical therapy or fail "maximal medical therapy." Finally, medical therapy cannot correct the anatomic irregularities of type I hiatal hernia. Given that most patients with type I hiatal hernias do not have reflux, it is likely that additional defects are present in those that do. In those patients, proton pump inhibitors may reduce the acid secretion but not affect adequate reduction in bile reflux, which may be harmful as well. Patients with severe reflux symptoms that are not controlled with medical therapy or complications of reflux (stricture, bleeding, and obstruction) that are not reversed with medical therapy are frequently referred for interventions to control these problems.

Endoscopic Therapy

New endoscopic techniques have recently been approved for treatment of gastroesophageal reflux by the Food and Drug Administration. At least three approaches have been described. One procedure utilizes endoscopic suturing to create a valve in the gastroesophageal junction to reduce reflux. The second is designed to thicken the same area with the administration of radiofrequency energy to the gastroesophageal junction in a manner designed to create a scar at the gastroesophageal junction. The third strategy is to inject material into the submucosa of the gastroesophageal junction in order to re-create a valve. All of these procedures are successful at least in the short term. However, the exact mechanism of action of each method is not certain. Most certainly, there is no anatomic repair of the hiatus or any hernia with these anti-reflux procedures. Although short-term results are promising and many patients are symptom- and medication-free, the long-term efficacy of these treatments is unclear. Additional long-term data are needed to better define patient selection criteria and make valid comparisons to both medical and surgical therapies.

Surgical Therapy

A variety of surgical approaches have been described to treat reflux and the reflux-associated symptoms of type I hiatal hernias. Each of these surgical approaches has been designed to approximate the normal anatomy of the gastroesophageal junction and increase the tone of the lower esophageal sphincter, in addition to repairing the hernia and hiatal defect. These techniques attempt to achieve these goals and prevent reflux by better approximating the angle of the gastroesophageal junction, by buttressing the lower esophageal sphincter or by "lengthening" the distal esophagus to create an intra-abdominal lower esophageal sphincter. A variety of operations have been described, but three main surgical approaches are currently in practice: (1) Belsey–Mark IV; (2) Nissen fundoplication; and (3) Hill repair.

A thorough evaluation of each patient with reflux symptoms is required to tailor the individual operation. Specifically, patients should undergo gastric pH probe testing, esophageal manometry, and barium swallow to demonstrate both the hiatal defect and esophageal motility prior to operative intervention.

Belsey–Mark IV

The Belsey–Mark IV operation was developed in the 1950s in England by Sir Ronald Belsey. The hallmarks of this anti-reflux procedure include restoring an intra-abdominal esophagus and creating and maintaining a narrow diameter in the distal esophagus to restore lower esophageal sphincter tone. The Belsey–Mark operation is performed through a left sixth interspace thoracotomy. In experienced hands, this procedure is effective in controlling the symptoms of 85% of patients at 5 years.

Nissen Fundoplication

Nissen fundoplication was originally developed by Rudolph Nissen in 1936 to protect the suture line after gastric resections. The Nissen procedure can be performed through a laparotomy, thoracotomy, or via laparoscopy. The procedure involves wrapping the anterior and posterior walls of the gastric fundus around the esophagus to augment the lower esophageal sphincter and thereby prevent reflux. Numerous modifications of this procedure that change the extent of gastric fundus encircling the distal esophagus (e.g., Dor, Thal) have

been added in hopes of modulating the increase in lower esophageal tone. Minimally invasive laparoscopic versions of these procedures have become quite common in modern-day surgical practice. Up to 95% of patients are still symptom free at 5 years.

Hill Repair

The Hill operation is designed to prevent reflux by re-creating the angle of the gastroesophageal junction with respect to the diaphragm (i.e., angle of His) in order to augment and thereby restore competency to the lower esophageal sphincter. This procedure can be performed through both abdominal and thoracic approaches. Laparoscopic and thoracoscopic approaches have been described as well. Long-term follow-up results of this procedure reveal that 75 to 85% of patients are asymptomatic at 5 years.

In summary, hiatal hernias can be classified into four types. Type I hernias are common and can be managed medically, endoscopically, or surgically. Type II, III, and IV hernias are more serious and once diagnosed often require prompt surgical attention in order to prevent the dire consequences of gastric strangulation.

See Also the Following Articles

Dysphagia • Esophageal Surgery • Gastroesophageal Reflux Disease (GERD) • Hernias • Laparoscopy • Minimally Invasive Surgery • Proton Pump Inhibitors

Further Reading

Ferraro, P., and Duranceau, A. (2001). Medical management of gastro-esophageal reflux disease. *Chest Surg. Clin. N. Am.* **11**, 517–522.

Katz, P. (2002). Gastro-esophageal reflux disease: New treatments. *Rev. Gastroenterol. Disord.* **2**, 66–74.

Luketich, J., Fernando, H., Christie, N., Buenaventura, P., Ikramuddin, S., and Schauer, P. (2002). Outcomes after minimally invasive reoperation for gastro-esophageal reflux disease. *Ann. Thorac. Surg.* **74**, 328–331.

Negre, J., Markkula, H., Keyrilainen, O., and Matikainen, M. (1983). Nissen fundoplication: Results at 10 years follow-up. *Am. J. Surg.* **146**, 635–637.

Skinner, D., and Belsey, R. (1967). "Management of Esophageal Disease." W. B. Saunders Philadelphia, PA.

Hiccups (Singultus)

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

diaphragm Muscular partition that separates the abdomen and thorax; a primary muscle used in breathing.

glottis Opening in the larynx through which exchange of air occurs between the mouth and lungs.

larynx Organ for production of voice; situated in the airway at the base of the tongue.

myoclonus Rapid involuntary alternate muscular contractions and relaxations resulting from paroxysmal outflow from the central nervous system to the involved muscles.

phrenic nerve Motor nerve arising from the third, fourth, and fifth cervical segments of the spinal cord; innervates the diaphragm.

The term "hiccup" is derived from the sound made by the rhythmic involuntary spasmodic contraction of the diaphragm and rapid closure of the glottis, followed by sharp inhalation. The medical term for hiccup, "singultus," probably originated from the Latin *singult*, which is translated as "the act of catching one's breath while sobbing." Singultus reflexes usually occur in sequence, accounting for the characteristic movements and sounds.

INTRODUCTION

The phrenic nerves are the motor nerves to the diaphragm. Inappropriate firing in the motor component of the phrenic nerves accounts for the diaphragmatic spasm associated with hiccups. The afferent arm of the hiccup reflex consists mainly of sensory nerves from receptors in the airways and upper digestive tract.

Copyright 2004, Elsevier (USA). All rights reserved.

been added in hopes of modulating the increase in lower esophageal tone. Minimally invasive laparoscopic versions of these procedures have become quite common in modern-day surgical practice. Up to 95% of patients are still symptom free at 5 years.

Hill Repair

The Hill operation is designed to prevent reflux by re-creating the angle of the gastroesophageal junction with respect to the diaphragm (i.e., angle of His) in order to augment and thereby restore competency to the lower esophageal sphincter. This procedure can be performed through both abdominal and thoracic approaches. Laparoscopic and thoracoscopic approaches have been described as well. Long-term follow-up results of this procedure reveal that 75 to 85% of patients are asymptomatic at 5 years.

In summary, hiatal hernias can be classified into four types. Type I hernias are common and can be managed medically, endoscopically, or surgically. Type II, III, and IV hernias are more serious and once diagnosed often require prompt surgical attention in order to prevent the dire consequences of gastric strangulation.

See Also the Following Articles

Dysphagia • Esophageal Surgery • Gastroesophageal Reflux Disease (GERD) • Hernias • Laparoscopy • Minimally Invasive Surgery • Proton Pump Inhibitors

Further Reading

Ferraro, P., and Duranceau, A. (2001). Medical management of gastro-esophageal reflux disease. *Chest Surg. Clin. N. Am.* **11**, 517–522.

Katz, P. (2002). Gastro-esophageal reflux disease: New treatments. *Rev. Gastroenterol. Disord.* **2**, 66–74.

Luketich, J., Fernando, H., Christie, N., Buenaventura, P., Ikramuddin, S., and Schauer, P. (2002). Outcomes after minimally invasive reoperation for gastro-esophageal reflux disease. *Ann. Thorac. Surg.* **74**, 328–331.

Negre, J., Markkula, H., Keyrilainen, O., and Matikainen, M. (1983). Nissen fundoplication: Results at 10 years follow-up. *Am. J. Surg.* **146**, 635–637.

Skinner, D., and Belsey, R. (1967). "Management of Esophageal Disease." W. B. Saunders Philadelphia, PA.

Hiccups (Singultus)

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

diaphragm Muscular partition that separates the abdomen and thorax; a primary muscle used in breathing.

glottis Opening in the larynx through which exchange of air occurs between the mouth and lungs.

larynx Organ for production of voice; situated in the airway at the base of the tongue.

myoclonus Rapid involuntary alternate muscular contractions and relaxations resulting from paroxysmal outflow from the central nervous system to the involved muscles.

phrenic nerve Motor nerve arising from the third, fourth, and fifth cervical segments of the spinal cord; innervates the diaphragm.

The term "hiccup" is derived from the sound made by the rhythmic involuntary spasmodic contraction of the diaphragm and rapid closure of the glottis, followed by sharp inhalation. The medical term for hiccup, "singultus," probably originated from the Latin *singult*, which is translated as "the act of catching one's breath while sobbing." Singultus reflexes usually occur in sequence, accounting for the characteristic movements and sounds.

INTRODUCTION

The phrenic nerves are the motor nerves to the diaphragm. Inappropriate firing in the motor component of the phrenic nerves accounts for the diaphragmatic spasm associated with hiccups. The afferent arm of the hiccup reflex consists mainly of sensory nerves from receptors in the airways and upper digestive tract.

Copyright 2004, Elsevier (USA). All rights reserved.

The vagus nerves transmit sensory information from the digestive tract and airways to the central nervous system. Sensory afferent fibers in the phrenic nerves may also participate in the reflex. The phrenic nerves are mixed nerves containing about half as many sensory afferent fibers as motor fibers. Processing centers for the sensory information and the integrative centers for the reflex are located in the brain stem along with the centers for respiratory control. Phrenic motor innervation of the diaphragm and the motor innervation of the larynx (laryngeal nerves) make up the efferent arm of the reflex. The phrenic motor axons project from neuronal cell bodies that make up the phrenic nucleus in the cervical spinal cord. The laryngeal nerves, which are mixed motor and sensory, are derived from the vagus nerves.

ETIOLOGY

Hiccups lasting only a few minutes are commonly experienced by normal individuals. These transient bouts of hiccups are considered to be a form of myoclonus, like the isolated myoclonic "jerks" that can be observed in individuals during drowsiness or light sleep. Therapy is usually nothing more than home remedies. On the other hand, pathologic prolongation of an episode of hiccups becomes a clinical problem that compromises quality of life and requires therapeutic intervention. An episode lasting more than a few minutes is generally referred to as a "bout," and one lasting longer than 48 hours is considered to be "protracted." The term "intractable" is reserved for attacks lasting more than 1 month. The longest reported attack lasted 60 years.

Hiccups lasting for extended periods are related to a variety of etiologies. Prolonged hiccups may be related to irritation of the diaphragm, distension of the stomach, central nervous sensitization, a variety of tumors, hyponatremia, or other metabolic disorders. Hiccups in hypersensitive individuals may be initiated by sudden chilling, wet feet, cold showers, or rapid ingestion of carbonated liquids or ice-cold food or drink. Prolonged symptoms can sometimes be traced to inflammation of the phrenic nerve or a more generalized neuritis. Pressure on the upper airway, by an enlarged goiter in the neck, for instance, may cause hiccups. Tumors of the esophagus, placement of stents in the esophagus, diaphragmatic hernia, bronchitis, or asthma may involve irritation of the phrenic nerves that leads to chronic hiccups. Irritation of the diaphragm from diseases of the liver frequently sensitizes the hiccup reflex, as does irritation of central integrative centers by brain tumors, epilepsy, or meningitis. Intractable hiccups may accompany systemic toxemias, septicemia, shock, advanced arteriosclerosis, severe avitaminosis, chronic nephritis with uremia, or diabetes mellitus. Patients undergoing hemodialysis sometimes suffer from intractable hiccups.

Persistent hiccups associated with oral or intravenous corticosteroid administration have been reported, as have hiccups thought to occur secondary to glucocorticoid-like effects of the reproductive hormone progesterone. A case of anabolic steroid-induced hiccups lasting for 12 hours prior to medical attention was reported for an elite "power" weight lifter. Cases of this nature are postulated to reflect steroid-mediated stimulation of the hiccup reflex by lowering of the threshold for synaptic transmission in the integrative microcircuitry of the brain stem. A report of persistent hiccups as a consequence of thoracic epidural steroid injection as treatment in a patient with chronic back pain is reminiscent of the action postulated to occur in the brain. Hiccups and emesis are linked in patients undergoing cancer chemotherapy with cisplatin. Cisplatin therapy is usually accompanied by administration of the steroid dexamethasone, and this raises the question of whether the hiccups reflect the central actions of the steroid. Men undergoing cisplatin chemotherapy are reported to have a significantly higher incidence of hiccups compared to women undergoing the chemotherapeutic regimens.

TREATMENT

For the run-of-the-mill benign case of hiccups, home remedies include breath holding, sudden fright, mental distraction, rebreathing from a paper bag, eating dry granulated sugar, drinking cold liquids, or pressure at the back of the throat sufficient to evoke the gag reflex. The approach to the management of the patient with debilitating effects of prolonged hiccups consists of identification and treatment of the underlying cause and therapeutic interventions to achieve resolution.

Baclofen, which is an analogue of the inhibitory neurotransmitter γ-aminobutyric acid, has been used in the successful treatment of intractable hiccups in cases of brain stem lesions, infectious abscess near the diaphragm, and hiccups related to the placement of self-expandable stents in the esophagus. Drugs that block dopamine receptors are reported to be efficacious in treatment of severe hiccups. Nefopam is a centrally acting analgesic that is related structurally to antihistamines and antiparkinsonian drugs. Intravenous administration of nefopam has been reported to stop refractory postoperative hiccups and hiccups in individual patients with protracted episodes related to either lymphoblastic leukemia, esophageal infection secondary to immunodeficiency syndrome, or central vascular malformation. Additional pharmacological approaches reported to be efficacious are intravenous administration of the local

anesthetic lidocaine, the psychotropic drug chlorpromazine, and the calcium entry blocker nifedipine. Intractable hiccups have been treated effectively with short-term general anesthesia. Acupuncture also is reported to be effective.

See Also the Following Articles

Swallowing • Vagus Nerve

Further Reading

Bilotta, F., and Rosa, G. (2000). Nefopam for severe hiccups. *N. Engl. J. Med.* **343**, 1973–1974.

Dickerman, R. D., and Jaikumar, S. (2001). The hiccup reflex arc and persistent hiccups with high-dose anabolic steroids: Is the brain stem the steroid-responsive locus? *Clin. Neuropharmacol.* **24**, 62–64.

Fischer, A. Q., and McLean, W. T., Jr. (1982). Intractable hiccups as presenting symptom of brainstem tumor in children. *Childs Brain* **9**, 60–63.

Hadjiyannacos, D., Vlassopoulos, D., and Hadjiconstantinou, V. (2001). Treatment of intractable hiccup in haemodialysis patients with baclofen. *Am. J. Nephrol.* **21**, 427–428.

Kolodzik, P. W., and Eilers, M. A. (1991). Hiccups (singultus): Review and approach to management. *Ann. Emerg. Med.* **20**, 565–573.

Muehlenberg, K., Weinerth, J., and Wiedmann, K. H. (2001). Short-term anesthesia to stop persistent hiccups. *Dtsch. Med. Wochenschr.* **126**, 35–26.

Rimirez, F. C., and Graham, D. Y. (1992). Treatment of intractable hiccup with baclofen: Results of a double-blind randomized, controlled, cross-over study. *Am. J. Gastroenterol.* **87**, 1789–1791.

Yang, W. (2001). Treatment of 60 cases of hiccup by puncturing points lateral to haiquan. *J. Tradit. Chin. Med.* **21**, 389–290.

Hirschsprung's Disease (Congenital Megacolon)

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

megacolon A colon that has abnormally large dimensions.

recto-anal reflex A reflex relaxation of contractile tone in the internal anal sphincter, evoked by distension of the rectum.

RET A gene that encodes for a receptor Tyrosine kinase.

taenia coli Three strips of longitudinal smooth muscle spaced equidistance around the circumference of the colon.

Herald Hirschsprung (1813–1916) was a Danish pediatric surgeon, who in 1888 published the first comprehensive report on extreme dilation of the colon in infants and young children. Megacolon of congenital origin is the most obvious manifestation of the disease that is now named after Hirschsprung. The megacolon is usually filled with impacted feces proximal to a constricted rectosigmoid segment. Hypertrophy of the musculature occurs in the grossly dilated colon and is a constant finding in children with the disease and in mice that are used as models for studying the disease.

INTRODUCTION

The earliest postulates for Hirschsprung's disease focused on the dilated region of the colon. It was logical at the time to assume that the musculature of the distended colon was paralyzed and incapable of forward propulsion of feces. Now, it is known that propulsive motility in the megacolon is normal and that the defect is localized to the distal constricted segment. The megacolon develops as feces accumulate due to blockade of passage through the terminal constricted segment (see Fig. 1). This proved to be the case because the megacolon and all other symptoms of Hirschsprung's disease disappeared after surgical resection of the narrowed segment. The same resolution of symptoms occurs after surgical diversion of the fecal stream through a colostomy. Congenital absence or reduction in numbers of ganglion cells in the enteric nervous system in the constricted segment is the underlying pathology of the

Copyright 2004, Elsevier (USA). All rights reserved.

anesthetic lidocaine, the psychotropic drug chlorpromazine, and the calcium entry blocker nifedipine. Intractable hiccups have been treated effectively with short-term general anesthesia. Acupuncture also is reported to be effective.

See Also the Following Articles

Swallowing • Vagus Nerve

Further Reading

Bilotta, F., and Rosa, G. (2000). Nefopam for severe hiccups. *N. Engl. J. Med.* **343**, 1973–1974.

Dickerman, R. D., and Jaikumar, S. (2001). The hiccup reflex arc and persistent hiccups with high-dose anabolic steroids: Is the brain stem the steroid-responsive locus? *Clin. Neuropharmacol.* **24**, 62–64.

Fischer, A. Q., and McLean, W. T., Jr. (1982). Intractable hiccups as presenting symptom of brainstem tumor in children. *Childs Brain* **9**, 60–63.

Hadjiyannacos, D., Vlassopoulos, D., and Hadjiconstantinou, V. (2001). Treatment of intractable hiccup in haemodialysis patients with baclofen. *Am. J. Nephrol.* **21**, 427–428.

Kolodzik, P. W., and Eilers, M. A. (1991). Hiccups (singultus): Review and approach to management. *Ann. Emerg. Med.* **20**, 565–573.

Muehlenberg, K., Weinerth, J., and Wiedmann, K. H. (2001). Short-term anesthesia to stop persistent hiccups. *Dtsch. Med. Wochenschr.* **126**, 35–26.

Rimirez, F. C., and Graham, D. Y. (1992). Treatment of intractable hiccup with baclofen: Results of a double-blind randomized, controlled, cross-over study. *Am. J. Gastroenterol.* **87**, 1789–1791.

Yang, W. (2001). Treatment of 60 cases of hiccup by puncturing points lateral to haiquan. *J. Tradit. Chin. Med.* **21**, 389–290.

Hirschsprung's Disease (Congenital Megacolon)

JACKIE D. WOOD

The Ohio State University College of Medicine and Public Health

megacolon A colon that has abnormally large dimensions.

recto-anal reflex A reflex relaxation of contractile tone in the internal anal sphincter, evoked by distension of the rectum.

RET A gene that encodes for a receptor Tyrosine kinase.

taenia coli Three strips of longitudinal smooth muscle spaced equidistance around the circumference of the colon.

Herald Hirschsprung (1813–1916) was a Danish pediatric surgeon, who in 1888 published the first comprehensive report on extreme dilation of the colon in infants and young children. Megacolon of congenital origin is the most obvious manifestation of the disease that is now named after Hirschsprung. The megacolon is usually filled with impacted feces proximal to a constricted rectosigmoid segment. Hypertrophy of the musculature occurs in the grossly dilated colon and is a constant finding in children with the disease and in mice that are used as models for studying the disease.

INTRODUCTION

The earliest postulates for Hirschsprung's disease focused on the dilated region of the colon. It was logical at the time to assume that the musculature of the distended colon was paralyzed and incapable of forward propulsion of feces. Now, it is known that propulsive motility in the megacolon is normal and that the defect is localized to the distal constricted segment. The megacolon develops as feces accumulate due to blockade of passage through the terminal constricted segment (see Fig. 1). This proved to be the case because the megacolon and all other symptoms of Hirschsprung's disease disappeared after surgical resection of the narrowed segment. The same resolution of symptoms occurs after surgical diversion of the fecal stream through a colostomy. Congenital absence or reduction in numbers of ganglion cells in the enteric nervous system in the constricted segment is the underlying pathology of the

Copyright 2004, Elsevier (USA). All rights reserved.

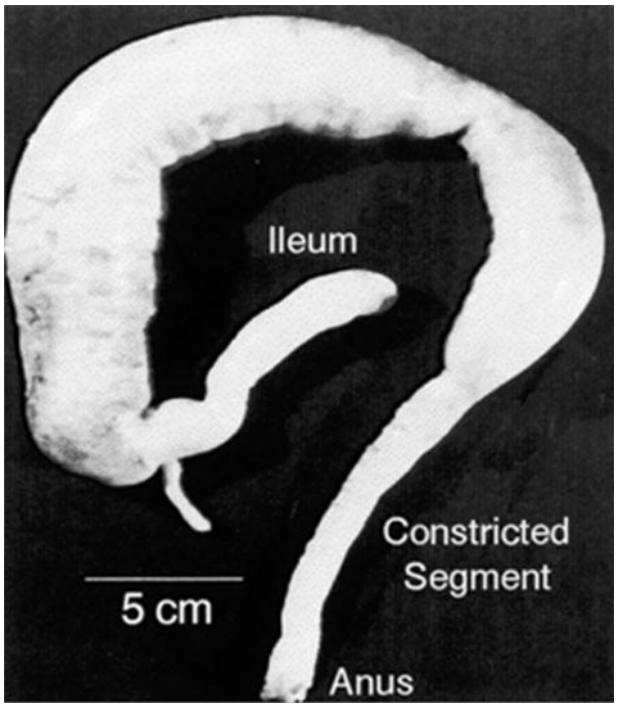


FIGURE 1 Colon from a 1-month-old infant with Hirschsprung's disease. The characteristic megacolon appears proximal to a constricted terminal segment. Histologic evaluation found a normal appearing enteric nervous system in the megacolon and absence of enteric neurons in the constricted terminal segment.

disease. Recognition of aganglionosis as the underlying factor in Hirschsprung's disease is the basis for use of the synonymous term aganglionic megacolon.

Hirschsprung's disease is usually suspected in cases of constipation during early childhood and must be ruled out by diagnostic testing. Failure of a newborn to pass the meconium within 3 to 4 days after birth also raises concern about the possibility for Hirschsprung's disease. Later in life, acquired megacolon and enteromegaly at higher levels of the gastrointestinal tract may similarly reflect pathologic loss of neurons from the enteric nervous system. Damage to the enteric nervous system can result from autoimmune attack directed against enteric neurons in patients with small cell lung carcinoma (i.e., paraneoplastic syndrome) and in patients infected with the blood-borne parasite *Trypanasoma cruzi* (i.e., Chagas' disease). The prominent histopathology in both congenital and acquired forms of the disease is distension and muscular hypertrophy proximal to an obstructive constricted segment for which enteric neurons are absent or grossly reduced in number.

In Hirschsprung's disease, both the myenteric and the submucosal plexuses are dysplastic in the distal constricted segment. The aganglionic constricted segment varies in length among different patients. The segment may involve only the distal 4 to 5 cm of rectum or it may involve the entire rectum, sigmoid colon, and descending colon. Occasionally, the entire colon is involved and in some cases the distal ileum is as well.

ANIMAL MODELS

Basic research aimed at elucidating the pathophysiology of Hirschsprung's disease has profited from animal models in which aganglionic megacolon can be induced experimentally and from mouse models in which aganglionosis reflects a congenital genetic defect. Attempts have been made to produce aganglionosis in dogs and rodents by destroying the enteric nervous system with the application of toxic substances to the intestinal surface or by restricting the oxygen supply to the intestine. Nevertheless, the confirmed and most thoroughly studied animal model of congenital megacolon occurs in mice. In 1966, researchers at The Jackson Laboratories, in Bar Harbor, Maine, reported the expression of two autosomal recessive genes in mice, both of which produced phenotypes with reduced pigmentation in the fur and decreased numbers of ganglion cells in the distal large intestine. One strain was named "piebald lethal" and the other "lethal spotting." The homozygotes ($S^{L}S^{L}$) of the piebald lethal strain have black eyes and white fur with an occasional patch of black pigment (Fig. 2). The piebald lethal mice always develop megacolon (Fig. 3) and die at the ages of a few days to several months. The heterozygous parents and siblings have black eyes and all black coat color.

The abnormal distribution of the enteric nervous system in the large intestine of the piebald mouse model is comparable with that in Hirschsprung's disease. In normal mouse large intestine, ganglia are present to within approximately 2 mm of the anus. In the abnormal mice, ganglia are absent from both the myenteric and submucosal plexuses in the distal 1 to 2.5 cm of the large intestine.

HEREDITARY FACTORS

Congenital megacolon in the mouse model is clearly a recessive trait that is not linked to sex chromosomes. Hereditary factors also seem to be involved in Hirschsprung's disease in humans. The incidence of Hirschsprung's disease in the general population of infants and children is 1 : 5000 (0.02%) and it occurs in siblings in 3.6% of cases. The incidence of Hirschsprung's disease, of the short-segment type where aganglionosis extends no higher than the sigmoid colon, is 5 to 10 times higher

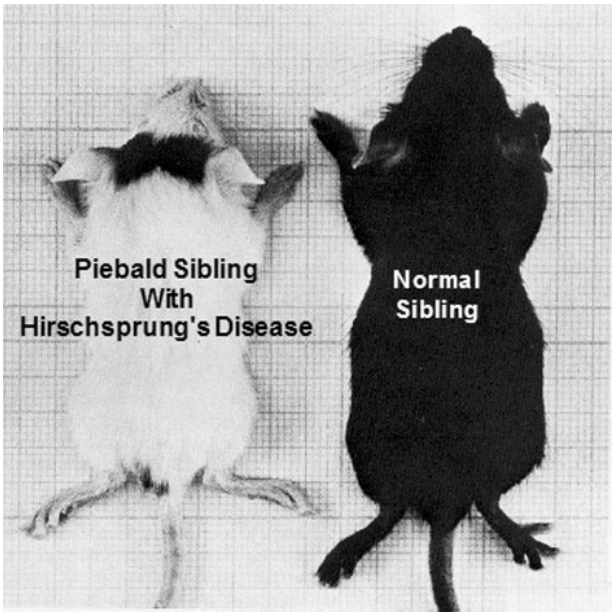


FIGURE 2 Piebald mouse model for Hirschsprung's disease. The two mice are siblings. The sibling on the right is normal and the one on the left is homozygous for the recessive gene that results in piebald pigmentation and aganglionic megacolon.

in human males than in females. Sexual differences are not apparent for the long-segment type in which the aganglionosis extends beyond the sigmoid colon.

The tools of molecular biology have enabled new insight into gene mutations that underlie the development of congenital forms of aganglionic megacolon in both humans and the mouse models. Proper functioning of two genes is now known to be necessary for the migration of enteric neuronal precursors from the embryonic neural crest to colonize the developing gut and for establishment and differentiation of the precursors into the kinds of neurons that link up synaptically to form the neural networks of the enteric nervous system. Mutations in two genes, one of which is called RET and the other of which is an endothelin gene, lead to aganglionic megacolon in mice. Nevertheless, the details of how the enteric nervous system reaches its final stages of development normally and what goes wrong when development fails in the aganglionic region of Hirschsprung's disease remain an unsolved scientific mystery.

PATHOPHYSIOLOGY

An early and often-cited hypothesis attributed failure of relaxation of the circular muscle coat in the aganglionic segment to denervation supersensitivity. Current evidence refutes this hypothesis. No increased sensitivity of the musculature of the constricted segment to the neurotransmitters acetylcholine or norepinephrine can be demonstrated either in the mouse model or in surgically resected diseased segments from children with Hirschsprung's disease.

The current hypothesis states that the behavior of the constricted terminal segment in Hirschsprung's disease directly reflects the loss of inhibitory motor neurons to the musculature. As a general rule, any condition that results in ablation of enteric inhibitory motor neurons results in contracture and achalasia of the intestinal circular muscle. Most evidence suggests that enteric inhibitory motor neurons are tonically firing and releasing their neurotransmitters at the neuromuscular junctions. Ablation of the inhibitory motor neurons releases the circular muscle from the inhibitory influence and this permits myogenic excitation and conduction to take place.

The physiology of neuromuscular relations in the large intestine predicts that spasticity and achalasia will accompany any condition in which inhibitory motor neurons are destroyed. Without inhibitory control, the self-excitable syncytium of nonsphincteric regions of the intestinal circular muscle will contract continuously and thereby behave as an obstruction to the passage of feces. This happens because each and every cycle of the myogenic pacemaker is freed to trigger

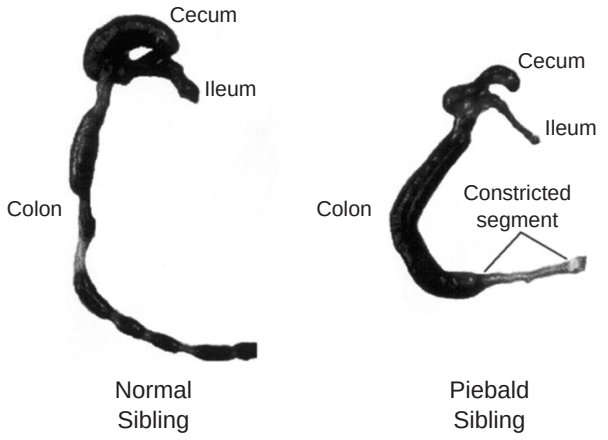


FIGURE 3 Large intestine of the piebald mouse model for Hirschsprung's disease. The two large intestines are from siblings. The one on the right is from the sibling with piebald coloration and shows the characteristics of the large intestine found in Hirschsprung's disease. The colon of the piebald sibling is grossly distended with impacted feces proximal to an empty and constricted terminal segment. The colon on the left is from the normal sibling. Histologic examination found a normally appearing enteric nervous system in the megacolon of the piebald sibling and absence of enteric neurons in the constricted terminal segment. The enteric nervous system was normal throughout the large intestine of the normal sibling.

contractions that propagate without any distance or directional control. Contractions spreading in opposite directions in the uncontrolled syncytium collide randomly, resulting in fibrillatory-like muscular activity in the affected intestinal segment.

In general, the function of enteric inhibitory motor neurons, aside from the various sphincters, is continuous suppression of activity of the inherently excitable musculature. The inhibitory neurons have the added function of mediating relaxation of contractile tone in sphincters. They are responsible for the internal anal sphincter relaxation required for defecation. The primary difference in the organization of the inhibitory outflow to the circular muscle of the intestine and the sphincters is that the inhibitory innervation of the bulk intestinal circular muscle is tonically active, whereas the inhibitory motor neurons to the sphincters are normally silent. In the intestine, myogenic contractions occur when ongoing firing of the inhibitory neurons is suppressed by the integrative microcircuits. In the sphincters, myogenic mechanisms maintain contractile tone. The sphincteric muscle is tonically contracted in the absence of inhibitory nervous input and the sphincter is relaxed when the normally silent inhibitory motor neurons are activated by the integrative microcircuits. This kind of physiology predicts that ablation of inhibitory motor neurons will lead to spasm of the intestinal circular muscle and to "achalasia" of the sphincteric musculature as found in Hirschsprung's disease and the mouse model of the disease.

Inflation of a balloon in the rectum triggers relaxation of the internal anal sphincter in humans and animals. This is called the recto-anal reflex. A normal recto-anal reflex requires that the enteric neural connections from the site of distension to the sphincter be intact and that the inhibitory motor neurons to the sphincter muscle be functional. Blockade or ablation of any of the neurons in the recto-anal reflex pathway eliminates the reflex. The inability to evoke the recto-anal reflex by balloon distension of the rectum is a diagnostic hallmark of Hirschsprung's disease.

See Also the Following Articles

Achalasia • Anal Sphincter • Basic Electrical Rhythm • Chagas' Disease • Colonic Motility • Colonic Obstruction • Constipation • Defecation • Disinhibitory Motor Disorder • Enteric Nervous System • Megacolon: Neuromuscular Enteric Abnormalities

Further Reading

Arey, J., and Valdes-Dapena, M. (1992). Embryology and developmental disorders. *In* "Pathology of the Gastrointestinal Tract" (S. Ming and H. Goldman, eds.), pp. 113–140. W. B. Saunders, Philadelphia, PA.

Gershon, M. (1999). Lessons from genetically engineered animal models. II. Disorders of enteric neuronal development: Insights from transgenic mice. *Am. J. Physiol.* **277**, G262–G267.

Madson, C. (1964). "Hirschsprung's Disease: Congenital Intestinal Aganglionosis." Charles C. Thomas, Springfield, IL.

Wood, J., and Cooke, H. (1985). Murine models for congenital megacolon: Hirschsprung's disease. *In* "Animal Models of Intestinal Disease" (C. Pfeiffer, ed.), pp. 181–195. CRC Press, Boca Raton, FL.

Histamine

GEORGE SACHS
University of California, Los Angeles

enterochromaffin-like cells Neuroendocrine cells in the gastric epithelium that control the peripheral regulation of acid secretion by releasing histamine as a paracrine stimulant.
gastrin Gastrointestinal hormone released from G cells in the gastric antrum in response to food.
histamine-2 (H2) receptor antagonists Drugs that competitively block the ability of histamine to interact with H2 receptors.
pituitary adenylate cyclase-activating peptide A neuropeptide present in the vagal nerve endings innervating the gastric mucosa.

Histamine released from neuroendocrine cells in the gastric mucosa is the major stimulant of acid secretion. A histamine-2 receptor subtype is highly expressed by the acid-secreting or parietal cells of the stomach and this receptor is the target for acid inhibition by histamine-2 subtype selective antagonists.

HISTORICAL INTRODUCTION

Whereas now it is universally accepted that histamine is a major stimulant of acid secretion as a consequence of the availability of pharmacological agents antagonizing its action on the stomach, it was only in the last quarter of the 20th century that this view was fully defensible.

The problem of regulation of acid secretion became of clinical relevance when the edict "no acid, no ulcer" was promulgated at the beginning of the 20th century. There had been several approaches to control of gastric acidity dating back to earliest written history with the consumption of baking soda by the Greeks. The Renaissance heralded the administration of atropine as "tincture belladonna" as the first nonselective muscarinic antagonist. Its side effects, such as dry mouth and mydriasis, precluded acceptance by patients. The end of the 19th century saw the beginning of gastric surgery as initially practiced by Billroth in Vienna. This progressed with time to more subtle approaches to control of acid secretion, culminating in highly selective vagotomy as a means of reducing stimulatory pathways. Hence, reduction of acid stimulation became a pharmacological and therapeutic target. But what were these pathways other than vagal?

In 1905, Edkins discovered that an extract of the gastric antrum was able to stimulate acid secretion and he called the extract gastrin. The reality of this extract was questioned by the leading English endocrinologists and the effect of gastrin and its mechanism were not solved until modern cellular methods were introduced.

Dale discovered histamine as a biological amine that performed a variety of functions, such as vasodilation. However, he did not describe the effect of histamine on acid secretion. This was left to a pupil of Pavlov, Popielski, who described the action of intravenous (iv) injection of histamine on dogs with gastric fistulas.

With the synthesis of histamine antagonists by Bovet in the 1950s, it was found that these agents, such as mepyramine, were relatively ineffective at blocking acid secretion. Two interpretations were viable: (1) that histamine was not really involved in physiological stimulation of acid secretion and (2) that these antagonists did not act on the site at which histamine acted in the stomach.

Gastrin became the major focus for animal physiologists as the key substance for stimulation of acid secretion with the purification and elucidation of the structure of gastrin in the early 1960s and the popularity of histamine as a major stimulant for secretion declined. However, in the 1970s, James Black and his team managed to synthesize and characterize new chemical entities that were able to inhibit acid secretion without affecting the other actions of histamine on blood vessels or mucus secretion. These were the histamine-2 (H2) receptor antagonists (H2RAs), the first being burimamide. With the ability to analyze the pharmacology of acid secretion by these selective receptor antagonists, it became clear that the action of gastrin on acid secretion was entirely histamine dependent. The action of cholinergic stimulation was much less sensitive to H2RA inhibition.

SOURCE OF GASTRIC HISTAMINE

As histamine became the dominant stimulatory ligand, the cell that provided the source of gastric histamine

 Copyright 2004, Elsevier (USA). All rights reserved.

became of interest. Evidently, histamine must be a paracrine stimulant since the pleiotropic effect of histamine on H1 receptors would not be tolerated.

At first, the speculation was that gastric mast cells were the source, but then the enterochromaffin-like (ECL) cell was discovered and its location at the base of the fundic glands made it the most likely source of gastric histamine. With the isolation of functional ECL cells, the peripheral pathway of stimulation of acid secretion was better defined. The major two stimulants of histamine release from these cells maintained in short-term culture were gastrin and pituitary adenylate cyclase-activating peptide (PACAP). Inhibitory ligands were somatostatin and galanin. Perhaps not surprisingly, no activation of this cell was found with muscarinic agonists, explaining the lack of sensitivity of cholinergic stimulation to H2RAs. Gastrin is released from the G cell of the gastric antrum and PACAP from the postganglionic fibers of the myenteric plexus, the former agent acting as an endocrine stimulant and the latter as a neural stimulant. A model of the pathways of stimulation can be seen in Fig. 1.

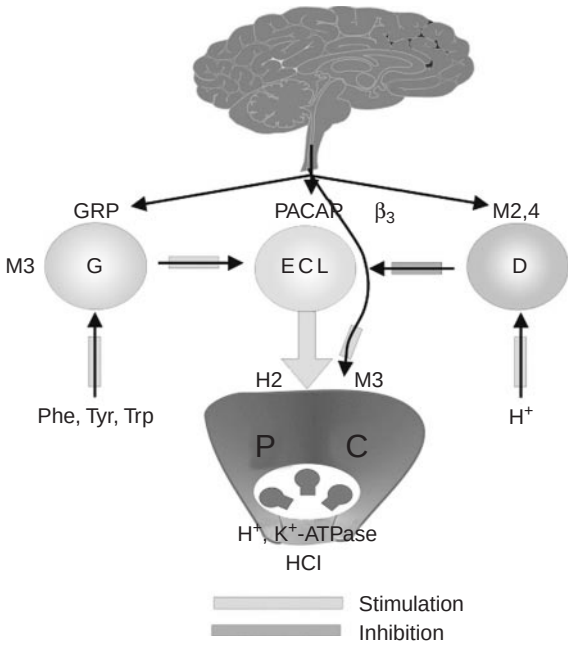


FIGURE 1 A simplified model of some of the pathways stimulating acid secretion. The central nervous system activates the myenteric plexus with release of PACAP to stimulate the ECL cell and with release of acetylcholine to stimulate the parietal cell directly. Gastrin is released from the G cell by gastrin-releasing peptide and acetylcholine as well as by aromatic amino acids in food. Somatostatin is released by the D cell in response to lumenal acidification and muscarinic M2 or M4 receptor stimulation.

THE HISTAMINE-2 RECEPTOR

The development of pharmacological agents selective for this receptor subtype long before methods were available to clone and sequence different receptors was a major breakthrough not only in therapy but in gastric biology. This receptor is expressed in the stomach, uterus, and heart but appears to have dominant function only in the stomach, allowing safe administration of H2RAs for the treatment of acid-related diseases without significant side effects.

The receptor is a seven-transmembrane segment protein inserted on the basolateral surface of the parietal

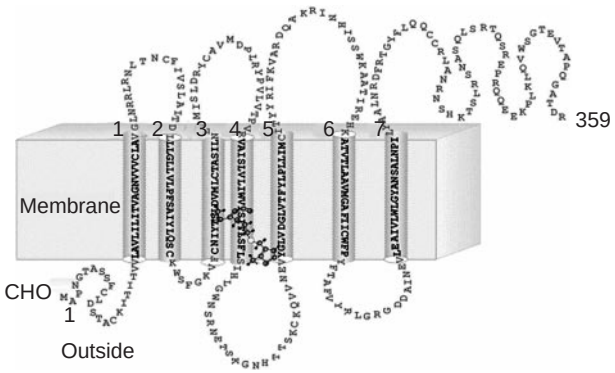


FIGURE 2 The amino acid sequence and secondary structure of the histamine-2 receptor with the antagonist docked close to the aspartyl (D) residue in TM3 and also interacting with TM5.

cell. Binding of histamine elevates cyclic AMP (cAMP) and this is the major intracellular messenger for stimulation of acid secretion by the parietal cell. However, the same receptor appears to stimulate elevation of intracellular Ca^{2+} although the role of this second messenger is not well understood. The action of cAMP is mediated by the activation of a type of protein kinase that phosphorylates several proteins in the stimulatory cascade. The action of Ca^{2+} may be mediated by a calmodulin-activated kinase.

The sequence of this receptor is known and the binding site of histamine and its antagonists is thought to involve a location between transmembrane 3 (TM3) and TM5. A model of the receptor with the region of binding of histamine receptor antagonists is shown in Fig. 2.

STRUCTURE OF HISTAMINE-2 RECEPTOR ANTAGONISTS

After the introduction of cimetidine (Tagamet), several other antagonists, such as ranitidine (Zantac) and

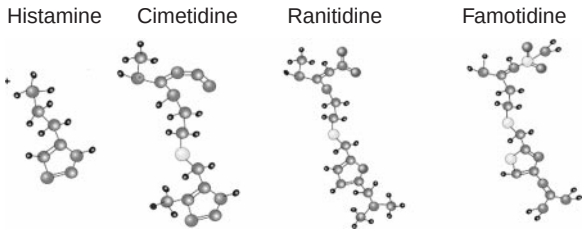


FIGURE 3 The structure of histamine compared to cimetidine, ranitidine, and famotidine. All have a protonatable amine similar to histamine, suggesting docking of a positive charge to the aspartyl in TM3. Famotidine is the most potent of these agents.

famotidine (Pepcid), were also put on the market. The structures of some of these compounds are shown in Fig. 3. They all have a similar profile of action, being able to inhibit nighttime acid secretion well but inhibiting daytime acid secretion with less efficacy. This indicates that nighttime secretion, which is low but highly acidic since there is no gastric buffering due to food, is mediated almost entirely by histamine, whereas daytime acid secretion, mediated mainly by gastrin release, is also dependent on a histamine-independent pathway. Characteristic of their action is tolerance, wherein their efficacy drops by approximately 50% after a few days of administration. The mechanism of this tolerance is not known. It is not due to receptor up-regulation since even 10-fold higher doses do not overcome the tolerance and there are no changes in gastrin levels or cAMP. Perhaps the muscarinic pathway for stimulation is enhanced.

MECHANISM OF ACTION

H2RAs are known to compete with histamine for binding to the receptor but their action is also that of inverse agonists in that they are able to inhibit the activity of the receptor's activation of adenylate cyclase in the absence of agonist. This inverse agonism probably explains their good efficacy with continuous iv administration.

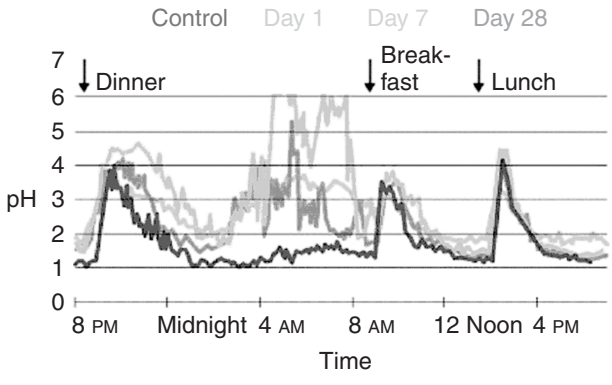


FIGURE 4 Median diurnal intragastric pH profiles following the administration of a 300 mg dose of ranitidine at 8 PM, showing excellent inhibition at night but little effect during the day. Also noteworthy is the tolerance, wherein even nighttime efficacy is much reduced by the seventh day of treatment. Reprinted from Teysson *et al.* (1995), with permission.

CLINICAL PROFILE

H2RAs and proton pump inhibitors are the mainstay of treatment of a variety of acid-related diseases, such as peptic ulcer and erosive and nonerosive esophageal reflux disease. They are considered to be almost entirely safe medications and all are available in over-the-counter formulations although at less than the therapeutic dose. A fairly typical profile of intragastric pH is shown for the standard nighttime dosage of a H2RA such as ranitidine (see Fig. 4).

See Also the Following Articles

Enterochromaffin-like (ECL) Cells • Gastric Acid Secretion • Gastrin • H2-Receptor Antagonists • Mast Cells • Parietal Cells • Pituitary Adenylate Cyclase Activating Peptide (PACAP)

Further Reading

Barger, G., and Dale, H. (1910). The presence in ergot and physiological action of B-imidazolylethylamine. *J. Physiol.* **40**, 38–40.

Black, J. W., Duncan, W. A. M., Durant, C. J., Ganellin, C. R., and Parsons, M. E. (1972). Definition and antagonism of histamine H2 receptors. *Nature* **236**, 385–387.

Bovet, D. (1950). Introduction to antihistamine agents and antergan derivatives. *Ann. N. Y. Acad. Sci.* **50**, 1089–1126.

Chew, C. S. (1989). Intracellular activation events for parietal cell hydrochloric acid secretion. *In* "Handbook of Physiology: The Gastrointestinal System III" (S. G. Schultz, and J. G. Forte, eds.), pp. 255–266. American Physiological Society, Bethesda, MD.

Chew, C. S. (1985). Parietal cell protein kinases: Selective activation of type I cAMP-dependent protein kinase by histamine. *J. Biol. Chem.* **260**, 7540–7550.

Chew, C. S., and Brown, M. R. (1986). Release of intracellular Ca^{2+} and elevation of inositol trisphosphate by secretagogues in parietal and chief cells isolated from rabbit gastric mucosa. *Biochim. Biophys. Acta* **888**, 116–125.

Edkins, J. (1906). The chemical mechanism of gastric secretion. *J. Physiol.* **34**, 133–144.

Fahrmann, M., Heinzmann, A., and Seidler, U. (2002). CaMKII is activated and translocated to the secretory apical membrane during cholinergically conveyed gastric acid secretion. *Cell. Signal.* **14**, 161–168.

Gantz, I., Schaffer, M., DelValle, J., Logsdon, C., Campbell, V., Uhler, M., and Yamada, T. (1991). Molecular cloning of a gene encoding the histamine H2 receptor. *Proc. Natl. Acad. Sci. USA* **88**, 5937.

Gregory, R., and Tracey, H. (1959). The preparation and properties of gastrin. *J. Physiol.* **149**, 70–71.

Hakanson, R., Owman, C. H., Sporong, B., and Sundler, F. (1971). Electron microscopic identification of the histamine-storing

argyrophil (enterochromaffin-like) cells in the rat stomach. *Z. Zellforschung Mikroskop. Anat.* **122**, 460.
Popielski, L. (1919). β-Imidazolyläthylamin und die Organextrakte. I. β-Imidazolyäthylamin als mächtiger erreger der Magendrüsen. *Pfluegers Arch. Gesamte Physiol. Menschen Tiere* **178**, 214–259.
Prinz, C., Kajimura, M., Scott, D. R., Mercier, F., Helander, H. F., and Sachs, G. (1993). Histamine secretion from rat enterochromaffin-like cells. *Gastroenterology* **105**, 449.
Sachs, G., Zeng, N., and Prinz, C. (1997). Physiology of isolated gastric endocrine cells. *Annu. Rev. Physiol.* **59**, 243–256.

H2-Receptor Antagonists

MITCHELL L. SCHUBERT
Virginia Commonwealth University's Medical College of Virginia and McGuire Veterans Affairs Medical Center, Richmond

erosion A shallow ulcer confined to the mucosal layer.
gastric H^+,K^+-ATPase An ATP-hydrolyzing enzyme responsible for catalyzing the exchange of lumenal K^+ for cytoplasmic H^+ by parietal cells, bringing about gastric lumenal acidification.
gastroesophageal reflux Backflow of acidic gastric contents into the esophagus, producing a spectrum of clinical symptoms and manifestations ranging from mild episodic heartburn and regurgitation without macroscopic esophagitis to chronic inflammation and ulceration; severe cases may involve stricture and hemorrhage.
gynecomastia Excessive development of the male mammary glands.
Helicobacter pylori A gram-negative urease-producing bacterium that causes an active chronic gastritis and is an important etiological factor in the development of gastric and duodenal ulcers.
nonsteroidal anti-inflammatory drugs Medications that reduce inflammation by inhibiting the enzymatic activity of cyclooxygenase, a key enzyme in the biosynthetic pathway leading to the formation of prostaglandins. Examples include aspirin, ibuprofen, naproxen, and diclofenac.
peptic ulcer A loss of tissue, extending through the mucosa into the submucosa, in the esophagus, stomach, or duodenum, due to acidic gastric secretions.
pharmacokinetics The disposition of drugs within the body in relation to their absorption, distribution, metabolism, and elimination.

Histamine receptors have been classified into four major subclasses: H1, H2, H3, and H4. H1 receptors are widely distributed and mediate the actions of histamine in the allergic response and as a bronchial constrictor. H2 receptors are present primarily in the stomach, where they regulate gastric acid secretion, and to a lesser extent in the heart, central nervous system, reproductive system, and on lymphocytes. Histamine-2 (H2)-receptor antagonists are drugs that competitively block the ability of histamine to interact with H2-receptors. H3-receptors were originally identified as inhibitory autoreceptors on histamine-containing nerve terminals in the central nervous system. In the stomach, H3-receptors augment acid secretion by suppressing somatostatin secretion. In the antrum, suppression of somatostatin leads to stimulation of gastrin, the main hormonal stimulant of acid secretion, whereas in the fundus, it leads to stimulation of histamine and acid secretion. Most recently, an H4-receptor has been identified in intestine, spleen, thymus, and immune cells. The H4-receptor shares 40% homology and overlapping pharmacology with the H3 receptor; its physiologic function is not known. This article focuses on the role of blocking H2 receptors in the stomach. The identification of the H2 receptor on the basolateral membrane of the parietal cell by Black in 1972 and the subsequent development of safe and effective H2-receptor antagonists have revolutionized the treatment of gastrointestinal acid-related disorders.

INTRODUCTION

Until the development of histamine-2 (H2)-receptor antagonists, the only treatments available for patients with acid-related disorders were bed rest, bland diets, milk drips, antacids, and surgery. Ulcers and ulcer pain

Copyright 2004, Elsevier (USA). All rights reserved.

Hyperemesis Gravidarum

MAKAU LEE
University of Mississippi Medical Center, Jackson

hyperemesis gravidarum Severe, intractable vomiting during early pregnancy resulting in fluid and electrolyte disturbances and/or nutritional deficiencies.

Hyperemesis gravidarum (also referred to as pernicious vomiting of pregnancy) is severe, intractable vomiting during early pregnancy resulting in fluid and electrolyte disturbances and/or nutritional deficiencies. In most cases, the vomiting begins soon after the first missed menstrual period, and the symptoms generally subside during the third month of pregnancy. The incidence of hyperemesis gravidarum averages 3.5 cases for every 1000 deliveries, and it is independent of parity, race, ethnicity, or psychological issues. The incidence of hyperemesis gravidarum has been noted to decrease significantly during wartime. Epidemiological studies have shown that hyperemesis gravidarum has no adverse effect on pregnancy outcome and that patients with hyperemesis gravidarum do not have an increased incidence of toxemia of pregnancy, stillbirths, miscarriages, growth retardation, congenital abnormalities, or prematurity. Of note, several studies have shown that women with hyperemesis gravidarum give birth to a higher proportion of female newborns.

CLINICAL FEATURES AND PATHOPHYSIOLOGIC CONSIDERATIONS

Women with multiple fetuses or with molar pregnancies (i.e., hydatidiform mole) tend to have extremely high concentrations of human chorionic gonadotropin (hCG) and have a higher incidence of hyperemesis gravidarum; however, the causal relationship between elevated hCG levels and hyperemesis gravidarum has not been firmly established. Patients with hyperemesis gravidarum also have abnormal thyroid function tests, and they may exhibit signs and symptoms of hyperthyroidism. Because hCG has intrinsic thyroid-stimulating activity and hyperthyroidism is a known cause of vomiting, these observations deserve further investigations. Mild abnormalities of liver chemistry tests are often present. The role of psychological derangements in the pathogenesis of hyperemesis gravidarum is controversial.

MANAGEMENT

The metabolic consequences of hyperemesis gravidarum are generally very severe, and if untreated, the mortality rate is high. Fluid and electrolyte disturbances, dehydration, and weight loss may be marked; and rare complications of intractable vomiting (such as Mallory–Weiss tear) may occur. The physician should exclude other causes of nausea and vomiting during pregnancy, such as preeclampsia (which usually occurs in later stages of pregnancy), urinary tract infection, appendicitis, hepatitis, intrahepatic cholestasis, gallstone disease and its complications, and acute fatty liver of pregnancy, which also typically occurs later (during the third trimester). Once the diagnosis of hyperemesis gravidarum is established, treatment consists of fluid and electrolyte replacement, parenteral hyperalimentation in cases complicated by nutritional deficiencies, supportive psychotherapy, and behavior modification.

Acknowledgment

Part of this article was previously published and adapted from Lee, M. and Feldman, M. (1998) Nausea and vomiting. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease," 6th Ed. by permission of W. B. Saunders.

See Also the Following Articles

Emesis • Hyperthyroidsm • Liver Disease, Pregnancy and • Mallory–Weiss Tear • Nausea • Pregnancy and Gastrointestinal Disease

Further Reading

del Mar Melero-Montes, M., and Jick, H. (2000). Hyperemesis gravidarum and the sex of the offspring. *Epidemiology* **12**, 123–124.

James, W. H. (2001). The associated offspring sex ratios and cause(s) of hyperemesis gravidarum. *Acta Obstet. Gynecol. Scand.* **80**, 378–379.

Lee, M., and Feldman, M. (1998). Nausea and vomiting. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., pp. 117–127. W. B. Saunders Company, Philadelphia.

Zimmermann, E. M., and Christman, G. M. (1999). Approach to gastrointestinal disease in the female patient. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., pp. 1059–1080. Lippincott Williams & Wilkins, Philadelphia.

 Copyright 2004, Elsevier (USA). All rights reserved.

Hyperlipidemia

MANISHA CHANDALIA AND NICOLA ABATE
University of Texas Southwestern Medical Center, Dallas

Dyslipidemia Abnormalities of plasma lipoproteins that include both higher and lower levels of certain lipoproteins.
Hyperlipidemia Elevated concentration of plasma lipoproteins.
Lipoproteins Particles carrying lipids in blood. Lipoproteins consist of esterified or unesterified cholesterol, triglycerides, phospholipids, and protein.
Lipoprotein X An abnormal lipoprotein that is rich in free cholesterol and phospholipids. Accumulation of lipoprotein X is noted in cholestatic liver disease. Its potential as an atherogenic particle is as yet unclear.
Lp (a) A modified form of LDL in which a large glycoprotein [apo (a)] is covalently bound to the apoB. Serum levels of Lp(a) are genetically defined. Its association with atherosclerosis is not completely established.

Hyperlipidemia refers to elevated blood levels of lipids, including cholesterol and triglycerides, due to various underlying causes. The clinically significant consequences of hyperlipidemia are two main life-threatening conditions: atherosclerosis and pancreatitis. Cholesterol and triglycerides are insoluble in plasma and circulating lipids are bound to apolipoproteins in order to be transported in plasma to various tissues where lipids are either utilized for cell function, energy utilization, steroid hormone production, and bile acid formation or deposited as a form of energy storage.

CLASSIFICATION OF LIPOPROTEINS

A variety of lipoproteins of different chemical composition can be identified in plasma. Based on their density, lipoproteins are commonly described as very-low-density lipoproteins (VLDLs), intermediate-density lipoproteins (IDLs), low-density lipoproteins (LDLs), high-density lipoproteins (HDLs), and chylomicrons. Table I is a classification of hyperlipidemias based on these discrete forms of lipoproteins.

Chylomicrons

Chylomicrons are very large particles that carry dietary lipid. They are associated with a variety of apolipoproteins, including A-I, A-II, A-IV, B-48, C-I, C-II, C-III, and E.

Very-Low-Density Lipoprotein

Very low-density lipoprotein carries endogenous triglycerides and to a lesser degree cholesterol. The major apolipoproteins associated with VLDL are B-100, C-I, C-II, C-III, and E.

Intermediate-Density Lipoprotein

Intermediate-density lipoprotein carries cholesterol esters and triglycerides. It is associated with apolipoproteins B-100, C-III, and E.

Low-Density Lipoprotein

Low-density lipoprotein carries cholesterol esters and is associated with apolipoprotein B-100.

High-Density Lipoprotein

High-density lipoprotein also carries cholesterol esters. It is associated with apolipoproteins A-I, A-II, C-I, C-II, C-III, D, and E.

Abnormalities of plasma levels of VLDL, IDL, LDL, HDL, or chylomicrons will lead to hypercholesterolemia, hypertriglyceridemia, or mixed hyperlipidemia depending on the lipid content of the specific lipoprotein abnormality. From a clinical standpoint, it is useful to evaluate hyperlipidemias/hyperlipoproteinemia on the basis of their potential effects, i.e., atherosclerosis and pancreatitis. Abnormalities in the metabolism of VLDL, IDL, and LDL [apolipoprotein B-100 (apoB-100)-containing lipoproteins] and HDL (apolipoprotein A-containing lipoproteins) are associated with risk of atherosclerosis. Abnormalities in chylomicron (apolipoprotein B-48-containing lipoproteins) metabolism are associated with risk of pancreatitis.

METABOLISM OF LIPOPROTEINS

Metabolism of Apolipoprotein B-100-Containing Lipoproteins

The metabolism of apoB-100-containing particles is depicted in Fig. 1. The initial apoB-100-containing

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Fredrickson Classification of Lipid Disorders

Phenotype	Lipoproteins increased	Lipid abnormality
I	Chylomicrons	↑↑Triglycerides to >99th percentile
IIa	LDL	↑↑Total cholesterol to >90th percentile
IIb	LDL, VLDL	↑Total cholesterol & triglycerides
III	VLDL remnants (IDL), chylomicrons	↑Total cholesterol & triglycerides
IV	VLDL	↑Triglycerides
V	VLDL, chylomicrons	↑↑Triglycerides to >99th percentile

lipoprotein is VLDL. These particles are synthesized by the hepatocyte. They contain a core of triglycerides (60% by mass) and cholesterol esters (20% by mass). In addition to apoB-100, the surface apolipoproteins for VLDL include apoC-II, which acts as a cofactor for lipoprotein lipase (LPL), apoC-III, which inhibits this enzyme, and apoE, which serves as ligand for the apolipoprotein B/E (LDL) receptor in the liver. Some newly synthesized VLDL particles are in fact taken up again by the liver through this mechanism, soon after secretion. The triglyceride core of nascent VLDL particles is hydrolyzed by LPL in the microcirculation. During lipolysis, the core of the VLDL particle is reduced, generating VLDL remnant particles (also called IDL). Some of the phospholipid, unesterified cholesterol, and apolipoproteins A, C, and E of VLDL remnants are transferred to HDL. VLDL remnants can either be cleared from the circulation by the apoB/E (LDL) receptor or be further depleted in triglycerides by lipases and become LDL particles. LDL particles contain a core of cholesterol esters and lesser amounts of triglyceride. LDL particles should be taken up by the LDL (apoB/E) receptor in the hepatocyte. However, if the LDL particle becomes modified LDL, i.e., glycated or oxidized, the scavenger LDL receptor of macrophages will recognize these particles and increase the formation rate of foam cells in the arterial intima, leading to the formation of atherosclerotic lesions. This process is physiologically counteracted by the apoA-containing lipoproteins (HDL).

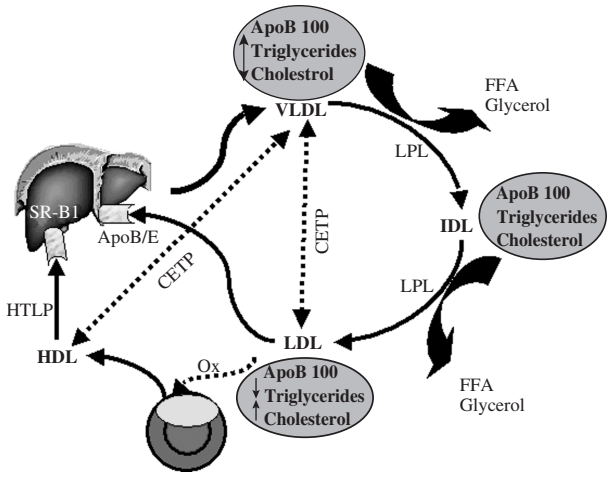


FIGURE 1 Metabolism of apolipoprotein B-100-containing lipoproteins: LPL, lipoprotein lipase; FFA, free fatty acid; Ox, oxidation; HTLP, hepatic lipase; SR-B1, scavenger receptor class B type 1; CETP, cholesterol ester transfer protein. See text for details.

Metabolism of Apolipoprotein A-Containing Lipoproteins

Figure 2 shows the metabolism of apoA-containing lipoproteins. The initial apoA-containing lipoprotein is the nascent HDL. This discoidal lipoprotein is synthesized in the gut and liver. It is rapidly hydrolyzed in the circulation unless enriched in cholesterol esters. Cholesterol ester uptake can occur from the macrophage/foam cell through a series of reactions that involve cholesterol efflux from the cytoplasm through the plasma membrane and rapid esterification of free cholesterol outside the cell. The transfer of cholesterol across the plasma membrane is mediated by the cholesterol efflux regulatory protein (CERP). Lecithin cholesterol acetyl transferase (LCAT) catalyzes the transfer of fatty acids from the phospholipid core of the nascent HDL to the free cholesterol coming out from the macrophage. Nascent HDL particles are therefore progressively enriched with cholesterol, becoming mature HDL particles. Mature HDL is recognized by scavenger receptor class B type 1 (SR-B1) receptor in the hepatocyte, adrenals, and gonads. HDL particles therefore mediate reverse cholesterol transport from the peripheral cells to the liver and to steroidogenic organs. This function seems largely to mediate the observed protective effects of HDL on atherosclerotic plaque formation. Other mechanisms have been also proposed, including antioxidant action of HDL to prevent LDL modification and recognition by the scavenger receptor of macrophages.

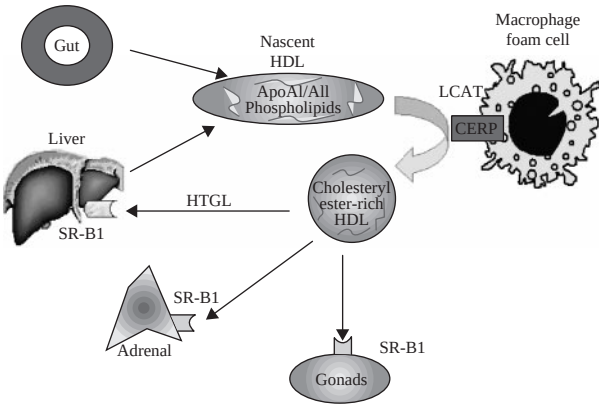


FIGURE 2 Metabolism of apoA-containing lipoproteins. LCAT, lecithin cholesterol acetyl transferase; CERP, cholesterol efflux regulatory protein; SR-B1, scavenger receptor class B type 1.

Metabolism of Apolipoprotein B-48-Containing Lipoproteins

The initial apoB-48-containing lipoproteins are chylomicrons. Figure 3 depicts schematically the metabolism of these particles. These particles are synthesized in the enterocytes and contain lipids derived from intestinal absorption. The main lipid component of chylomicrons is triglycerides. Dietary and biliary cholesterol is also incorporated into chylomicrons. These particles are exposed to LPL in the microcirculation and become progressively depleted of triglycerides. Triglycerides are utilized by the peripheral cells either as a direct source of energy or as an energy store in the adipose tissue. The result of triglyceride depletion is the formation of chylomicron remnants, which are eventually taken up by hepatocytes through the hepatic chylomicron remnant receptors. No chylomicrons should be detected under postabsorptive conditions.

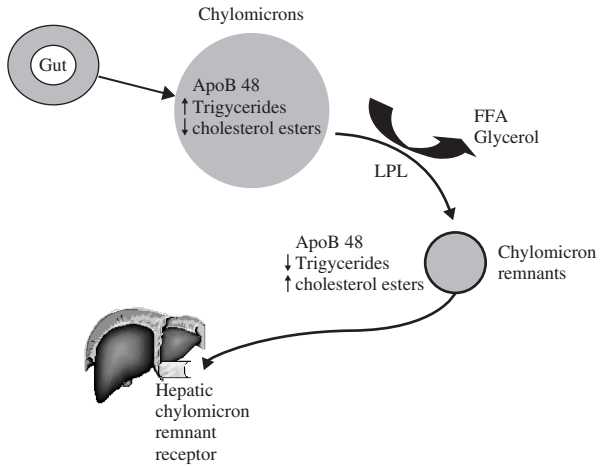


FIGURE 3 Metabolism of apoB 48-containing lipoproteins. LPL, lipoprotein lipase; FFA, free fatty acid.

PATHOPHYSIOLOGY OF HYPLIPOPROTEINEMIA

Apolipoprotein B-100-Containing Lipoproteins

Increased VLDL may occur from increased production by the hepatocyte and/or decreased clearance. Increased production of VLDL is observed in genetic conditions, such as familial combined hyperlipidemia. Secondary causes of increased production of VLDL are more common and include excessive intake of fats, calories, and alcohol, diabetes, insulin resistance, nephrotic syndrome, hypothyroidism, and obesity. Decreased clearance of VLDL is observed in genetic conditions, such as LPL deficiency and apoC-II deficiency. Secondary causes of decreased VLDL clearance include insulin resistance, type 2 diabetes, and hypothyroidism. Theseconditions decrease LPL activity either directly or through decreased production of apoC-II and increased production of apoC-III.

IDL typically increases in the presence of homozygosity for the isoform apoE-2. This isoform of apoE has the least affinity for the LDL receptor and decreases uptake of VLDL remnants/IDL. Usually, this hyperlipoproteinemia does not manifest unless a concomitant secondary cause, such as hypothyroidism, diabetes, or postmenopausal state, is present. This condition is clinically associated with palmar xanthomas and severe atherosclerosis.

LDL increases when defective clearance through the LDL receptor in the liver occurs. Defective receptor-mediated clearance of LDL is seen in familial hypercholesterolemia and polygenic hypercholesterolemias. Secondary causes of defective receptor-mediated clearance of LDL are excessive dietary intake of saturated fats and obesity. Increased LDL is the most powerful risk factor for coronary artery disease. It may also manifest with tendon xanthomas (in familial hypercholesterolemia) and xanthelasma.

Apolipoprotein A-Containing Lipoproteins

Low HDL may occur because of reduced production of HDL or increased catabolism through the SR-B1 receptor. However, low-HDL cholesterol is also seen in the condition of enhanced exchange of cholesterol between HDL and VLDL through the cholesterol ester transfer protein (CETP) when hypertriglyceridemia is present. As shown in Fig. 1, hypertriglyceridemia

through increased VLDL accelerates the transfer of cholesterol esters from HDL into VLDL in exchange for triglycerides. This explains the frequent association between hypertriglyceridemia and low-HDL cholesterol. Decreased production of HDL may occur because of a genetic mutation affecting CERP function (ABCA1 mutation in Tangier disease and familial low HDL) or a genetic mutation affecting LCAT activity.

Increased clearance of HDL particles may occur because of increased activity of hepatic lipase, an enzyme that mediates the maturation of HDL and facilitates the recognition of HDL by SR-B1 receptor. Increased activity of hepatic lipase is seen in diabetes and insulin resistance. Genetic polymorphisms responsible for decreased activity of HDL have been reported and correlate with increased levels of HDL.

Apolipoprotein B-48-Containing Lipoproteins

Increased chylomicron concentrations, leading to the presence of chylomicrons in postabsorptive conditions, may occur because of genetic conditions leading to defective triglyceride clearance, including LPL deficiency and apoC-II deficiency. Decreased clearance of chylomicrons may also occur secondary to excessive dietary triglyceride intake, hypothyroidism, obesity, diabetes, chronic renal failure, estrogens, and alcohol. Postabsorptive chylomicronemia causes eruptive xanthomas and acute pancreatitis.

HYPERCHOLESTEROLEMIA IN PRIMARY BILIARY CIRRHOSIS

Hypercholesterolemia is a common feature in primary biliary cirrhosis (PBC) and other forms of cholestatic liver disease. The lipid profile in PBC typically shows elevated total and LDL cholesterol and normal or mildly elevated triglycerides. HDL cholesterol may be elevated in the early stages of PBC and may fall markedly as the disease progresses to cirrhosis. Some of the excess LDL in PBC is composed of an abnormal lipoprotein particle called lipoprotein X, which is seen in cholestatic liver disease. Lipoprotein X is rich in free cholesterol and phospholipids. The total cholesterol concentration may be between 200 and 300 mg/dl; however, higher values have been reported in the literature. The lipid abnormalities may be identified during routine screening. However, classic presentation of patients with PBC is xanthomata and xanthelasma. Xanthomas correlate with cholesterol levels and develop on the elbow, palms, soles, buttocks, knees, back, and chest. The xanthomas are seen usually in patients with cholesterol levels >600 mg/dl. Due to their locations, xanthomas can cause significant difficulties in patients (Fig. 4). Lowering of cholesterol leads to an improvement in xanthomas. For patients with pain and disability due to xanthomas on the palm and sole, LDL aphaeresis may be required for relief. There is at present no conclusive evidence that hypercholesterolemia due to PBC is associated with increased atherogenesis and risk for coronary heart disease (CHD). A lipid profile, however, should be obtained in patients with PBC and decisions regarding treatment should be based on the risk for CHD. The efficacy of drug therapy is uncertain in PBC. However, statins and bile acid-binding resins have been tried in patients with variable results. LDL aphaeresis has been successfully used to treat symptomatic xanthomas on palms and xanthomatous neuropathy.

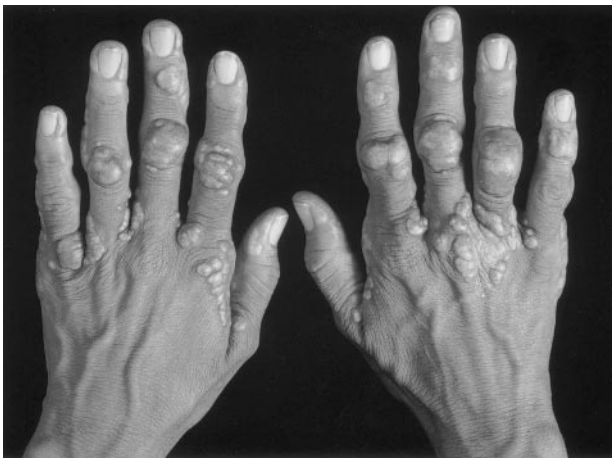


FIGURE 4 Xanthomata in a patient with primary biliary cirrhosis (PBC).

TREATMENT OF HYPERLIPIDEMIA

Coronary heart disease is a major cause of morbidity and mortality in the western world. Both high-LDL cholesterol and low-HDL cholesterol affect the risk of CHD. Positive risk factors for CHD include the following: (1) cigarette smoking; (2) hypertension [blood pressure $\geq 140/90$ mm Hg or on antihypertensive medications]; (3) low-HDL cholesterol (<40 mg/dl); (4) family history of premature CHD; and (5) age (men ≥ 45 years and women ≥ 55 years).

Recently, numerous clinical trials have demonstrated that lowering LDL cholesterol results in a significant reduction in CHD events and mortality. Adult Treatment Panel III (ATP III) guidelines outline identifying subjects who are at high risk of CHD and are candidates for treatment. Table II shows LDL cholesterol goals and cut-off points for intervention.

TABLE II LDL Cholesterol Goal and Cutpoint for Therapy

Risk category	LDL goal (mg/dl)	LDL level at which to initiate TLC (mg/dl)	LDL level at which to consider drug therapy (mg/dl)
CHD or CHD risk equivalents (10-year risk >20%)	<100	≥100	≥130 (100–129: drug therapy optional)
2 or more risk factors (10-year risk ≤20%)	<130	≥130	10-year risk 10–20%: ≥130 10-year risk <10%: ≥160
0–1 risk factor	<160	≥160	≥190 (160–189: drug therapy optional)

Note. Adapted from Adult Treatment Panel III at http://www.nhlbi.nih.gov, with permission. TLC, therapeutic life style changes; CHD, coronary heart disease; CHD risk equivalent: other forms of atherosclerotic diseases, diabetes, and multiple risk factors that confer a 10-year CHD risk > 20%.

Therapeutic Lifestyle Changes

Lifestyle modification in patients with hyperlipidemia forms a very important backbone of the treatment plan. Therapeutic lifestyle changes comprise diet, weight management, and increased physical activity. Table III shows the nutrient composition recommended by ATP III of the National Cholesterol Education Panel. Dietary modifications may lead to a reduction in LDL cholesterol of 8 to 15%. A low-fat diet is a very important component of treatment for hypertriglyceridemia. In addition to improving the lipid panel, therapeutic lifestyle changes can help with weight reduction, improving insulin resistance, and blood pressure. Other dietary components, such as fish oil, plant stenol incorporated in margarine, and soy protein, have been shown to improve lipid levels and their inclusion in the diet should be mentioned in the advice given regarding diet. Regular exercise, such as a brisk walk for 30 min, has been shown to improve cardiovascular fitness and should be incorporated in the treatment plan.

TABLE III Nutrient Composition of Therapeutic Lifestyle Changes Diet

Nutrient	Recommended intake
Saturated fat[a]	Less than 7% of total calories
Polyunsaturated fat	Up to 10% of total calories
Monounsaturated fat	Up to 20% of total calories
Total fat	25–35% of total calories
Carbohydrate[b]	50–60% of total calories
Fiber	20–30 g/day
Protein	Approximately 15% of total calories
Cholesterol	Less than 200 mg/day
Total calories (energy)	Balance energy intake and expenditure to maintain desirable body weight/prevent weight gain

Note. Adapted from Adult Treatment Panel III at http://www.nhlbi.nih.gov, with permission.

[a] Transfatty acid intake should be kept low.

[b] Carbohydrates should be derived from foods rich in complex carbohydrates including grains, fruits, and vegetables.

Pharmacotherapy of Hyperlipidemia

Statins

This class of drugs includes lovastatin, simvastatin, pravastatin, atorvastatin, and fluvastatin. The statins are competitive inhibitors of the rate-limiting enzyme in cholesterol synthesis, HMG coenzyme A (coA) reductase. By inhibiting HMG coA reductase, they decrease the intracellular cholesterol pool and this results in the up-regulation of the LDL receptor. This up-regulation of the LDL receptor promotes clearance of the circulating particles that have apoB or apoE on the surface and lead to increased clearance of cholesterol-rich LDL and to lesser extent VLDL and IDL. As a class, these drugs are most potent in reducing LDL cholesterol. In addition, they can reduce triglycerides and raise HDL cholesterol. Statins are efficacious in patients with isolated elevations in LDL cholesterol or combined hyperlipidemia with elevated LDL cholesterol and triglycerides. The reduction in LDL cholesterol with statins ranges from 24 to 60%. Numerous clinical trials have shown that statins are effective in lowering LDL cholesterol safely and lead to a significant reduction in CHD events and mortality. The predominant route of excretion is through the bile after hepatic transformation. Major side effects are liver dysfunction and myopathy. All statins have been associated with transaminase elevation, although the incidence of an elevation greater than three times the upper limit of normal is generally <2%. Statins can also induce myopathy, especially in combination with other drugs that are cleared through the cytochrome P450 system. Although significant myopathy is uncommon, definite rhabdomyolysis occurs in ~0.1% of patients who receive statin therapy.

Bile Acid Sequestrants

The drugs available in this class are cholestyramine, colestipol, and colesevelam. The primary action of bile acid sequestrants is to interfere with intestinal reabsorption of bile acids. This activity interrupts the enterohepatic circulation of bile acids, increases the fecal excretion of cholesterol, and depletes the cholesterol pool in the hepatocyte. The hepatocellular depletion of cholesterol stimulates the up-regulation of LDL receptors, which results in increased uptake of LDL and IDL, thus lowering LDL cholesterol. A maximum reduction of 15 to 30% in LDL cholesterol may be seen with these agents. A disadvantage of these drugs is that cholesterol depletion in the liver results in increased cholesterol synthesis, which partially negates the cholesterol-lowering efficacy of these drugs. However, statin can be combined with these drugs to prevent this effect and can be used in severe hypercholesterolemias as a combination therapy. One other disadvantage of bile acid sequestrants is exacerbation of hypertriglyceridemia due to increased hepatic VLDL production. Bile acid sequestrants are contraindicated in patients who have serum triglyceride concentrations >400 mg/dl. The major side effects are usually limited to the gastrointestinal tract as bile acid sequestrants are not absorbed systemically. Common side effects are abdominal bloating, belching, flatulence, and constipation. Bile acid resins may bind to other medications, causing decreased absorption and efficacy of these medications. Therefore, other medications should be taken either 1 h before or 4 h after bile acid resins.

Fibrates

This class of drugs includes gemfibrozil, fenofibrate, clofibrate, bezafibrate, etofibrate, and ciprofibrate. The first three are currently available in the United States. However, since the use of clofibrate has been associated with an increased incidence of cholangiocarcinoma and other gastrointestinal cancers, only gemfibrozil and fenofibrate are routinely used for the treatment of dyslipidemia. Gemfibrozil and fenofibrate markedly lower plasma triglyceride concentrations (up to 50%), increase plasma HDL cholesterol concentrations (up to 25%), and have a variable effect on plasma LDL-C concentrations. Recently, it has become evident that many of the effects of fibrates on lipids are mediated by the activation of peroxisome proliferator-activated receptor-γ (PPAR-γ, a component of the subfamily of nuclear receptors known as PPARs). Fibrates bind to and activate PPAR-γ, which then forms a heterodimer with another nuclear receptor, the retinoid X receptor. The heterodimer binds to a specific peroxisome proliferator-response element and the result is a change in the transcription of target genes. Target genes include lipoprotein lipase, apoC-III, apoA-I, and apoA-II. The expression of lipoprotein lipase, apoA-I, and apoA-II has been found to increase with fibrates, whereas the expression of apoC-III has been found to decrease. As a result, fibrates induce the following: (1) an increase in lipoprotein lipase activity (resulting from PPAR-γ activation and a related increase in the production of LPL and a decrease in the production of apoC-III, an inhibitor of LPL) and (2) an increased production of HDL (resulting from PPAR-γ activation and a related increase in the production of apoA-I and apoA-II). The stimulation of LPL activity has the effect of increasing the catabolism of triglycerides in lipoproteins, such as VLDL and chylomicrons. Accelerated catabolism of VLDL and chylomicron to IDL and chylomicron remnants, respectively, induces decreased plasma triglyceride concentrations and may also lead to a higher production of LDL particles. If there is no adequate compensatory mechanism to increase the catabolism of LDL particles, the accelerated conversion of VLDL to LDL may result in the elevation of serum LDL-C concentrations. Fibrates, particularly fenofibrate, also have been shown to reduce VLDL production in the liver. This mechanism contributes to a decrease in plasma triglyceride concentrations and at the same time explains the LDL-C-lowering effect seen in some patients treated with fenofibrate. Reduction of triglyceride-containing particles also has the effect of decreasing the exchange of triglycerides for cholesterol esters between VLDL and LDL or HDL, a process physiologically mediated by CETP. This mechanism contributes to an increase in HDL-C levels and to a change in the quality of LDL from small–dense LDL to less atherogenic, large–buoyant LDL-C in patients treated with fibrates. Increased serum concentrations of HDL-C are also mediated by the increased synthesis of apoA-I and apoA-II, and hence HDL particle production. Therefore, both the quality and the number of HDL particles change with the use of fibrates. The effects of fibric acid derivatives on lipoprotein levels differ widely according to the starting lipoprotein profile, the presence or absence of a genetic hyperlipoproteinemia, and the associated environmental influences. Among all patients, hyperlipidemic subjects with apoE-2/E-2 monozygosity respond best to gemfibrozil therapy. Elevated triglyceride and cholesterol levels may be dramatically lowered and tuberoeruptive and palmar xanthomas may resolve completely. Gemfibrozil treatment in patients with mild hypertriglyceridemia [triglycerides (TG) <400 mg/dl] usually produces a decrease in triglyceride levels of 50% or more, an increase in HDL-C of 15 to 25%, and either no change or

an increase in LDL-C levels, particularly in subjects with familial combined hyperlipidemia. Fenofibrate, bezafibrate, and ciprofibrate lower VLDL levels to a degree similar to that produced by gemfibrozil, but they also decrease LDL-C levels by 15 to 20%. In patients with more marked hypertriglyceridemia (TG of 400 to 1000 mg/dl), a similar fall in triglycerides occurs. However, these patients frequently experience an increase in LDL-C of 10 to 30%. In contrast, treatment of patients with heterozygous familial hypercholesterolemia usually produces a decrease in LDL-C levels of 10% with gemfibrozil and a decrease of 20 to 30% with fenofibrate. The effect of fibrates on lipoproteins has been associated with a reduction in cardiovascular risk in both primary and secondary intervention trials.

Nicotinic Acid

Nicotinic acid is a water-soluble B-complex vitamin that has a significant hypolipidemic activity at high doses (1−3 g/day). The hypolipidemic properties of nicotinic acid are unrelated to its role as a vitamin. Nicotinic acid is used to increase serum HDL-C levels (up to 30%), reduce triglycerides (20 to 80%), and reduce LDL-C (10 to 15%). Nicotinic acid is the only hypolipidemic agent that is effective in reducing lipoprotein-a concentrations (by approximately 30%). It appears that the hypolipidemic properties of nicotinic acid are related to inhibition of peripheral lipolysis. This effect reduces free fatty acid mobilization from peripheral tissues and reduces free fatty acid delivery to the liver. As a result, triglyceride synthesis in hepatocytes is diminished. Since hepatic triglyceride synthesis is an essential step for the synthesis and transport of VLDL particles from the hepatocyte into the systemic circulation, the inhibition of lipolysis in the peripheral cells by nicotinic acid induces a reduction in the assembly and secretion of VLDL particles into the bloodstream. Reduced VLDL production explains the decreased plasma triglyceride concentrations in patients treated with nicotinic acid. Since LDL particles derive from VLDL metabolism, another effect of nicotinic acid-induced inhibition of VLDL output is a reduction of serum LDL-C concentrations. Similarly to what was described for fibrates above, the decreased VLDL synthesis induces a reduction in CETP activity, which, in turn, reduces the loss of cholesterol esters from the core of LDL particles and changes the quality of LDL particles from small−dense LDL to large−buoyant LDL and less atherogenic LDL particles. The same effects on CETP activity will contribute to increased HDL-C levels in patients treated with nicotinic acid. Another mechanism that may contribute to the significant effect of nicotinic acid on HDL-C is the reduced clearance of HDL particles and increased synthesis of apoA-II. The mechanisms underlying these effects of nicotinic acid are not completely understood. The numerous clinical trials using nicotinic acid in monotherapy or in combination therapy have confirmed the effectiveness of this drug in decreasing cardiovascular risk in patients with high triglycerides, low HDL-C levels, and elevated LDL-C. Nicotinic acid causes numerous side effects, the major one being flushing and associated pruritus that usually involve the face and upper part of the body. The interval between the intake of nicotinic acid and the onset of these side effects may be variable. With persistent use of the drug, the frequency and intensity of flushing decrease in 70 to 80% of patients within 1 to 2 weeks. Taking one aspirin per day may alleviate the symptoms in some patients, suggesting that this effect is mediated by prostaglandins. Gastrointestinal symptoms, such as dyspepsia, vomiting, and diarrhea, also occur, but these effects are reduced if nicotinic acid is taken with a meal. Nicotinic acid may precipitate peptic ulcer disease. Abnormalities of hepatic function are an important and potentially serious complication of high-dose nicotinic acid therapy. These abnormalities generally occur in patients taking more than 2 g of nicotinic acid per day. Increases in serum transaminases are the most common abnormalities in hepatic function, but jaundice may occur and hepatic failure has been observed with sustained-release preparations. With mild changes in liver function tests, reducing the dose of nicotinic acid frequently will result in normalization of tests without the need to discontinue the drug. The combined use of nicotinic acid and statins may increase the risk of myopathy and rhabdomyolysis. Finally, nicotinic acid can cause hyperglycemia and, if used in diabetics with dyslipidemia, requires careful monitoring of blood glucose levels.

Cholesterol Absorption Inhibitors

Sitostanol/sitosterol and ezetimibe interfere with cholesterol absorption at the brush border of enterocytes. Sitostanol (Benechol) and sitosterol (Take Control) used with the meals displace cholesterol from the micellar solution required for transport of cholesterol across the plasma membrane of enterocytes. Sitosterol and sitostanol are reexcreted from the enterocyte through the ABCG5/G8 cholesterol transporter protein and not absorbed in significant quantities. Ezitimibe has been recently developed as the first agent acting as cholesterol transport blocker. It recirculates through the entero-hepatic circulation and blocks a putative cholesterol transporter located at the brush border of the enterocyte plasma membrane. Absorption studies

have shown decreased cholesterol absorption and increased synthesis as a response of the hepatocytes to negative cholesterol balance. Ezetimibe promises to be a safe adjunct to other hypocholesterolemic agents to further decrease LDL-C in patients at high risk for coronary artery disease. It also seems to be a promising alternative for patients who cannot tolerate statins or other hypocholesterolemic agents.

See Also the Following Articles

Apoproteins • Barrier Function in Lipid Absorption • Cholestatic Diseases, Chronic • Cholesterol Absorption • Lipoproteins

Further Reading

Clinical Guidelines on Cholesterol Management in Adults (ATP III) (2001). Full text of the guidelines available at http://www.nhlbi.nih.gov.

Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (2001). *J. Am. Med. Assoc.* **285**, 2486–2497.

Ginsberg, H., and Goldberg, I. J. (2001). Disorders of lipoprotein metabolism. *In* "Harrison's Principles of Internal Medicine" (E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and L. Jameson, eds.), 15th Ed., pp. 2245–2256. McGraw-Hill, New York.

Rose, B. D. (2002). Hyperlipidemia. Available at www.uptodate.com.

Hyperparathyroidism

ASRA KERMANI

Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

carcinoid An endocrine tumor consisting of enterochromaffin cells or enterochromaffin-like cells.

MEN A syndrome associated with multiple endocrine neoplasms; three types are recognized (1, 2a, and 2b) and types 1 and 2a are associated with hyperparathyroidism.

primary hyperparathyroidism Inappropriately excessive parathyroid hormone production resulting in hypercalcemia.

Routine chemistry tests have made primary hyperparathyroidism the most common cause of hypercalcemia detected in the outpatient setting. Prior to the widespread availability of laboratory analysis, hyperparathyroidism was an uncommon disease that usually was discovered when a patient presented with a clinical manifestation. It has now predominantly become an asymptomatic diagnosis. Primary hyperparathyroidism is a diagnosis commonly seen in adults of all ages but is uncommon in children. The female:male ratio is 3 : 1 with the peak incidence between the third and sixth decades. The most common cause is a solitary parathyroid adenoma, accounting for approximately 80% of all causes. Multiple parathyroid adenomas constitute the next most frequent diagnosis; this entity is commonly seen in the multiple endocrine neoplasia syndromes (MEN), both type 1 and type 2A. Older literature suggested that the parathyroid glands in MEN syndromes were "hyperplastic" but many are now known to be monoclonal in origin. Very rarely seen as a cause of hyperparathyroidism is parathyroid carcinoma (<1%).

GASTROINTESTINAL SYMPTOMS AND SIGNS

Most of the symptoms of hyperparathyroidism, if any are present, pertain to the degree of hypercalcemia and the rate of increase of serum calcium levels. Gastrointestinal symptoms occur in approximately 35% of patients. These symptoms include anorexia, nausea, vomiting, heartburn, constipation, weight loss, and dyspepsia. Constipation occurs in up to 30% of patients with primary hyperparathyroidism. This is thought to occur as a result of decreased smooth muscle tone of the gut. It is believed that this is not a myopathy but actually a result of abnormal activity of the autonomic

 Copyright 2004, Elsevier (USA). All rights reserved.

have shown decreased cholesterol absorption and increased synthesis as a response of the hepatocytes to negative cholesterol balance. Ezetimibe promises to be a safe adjunct to other hypocholesterolemic agents to further decrease LDL-C in patients at high risk for coronary artery disease. It also seems to be a promising alternative for patients who cannot tolerate statins or other hypocholesterolemic agents.

See Also the Following Articles

Apoproteins • Barrier Function in Lipid Absorption • Cholestatic Diseases, Chronic • Cholesterol Absorption • Lipoproteins

Further Reading

Clinical Guidelines on Cholesterol Management in Adults (ATP III) (2001). Full text of the guidelines available at http://www.nhlbi.nih.gov.

Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (2001). *J. Am. Med. Assoc.* **285**, 2486–2497.

Ginsberg, H., and Goldberg, I. J. (2001). Disorders of lipoprotein metabolism. *In* "Harrison's Principles of Internal Medicine" (E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and L. Jameson, eds.), 15th Ed., pp. 2245–2256. McGraw-Hill, New York.

Rose, B. D. (2002). Hyperlipidemia. Available at www.uptodate.com.

Hyperparathyroidism

ASRA KERMANI

Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

carcinoid An endocrine tumor consisting of enterochromaffin cells or enterochromaffin-like cells.

MEN A syndrome associated with multiple endocrine neoplasms; three types are recognized (1, 2a, and 2b) and types 1 and 2a are associated with hyperparathyroidism.

primary hyperparathyroidism Inappropriately excessive parathyroid hormone production resulting in hypercalcemia.

Routine chemistry tests have made primary hyperparathyroidism the most common cause of hypercalcemia detected in the outpatient setting. Prior to the widespread availability of laboratory analysis, hyperparathyroidism was an uncommon disease that usually was discovered when a patient presented with a clinical manifestation. It has now predominantly become an asymptomatic diagnosis. Primary hyperparathyroidism is a diagnosis commonly seen in adults of all ages but is uncommon in children. The female:male ratio is 3 : 1 with the peak incidence between the third and sixth decades. The most common cause is a solitary parathyroid adenoma, accounting for approximately 80% of all causes. Multiple parathyroid adenomas constitute the next most frequent diagnosis; this entity is commonly seen in the multiple endocrine neoplasia syndromes (MEN), both type 1 and type 2A. Older literature suggested that the parathyroid glands in MEN syndromes were "hyperplastic" but many are now known to be monoclonal in origin. Very rarely seen as a cause of hyperparathyroidism is parathyroid carcinoma (<1%).

GASTROINTESTINAL SYMPTOMS AND SIGNS

Most of the symptoms of hyperparathyroidism, if any are present, pertain to the degree of hypercalcemia and the rate of increase of serum calcium levels. Gastrointestinal symptoms occur in approximately 35% of patients. These symptoms include anorexia, nausea, vomiting, heartburn, constipation, weight loss, and dyspepsia. Constipation occurs in up to 30% of patients with primary hyperparathyroidism. This is thought to occur as a result of decreased smooth muscle tone of the gut. It is believed that this is not a myopathy but actually a result of abnormal activity of the autonomic

 Copyright 2004, Elsevier (USA). All rights reserved.

nervous system. Constipation can also be potentiated by renal free water loss from hypercalcemia, which leads to dehydration.

Approximately 60% of patients with primary hyperparathyroidism suffer from heartburn secondary to gastroesophageal reflux. Reduced lower esophageal sphincter pressure has been documented in hyperparathyroid patients with hypercalcemia. Gastroesophageal reflux normalizes after successful parathyroidectomy and therefore might be related to hypercalcemia.

A rare but potentially life-threatening complication is acute primary hyperparathyroidism, an entity characterized by marked hypercalcemia, exceeding 15–17 mg/dl. Patients complain of severe abdominal pain, nausea, vomiting, and volume depletion and symptoms may progress to altered mentation, coma, and seizures. Peptic ulceration, occasionally with perforation, and pancreatitis occur commonly in this situation. Emergent parathyroidectomy, along with supportive care, is curative.

ASSOCIATIONS WITH GASTROINTESTINAL DISEASES

Pancreatitis

Primary hyperparathyroidism is associated with acute or chronic pancreatitis in approximately 3% of patients who underwent surgery for hyperparathyroidism. The incidence of pancreatitis was related to high serum calcium levels. Patients undergoing an operation for secondary hyperparathyroidism, who therefore were hypocalcemic or normocalcemic, did not have pancreatitis, suggesting that hypercalcemia is an important mediator of pancreatic injury in primary hyperparathyroidism.

There are several reports of primary hyperparathyroidism diagnosed in the setting of acute pancreatitis. One hypothesis for this occurrence is that hypercalcemia results in calcium deposition within the pancreatic duct, resulting in obstruction. Activation of trypsinogen by calcium also occurs in acute pancreatitis. In severe pancreatitis of any etiology, extensive formation of calcium–fatty acid soaps might render the patient hypocalcemic. Therefore, normocalcemia in the setting of acute pancreatitis may be a clue to the presence of hyperparathyroidism.

Peptic Ulceration

Whereas the association between peptic ulcer disease and hyperparathyroidism is clearly present in multiple endocrine neoplasia (MEN 1) syndrome discussed below, the case is not as clear in sporadic primary hyperparathyroidism. Peptic ulceration has been reported in approximately 15–20% of patients with primary hyperparathyroidism. This prevalence is only slightly greater than that in the general population. Although such increased reports of primary hyperparathyroidism in patients with documented peptic ulcer may be causal, these findings may reflect an increased search for it. The two conditions are highly prevalent in unselected individuals and therefore might occur coincidently in some. One hypothesis that attempts to explain peptic ulceration as a consequence of primary hyperparathyroidism relates to hypergastrinemia. Elevated circulating calcium levels stimulate gastrin secretion in humans. The best clinical evidence is the MEN 1 syndrome, discussed below.

Some reports indicated an increase in basal acid output in up to one-third of patients with primary hyperparathyroidism but there was significant overlap with normals. In another study, patients with primary hyperparathyroidism were studied after successful parathyroid adenoma removal. There was no change in the mean fasting gastrin level, gastrin response to feeding, mean basal acid output, or mean peak acid output.

Multiple Endocrine Neoplasia Syndromes

MEN 1 is an autosomal dominant disorder characterized by multiple parathyroid adenomas in 90–100% of cases, anterior pituitary adenoma in 10–60%, and entero-pancreatic endocrine tumors such as gastrinoma in 30–75%. The syndrome is difficult to define, as there may be permutations of 20 or more tumors. Therefore, an individual with two of the three above criteria may be defined as having MEN 1. The most frequent endocrine problem in MEN 1 is primary hyperparathyroidism, occurring in almost 100% of patients by age 50. The age of onset may be as early as in the twenties, which is 30 years earlier than sporadic primary hyperparathyroidism. Three or four parathyroid glands are frequently involved by tumor in MEN 1.

Gastrinomas occur in approximately 40% of patients with MEN 1 and are associated with Zollinger-Ellison syndrome (ZES). ZES results in widespread peptic ulceration due to hypergastrinemia and resulting excess acid production. Peptic ulceration in MEN 1 is potentiated by hypercalcemia from primary hyperparathyroidism, leading in turn to increased gastrin release from the gastrinoma. The converse is also true: removal of the parathyroids lowers gastrin production from gastrinoma. Most gastrinomas associated with MEN 1 have malignant potential and over half of them are metastatic at the time of diagnosis.

MEN 2A is also an autosomal dominant disorder resulting in medullary carcinoma of the thyroid in 90% of cases, unilateral or bilateral pheochromocytoma in 50%, and multiple parathyroid tumors in 20–30%. Unlike the aggressive course of patients with primary hyperparathyroidism in MEN 1, the hyperparathyroidism of MEN 2A patients is mild and often asymptomatic. Usually fewer than four parathyroid glands are affected in this condition.

Carcinoid

Foregut carcinoids may occur as part of MEN 1 syndrome. Therefore, if a diagnosis of carcinoid is made, patients should be screened for primary hyperparathyroidism.

Rode *et al.* found gastric carcinoid in three cases of primary hyperparathyroidism. Two of these had parathyroid hyperplasia and one had a parathyroid adenoma. All carcinoids were located in the nonantral part of the stomach. Interestingly, all three carcinoids showed immunoreactivity for gastrin. None of these patients had other features of MEN.

Gallstones

The prevalence of gallstones in patients with primary hyperparathyroidism ranges from 25 to 35%. It is unclear whether there is a real association.

See Also the Following Articles

Duodenal Ulcer • Gastric Ulcer • Gastrinoma • Multiple Endocrine Neoplasia (MEN) • Pancreatitis, Acute

Further Reading

Attie, J. N. (1997). Primary hyperparathyroidism. *In* "Current Therapy in Endocrinology and Metabolism," 6th Ed., pp. 557–564. Mosby-Year Book, Inc., St. Louis.

Brandi, M. L., Gagel, R. F., Angeli, A., *et al.* (2001). Consensus guidelines for diagnosis and therapy of MEN type 1 and type 2. *J. Clin. Endocrinol. Metab.* **86**, 5658–5671.

Carnaille, B., Oudar, C., Pattou, F., *et al.* (1998). Pancreatitis and primary hyperparathyroidism: Forty cases. *Aust. N. Z. J. Surg.* **68**, 117–119.

Corleto, V. D., Minisola, S., Moretti, A., *et al.* (1999). Prevalence and causes of hypergastrinemia in primary hyperparathyroidism: A prospective study. *J. Clin. Endocrinol. Metab.* **84**, 4554–4558.

Gurbuz, A. T., and Peetz, M. E. (1996). Giant mediastinal parathyroid cyst: An unusual cause of hypercalcemic crisis–Case report and review of the literature. *Surgery* **120**, 795–800.

Mowschenson, P. M., Rosenberg, S., and Pallotta, J. (1982). Effect of hyperparathyroidism and hypercalcemia on lower esophageal sphincter pressure. *Am. J. Surg.* **143**, 36–39.

Paloyan, D., Simonowitz, D., Paloyan, E., *et al.* (1982). Pancreatitis associated with primary hyperparathyroidism. *Am. Surg.* **48**, 366–368.

Rode, J., Dhillon, A. P., Cotton, P. B., *et al.* (1987). Carcinoid tumour of stomach and primary hyperparathyroidism: A new association. *J. Clin. Pathol.* **40**, 546–551.

Wise, S. R., Quigley, M., Saxe, A. W., *et al.* (1990). Hyperparathyroidism and cellular mechanisms of gastric acid secretion. *Surgery* **108**, 1058–1063.

Hyperthyroidism

ASRA KERMANI
Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

hyperthyroidism (**thyrotoxicosis**) Variety of disorders resulting in excessive thyroid hormone levels and effects throughout the body.

tachygastria Faster than the normal (three/minute) electrical slow-wave activity generated by the smooth muscle layer of the stomach.

The most common cause of thyrotoxicosis is Graves' disease, an autoimmune disorder characterized by thyroid-stimulating antibodies that bind to the thyroid-stimulating hormone receptor on the thyroid gland, resulting in overproduction of thyroid hormone. Graves' disease may occur at any age but it is most frequently seen in the third and fourth decades; there is gender prevalence, with a female:male ratio ranging from 7 : 1 to 10 : 1. Histologically, there is predominantly T lymphocytic infiltration of the thyroid gland. It is hypothesized that stimulated T lymphocytes induce B lymphocytes to produce antibodies to the thyroid-stimulating hormone (TSH) receptor. Because of a variety of human leukocyte antigen associations, several other autoimmune diseases are associated with Graves' disease. Toxic multinodular goiter is a frequent cause of hyperthyroidism in the elderly, usually appearing late in the course of a nontoxic multinodular goiter. Other causes of thyrotoxicosis are adenoma, various forms of transient thyroiditis, exogenous thyroid hormone ingestion, and iodine ingestion in populations with endemic iodine deficiency. Rare causes include pituitary tumor secretion of thyroid-stimulating hormone, pituitary resistance to thyroid hormone, and trophoblastic disease (human chorionic gonadotropin and thyroid-stimulating hormone share the same α subunit).

GASTROINTESTINAL SYMPTOMS AND SIGNS

Constitutional Symptoms and Signs

The classical gastrointestinal manifestation of hyperthyroidism is weight loss in spite of preserved or increased appetite. Although some patients with mild disease might actually gain weight if their caloric intake is in excess of their metabolism, the classic patient with severe thyrotoxicosis is unable to keep up with the increased catabolism by increased food intake. Reduced intestinal absorption of nutrients also contributes to weight loss. Apathetic hyperthyroidism is an entity characterized by atypical manifestations, particularly in the elderly, and anorexia is a common presenting feature in this population.

Motility Disturbances

Hyperdefecation is common, and is defined as greater than two bowel movements a day. Diarrhea is rare. Patients with chronic constipation who become thyrotoxic may develop normal bowel movements. Acute or vague chronic abdominal discomfort may supervene in some patients, and hyperthyroidism is a rare cause of "acute abdomen." The cause for the abdominal pain is unknown, but corrects with treatment of hyperthyroidism. Dysphagia as a result of esophageal motor dysfunction has been also been reported.

Hepatic Symptoms and Signs

Pruritus and jaundice may occur as a result of cholestatic hepatitis in Graves' disease in patients who have no evidence of autoimmune hepatitis or congestive heart failure. This entity usually resolves completely with treatment of thyrotoxicosis. Hepatomegaly and splenomegaly may also be manifestations of Graves' disease.

Malabsorption

D-Xylose absorption by the intestine is normal in patients with hyperthyroidism. Lactose malabsorption may occur in Graves' disease and may sometimes correct following successful treatment. Steatorrhea occurs in more than 25% of hyperthyroid patients and can contribute to weight loss. Reasons include rapid intestinal transit, leading to reduced contact time with the small bowel mucosal surface and poor absorption, and possibly ingestion of large amounts of fat. Minor abnormalities in pancreatic and biliary secretion may be present but they are insufficient to cause steatorrhea. Jejunal calcium absorption can also be reduced in

 Copyright 2004, Elsevier (USA). All rights reserved.

thyrotoxicosis and can contribute to weight loss and a negative calcium balance, promoting osteoporosis.

Gastric Emptying

Hyperthyroidism does not significantly affect gastric emptying of solids. Pfaffenbach *et al*. studied 23 untreated hyperthyroid patients who also had dyspepsia. Dyspepsia included retrosternal pain, epigastric pain, epigastric fullness, belching, nausea and vomiting, and/or abdominal pain unexplained by other known causes. Electrogastrography (EGG) showed tachygastria and delayed gastric emptying. EGG abnormalities did not correlate significantly with dyspeptic symptoms, but tachygastria improved with treatment of hyperthyroidism.

Orocecal Transit Time

Reports find that mouth-to-cecal stool transit in thyrotoxicosis is accelerated by 50%. Studies uniformly find that hyperthyroid patients have more rapid orocecal time, which normalizes after treatment of hyperthyroidism. This increased bowel motility may explain diarrhea in thyrotoxic patients.

Anorectal Physiology

Hyperthyroid patients exhibit lower maximum resting pressure and maximum squeeze pressure on anal manometry, with no difference in maximum tolerable rectal volumes. They also have a significantly lower threshold sensation for impending evacuation compared with controls, which could conceivably contribute to hyperdefecation.

ASSOCIATIONS WITH GASTROINTESTINAL DISEASES

Immunologic Associations

Hyperthyroidism may be associated with celiac disease, hepatitis C virus, and inflammatory bowel disease (IBD). Autoimmune thyroid diseases, including Graves' disease and Hashimoto's thyroiditis, are seen more frequently in patients with celiac disease. Female patients with hepatitis C virus infection have been reported to have an increased prevalence of thyroid microsomal antibody titers. Titers may increase with time with development of hyperthyroidism. Nonspecific IBD has been observed to be associated in various ways with Graves' disease.

Liver Dysfunction

Nonspecific Abnormalities

Severe thyrotoxicosis, including thyroid storm, is frequently associated with nonspecific liver function test abnormalities. The serum total protein level may be low, and there may be elevations in alanine aminotransferase (ALT) and alkaline phosphatase (most of the latter, however, arises from bone). Finally, in severe thyrotoxicosis, hypoxemia has been invoked as a mechanism for hepatic dysfunction, because of unaltered splanchnic oxygen delivery in the face of increased consumption, but this hypothesis is controversial.

Cholestatic Hepatitis

A particular type of cholestatic hepatitis with marked hyperbilirubinemia and mild to moderate transaminase elevation has been described. This is characterized by liver biopsy findings of intrahepatic steatosis, portal infiltration by lymphocytes, and Kupffer cell hyperplasia. There is no biliary ductal dilatation. A different histological picture but identical clinical phenotype is seen with granulomatous hepatitis. Liver biopsy shows intralobular granulomas and a mixed cholestatic and cytolytic hepatitis, in the absence of other causes, including primary biliary cirrhosis and sarcoidosis.

Autoimmune Hepatitis

Autoimmune hepatitis and Graves' disease may also coexist, with histological findings showing mononuclear cell infiltration and portal fibrosis. High globulin fraction may be seen in autoimmune and/or chronic viral hepatitis.

Heart Failure

Right-sided heart failure is a common cause of hepatomegaly in Graves' disease, resulting from chronic passive congestion of the liver.

Drug Toxicity

Drug treatment of hyperthyroidism is an important cause of hepatotoxicity. In most cases, asymptomatic elevations of transaminases are the abnormality. The two most common agents are propylthiouracil and methimazole. Methimazole can result in cholestatic hepatitis with jaundice, often in the first few months of treatment. Gradual resolution usually occurs with discontinuation of this drug. Propylthiouracil, however, may present a wider clinical spectrum of hepatic damage, including a viral hepatitis-like picture, cholestatic hepatitis, or fulminant hepatic necrosis. The mechanism

appears to be a hypersensitivity reaction, which can occur either acutely or as late as 14 months after initiation of therapy. Fulminant hepatic failure is an exceedingly rare but potentially fatal outcome related to propylthiouracil use, requiring orthotopic liver transplantation.

THYROID FUNCTION TESTING PRIOR TO TREATMENT WITH INTERFERON α FOR CHRONIC HEPATITIS C INFECTION

Treatment with interferon α (IFNα) for chronic hepatitis C infection results in thyroid dysfunction in 3–14% of patients. Consequences include thyroiditis, hyperthyroidism, or hypothyroidism; many patients develop antithyroid antibodies. Predictors for thyroid dysfunction are female gender, subclinical thyroid disease prior to interferon administration, previously manifest thyroid disease, or other autoimmune disease. All patients should have thyroid function testing prior to initiation of IFNα therapy and be periodically monitored during the course of therapy. Preexisting thyroid disease is not a contraindication to interferon therapy, however.

See Also the Following Articles

Celiac Disease • Colitis, Ulcerative • Crohn's Disease • Hepatitis C • Hepatotoxicity, Drug-Induced

Further Reading

Branski, D., Levy, J., Globus, M., *et al.* (1984). Dysphagia as a primary manifestation of hyperthyroidism. *J. Clin. Gastroenterol.* **6**(5), 437–440.

Cathebras, P. J., Mosnier, J. F., Gouilloud, S., *et al.* (1995). Hepatic granulomatosis in a patient with Graves' disease. *Eur. J. Gastroenterol. Hepatol.* **7**(9), 905–908.

Deen, K. I., Seneviratne, S. L., and de Silva, H. J. (1999). Anorectal physiology and transit in patients with disorders of thyroid metabolism. *J. Gastroenterol. Hepatol.* **14**(4), 384–387.

Fong, T. L., McHutchison, J. G., and Reynolds, T. B. (1992). Hyperthyroidism and hepatic dysfunction. A case series analysis. *J. Clin. Gastroenterol.* **14**(3), 240–244.

Gurlek, A., Cobankara, V., and Bayraktar, M. (1997). Liver tests in hyperthyroidism: Effect of antithyroid therapy. *J. Clin. Gastroenterol.* **24**(3), 180–183.

Huang, M. J., and Liaw, Y. F. (1995). Clinical associations between thyroid and liver diseases. *J. Gastroenterol. Hepatol.* **10**(3), 344–350.

Huang, M.-J., Tsai, S.-L., Huang, B-.Y., *et al.* (1999). Prevalence and significance of thyroid autoantibodies in patients with chronic hepatitis C virus infection: A prospective controlled study. *Clin. Endocrinol. (Oxf.)* **50**(4), 503–509.

Jonderko, K., Jonderko, G., Marcisz, C., *et al.* (1997). Gastric emptying in hyperthyroidism. *Am. J. Gastroenterol.* **92**(5), 835–838.

Marazuela, M., Garcia-Buey, L., Gonzalez-Fernandez, B., *et al.* (1996). Thyroid autoimmune disorders in patients with chronic hepatitis C before and during interferon-alpha therapy. *Clin. Endocrinol. (Oxf.)* **44**(6), 635–642.

Pfaffenbach, B., Adamek, R. J., Hagelmann, D., *et al.* (1997). Effect of hyperthyroidism on antral myoelectrical activity, gastric emptying and dyspepsia in man. *Hepatogastroenterology* **44**(17), 1500–1508.

Ruiz, J. K., Rossi, G. V., Vallejos, H. A., *et al.* (2003). Fulminant hepatic failure associated with propylthiouracil. *Ann. Pharmacother.* **37**(2), 224–228.

Tobin, M. V., Fisken, R. A., Diggory, R. T., *et al.* (1989). Orocaecal transit time in health and in thyroid disease. *Gut* **30**, 26–29.

Wegener, M., Wedmann, B., Langhoff, T., *et al.* (1992). Effect of hyperthyroidism on the transit of a caloric solid–liquid meal through the stomach, the small intestine, and the colon in man. *J. Clin. Endocrinol. Metab.* **75**(3), 745–749.

Hypothyroidism

ASRA KERMANI
Veterans Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

hypothyroidism A syndrome characterized by inadequate thyroid hormone production or replacement.
myxedema Synonymous with hypothyroidism, but generally used to connote a more severe (advanced) case with accumulation of mucopolysaccharide material in soft tissues, body cavities, and organs.

The most common cause of hypothyroidism in developed countries is primary thyroid failure as in Hashimoto's thyroiditis; iodine deficiency prevails as the leading cause in the developing world. Hashimoto's thyroiditis is an autoimmune disorder associated with circulating antithyroid antibodies. There are similar proportions of B-cell and T-cell lymphocytes infiltrating the thyroid gland. It occurs with a frequency of 3.5 cases per 1000 people per year in women and 0.8 cases per 1000 people per year in men. Other frequent causes of hypothyroidism include post-radioiodine or postsurgical ablation of the thyroid gland for overproduction of thyroid hormone as in Graves' disease or toxic goiter or hypothalamic/pituitary disease with inadequate thyroid-releasing or thyroid-stimulating hormone secretion. Because thyroid hormone has profound effects on metabolism and muscle contractility, the gastrointestinal tract is significantly affected by its deficiency. As Hashimoto's thyroiditis is the commonest cause of hypothyroidism in developed countries, other autoimmune diseases often coexist with it.

GASTROINTESTINAL SYMPTOMS AND SIGNS

Constitutional Symptoms

Most symptoms referable to the gastrointestinal tract are nonspecific. However, there are certain clues that, in conjunction with other systemic signs and symptoms, are characteristic. Weight gain despite reduced appetite is common. The weight gain is mild and is a result of fluid retention from mucopolysaccharide accumulation within soft tissues. A yellowish skin color occurs in hypothyroidism as a result of carotenemia, from impaired conversion of carotene to vitamin A.

Dysphagia/Compression

Although dysphagia obviously can occur in the setting of a large goiter compressing the pharynx, it is also seen in hypothyroid patients with esophageal dysmotility. Patients with myxedema can rarely develop megaesophagus. A large substernal goiter (independent of function) may cause superior vena cava syndrome, resulting in esophageal varices in up to 12% of individuals. This is a rare cause of variceal hemorrhage.

Hypomotility

Approximately 10% of patients with hypothyroidism experience nonspecific abdominal pain. Many patients complain of bloating and flatulence, which is thought to be a consequence of reduced intestinal peristalsis. Constipation is frequent and may progress to severe ileus and fecal impaction. Hypothyroidism is a rare cause of megacolon from intestinal atony. Although constipation prevails, bacterial overgrowth in the setting of intestinal hypomotility has been reported, resulting in weight loss and diarrhea. Hypothyroidism is also one of the causes of an elevated CEA (carcinoembryonic antigen) level, which, if seen in a patient with abdominal distension, can mimic colon cancer.

Gastrointestinal Motility Changes

Some studies document significant delays in orocecal transit time in hypothyroid patients, suggesting that small bowel hypomotility is an early sign of hypothyroidism. Other studies have not confirmed this finding.

Anorectal Function

The anal sphincter mechanism is unaltered in hypothyroidism. The threshold sensation for evacuation is higher than in controls, which may provide a further explanation for constipation in hypothyroid patients.

 Copyright 2004, Elsevier (USA). All rights reserved.

Myxedema Ascites

Myxedema ascites accounts for less than 1% of all causes of ascites. It usually is chronic. The ascitic fluid is protein-rich (>2.5 g/dl), has a moderate white blood cell count that is predominantly lymphocytic, and has a serum ascites albumin gradient (SAAG) of greater than 1.1 g/dl. Some authors dispute the increased SAAG as being integral to the definition of myxedema ascites and invoke congestive heart failure, which often occurs concomitantly, as the cause of the increased SAAG. Like similar pleural and pericardial effusions, there is complete resolution with thyroid hormone replacement. A solitary case of chylous ascites and chylous pleural effusions has been described.

ASSOCIATIONS WITH GASTROINTESTINAL DISEASES

Immunologic Associations

Celiac disease, an autoimmune disease characterized by gluten sensitivity and small bowel villous atrophy, is seen more frequently in association with Hashimoto's thyroiditis and Graves' disease.

Hepatitis C viral (HCV) infection is associated with an increased prevalence of thyroid microsomal antibody. Furthermore, 3% of HCV positive euthyroid patients treated with interferon-α may develop antithyroid antibodies and become hypothyroid. Other autoantibodies or autoimmune diseases were risk factors. Thyroid abnormalities were reversible after interferon-α therapy was discontinued.

Pernicious anemia may occur as part of a polyglandular autoimmune syndrome, with coexistent Hashimoto's thyroiditis. In sporadic cases of Hashimoto's disease, parietal cell antibodies occur in approximately 25% of patients. Atrophic gastritis is also more common in Hashimoto's thyroiditis and may result in achlorhydria or hypochlorhydria.

Primary biliary cirrhosis (PBC), a disease of autoimmune destruction of interlobular bile ducts, may coexist with hypothyroidism. The prevalence of thyroid failure ranges from 12 to 22%. Thyroid disease may predate the onset of PBC by years. Patients with hypothyroidism who develop cholestatic liver disease should be screened for PBC.

Autoimmune hepatitis is a rare association and has been reported in polyglandular autoimmune syndromes. Nonspecific elevations of serum transaminase and alkaline phosphatase that correct with treatment of hypothyroidism have also been reported.

Sjogren's syndrome is occasionally associated with Hashimoto's disease, which leads to diminished salivary volumes.

Nutritional Deficiencies

Although overall nutrient absorption is normal in hypothyroidism, malabsorption or deficiencies of micronutrients may occur. Iron deficiency anemia is frequently seen in premenopausal women with hypothyroidism as a result of menorrhagia, but may also occur as a result of impaired iron absorption. Folate deficiency may also rarely result from impaired intestinal absorption.

See Also the Following Articles

Celiac Disease • Cholestatic Diseases, Chronic • Dysphagia • Hepatitis C • Malabsorption • Pernicious Anemia • Small Intestinal Motility

Further Reading

Deen, K. I., Seneviratne, S. L., and de Silva, H. J. (1999). Anorectal physiology and transit in patients with disorders of thyroid metabolism. *J. Gastroenterol. Hepatol.* **14**, 384–387.

Huang, M. J., and Liaw, Y. F. (1995). Clinical associations between thyroid and liver diseases. *J. Gastroenterol. Hepatol.* **10**, 344–350.

Marazuela, M., Garcia-Buey, L., Gonzalez-Fernandez, B., *et al.* (1996). Thyroid autoimmune disorders in patients with chronic hepatitis C before and during interferon-α therapy. *Clin. Endocrinol.* **44**, 635–642.

Rahman, Q., Haboubi, N. Y., Hudson, P. R., *et al.* (1991). The effect of thyroxine on small intestinal motility in the elderly. *Clin. Endocrinol.* **35**, 443–446.

Sharma, S., Longo, W. E., Baniadam, B., *et al.* (1995). Colorectal manifestations of endocrine disease. *Dis. Colon Rectum* **38**, 318–323.

Tobin, M. V., Fisken, R. A., Diggory, R. T., *et al.* (1989). Orocaecal transit time in health and in thyroid disease. *Gut* **30**, 26–29.

Ileal Brake

HENRY C. LIN AND NANCY LING
Cedars-Sinai Medical Center and David Geffen School of Medicine, University of California, Los Angeles

bioavailability The amount of an active agent, such as a drug, that is available at the targeted site after administration.
immunoneutralization Technique to abolish the effect of a circulating peptide by administering antibodies directed against the peptide.
peptide YY A distal gut peptide that is released in response to fat and is involved in the ileal brake and jejunal brake responses.
premeal A portion of food consumed shortly before the actual meal.

The ileal brake is one of the primary nutrient-triggered inhibitory feedback mechanisms that control transit of a meal through the gastrointestinal (GI) tract. It is activated when end products of digestion, such as fatty acids, come into contact with the distal small intestine. This braking response initiates a feedback reflex that inhibits upper GI motility to slow gastric emptying and transit through the small intestine. By controlling the speed of transit of a meal so that there is adequate time for assimilation of nutrients, the ileal brake works to optimize digestion and absorption.

COMPANION BRAKING MECHANISMS OF THE ILEAL BRAKE

The ileal brake works in concert with other transit controls, in particular, the jejunal brake, a similar transit regulatory mechanism located in the proximal end of the small intestine. However, the ileal brake is more potent than the jejunal brake. This difference allows the small intestine to vary its transit control response. For example, after a meal containing a difficult-to-assimilate food, nutrients would remain in the lumen to spread from the area of the proximal to the distal small intestine. Both the jejunal brake and the ileal brake will be activated in turn. The greater inhibitory feedback on transit that is then generated would work to ensure that there is adequate time for the assimilation of food. In addition to the jejunal brake, a duodenal brake and a colonic brake have also been described.

MEDIATORS OF THE ILEAL BRAKE

The precise neural circuitry mediating the ileal brake is an area of active research. Both gut peptides and nerves are involved in this response, as shown by the abolishment of the fat-induced ileal brake when either the distal gut peptide, peptide YY, was immunoneutralized or the inhibitory effect of an enteric opioid nerve was blocked.

CLINICAL EFFECTS OF ALTERED ILEAL BRAKE FUNCTION

The ileal brake response may be reduced or lost due to disease or intestinal resective surgery. When the control of GI transit is disrupted, the movement of a meal becomes accelerated. As a result, symptoms such as diarrhea, nausea, bloating, and abdominal pain develop. Consequences of such rapid transit may include maldigestion and malabsorption, leading to malnutrition and reduced bioavailability of oral medications. In the absence of the ileal brake, transit may become so markedly accelerated that end products of fat digestion are not adequately available. As a result, the remaining transit controls, such as the jejunal brake, are unable to compensate for the loss of the ileal brake response.

ACTIVATING THE ILEAL BRAKE AND JEJUNAL BRAKE AS TREATMENT

Intestinal transit is accelerated in many diarrheal conditions and rapid gastric emptying disorders. Since end products of fat digestion serve to trigger the ileal brake and the jejunal brake, a nutrient-based treatment using a premeal containing oleic acid has been shown to slow transit and reduce symptoms in these patients.

See Also the Following Articles

Gastric Emptying • Gastric Motility • Pancreatic Polypeptide Family

 Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Lin, H. C., Van Citters, G., Heimer, F., and Bonorris, G. (2001). Slowing of gastrointestinal transit by fatty acids: A novel, nutrient-based treatment for diarrhea. *Digest. Dis. Sci.* **46**, 223–229.

Lin, H. C., Zhao, X.-T., and Wang, L.-J. (1996). Jejunal brake: Inhibition of intestinal transit by fat in the small intestine. *Digest. Dis. Sci.* **41**, 326–329.

Lin, H. C., Zhao, X.-T., Wang, L.-J., and Wong, H. (1996). Fat-induced ileal brake depends on peptide YY. *Gastroenterology* **110**, 1491–1495.

Read, N. W., McFarlane, A., Kinsman, R. I., Bates, T. E., Blackhall, N. W., Farrar, G. B., Hall, J. C., Moss, G., Morris, A. P., O'Neill, B., *et al.* (1984). Effect of infusion of nutrient solutions into the ileum on gastrointestinal transit and plasma levels of neurotensin and enteroglucagon. *Gastroenterology* **86**, 274–280.

Spiller, R. C., Trotman, I. F., Higgins, B. E., Ghatei, M. A., Grimble, G. K., Lee, Y. C., Bloom, S. R., Misiewicz, J. J., and Silk, D. B. (1984). The ileal brake—Inhibition of jejunal motility after ileal fat perfusion in man. *Gut* **25**, 365–374.

Zhao, X.-T., Wang, L.-J., and Lin, H. C. (2000). Slowing of intestinal transit by ileal fat depends on naloxone-blockable, efferent opioid pathway. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G866–G870.

Ileoanal Pouch

RONALD BLEDAY*,† AND ANTHONY RAMSANAHIE†

** Harvard Medical School and † Brigham and Women's Hospital*

anal transitional zone Anal canal segment that lies between the uninterrupted columnar mucosa proximal and the squamous epithelium below.

Brooke ileostomy Surgical procedure described by Bryan Brooke in which the ileum is passed through an opening in the abdominal wall, with several centimeters protruding and its distal end everted.

cystectomy Surgical removal of the urinary bladder.

familial polyposis coli General neoplastic disorder of the intestine, presenting most commonly as multiple polyps in the colon.

Anileoanal pouch is produced in a procedure whereby the colon and rectum are removed but the anus and anal muscles are retained. A reservoir, or pouch, is then created, using the ileum such that continence of fecal matter is preserved.

INTRODUCTION

The ileal pouch reservoir was first described more than 70 years ago. It was constructed from an isolated segment of ileum shaped into a doubly folded loop after rectal excision for carcinoma. The procedure has evolved from the use of a straight ileal–anal anastomosis to the use of variously shaped reservoirs (J, S, or W pouches). The rationale for the pouch came about as an idea for pooling the intestinal content in the ileal reservoir; it was experimentally described by Valiente and Bacon in 1955 for rectal replacements in dogs. The procedure for clinical use in humans, introduced by Professor Nils Kock in 1969, was initially used as a bladder replacement following cystectomy. Later, the ileal anal pouch was used as the replacement for the resected rectum, as by Parks and Nicholls. These early experiences with ileal pouch formation, along with the endoanal sleeve technique devised by Alan Parks for cancer operations, have motivated surgeons to perform these procedures such that a pouch procedure restorative has become an alternative to a permanent ileostomy after proctocolectomy for ulcerative colitis (UC) and familial adenomatous polyposis (FAP).

PREOPERATIVE PREPARATIONS

As in other major operations, the patient's medical status is optimized for the procedure to ensure low or no morbidity. Correction of anemia, fluid depletion, and

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Lin, H. C., Van Citters, G., Heimer, F., and Bonorris, G. (2001). Slowing of gastrointestinal transit by fatty acids: A novel, nutrient-based treatment for diarrhea. *Digest. Dis. Sci.* **46**, 223–229.

Lin, H. C., Zhao, X.-T., and Wang, L.-J. (1996). Jejunal brake: Inhibition of intestinal transit by fat in the small intestine. *Digest. Dis. Sci.* **41**, 326–329.

Lin, H. C., Zhao, X.-T., Wang, L.-J., and Wong, H. (1996). Fat-induced ileal brake depends on peptide YY. *Gastroenterology* **110**, 1491–1495.

Read, N. W., McFarlane, A., Kinsman, R. I., Bates, T. E., Blackhall, N. W., Farrar, G. B., Hall, J. C., Moss, G., Morris, A. P., O'Neill, B., *et al.* (1984). Effect of infusion of nutrient solutions into the ileum on gastrointestinal transit and plasma levels of neurotensin and enteroglucagon. *Gastroenterology* **86**, 274–280.

Spiller, R. C., Trotman, I. F., Higgins, B. E., Ghatei, M. A., Grimble, G. K., Lee, Y. C., Bloom, S. R., Misiewicz, J. J., and Silk, D. B. (1984). The ileal brake—Inhibition of jejunal motility after ileal fat perfusion in man. *Gut* **25**, 365–374.

Zhao, X.-T., Wang, L.-J., and Lin, H. C. (2000). Slowing of intestinal transit by ileal fat depends on naloxone-blockable, efferent opioid pathway. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G866–G870.

Ileoanal Pouch

RONALD BLEDAY*,† AND ANTHONY RAMSANAHIE†

** Harvard Medical School and † Brigham and Women's Hospital*

anal transitional zone Anal canal segment that lies between the uninterrupted columnar mucosa proximal and the squamous epithelium below.

Brooke ileostomy Surgical procedure described by Bryan Brooke in which the ileum is passed through an opening in the abdominal wall, with several centimeters protruding and its distal end everted.

cystectomy Surgical removal of the urinary bladder.

familial polyposis coli General neoplastic disorder of the intestine, presenting most commonly as multiple polyps in the colon.

Anileoanal pouch is produced in a procedure whereby the colon and rectum are removed but the anus and anal muscles are retained. A reservoir, or pouch, is then created, using the ileum such that continence of fecal matter is preserved.

INTRODUCTION

The ileal pouch reservoir was first described more than 70 years ago. It was constructed from an isolated segment of ileum shaped into a doubly folded loop after rectal excision for carcinoma. The procedure has evolved from the use of a straight ileal–anal anastomosis to the use of variously shaped reservoirs (J, S, or W pouches). The rationale for the pouch came about as an idea for pooling the intestinal content in the ileal reservoir; it was experimentally described by Valiente and Bacon in 1955 for rectal replacements in dogs. The procedure for clinical use in humans, introduced by Professor Nils Kock in 1969, was initially used as a bladder replacement following cystectomy. Later, the ileal anal pouch was used as the replacement for the resected rectum, as by Parks and Nicholls. These early experiences with ileal pouch formation, along with the endoanal sleeve technique devised by Alan Parks for cancer operations, have motivated surgeons to perform these procedures such that a pouch procedure restorative has become an alternative to a permanent ileostomy after proctocolectomy for ulcerative colitis (UC) and familial adenomatous polyposis (FAP).

PREOPERATIVE PREPARATIONS

As in other major operations, the patient's medical status is optimized for the procedure to ensure low or no morbidity. Correction of anemia, fluid depletion, and

Copyright 2004, Elsevier (USA). All rights reserved.

electrolyte and acid–base disorders and nutritional assessment are mandatory. Many patients require total parenteral nutrition (TPN) and bowel rest because eating may worsen colitis symptoms. The simplest surgical corrective procedure should be performed in such severely compromised patients. Colon preparation may not be necessary and may be poorly tolerated, because the majority of these patients have severe diarrhea. The type of bowel preparation will depend on the patient's preoperative condition (bowel obstruction, frequent diarrhea); bowel preparations that are too aggressive may further compromise the patient. A double antibiotic prophylaxis has been shown to be of benefit in immunocompromised patients and in patients on whom surgery is being performed below the peritoneal reflection. Antibiotics (neomycin and erythromycin base regimen) are given orally 1 day prior to surgery and are given intravenously at induction of anesthesia in order to decrease the incidence of infectious complications.

Often these procedures require a temporary diverting ileostomy. A staged procedure would require a marked site for the loop ileostomy preoperatively, and the patient should have been seen by a stomatherapist. Patients on high-dose steroids preoperatively will require parenteral steroids in the perioperative period until their baseline physiological dose has being reached postoperatively.

SURGICAL TECHNIQUES

Position

Most patients can have the total procedure performed in the lithotomy position However, some patients require a mucosectomy of the very last few centimeters of the rectal mucosa. To facilitate the procedure, the prone jackknife position can be used at the start of the procedure for the rectal mucosectomy performed per anus. This position allows for greater exposure than is possible with the lithotomy position. After the mucosectomy, the patient is turned to the lithotomy position for the abdominal phase of the operation.

Creation of the Ileoanal Pouch

Total abdominal proctocolectomy is carried out as for colectomy, with the colon mobilized from the ileocecal to the rectosigmoid junction. It is important to dissect as close to the rectal wall as possible to avoid injury to the pelvic autonomic nerves. To transect the distal rectum/proximal anus, a stapling device is placed in the pelvis. A finger or sigmoidoscope ascertains the position of the stapler in relation to the dentate line. The distal rectum/upper anus is stapled off, leaving no more than 3 cm of distal rectum above the dentate line (it is, in fact, preferable to leave as little mucosa as possible while preserving all of the anal sphincters).

Once the rectum is removed, the ileal pouch is created. The ileal mesentery is freed from the retroperitoneum up to the lower duodenal margin, to increase the length of the mesentery so that it can reach the pelvis. In general, the distal ileum at the apex of the pouch that is to be created needs to be able to reach to the lower border of the pubis. The ileum is then folded into the general shape of the pouch to be formed. Ileoanal pouches are usually made into two shapes. The J pouch is made out of the distal ileum folded back onto itself once, whereas the S pouch is made by folding the distal ileum twice. The J-pouch is usually made 15 cm in length and can be constructed with a bowel anastomotic stapling device. The S pouch must be hand sewn. Because of its simplicity and the speed at which it can be made using a stapler, the J pouch is made more often (see Figs. 1 and 2).

To anastomose the pouch to the distal rectum /anus, a circular stapling device is usually used. The anvil of the circular stapler is sutured into the apex of the pouch through the enterotomy made for the linear cutting stapling device, using a pursestring suture. To perform the anastomosis, the circular stapler with the trocar is placed into the anorectal stump and the trocar is advanced, piercing the stapled upper anus either near or directly through the staple line. The trocar is removed and the anvil in the pouch is brought down to the accepting socket of the circular device. The bowel ends are apposed and stapling device is fired, ensuring that the pouch is not twisted.

Mucosectomy and Cuff Formation

In some patients, the situation requires removal of the mucosa of the last 0–3 cm of distal rectum. This procedure is called a mucosectomy because only the mucosa is removed, with the anal muscles and residual muscularis propria of the rectal cuff preserved. With the patient in either the lithotomy or prone position, the dissection starts just above the dentate line. The mucosa is raised off the anal muscles using saline or a weak epinephrine solution. The mucosa is then stripped away from the internal anal sphincter and muscularis propria. All mucosa is removed. Hemostasis is ensured with cauterization.

The ileal pouch is anastomosed in a series of handsewn sutures using a long-term absorbable braided suture. Eight to 12 full-thickness sutures are placed

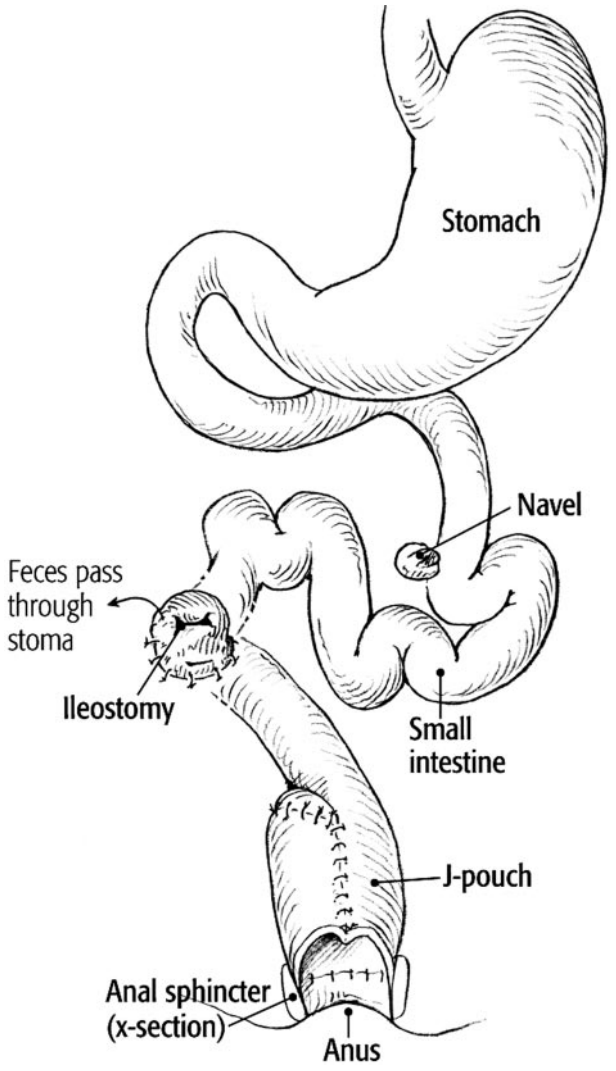


FIGURE 1 A J pouch and temporary stoma.

through the pouch wall and the internal sphincter, and out through the dentate line mucosa to complete the anastomosis. The anastomosis is then checked digitally for any gaps.

Other Pouches

Though the S pouch was the first clinically applied pelvic pouch, the J pouch is now most commonly used. The S-shaped pouch is formed by folding three ileal limbs 10 cm in length. On the distal limb, an extra 1 to 2 cm is left to project into the anal canal. This small spout is then anastomosed end to end to the anal canal. The S pouch causes greater difficulty in evacuation and takes a longer time to construct. A W pouch has been described as two J pouches placed together. The terminal end of the small intestine is folded into four loops, 12 cm long, forming the W pouch. A side-to-end anastomosis is made directly between the anal canal and reservoir. The W pouch and the J pouch have been shown to have equivalent function that, but not enough data have been collected to show that there is any value added to J pouch.

In addition to the J, W, and W types of reservoirs, several other variations have been described. The best known is the Kock, or K, pouch. This is an abdominal pouch that consists of a reservoir constructed from the small intestine, and a nipple valve that keeps the contents of the reservoir inside the body, permitting entry of an external catheter to drain the pouch when desired. Occasionally the Kock pouch can be converted to a pelvic pouch after a period of pouch dilatation, with anastomosis to the anus allowing restoration of intestinal continuity. It is an alternative for patients who have a standard (Brooke) ileostomy, or who have poor results with an ileoanal operation.

These operations can be carried out either in one, two, or three stages. The most common approach is the two-stage procedure. In the first stage, the colon and rectum are removed, the ileoanal pouch is created, and the surgeon creates a diversion ileostomy to protect the neorectal pouch and ileoanal anastomosis. This procedure is preferable to the one-stage procedure because it is associated with a lower rate of pelvic sepsis. It is especially important to limit pelvic septic complications in women of childbearing age, because the incidence of infertility after pelvic sepsis is high. The ileostomy helps ensure that any clinical problems are not worsened by a small anastomotic leak.

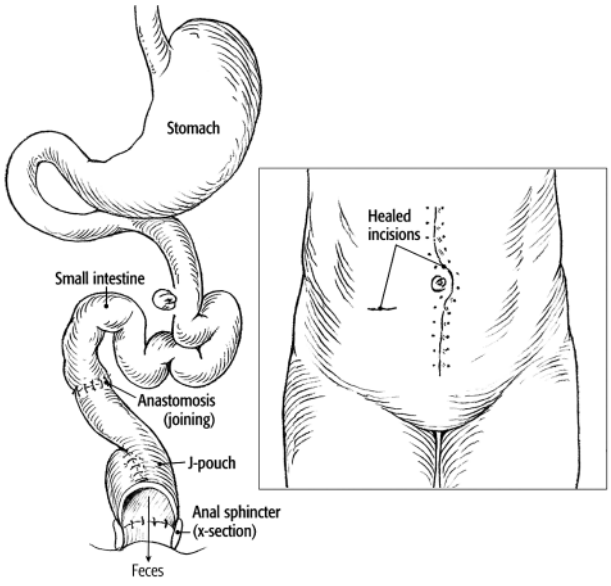


FIGURE 2 A J pouch reconnection.

Hand-Sewn vs. Stapled Anastomosis

The hand-sewn anastomosis with rectal mucosectomy was first employed for the ileoanal J pouch. However, a modification of the hand-sewn anastomosis was utilized after the introduction of the double-stapled technique for pelvic procedures. The stapled technique has resulted in a shorter operative time compared to the hand-sewn operation. It is also suggested that this technique improves continence by preserving the anal transition zone (ATZ) and by reducing the injury sustained on the anal sphincter mechanism by the prolonged dilatation required for completion of the mucosectomy. Some surgeons, however, suggest that the hand-sewn procedure is preferred, because the retained ATZ may be a source of diseased mucosa. Second, it is questionable that the early functional outcome of the stapling technique is better than the hand-sewn procedure over time. Data from retrospective series are inconclusive. The stapled anastomosis with the retained ATZ provides better nocturnal continence and better discrimination of flatus versus stool. Unless there are multiple polyps in the distal 3 cm of mucosa in a patient with FAP, or severe dysplasia in a patient with colitis, it is preferable to perform the stapled anastomosis. An attempt should be made to leave only 1–2 cm of mucosa above the dentate line in order to minimize any potential symptoms or problems associated with longer amounts of mucosa.

POSTOPERATIVE CARE

In the immediate postoperative period, observations of the vitals signs are monitored; hemoglobin and electrolytes are checked and corrected as necessary. The urinary catheter is removed after 3–5 days. Patients also should have prophylaxis against deep venous thrombosis. Watery stool usually begins within 48–72 hours after operation and can be as frequent as 10–15 times a day. Loperamide or codeine can be given to decrease the frequency of stool passage. Infection, ileus, and obstruction are the most common serious complications that need to be watched for in the first 30 days postoperatively.

After takedown of the temporary ileostomy (usually between 8 and 12 weeks after the first stage), soiled perianal skin is rinsed and kept dry to avoid skin ulcers and infections. Oral intake is allowed after 1 day, with a gradual increase in solid food. Advice is given about diet and the use of medications to slow the frequency of bowel movements if needed. Patients are followed for 1–3 months until the frequency of bowel movements settles down to six times per 24 hours. Patients are then followed on an annual basis.

COMPLICATIONS

The mortality after pouch–anal anastomosis is very low in all reported series (approximately 0.5%). General complications such as wound and chest infections, urinary retention, and deep venous thrombosis must be monitored and treated. Early complications specific to the procedure include cuff abscesses, anastomosis dehiscence, and intestinal obstruction. Metabolic disturbances due to dehydration can occur because the evacuation rate in the early days can be high. Late complications include a condition called pouchitis. Its incidence is 10–30% and it is most often treated with antibiotics or probiotics. If a patient develops severe and debilitating pouchitis or anal fistulas, Crohn's disease should be considered. Fistula formation can occur from the pouch or from the anastomosis to the urinary tract, vagina, perineum (vulva), or anterior abdominal wall.

Recurrent colitis can rarely cause disease in the residual mucosa. Polyps can sometimes occur in patients with FAP. There is a small risk for progression to malignancy both in patients who have had a mucosectomy and in those who have not, therefore all patients require annual surveillance.

QUALITY OF LIFE

The restorative proctocolectomy with ileopelvic reservoir is a safe procedure, but the incidence of postoperative morbidity can be high. Studies in the form of questionnaires with a scoring system have assessed the functional outcome and the effects on the lifestyle of patients who have undergone these procedures. For instance, quality of life (QOL) may be better in the patient group that had the procedure done for ulcerative colitis compared to patients with FAP. This may be due to the fact that, before the procedure, patients with ulcerative colitis are more symptomatic as compared to patients with FAP. In general, most patients attain a high QOL postoperatively. They show improvement in QOL compared to their preoperative condition and they also have a higher QOL than do those with a permanent ileostomy.

CONTRAINDICATIONS TO POUCH FORMATION

A contraindication to ileoanal pouch procedure, as in any operation to preserve anorectal continence, is absence preoperatively of an efficient anal sphincter complex. Therefore, the presence of fecal incontinence or lax anal tone should be screened for preoperatively. In

many major series, patients with Crohn's disease generally seem to do badly, with ileoanal pouch anastomosis complicated with fistulas and abscesses. In patients without anal or small bowel disease, there is evidence that about 80% of patients have a satisfactory outcome after creation of an ileoanal pouch.

Patients selected should be highly motivated, because complications often occur and may frequently require further operations. Therefore, patients should be fully informed of the potential risks and benefits of the procedure. Contact with other patients who have had the procedure, whatever the outcome, should be encouraged. Age does not seem to be a contraindication as long as there is no significant sphincter dysfunction and no serious comorbidity, because the operation is very lengthy.

See Also the Following Articles

Colitis, Ulcerative • Familial Adenomatous Polyposis (FAP) • Fistula • Parenteral Nutrition • Pouchitis

Further Reading

Cohen, Z., McLeod, R. S., Wesley, S., *et al.* (1992). Continuing evolution of the pelvic pouch procedure. *Ann. Surg.* **216**(No. 4), 506–512.

Kock, N. G., Hulton, L., and Myrvold, H. E. (1989). Ileoanal anastomosis with interposition of the ileal "Kock pouch." *Dis. Colon Rectum* **32**, 1050–1054.

Koher, L. W., Pemberton, J. H., Zinsmeister, A. R., and Kelly, K. A. (1991). Quality of life after proctocolectomy: A comparison of Brooke ileostomy, Kock pouch and ileal pouch anastomosis. *Gastroenterology* **101**, 679–684.

Parks, A. G., and Nicholls, R. J. (1978). Proctocolectomy without ileostomy for ulcerative colitis. *Br. Med. J.* **2**, 85–88.

Ravitch, M. M., and Sabiston, D. C. (1947). Anal ileostomy with preservation of the sphincter: A proposed operation in patients requiring total colectomy for benign lesions. *Surg. Gynecol. Obstet.* **84**, 1095–1099.

Sagar, P. M., Lewis, W., Holdsworth, P. J., Johnston, D., Mitchell, C., and Mac Fie, J. (1993). Quality of life after restorative proctocolectomy with a pelvic ileal reservoir compares favorably with medically treated colitis. *Dis. Colon Rectum* **36**, 584–592.

Thomas, P. E., and Taylor, T. V. (1991). "Pelvic Pouch Procedures." Butterworth-Heinemann, U. K.

Utsunomiya, J. (1994). Total colectomy, mucosal proctectomy, and ileoanal anastomosis. *In* "Operative Colorectal Surgery" (G. E. Block and A. R. Moossa, eds.), pp. 165–191. W. B. Saunders, Philadelphia.

Valiente, M. A., and Bacon, H. E. (1955). Construction of a pouch using "pantaloon" technique for pull-through of ileum following total colectomy: Report of experimental works and results. *Am. J. Surg.* **90**, 742–750.

Williams, N. S. (1989). Restorative proctocolectomy is the first choice elective surgical treatment for ulcerative colitis. *Br. J. Surg.* **76**, 1109–1110.

Immunodeficiency

MATTHIEU ALLEZ* AND LLOYD MAYER†
*Hopital Saint-Louis, Paris and †Mount Sinai Medical Center, New York

agammaglobulinemia Total absence of antibody.
antibody/immunoglobulin Protein elicited in response to antigen; there are five primary forms, IgG, IgA, IgM, IgE, and IgD. There are four subclasses of IgG (IgG1–IgG4) and two subclasses of IgA (IgA1 and IgA2).
antigen/immunogen Any compound that can elicit an immune response.
B lymphocyte Precursor of an antibody-producing cell (plasma cell); has a surface antigen receptor that can recognize antigens in solution.
immunodeficiency Any state in which the host's immune system (antibodies, T cells) is not functioning at optimal capacity.
T lymphocyte Cell of the immune system that regulates immunity; recognizes only foreign protein/antigen in the context of major histocompatibility antigens.

Primary immunodeficiencies are caused by inherited defects of the immune system. Most patients manifest their immunodeficiency by developing infections involving the sinuses and the lungs. Gastrointestinal manifestations, however, are particularly frequent, and in some cases may represent the main complaint in patients with primary immunodeficiency. The most common primary immunodeficiencies are characterized by an impaired ability to produce specific antibodies. These antibody deficiencies are often diagnosed in childhood or adolescence, but are sometimes apparent only in adulthood. Primary T cell immunodeficiencies are rarer and are generally diagnosed early in life.

CLASSIFICATION OF PRIMARY IMMUNODEFICIENCIES

Primary Antibody Deficiencies

Common Variable Immunodeficiency

Common variable immunodeficiency (CVI) is the most common clinically significant antibody deficiency (Table I) of adults, with a prevalence estimated from 1 in 50,000 in Sweden. CVI is a heterogeneous group of immunodeficiency disorders, characterized by decreased serum immunoglobulin levels and heterogeneous clinical features. Patients with CVI usually develop symptoms either as adolescents or as young adults and present most commonly with recurrent sinopulmonary infections and gastrointestinal disorders. Frequently, however, the diagnosis may not be obvious until late adult life. Respiratory infections often involve encapsulated bacteria (*Haemophilus influenzae, Streptococcus pneumoniae*). CVI patients may develop bronchiectasis or restrictive or obstructive lung disease. Immunoglobulin G (IgG) levels are typically less than 300 mg/dl with concurrent low IgA and IgM levels. Circulating B cells are generally not decreased in number, but they may exhibit defects in growth and differentiation. The treatment of CVI is intravenous (iv) immunoglobulin replacement and antibiotics, which may prevent the recurrence of infections and the development of irreversible lesions in the lungs. However, iv immunoglobulin replacement does not seem to prevent the gastrointestinal disorders observed in CVI patients.

Selective IgA Deficiency

Selective IgA deficiency is the most common immunodeficiency, present in 1 in 300–700 Europeans and North Americans. Levels of serum IgA are less than 5 mg/dl, with normal or increased concentrations of other serum immunoglobulins. The number of circulating B cells is normal. In most individuals, there is a decreased level of secretory IgA at mucosal surfaces, which can be compensated for by an increased production of secretory IgM. A majority of IgA-deficient patients are asymptomatic and need no treatment. Others may have allergic manifestations, recurrent sinopulmonary infections, or autoimmune disorders. These

TABLE I Classification of Primary Immunodeficiencies

Primary antibody deficiencies	Primary T cell immunodeficiencies
X-Linked agammaglobulinemia	Severe combined immunodeficiency
Selective IgA deficiency	DiGeorge syndrome
Selective deficiency of IgG subclasses	Ataxia–telangiectasia
Common variable immunodeficiency	Wiskott–Aldrich syndrome

Copyright 2004, Elsevier (USA). All rights reserved.

symptomatic patients often have concurrent IgG2 deficiency. Only patients with severe sinopulmonary infections and those with concomitant IgG defects are candidates for iv immunoglobulin replacement therapy. A small subset of these patients have anti-IgA antibodies, which may increase the risk for anaphylactic reactions to iv immunoglobulins or other blood products.

X-Linked Agammaglobulinemia (Bruton's Disease)

X-Linked agammaglobulinemia (XLA) is a rare immunodeficiency, with an incidence of approximately 1 in 10^5 births. This deficiency is characterized by a failure of B cells to mature beyond the pre-B cell stage and the consequent inability to produce all classes of immunoglobulins. The inherited defect is mapped to the gene encoding Bruton's tyrosine kinase (Btk), located on the long arm of the X chromosome. Defects in this intracellular tyrosine kinase block the maturation of pre-B cells. XLA patients (young males) typically present with recurrent bacterial infections, especially with encapsulated bacteria such as *S. pneumoniae* and *H. influenzae*. The treatment of XLA is iv immunoglobulin replacement.

Primary T Cell Deficiencies

DiGeorge Syndrome

Patients with DiGeorge syndrome have thymic hypoplasia with or without parathyroid glands (hypocalcemia and tetany), cleft palate, and other facial anomalies (due to a defect in the migration of the third and fourth branchial pouches). The T cell deficiency is variable from patient to patient (depending on the degree of thymic aplasia) and may predispose patients to opportunistic infections, such as *Pneumocystis carinii*, candidiasis, and viral infections.

Ataxia–Telangiectasia

Ataxia–telangiectasia is inherited in an autosomal recessive pattern. The first symptoms, cerebellar ataxia and telangiectasias (of the conjunctiva, nose, ears, or shoulders), appear in young patients between 2 and 5 years of age. They have thymic hypoplasia and impaired T cell function. IgA deficiency is found in 50% of patients. These patients have a high risk of malignancy.

Wiskott–Aldrich Syndrome

Wiskott–Aldrich syndrome (WAS) is a rare X-linked recessive disorder. Patients usually present with eczema, thrombocytopenia (and hemorrhage), impaired T cell function, and recurrent opportunistic infections. The genetic defect is a mutation in the WAS protein, which is involved in intracellular signaling and actin/cytoskeletal rearrangements within the lymphocyte. Bone marrow transplantation is the treatment of choice for long-term survival.

Severe Combined Immunodeficiency

Severe combined immunodeficiency (SCID) is a heterogeneous set of congenital immune disorders in which there is deficiency of both T and B cell development and function. Patients usually present with severe recurrent bacterial or viral infections during the first year of life. These patients have few or no circulating lymphocytes. SCID patients are susceptible to both bacterial and opportunistic infections. The standard treatment for SCID is bone marrow transplantation.

GASTROINTESTINAL MANIFESTATIONS

Primary Antibody Deficiencies

Gastrointestinal manifestations are quite frequent in patients with CVI; 40–60% experience chronic diarrhea and many have malabsorption. These manifestations may reflect gastrointestinal infections, particularly by *Giardia lamblia*, but also diverse inflammatory disorders and malignant conditions. CVI patients may have lymphoproliferative disorders, such as diffuse nodular lymphoid hyperplasia (NLH) of the small bowel, which may represent a premalignant condition. Most patients with IgA deficiency do not have gastrointestinal manifestations, but when symptoms occur they are similar to those observed in CVI.

Infection with *G. lamblia* may cause chronic diarrhea with malabsorption in patients with antibody deficiencies, particularly with CVI. Symptoms frequently consist of an abrupt onset of watery diarrhea with abdominal cramping, bloating, and anorexia. The diagnosis is made by stool examination, demonstrating the presence of cysts or trophozoites, but frequently requires duodenal biopsies or aspirates. Duodenal biopsies may show mucosal damage with villous atrophy. The course of the infestation, which is usually short in immunocompetent patients, may last several months in CVI patients. Treatment with metronidazole reverses the symptoms and mucosal damage and is often used as a therapeutic test for *Giardia* infection. Treatment with iv immunoglobulins may reduce the risk of recurrent giardiasis, although there are limited data to support this and T cell immunity appears to be more important in the clearance of this infection. Prolonged *G. lamblia* infestation may also be observed in IgA deficiency, but is much less frequent than in CVI.

Patients with CVI have also a slightly increased prevalence of infections with *Salmonella* and *Campylobacter jejuni*. Cryptosporidiosis has also been described in CVI patients, although the incidence is low. One other real consideration is the role that *Clostridium difficile* may play in recurrent diarrhea in these patients, especially considering the amount of antibiotics used in treatment of infections.

Some patients with CVI present with atypical inflammatory gastrointestinal disease that is unrelated to *G. lamblia* and results in diarrhea, malabsorption, and weight loss. Villous atrophy on duodenal biopsies may be observed in some cases, similar to the lesions observed in celiac disease. However, antibodies against gliadin and endomysium are generally not found in the serum of these patients. More importantly, gluten-free diets seem to have no effect on the clinical course or on the villous atrophy in the majority of patients. Among the CVI patients in a series at The Mount Sinai Medical Center in New York, six had a villous atrophy that did not improve with a gluten-free diet.

After intestinal infection has been ruled out, steroid and/or immunomodulator (6-mercaptopurine/azathioprine) therapy may be considered in these patients. Such therapy should be used with caution in these patients for obvious reasons. However, the overwhelming majority of patients tolerate immunomodulators quite well while they are receiving iv immunoglobulins. The doses used to control inflammation are never so high as to further compromise immune function (1.5–2.5 mg/kg for 6-mercaptopurine and azathioprine, respectively). In contrast, severe adverse events have been reported in CVI patients receiving steroids or methotrexate (detailed later).

There is a stronger association between IgA deficiency and celiac disease. Of patients with celiac disease, 2% have IgA deficiency. Intestinal biopsy findings in these patients are similar to those seen in classical celiac disease except for an increase in IgM plasma cells in the lamina propria. In contrast to CVI, patients with IgA deficiency and celiac disease may respond normally to treatment with a gluten-free diet.

There is an increased frequency of atrophic gastritis and/or pernicious anemia in patients with CVI. Pernicious anemia may be even more frequent in IgA-deficient patients. The diagnosis is made at an earlier age (20–40 years old) than in immunocompetent patients (60 years old). In contrast to pernicious anemia in immunocompetent individuals, antibodies directed against parietal cells and intrinsic factor are not found in the serum. Histological findings are similar, except for the absence of plasma cells in the lamina propria of CVI patients. These findings suggest that pernicious anemia in this setting is not antibody mediated. CVI patients have an increased risk, compared to the general population, of developing gastric adenocarcinoma, which may be related to the increased frequency of atrophic gastritis (see later).

There is also an increased frequency of inflammatory bowel disease in CVI patients. Many cases of ulcerative colitis and Crohn's disease have been reported. In the Mount Sinai series, among 248 patients with CVI, 16 (6%) had inflammatory bowel disease (IBD), 7 had ulcerative colitis, and 9 had Crohn's disease. Hermaszewski *et al.* reported 10 cases of IBD among 240 patients with CVI, and 3 cases among 44 patients with agammaglobulinemia. Teahon *et al.* reported 12 patients with CVI or agammaglobulinemia when chronic diarrhea was explored by upper and/or lower endoscopy and biopsy; 4 of 8 patients had villous atrophy on duodenal biopsies and 8 of 10 had microscopic colitis. Two of these patients had multiple stenoses in the small bowel, suggestive of Crohn's disease. Cases of inflammatory bowel disease have been reported in IgA-deficient patients, but the frequency of IBD does not seem to be increased. IBD in patients with CVI seems to respond normally to conventional therapeutic approaches. Steroids and immunomodulatory agents (such as 6-mercaptopurine/azathioprine) can be used in these patients. Patients with CVI-associated IBD may experience long-term remissions with these therapies. However, as is the case for refractory sprue, data from the Mount Sinai series suggest that specific types of immunosuppression in CVI patients may result in complications. Indeed, four patients receiving steroids had major complications: *Pneumocystis carinii* pneumonia, a *Nocardia* brain abscess, a progressive multifocal leukoencephalopathy, and a severe anaerobic leg infection leading to amputation. Two patients receiving methotrexate for autoimmune disease (both had a prior history of Hodgkin's disease) developed either an additional malignancy or recurrent disease.

Nodular lymphoid hyperplasia, defined as multiple discrete mucosal nodules confined to the lamina propria and superficial submucosa, is found frequently in CVI patients. The nodules represent hyperplastic lymphoid tissue with prominent germinal centers. About 10–20% of CVI patients have diffuse intestinal NLH. NLH can affect the ileum and/or the colon of immunocompetent children and young adults, but the diffuse form of NLH is mainly observed in patients with primary immunodeficiencies. The exact meaning of this manifestation in these patients is unknown, but could reflect a compensatory B cell proliferative response to an increase in the precursor pool of antibody-producing cells. There have been several reports of patients with

diffuse NLH, both immunodeficient and immunocompetent, who developed small intestine lymphomas (see later), and several reports in the literature suggest that diffuse NLH could be a premalignant condition. Diffuse NLH is seen less frequently in selective IgA deficiency than in CVI patients.

Patients with CVI are reported to have an increased prevalence of, and mortality due to, malignant neoplasms. The immunodeficiency cancer registry has reported 120 cases of cancer among CVI patients. The predominant tumors were non-Hodgkin lymphomas (55 cases, 48%), followed by gastrointestinal carcinomas (30 cases, 25%), predominantly of the stomach (19/30), and Hodgkin's disease (8 cases, 7%). The relative risk of non-Hodgkin lymphoma has been reported to range from 47-to 438-fold in the general population. From 1.4 to 7% of CVI patients may develop a lymphoma, which occurs typically after 40 years of age and more often in women. These lymphomas are generally extranodal and of a B cell type, and may be associated with Epstein–Barr virus (EBV) infection. As mentioned earlier, some small intestinal lymphomas described in CVI patients are associated with diffuse NLH. Several cases of T cell lymphomas have been described in CVI patients. A majority of these lymphomas are located in the small bowel and are associated with nodular lymphoid hyperplasia. Some cases of colonic lymphomas have been reported simulating, in one case, cryptogenic colitis.

The increased risk of cancer in patients with CVI includes an increased frequency of gastric adenocarcinoma (30-fold). Three cases were reported in a series of 240 patients with CVI, and two cases in another series of 248 patients. As mentioned earlier, the increased risk of gastric adenocarcinoma in CVI patients could be linked to the increased frequency of atrophic gastritis in those patients. This risk could justify an endoscopic survey of these patients, particularly in cases of atrophic gastritis. In a recent study, there was an increased prevalence of *Helicobacter pylori* infection in CVI patients with dyspepsia, and these patients had an increased frequency of multifocal gastritis. More interestingly, this study suggested that *H. pylori* may play a role in gastric carcinogenesis in patients with CVI. Last, the incidence of colonic cancer is not increased in CVI patients but has been reported in young patients with agammaglobulinemia.

Primary T Cell Deficiencies

The clinical manifestations observed in patients with ataxia–telangiectasia are mainly related to the deficiency in IgA and IgG2. Thus, the gastrointestinal manifestations are similar to those observed in patients with antibody deficiencies. The occurrence of chronic diarrhea and/or of malabsorption suggests the possibility of an infection with *G. lamblia*. Malignant disorders are particularly increased in these patients. Indeed, patients with ataxia–telangiectasia have a risk of cancer, mainly lymphomas, 78-fold higher when compared to the general population. This may relate to the defects in DNA repair mechanisms seen in these patients. However, the frequency of cancers or lymphoma in the gastrointestinal tract has not been specified in the literature.

Gastrointestinal manifestations are rare in patients with Wiskott–Aldrich syndrome. There is also an increased risk of cancer in this population, with an incidence estimated at 2% per year. Intestinal hemorrhage may occur because of thrombocytopenia. Interestingly, the mouse model of WAS is associated with colitis and there have been some reports of colitis in patients with this disease.

The frequency of digestive manifestations does not seem increased in patients with DiGeorge syndrome, except for intestinal infections. Oral candidiasis is frequent and viral infections causing diarrhea and necrotizing enterocolitis have been reported.

Children with severe combined immunodeficiency (SCID) often present with intractable diarrhea with villous atrophy, which is resistant to medical treatment. These children may present with oral candidiasis and viral infections; rotavirus, adenovirus, or picornavirus. After bone marrow transplantation, graft-versus-host disease with wasting and diarrhea may develop.

CONCLUSION

Patients with primary immunodeficiencies, particularly with antibody deficiencies, have frequent gastrointestinal manifestations. In some cases, gastrointestinal manifestations may represent the only complaint. Thus, physicians should be aware of these manifestations and be able to recognize these immunodeficiencies, which can first present in adulthood. Furthermore, progress in the management of severe immunodeficiencies allows more patients to survive into adulthood.

See Also the Following Articles

Atrophic Gastritis • Celiac Disease • Colitis, Ulcerative • Crohn's Disease • Giardiasis • Pernicious Anemia

Further Reading

Allez, M., Mariette, X., and Lemann, M. (2000). Manifestations digestives des déficits immunitaires primitifs. *In* "Traité de

Gastroentérologie" (J. C. Rambaud, ed.), Chap. 73, pp. 815–824. Flammarion, Paris.

Ament, M. E. (1985). Immunodeficiency syndromes and the gut. *Scand. J. Gastroenterol.* **114**, 127–135.

Beard, L. J., Ferrante, A., Oxelius, V. A., and Maxwell, G. M. (1986). IgG subclass deficiency in children with IgA deficiency presenting with recurrent or severe respiratory infections. *Pediatr. Res.* **20**, 937–942.

Burrows, P. D., and Cooper, M. D. (1997). IgA deficiency. *Adv. Immunol.* **65**, 245–276.

Cataldo, F., Marino, V., Bottaro, G., *et al.* (1997). Celiac disease and selective immunoglobulin A deficiency. *J. Pediatr.* **131**, 306–308.

Cunningham-Rundles, C., and Bodian, C. (1999). Common variable immunodeficiency: Clinical and immunological features of 248 patients. *Clin. Immunol.* **92**, 34–48.

Cunningham-Rundles, C., Siegal, F. P., Cunningham-Rundles, S., *et al.* (1987). Incidence of cancer in 98 patients with common varied immunodeficiency. *J. Clin. Immunol.* **7**, 294–299.

Elenitoba-Johnson, K. S., and Jaffe, E. S. (1997). Lymphoproliferative disorders associated with congenital immunodeficiencies. *Semin. Diagnost. Pathol.* **14**, 35–47.

Filipovitch, A. H., Mathur, A., Kamat, D., *et al.* (1994). Lymphoproliferative disorders and other tumors complicating immunodeficiencies. *Immunodeficiency* **5**, 91–112.

Heneghan, M. A., Stevens, F. M., Cryan, E. M., *et al.* (1997). Celiac sprue and immunodeficiency states: A 25-year review. *J. Clin. Gastroenterol.* **25**, 421–425.

Hermans, P. E., Diaz-Buxo, J. A., and Stobo, J. D. (1976). Idiopathic late-onset immunoglobulin deficiency. Clinical observations in 50 patients. *Am. J. Med.* **61**, 221–237.

Hermaszewski, R. A., and Webster, A. D. B. (1993). Primary hypogammaglobulinemia: A survey of clinical manifestations and complications. *Q. J. Med.* **86**, 31–42.

Kinlen, L. J., Webster, A. B. D., Bird, A. G., *et al.* (1985). Prospective study of cancer in patients with hypogammaglobulinemia. *Lancet* **1**, 263–265.

Lai Ping So, A., and Mayer, L. (1997). Gastrointestinal manifestations of primary immunodeficiency disorders. *Semin. Gastrointest. Dis.* **8**, 22–32.

Morrell, D., Cromartie, E., and Swift, M. (1986). Mortality and cancer incidence in 263 patients with ataxia–telangiectasia. *J. Natl. Cancer Inst.* **77**, 89–92.

Perry, G. S., Spector, B. D., and Schuman, L. M. (1984). The Wiskott–Aldrich syndrome in the United States and Canada (1892–1979). *J. Pediatr.* **97**, 72–78.

Sander, C. A., Medeiros, L. J., Weiss, L. M., Yano, T., Sneller, M. C., and Jaffe, E. S. (1992). Lymphoproliferative lesions in patients with common variable immunodeficiency syndrome. *Am. J. Surg. Pathol.* **16**, 1170–1182.

Spickett, G. P. (2001). Current perspectives on common variable immunodeficiency (CVID). *Clin. Exp. Allergy* **31**, 536–542.

Teahon, K., Webster, A. D., Price, A. B., Weston, J., and Bjarnsaon, I. (1994). Studies on the enteropathy associated with primary hypogammaglobulinemia. *Gut* **35**, 1244–1249.

Twomey, J. J., Jordan, P. H., Laughter, A. H., *et al.* (1970). The gastric disorder in immunoglobulin-deficient patients. *Ann. Intern. Med.* **72**, 499–504.

Zullo, A., Romiti, A., Rinaldi, V., Vecchione, A., Tomao, S., Aiuti, F., Frati, L., and Luzi, G. (1999). Gastric pathology in patients with common variable immunodeficiency. *Gut* **45**, 77–81.

Immunoglobulins

MICHAEL E. LAMM
Case Western Reserve University, Ohio

polymeric immunoglobulin receptor Transmembrane protein expressed on the basolateral surface of epithelial cells lining mucous membranes; binds polymeric IgA and also IgM and transcytoses them into the lumen.

secretory component External domain of the polymeric immunoglobulin receptor; remains attached to the secreted antibody.

The immunoglobulins of the gastrointestinal tract and their secretion reflect the special features of this branch of the mucosal immune system. Immunoglobulins of the gastrointestinal tract and antibodies in serum, which reflect systemic immunity, are markedly different.

DOMINANCE OF IgA

Although every class of immunoglobulin (Ig)—IgA, IgD, IgE, IgG, and IgM—is present in the wall of the gastrointestinal tract as a result of local synthesis and diffusion from serum, IgA is quantitatively and functionally the most important. IgA also predominates in the lumen as a result of both extensive local synthesis in the intestinal mucosa and a specific transport mechanism across the epithelium. Hence, this discussion focuses on IgA, although other immunoglobulin classes undoubtedly play significant roles under particular circumstances (e.g., IgA deficiency; allergy).

ORIGIN OF INTESTINAL IgA

Most of the IgA in the intestinal tract is produced locally by the abundant plasma cells in the lamina propria. Precursors of these cells arise in foci of organized lymphoid tissue, exemplified by the Peyer's patches, which underlie specialized epithelial cells, the M cells, that transport antigens intact from the lumen for processing and presentation. The net effect, due largely to the particular mix of local cytokines, is the start of a differentiation and maturation process on the part of B cells that results in a commitment to IgA production. The B cells then leave the Peyer's patches in the efferent lymph and reach the mesenteric lymph nodes. Next, they exit via the efferent lymph to reach the thoracic duct, from which they enter the blood and selectively traffic to the intestines, where they lodge in the lamina propria. Mechanistically, this selective homing reflects particular receptor–counterreceptor interactions between molecules on the B cell surface and the intestinal venule endothelium, as well as local chemotactic factors. In the lamina propria, the B cells complete their differentiation into plasma cells and secrete IgA, mostly in polymeric form. (In contrast, the IgA in serum is mostly monomeric.) The IgA polymers bind to the polymeric immunoglobulin receptor on the basolateral surface of the lining epithelial cells. The IgA–receptor complex is then endocytosed, transcytosed, and released as secretory IgA from the apical surface into the lumen with the IgA attached to the external domain of its receptor. The entire process results in the intestinal secretion of some 40 mg of IgA per kilogram of body weight per day. In addition to secretory IgA, the lumen contains smaller amounts of IgM, which is transported by the same active mechanism, as well as other immunoglobulins that diffuse passively from the lamina propria across the epithelium.

FUNCTIONS OF INTESTINAL IgA

IgA is thought to function in host defense mainly by providing an immune exclusion barrier in the luminal secretions. Here, IgA antibodies help prevent microbes and inert antigens from attaching to or penetrating the epithelium. In addition, during their ongoing transit through epithelial cells to the secretions, IgA antibodies are thought to be able to neutralize viruses that may be infecting the epithelium and transport antigens from the lamina propria through the epithelial cells into the luminal secretions, a kind of excretory immune system. In sum, luminal IgA antibodies should be capable of preventing infections, hence the rationale for mucosal vaccines, whereas the intraepithelial cell neutralization and excretory properties of IgA could assist in the recovery from infections.

A distinctive feature of IgA compared to other classes of antibody is its general noninflammatory tone. For example, unlike IgM and IgG, it activates

 Copyright 2004, Elsevier (USA). All rights reserved.

complement poorly, if at all, and unlike IgE, it does not promote the release of mediators of allergic inflammation. Thus, IgA antibodies can bind to ever-present antigens, such as those of food and the intestinal flora, without provoking inflammation.

See Also the Following References

Endomysial and Related Antibodies • Lymphocytes

Further Reading

Brandtzaeg, P., Farstad, I. N., Johansen, F.-E., Morton, H. C., Norderhaug, I. N., and Yamanaka, T. (1999). The B-cell system of human mucosae and exocrine glands. *Immunol. Rev.* **171**, 45–87.

Lamm, M. E. (1997). Interaction of antigens and antibodies at mucosal surfaces. *Annu. Rev. Microbiol.* **51**, 311–340.

Ogra, P. L., Mestecky, J., Lamm, M. E., Strober, W., Bienenstock, J., and McGhee, J. R. (eds.) (1999). "Mucosal Immunology," 2nd Ed. Academic Press, San Diego.

Imperforate Anus

ROBERT W. CHANG, STEVEN M. ANDREOLI, AND MORITZ M. ZIEGLER
Children's Hospital Boston

appendicostomy An opening that affords access to the large intestine through the tip of the vermiform appendix. It is usually attached to the anterior abdominal wall and is used for irrigation and bowel management in some older patients with bowel dysfunction following imperforate anus repair.

cloaca A common channel for fecal, urinary, and sexual material passage, typically seen in lower vertebrates. In humans, it denotes an embryologic abnormality in which the terminal hindgut fails to divide into the rectum, bladder, and genital primordium.

fistula An abnormal communication between two or more epithelial-lined organs or between an epithelial-lined organ and the surface of the body.

Imperforate anus encompasses a wide range of congenital anorectal malformations involving atresia of the anal canal. The spectrum includes simple perineal fistulas to the complex cloacal malformations. Each anatomic anomaly requires specialized treatment and portends an individual outcome.

INTRODUCTION

Imperforate anus and associated anorectal malformations are well-known entities in pediatric surgery and pose a significant challenge to the clinician. These defects are frequently associated with lifelong debilitating sequelae such as fecal and urinary incontinence and sexual inadequacy.

The pathogenesis of imperforate anus evolves as early as the sixth week of gestation when the urorectal septum migrates caudally to divide the cloaca into the urogenital sinus anteriorly and the anorectal canal posteriorly. Failure of the urorectal septum to form results in an abnormal connection between the gastrointestinal and genitourinary tracts, gender differences notwithstanding. The development of the perineum can also influence the positioning of the anal opening.

Imperforate anus has long been recognized as a surgically correctable defect. In 1835, Amussat sutured the rectal wall to the skin edges, which likely constituted the first known "anoplasty." For the first half of the 20th century, perineal anoplasty without colostomy for "low" malformations and newborn colostomy plus abdominoperineal resection for "high" malformations were preferred treatments. In 1953, Stephens performed the first objective study of human specimens. He proposed an initial sacral approach followed by abdominoperineal resection when necessary. The purpose of this approach was to preserve the puborectalis sling, which was thought to be a key determinant in a successful outcome. Many of the surgical developments of the following period held to these same ideals.

Copyright 2004, Elsevier (USA). All rights reserved.

complement poorly, if at all, and unlike IgE, it does not promote the release of mediators of allergic inflammation. Thus, IgA antibodies can bind to ever-present antigens, such as those of food and the intestinal flora, without provoking inflammation.

See Also the Following References

Endomysial and Related Antibodies • Lymphocytes

Further Reading

Brandtzaeg, P., Farstad, I. N., Johansen, F.-E., Morton, H. C., Norderhaug, I. N., and Yamanaka, T. (1999). The B-cell system of human mucosae and exocrine glands. *Immunol. Rev.* **171**, 45–87.

Lamm, M. E. (1997). Interaction of antigens and antibodies at mucosal surfaces. *Annu. Rev. Microbiol.* **51**, 311–340.

Ogra, P. L., Mestecky, J., Lamm, M. E., Strober, W., Bienenstock, J., and McGhee, J. R. (eds.) (1999). "Mucosal Immunology," 2nd Ed. Academic Press, San Diego.

Imperforate Anus

ROBERT W. CHANG, STEVEN M. ANDREOLI, AND MORITZ M. ZIEGLER
Children's Hospital Boston

appendicostomy An opening that affords access to the large intestine through the tip of the vermiform appendix. It is usually attached to the anterior abdominal wall and is used for irrigation and bowel management in some older patients with bowel dysfunction following imperforate anus repair.

cloaca A common channel for fecal, urinary, and sexual material passage, typically seen in lower vertebrates. In humans, it denotes an embryologic abnormality in which the terminal hindgut fails to divide into the rectum, bladder, and genital primordium.

fistula An abnormal communication between two or more epithelial-lined organs or between an epithelial-lined organ and the surface of the body.

Imperforate anus encompasses a wide range of congenital anorectal malformations involving atresia of the anal canal. The spectrum includes simple perineal fistulas to the complex cloacal malformations. Each anatomic anomaly requires specialized treatment and portends an individual outcome.

INTRODUCTION

Imperforate anus and associated anorectal malformations are well-known entities in pediatric surgery and pose a significant challenge to the clinician. These defects are frequently associated with lifelong debilitating sequelae such as fecal and urinary incontinence and sexual inadequacy.

The pathogenesis of imperforate anus evolves as early as the sixth week of gestation when the urorectal septum migrates caudally to divide the cloaca into the urogenital sinus anteriorly and the anorectal canal posteriorly. Failure of the urorectal septum to form results in an abnormal connection between the gastrointestinal and genitourinary tracts, gender differences notwithstanding. The development of the perineum can also influence the positioning of the anal opening.

Imperforate anus has long been recognized as a surgically correctable defect. In 1835, Amussat sutured the rectal wall to the skin edges, which likely constituted the first known "anoplasty." For the first half of the 20th century, perineal anoplasty without colostomy for "low" malformations and newborn colostomy plus abdominoperineal resection for "high" malformations were preferred treatments. In 1953, Stephens performed the first objective study of human specimens. He proposed an initial sacral approach followed by abdominoperineal resection when necessary. The purpose of this approach was to preserve the puborectalis sling, which was thought to be a key determinant in a successful outcome. Many of the surgical developments of the following period held to these same ideals.

Copyright 2004, Elsevier (USA). All rights reserved.

Pena and colleagues performed the first posterior sagittal approach for anorectal malformations in 1980 (posterior sagittal anorectoplasty, or PSARP). Its description, published in 1982, has been time-tested and has vastly changed the treatment for this group of anatomic defects. This technique allowed direct exposure of the anatomy and enabled correlation of perineal appearance with operative findings. This has led to a terminology and classification scheme with prognostic importance.

Imperforate anus is seen in 1 in every 4000–5000 newborns. The frequency of this defect is slightly higher in males, with the most common defect being an imperforate anus with a rectourethral fistula. In females, the most common defect is a rectovestibular fistula. Congenital absence of a fistula is rare, seen in only 5% of all cases. In families with a newborn with an imperforate anus, the risk of having a second child with this affliction is approximately 1%.

TYPES OF DEFECTS

Historically, anorectal malformations were classified using three main descriptors: low, intermediate, and high. These terms loosely referred to the level of rectal fistula. These terms were not standardized and, more importantly, did not address treatment implications or outcome. With the development of the posterior sagittal approach by Pena and colleagues, a new classification scheme has been popularized. It offers anatomic descriptors based on gender, is based on therapeutic and diagnostic facts, and is of prognostic significance.

Male

Perineal Fistula

In this defect, the lowest part of the rectum is anteriorly displaced (see Fig. 1). The fistula most often opens onto the perineum anterior to the location of the "normal anus," somewhere along the midline raphe, at the scrotum, or even at the base of the penis. The diagnosis is made primarily by perineal inspection. The descriptors "anal dimple," "bucket handle," and "anal membrane" are used to refer to the external appearance of the perineal area. These patients require no further testing. One hundred percent of these patients achieve satisfactory bowel control after repair.

Rectourethral Fistula

This is the most frequent variant seen in males (see Figs. 2 and 3). The fistula is located at or near the bulbar region of the urethra or near the prostatic urethra. Immediately superior, the rectum and urethra share a common wall. Between the often distended rectum and the skin lies a band of striated voluntary muscle. Lower urethral fistulas are associated with higher quality musculature, a well-developed sacrum, and a prominent

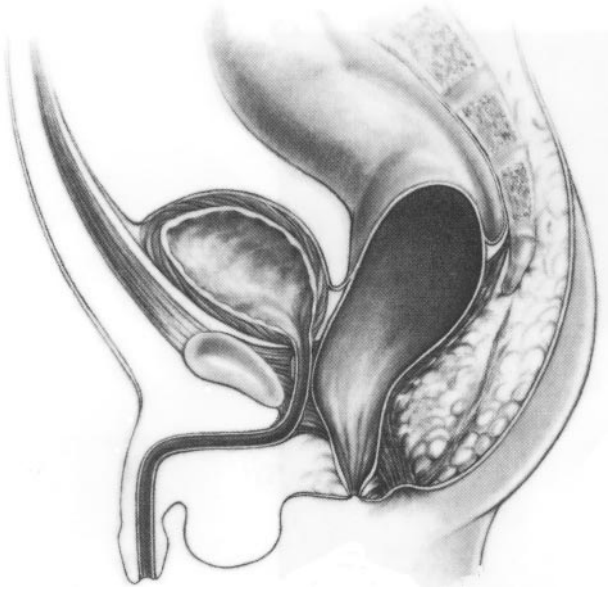


FIGURE 1 Low defect, perineal fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

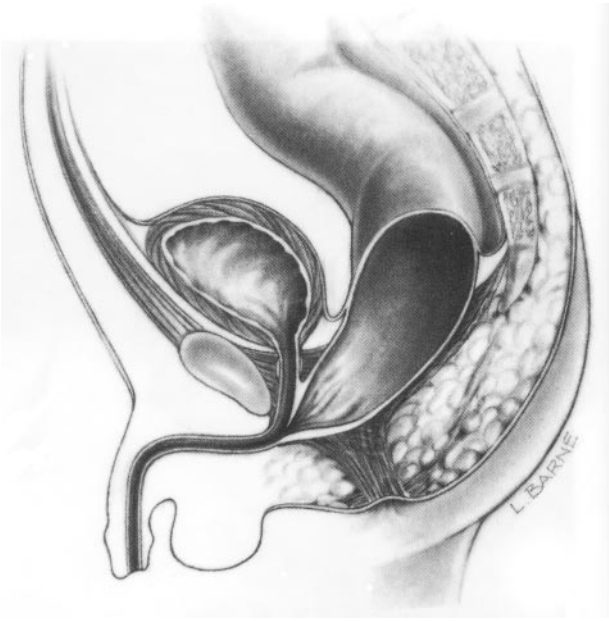


FIGURE 2 Rectourethral bulbar fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

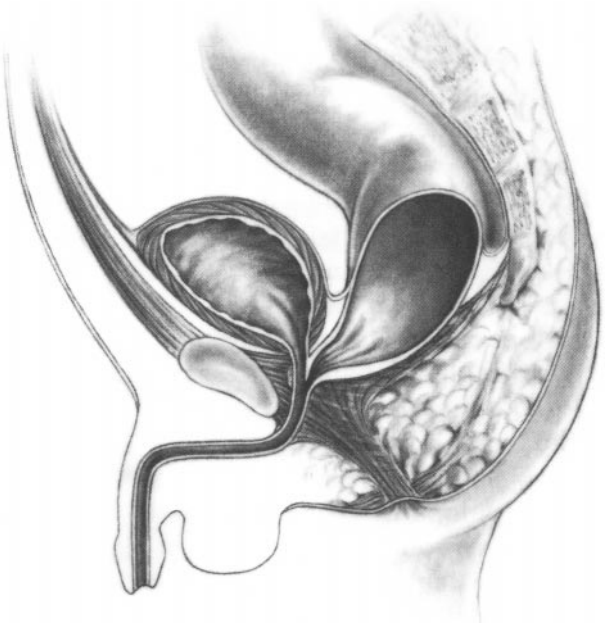


FIGURE 3 Rectourethral prostatic fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

anal dimple. Patients with this defect can frequently be observed to pass meconium or air (pneumaturia) through the urethra, an unequivocal sign.

Rectovesical Fistula

In this higher level defect, the rectum opens at the bladder neck (see Fig. 4). These patients have poor prognoses due to poor development of the levator muscle, muscle complex, and external sphincter. Ten percent of males with imperforate anus have this variant.

Anorectal Agenesis

This rare defect carries a good prognosis due to a well-developed sacrum and good muscle structures. The rectum usually ends 2 cm from the perineal skin. Approximately half of these patients suffer from Down's syndrome. This fact does not interfere with the high likelihood of achieving bowel control.

Rectal Atresia

This type is seen in approximately 1% of all anorectal malformations and has an excellent prognosis. The lumen of the rectum may be totally or partially interrupted. Patients usually possess all of the necessary elements to be continent. Usually, the anal canal is well developed and normal sensation is observed in the anorectum.

Female

Perineal Fistula

From a therapeutic and prognostic standpoint, this defect in females parallels the male situation (see Fig. 5). The rectum is located well within the sphincter mechanism and the rectum and vagina are well separated, with the anus being "anterior" to its normal location.

Vestibular Fistula

In this defect, the bowel opens immediately behind the hymen in the vestibule of the genital system (see Figs. 6–8). Immediately above the fistula, the rectum and vagina are separated by a thin, common wall. This condition is often initially misdiagnosed as a rectovaginal fistula, which is much more rare, at 1% of all defects. The correct diagnosis is usually derived by meticulous inspection of the newborn genitalia. These patients usually have well-developed muscles and a normal sacrum.

Anorectal Agenesis without Fistula

This very unusual defect carries the same implications for treatment and outcome as in males, close to 100% incidence of bowel control.

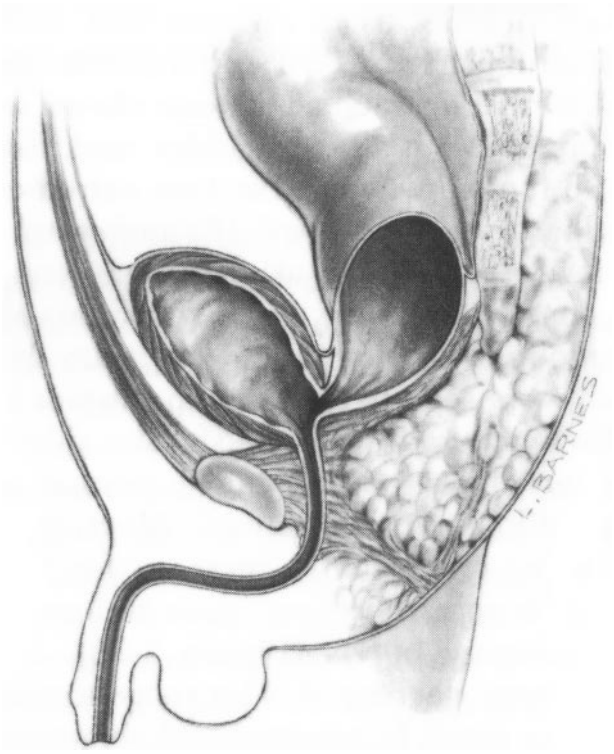


FIGURE 4 Rectobladder-neck fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

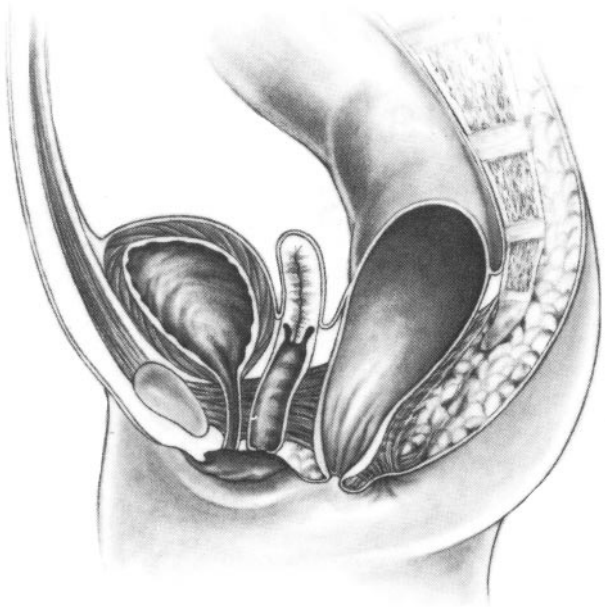


FIGURE 5 Perineal (cutaneous) fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

Persistent Cloaca

This condition is usually a clinical diagnosis and should be strongly suspected in a female newborn with imperforate anus and small-looking genitalia (see Fig. 9). In this group of anomalies, the rectum, vagina, and urinary tract meet and fuse into a single common channel. Careful inspection reveals a single perineal orifice usually hidden behind the labia. The length of the common channel varies, a factor that has prognostic significance. A common channel longer than 3.5 cm usually represents a complex defect and implies difficult mobilization of the vagina. A common channel less than 3.4 cm usually signifies that repair can be accomplished without need for laparotomy (PSARP only). There are numerous other complicated variants of this defect that pose significant challenges to the pediatric surgeon.

ASSOCIATED ANOMALIES

The two main associated abnormalities usually encountered with anorectal defects involve the skeletal and genitourinary systems. Spinal or sacral deformities seem to be frequently related to anorectal malformations. These occur in approximately one-third of all patients and include absence or asymmetry of vertebrae or accessory or hemi-vertebrae. Usually, one missing vertebra does not worsen outcome. However, more than two absent sacral vertebrae is a prognostic sign for poor continence and urinary control. A radiographic

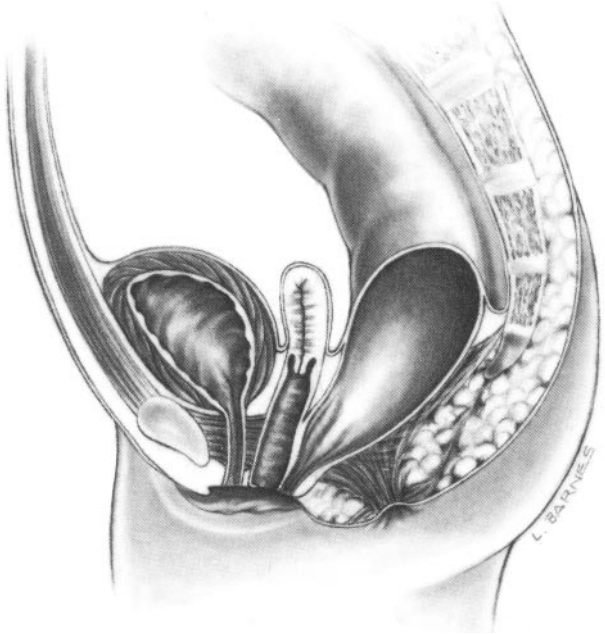


FIGURE 6 Vestibular fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

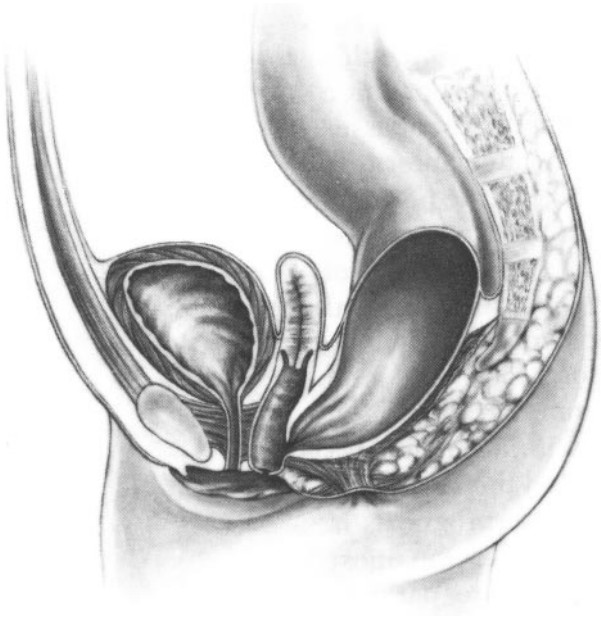


FIGURE 7 Low rectovaginal fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

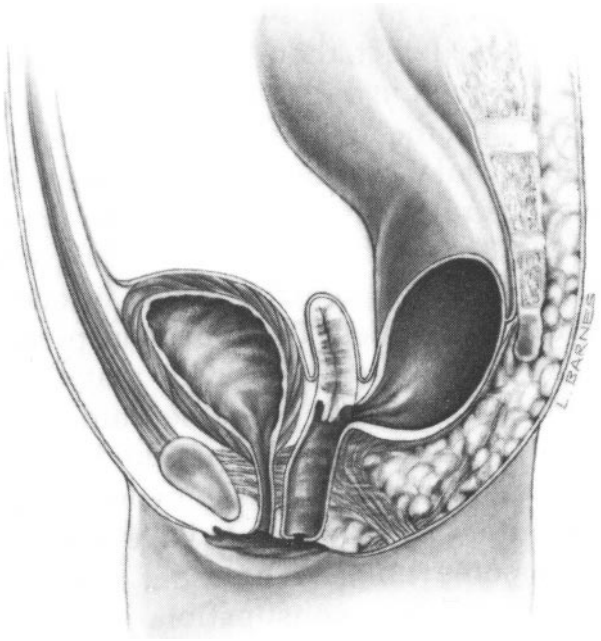


FIGURE 8 High rectovaginal fistula. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

sacral ratio has been developed for objective evaluation. It is calculated from a lateral film: a line drawn from upper and lower borders of the iliac crest is compared with a line drawn from the lowest part of the iliac crest to the lowest point on the sacrum. A ratio is made of this upper to lower portion, with the normal ratio being 0.7; poor prognosis for bowel control is predicted in patients with a ratio less than 0.3.

Concomitant genitourinary (GU) defects are seen in 20 to 54% of patients with anorectal malformations. The "higher" the malformation, the more frequent the association. Patients with rectovesical fistulas have a 90% incidence of a GU defect compared to a less than 10% incidence in perineal defects. Vesicoureteral reflux and hydronephrosis are the most common abnormalities, but other findings such as absent, dysplastic, or horseshoe kidneys must also be considered. Hydronephrosis, urosepsis, and metabolic acidosis are main sources of morbidity and mortality in affected newborns. Thus, a urologic examination, though always required, has a much higher priority in patients with higher defects. This includes a renal and abdominal ultrasound to look for hydronephrosis or other obstructive processes. A voiding cystourethrogram may be useful to localize the level of rectourethral fistula in males and it may also diagnose vesicoureteral reflux.

PRESENTATION AND DIAGNOSIS

The first step in management is to determine whether the newborn has a life-threatening anomaly. Next is the decision to perform a colostomy before definitive repair is addressed. Perineal inspection is usually sufficient to make the initial decision regarding establishment of a protective colostomy in up to 80% of boys. Initial diagnostic measures are essentially the same in females. Ninety to 95% of newborn females can be diagnosed by visual inspection alone. The presence of a single perineal orifice signifies a cloaca, which requires emergent urinary drainage/diversion and protective colostomy.

Occasionally, the diagnosis is not apparent at birth. Final evaluation is delayed 16–24 h to allow bowel distension and intralumenal pressure to increase so that air and meconium are forced through to the distal gut and potentially the urethra.

A prone cross-table lateral X ray is then obtained. The presence of a bowel pouch that is more than 1 cm from the anal skin is an indication for colostomy. On occasion, perineal ultrasound is used to define this level. The diagnostic workup must also include a high index of suspicion for esophageal atresia, which exists in 5% of patients. The completely diverting colostomy provides bowel decompression and urinary tract protection

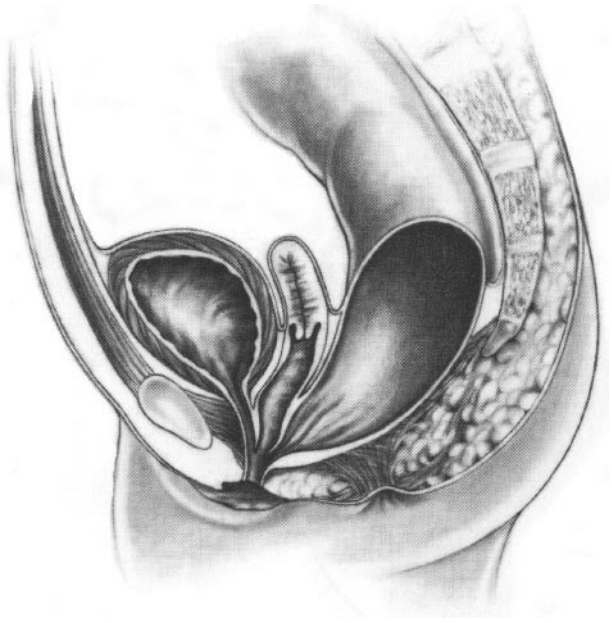


FIGURE 9 Typical cloaca. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

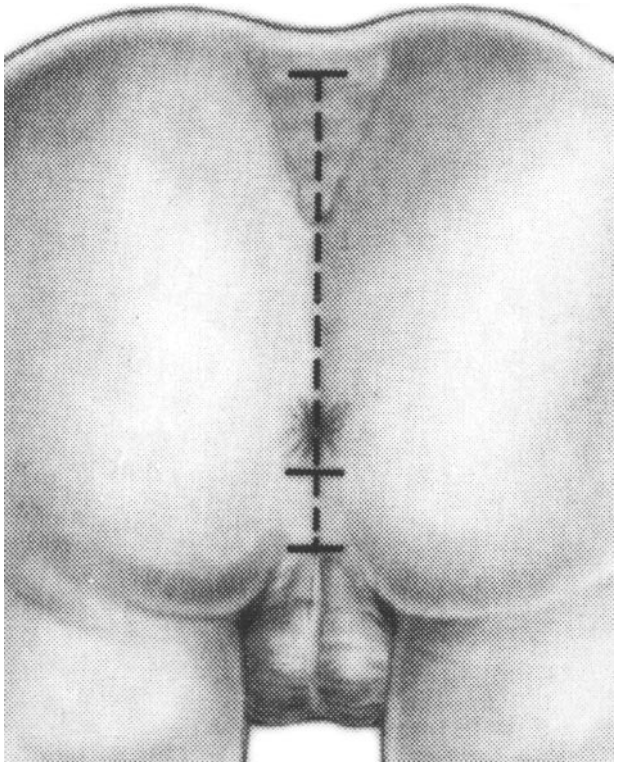


FIGURE 10 Posterior sagittal incision (for PSARP). Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

while waiting for the future repair. Risk of infection and dehiscence is decreased and mechanical cleansing is made easier. The procedure of choice is the divided descending colostomy. The operating surgeon must take care to provide enough length of rectosigmoid colon for adequate mobilization in the final repair. In addition, this procedure creates an additional stoma, which allows for an essential preoperative colostogram to be performed. The colostogram is the most accurate diagnostic study to determine the detailed anatomy of these defects, particularly the level of recto-vaginal or recto-urinary connection. Definitive repair is usually performed 4 to 6 weeks after colostomy placement. Though primary neonatal repair of even "high" imperforate anus anomalies has been performed, avoiding colostomy placement, its safety and efficacy are unproven.

Surgical Considerations

As mentioned above, previous corrective operations sought to preserve the puborectalis sling in order to preserve bowel control. Others have advocated endorectal dissection to avoid damaging the pelvic nerves. The standard of care today is the posterior sagittal anorectoplasty (PSARP) mentioned above (see Figs. 10–13). All anorectal malformations can be corrected by this approach with or without laparotomy. The patient is placed in the prone position with the pelvis elevated. The use of the electric muscle stimulator to elicit muscle contraction is essential; these contractions serve as a guideline for operating in the midline, leaving an equal amount of muscle on both sides. This is based on the principle that no important nerves or vessels cross the midline. In addition, a fine midline fascia divides the anatomy in two parts. As part of the PSARP, the voluntary striated muscles from both sides are separated rather than divided. This allows exposure and reduces the likelihood of damage to adjacent structures such as the vas deferens, ectopic ureters, prostatic tissue, urethra, and seminal vesicles. The PSARP proceeds by identifying, opening, and then dividing the rectourethral fistula followed by a "pull-through" of the terminal rectum into the normal, newly created anal position. The pelvic muscle complex (levator ani and parasagittal fibers of the external sphincter complex) is reconstructed around the rectum and a cutaneous anoplasty is performed where indicated by electrical stimulation to complete the pull-through.

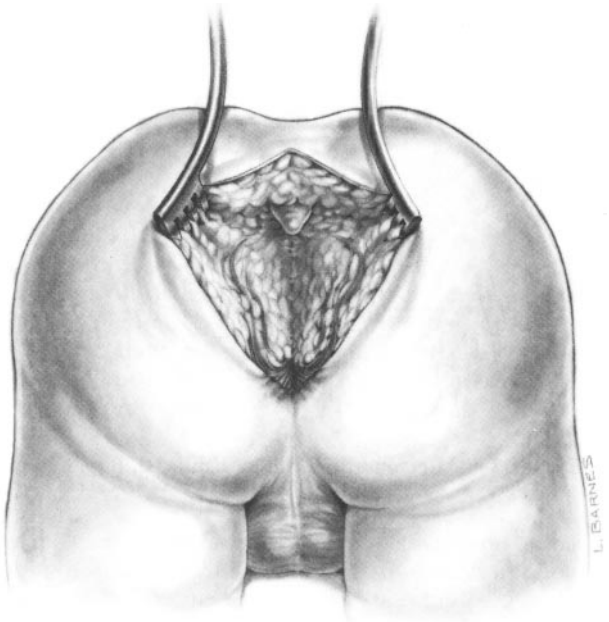


FIGURE 11 Incision continued through the center of the external sphincter showing the lowest part of the muscle complex (vertical fibers). Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

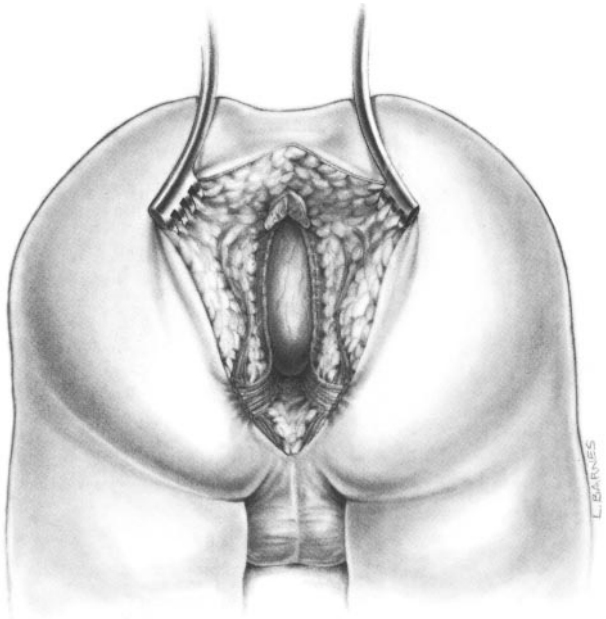


FIGURE 12 Divided levator muscle, muscle complex, and coccyx. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

Approximately 90% of defects in boys can be repaired via the PSARP without opening the abdomen. A posterior sagittal approach should never be attempted without a technically adequate distal colostogram to determine the exact position of the rectum and the fistula. Attempting repair without this vital information poses significant risk for potential nerve damage, damage to the seminal vesicles and prostate, and bladder denervation.

Most patients experience a smooth recovery period. Patients are usually discharged from the hospital 3 to 4 days after PSARP, with longer stays for those who underwent laparotomy. Two weeks after repair, the patient is brought to the office, where anal dilations are started. This is taught to the parents and is initially performed twice a day. It is essential to have adequate follow-up during this period as severe unmanageable stricturing can result from inadequate dilation. Once the rectum is dilated to the age-appropriate size, the colostomy is closed in the third and final stage.

Urinary incontinence occurs as an overflow phenomenon in two-thirds of cloaca patients with common channel longer than 3 cm and in 20% of patients with a common channel <3 cm. Intermittent catheterization usually keeps these patients dry. In males, incontinence after imperforate anus repair usually represents inadequate primary neuromuscular development, poor surgical technique, or an element of nerve damage.

POSTOPERATIVE MANAGEMENT

Most patients who undergo operative repair for imperforate anus suffer from some departure from normal bowel function. Fecal continence depends mainly on voluntary muscle structures, sensation, and bowel motility. An intact sensory feedback mechanism as well as a muscle contraction mechanism is necessary for the complex procedure of defecation. Most imperforate anus patients do not have the highly sensitive neural feedback in the anal canal that normal patients do, as discrimination between liquid stool and soft fecal material is often severely impaired. Bowel motility is perhaps the most important factor in achieving fecal continence. The main clinical manifestation in after PSARP is constipation, afflicting those with "lower" defects. Interestingly, patients with the best prognosis have the highest incidence. Constipation may be directly related to the

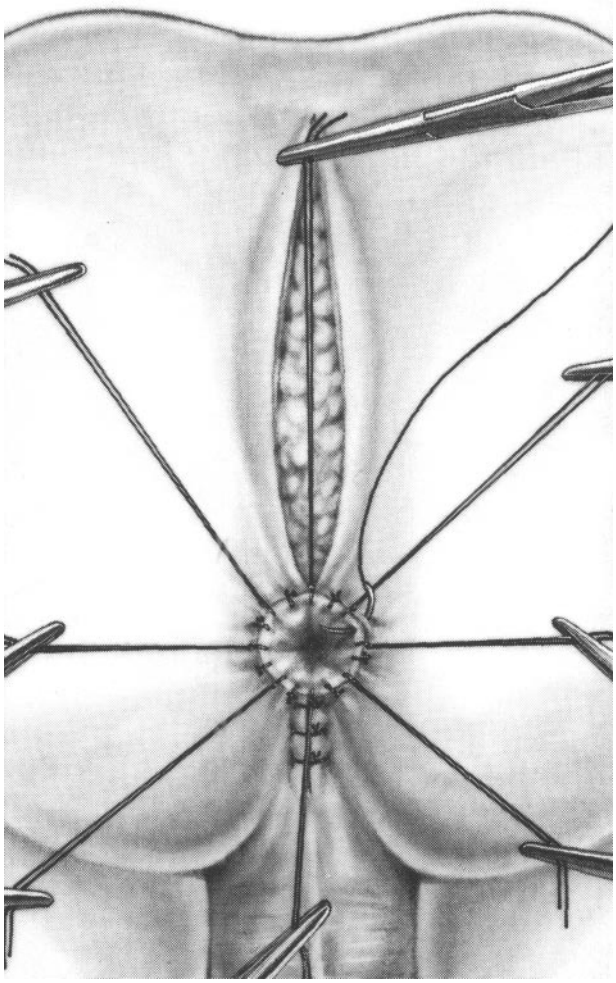


FIGURE 13 Anoplasty. Reprinted from Pena, A. (1992). "Atlas of Surgical Management of Anorectal Malformations." Springer-Verlag, New York, with permission. Copyright Springer-Verlag.

degree of rectal ectasia. Efforts to keep the rectosigmoid as decompressed as possible from the time the colostomy is established result in better ultimate bowel function. Creating and implementing a bowel management program is a trial-and-error process that must be tailored to each individual. For some, obstipation is relieved by frequent enemas. In older patients who are best managed by greater independence, appendicostomy has been performed, allowing older children to administer the treatments themselves. Still other patients may suffer from a hypermotility state and require anti-motility agents.

OUTCOMES

Pena and colleagues have published the largest series of imperforate anus cases. Of the 1192 patients, 75% had voluntary bowel movements by the age of 3. Of these, half occasionally soil. Thus, 37.5% overall are totally continent. Twenty-five percent of all patients are totally incontinent and require bowel management.

There have been two recent developments that may have benefit. Early PSARP has been performed without colostomy in the newborn period to allow infants to gain the experience of normal defecation without colostomy. Second, for the 10% of patients who require laparotomy with PSARP, a laparoscopic approach has recently been proposed to reduce posterior dissection required for accurate placement of the bowel into the muscle complex. Both modifications have uncertain efficacy and long-term data are required.

See Also the Following Articles

Anal Canal • Fecal Incontinence • Fistula

Further Reading

Georgeson, K. E. I., and Albanese, C. T. (2000). Laparoscopically assisted anorectal pull-through for high imperforate anus—A new technique. *J. Pediatr. Surg.* **35**, 927–931.

Pena, A. (2000). Imperforate anus and cloacal malformations. *In* "Pediatric Surgery" (K. W. Ashcraft, ed.), 3rd Ed., pp. 473–492. W. B. Saunders, Philadelphia, PA.

Pena, A., and Devries, P. A. (1982). Posterior sagittal anorectoplasty: Important technical considerations and new applications. *J. Pediatr. Surg.* **17**, 796–811.

Pena, A., and Hong, A. (2000). Advances in the management of anorectal malformations. *Am. J. Surg.* **180**, 370–376.

Pena, A., Guardino, K., Tovilla, J. M., Levitt, M. A., Rodriguez, G., and Torres, R. (1998). Bowel management for fecal incontinence in patients with anorectal malformations. *J. Pediatr. Surg.* **33**, 133–137.

Interstitial Cells of Cajal

SEAN M. WARD AND KENTON M. SANDERS
University of Nevada School of Medicine

electrical slow waves Phasic electrical activity underlying generation of spontaneous mechanical activity.
gastrointestinal stromal tumors The most common mesenchymal tumors in the gastrointestinal tract.
interstitial cells of Cajal Responsible for the generation of slow-wave activity and serve as intermediaries in enteric motor neurotransmission.
Kit Type III tyrosine kinase receptor expressed on populations of interstitial cells of Cajal throughout the gastrointestinal tract.
omental mesenchymal tumors Thought to represent gastrointestinal stromal tumors of the omentum.

The interstitial cells of Cajal are specialized cells distributed in specific locations within the tunica muscularis of the gastrointestinal tract; these cells serve as electrical pacemakers, as active propagation pathways for slow waves, and as mediators of enteric neurotransmission. Mutations in these cells are also thought to be responsible for the generation of tumors within the tunica muscularis and omentum of the gastrointestinal tract.

INTRODUCTION

Until recently, evidence supporting a functional role of the interstitial cells of Cajal (ICCs) in pacemaking and neurotransmission within the gastrointestinal tract was indirect. Electron microscopy has now revealed ICCs within the tunica muscularis in sites where electrical slow waves are thought to originate. Morphological studies have also revealed close relationships between ICCs and varicose nerve fibers, and immunohistochemical techniques have shown that ICCs possess receptors for several neurotransmitters and display responses to neurotransmitters. Using a variety of specific molecular markers and mutant animal models, more direct physiological evidence now supports the hypothesis that specific populations of these cells are responsible for the generation of electrical slow waves; discrete populations of ICCs also act as intermediates between enteric nerves and smooth muscle cells in motor neurotransmission within the gastrointestinal (GI) tract.

LOSS OF ICCs, SLOW WAVES, AND NEUROTRANSMISSION IN KIT MUTANT MICE

Several classes of ICCs of the mouse and other species express a type III tyrosine kinase receptor (Kit). Activation of Kit by its ligand, stem cell factor, or "steel" factor, is important in the development of ICCs because blockade of Kit with neutralizing antibodies or mutations in c-*kit* or *steel* impair the development of ICCs. Loss of ICCs in the myenteric plexus region (ICC-MY) of the small intestine, for example, causes abnormal electrical activity that includes total loss of electrical slow-wave activity. In the white spotting mutation (W/W^V), ICCs within the circular and longitudinal muscle layers (ICC-IM) of the stomach and lower esophageal and pyloric sphincters fail to develop. Cholinergic excitatory and nitrergic inhibitory motor neurotransmissions are compromised in tissues of these animals, providing functional evidence for the idea, originally proposed by Cajal, that ICCs mediate enteric neurotransmission.

ICCs ARE DISRUPTED IN PATIENTS WITH GI MOTILITY DISORDERS

The discovery that ICCs express Kit has allowed simple morphological identification of these cells with light microscopy. Labeling of Kit receptors or c-*kit* mRNA has provided an efficient means of identifying ICCs throughout the GI tracts of several species, including humans. Kit labeling combined with confocal laser imaging techniques has improved our understanding of the structure and distribution of ICC networks (see Fig. 1) and has enhanced our perception of the anatomical relationships between ICCs and with other cell types, such as enteric neurons and smooth muscle cells. Kit labeling of ICCs has allowed pathologists to determine the viability of ICCs in a variety of GI motility disorders. ICC networks are disrupted in several motility disorders, including achalasia, hypertrophic pyloric stenosis, chronic idiopathic intestinal pseudo-obstruction, ulcerative colitis, Crohn's disease, and

 Copyright 2004, Elsevier (USA). All rights reserved.

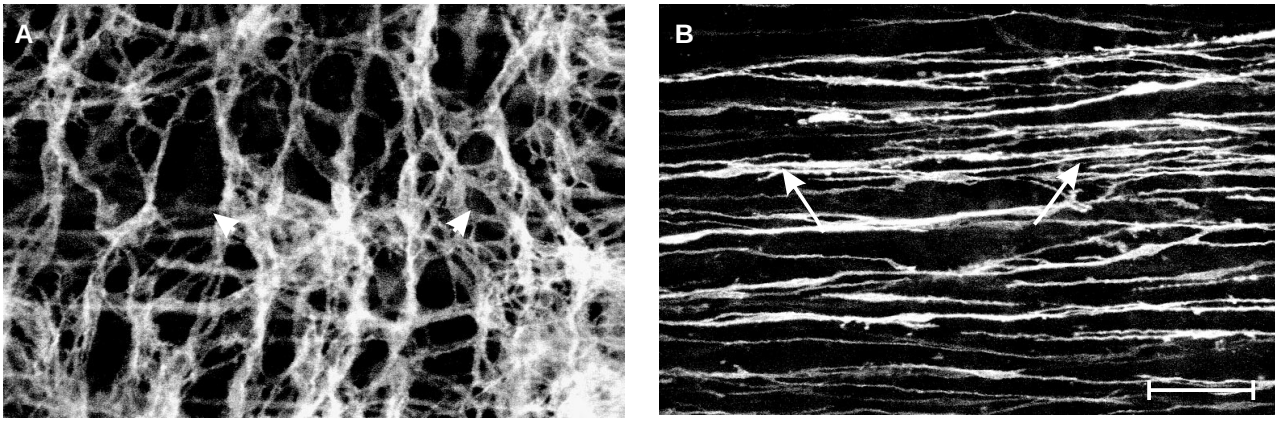


FIGURE 1 Distinct ICC populations in the murine gastric antrum revealed using whole-mount preparations and Kit immunohistochemistry. (A) ICCs are seen at the level of the myenteric plexus (arrowheads), between the circular and longitudinal muscle layers. These cells are the pacemaker cells responsible for the generation of electrical slow waves in the stomach. (B) Intramuscular ICCs are seen within the circular muscle layer (arrows) of the antrum. These cells are responsible for mediating enteric neurotransmission in the stomach (scale bar in panel B applies to both panels).

Hirschsprung's disease. Recent findings in the murine animal model and in human cases have also shown that ICCs are also disrupted in insulin-dependent diabetes.

MANIPULATION OF THE DEVELOPMENT OF ICCs

A powerful means of investigating the physiological role of ICCs is the ability to manipulate the development of these cells. Developmental studies have revealed that Kit expression within the tunica muscularis begins midway through gestation, and ICC networks are formed prior to birth throughout the GI tract. Development of specific populations of ICCs and electrical rhythmicity continues after birth. Blocking of the ICC networks in newborn animals with Kit neutralizing antibodies disrupts ICC networks, electrical rhythmicity, and enteric neurotransmission in the murine jejunum, suggesting that the Kit signaling pathway is critical for the development and maintenance of the ICC phenotype. Investigations using animal models have also revealed that ICCs do not die in some motility disorders, but rather they redifferentiate into a more smooth muscle-like phenotype. These observations provide the exciting possibility that ICCs may not be permanently lost in some motility disorders, and it may be possible to restore these cells and their function in some cases.

SLOW-WAVE PACEMAKING MECHANISM IN ICCs

Physiological studies of ICCs have demonstrated that pacemaker currents generated by ICCs result primarily from the activation of a nonspecific cation conductance in the plasma membrane. The nature of this current is not fully understood, but the current is voltage independent and may conduct Na^+ and Ca^{2+} in physiological ionic gradients. The mechanism that initiates slow-wave activity has been an ongoing subject of investigation for many decades. ICCs contain an abundance of mitochondria and it has been assumed by some that slow waves are tied to the metabolic activity of ICCs (e.g., the mechanism regulating the pacemaker current is metabolically controlled). Others have suggested that pacemaker activity depends on Ca^{2+} release from intracellular stores. A link between Ca^{2+} release, metabolic activity, and current induction in the plasma membrane has recently been made and it is now thought that there is a close physiological relationship between Ca^{2+} release from inositol 1,4,5-trisphosphate (IP_3) receptor-operated stores, Ca^{2+} uptake by mitochondria, and initiation of pacemaker currents in ICCs. Tissues of several species appear to have this mechanism in common. Thus, integrated Ca^{2+} handling by the endoplasmic reticulum (ER) and mitochondria are required for electrical pacemaking in the GI tract. See Fig. 2 for a summary of the current knowledge of ICC function in the GI tract.

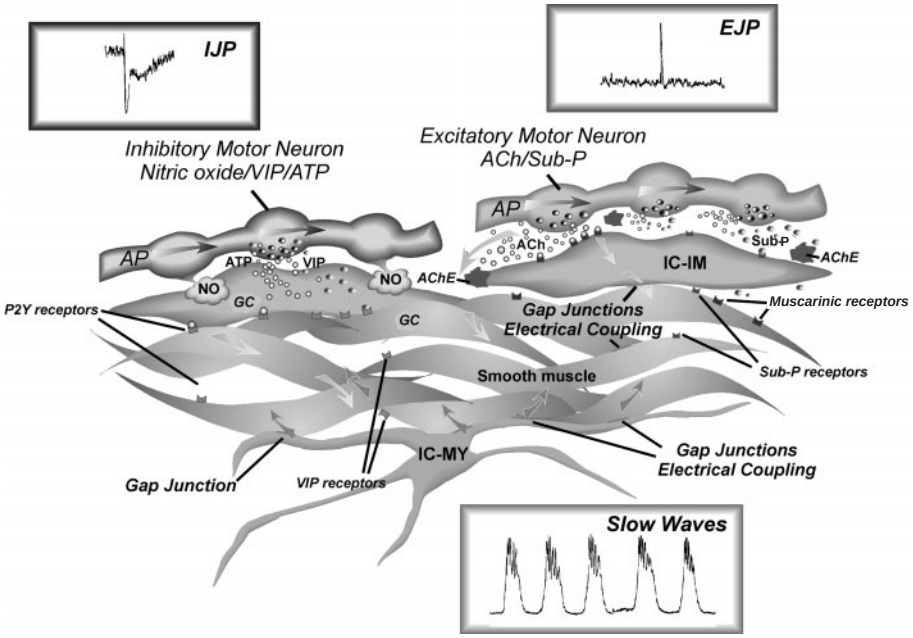


FIGURE 2 Model illustrating the role of ICCs in slow-wave generation and propagation and as mediators of enteric motor neurotransmission. Data summarize findings from several different animal species. IJP, Inhibitory junction potential; EJP, excitatory junction potential; AP, action potential; NO, nitric oxide; VIP, vasoactive intestinal peptide; GC, guanylyl cyclase; ACh, acetylcholine; AChE, acetylcholine esterase; Sub-P, substance P; ICC-IM, interstitial cells of Cajal of the circular and longitudinal muscle layers; IC-MY, interstitial cells (of Cajal) of the myenteric plexus region. Modified from Ward, S. M. (2000). Role of ICC in gastric motility and neuromuscular dysfunctions. *In* "Evolving Pathophysiological Models of Functional GI Disorders," p. 32. With permission of Health Education Alliance and International Foundation for Functional Gastrointestinal Disorders.

ICCs ARE POSSIBLE PRECURSORS TO STROMAL AND OMENTAL TUMORS

The discovery that ICCs express Kit has allowed pathologists to examine the origin of GI stromal tumors (GISTs) and omental mesenchymal tumors (OMTs). GISTs are the most common mesenchymal tumors in the GI tract and OMTs are thought to represent GISTs of the omentum. It has recently been found that both GISTs and OMTs express Kit and it has been proposed that both GISTs and OMTs originate from ICCs. Mutations of the c-*kit* gene (somatic gain-of-function gene) have been observed in solitary GISTs, and germ-line gain-of-function mutations of the c-*kit* gene have also been observed in familial and multiple GISTs. The gain-of-function mutations of the c-*kit* gene are considered to be a cause of the development of these human GI tumors.

ICCs at the level of the myenteric plexus in the stomach, small intestine, and colon and at the level of the submucosal plexus in the colon are the pacemaker cells in the gastrointestinal tract. These cells generate electrical slow waves. Pacemaker ICCs have unique ionic conductances that generate the inward currents underlying slow waves. The intracellular mechanisms responsible for the generation of the pacemaker "clock" involve Ca^{2+} release from IP_3 receptor-operated stores, Ca^{2+} uptake by mitochondria, and initiation of pacemaker currents. Smooth muscle cells do not have the slow-wave mechanism and cannot be induced to produce slow waves in the absence of ICCs. ICCs and the smooth muscle cells are electrically coupled via gap junctions. Slow waves conduct to neighboring smooth muscle cells and activate voltage-dependent Ca^{2+} channels. Influx of Ca^{2+} through these channels leads to muscle contraction.

ICCs are also necessary for the propagation of slow-wave activity around and along the phasic organs of the GI tract. Slow-wave amplitude is maintained in the ICC networks but decays in the smooth muscle syncytium when ICC networks are disrupted.

Morphological and electrophysiological data suggest that the majority of neuromuscular junctions in GI muscles are composed of excitatory and inhibitory enteric nerve terminals, intramuscular interstitial cells of Cajal, and smooth muscle cells. When action potentials invade varicosities, transmitters are released. The close apposition between nerve terminals and ICC-IMs facilitates rapid diffusion to ICC receptors. ICC-IMs are coupled to smooth muscle via gap junctions, and electrical responses elicited in ICCs are conducted to smooth muscle cells. In smooth muscle cells, excitatory depolarization responses, or excitatory junction potentials (EJPs), enhance excitability and increase Ca^{2+} influx; inhibitory hyperpolarization responses, or inhibitory junction potentials (IJPs), reduce excitability and block contraction. ICC-IMs express receptors for many neurotransmitters. Receptors for the inhibitory transmitters [nitric oxide (NO), soluble guanylyl cyclase (GC), vasoactive intestinal peptide (VIP), and ATP] and the excitatory transmitters [acetylcholine (ACh) and substance P (Sub-P)] are shown in Fig. 2.

Acknowledgments

The authors are very grateful for the artwork provided by Elizabeth A. H. Beckett and Geoffrey D. Sanders.

See Also the Following Articles

Duodenal Motility • Gastric Motility • Sensory Innervation

Further Reading

Hirota, S., Isozaki, K., Moriyama, Y., Hashimoto, K., Nishida, T., Ishiguro, S., Kawano, K., Hanada, M., Kurata, A., Takeda, M., Muhammad Tunio, G., Matsuzawa, Y., Kanakura, Y., Shinomura, Y., and Kitamura, Y. (1998). Gain-of-function mutations of *c*-kit in human gastrointestinal stromal tumors. *Science* **279**, 577–580.

Hirst, J. D. and Ward, S. M. (2003). Interstitial cells: Involvement in rhythmicity and neural control of gut smooth muscle. *J. Physiol.* (in press).

Sanders, K. M. (1996). A case for interstitial cells of Cajal as pacemakers and mediators of neurotransmission in the gastrointestinal tract. *Gastroenterology* **111**, 492–515.

Sanders, K. M., Torihashi, S., Ordog, T., Koh, S. D., and Ward, S. M. (1999). Development and plasticity of interstitial cells of Cajal. *Neurogastroenterol. Motil.* **11**, 311–338.

Ward, S. M., and Sanders, K. M. (2001). Interstitial cells of Cajal: Primary targets of enteric motor innervation. *Anat. Rec.* **262**, 125–135.

Ward, S. M., and Sanders, K. M. (2001). Physiology and pathophysiology of the interstitial cell of Cajal: From bench to bedside. III. I. Functional development and plasticity of interstitial cells of Cajal networks. *Am. J. Physiol.* **281**, G602–G611.

Ward, S. M., Ordog, T., Koh, S. D., Baker, S. A., Jun, J. Y., Amberg, G., Monaghan, K., and Sanders, K. M. (2000). Pacemaking mechanism driven by a mitochondrial clock. *J. Physiol.* **525**, 355–361.

Intestinal Atresia

MORITZ M. ZIEGLER*,† AND JENNIFER J. GARZA*
*Children's Hospital Boston and †Harvard Medical School

stenosis Incomplete obstruction, with a small opening secondary to a diaphragm or web.
web Thin sheet of tissue or large mucosal fold protruding into the lumen of a hollow organ, often causing obstruction.

An atresia is a complete obstruction caused by the filling of a natural opening; in an intestinal atresia, the obstruction is caused by pancreatic tissue or simply may be a muscular continuity of the duodenum without a lumen. There are three types of intestinal atresia: duodenal, jejunoileal, and colonic.

DUODENAL ATRESIA

Embryology

The embryologic cause of duodenal stenosis and atresia is thought to be a failure of recanalization. The formation of the hepatobiliary system and pancreas occurs in the third week of gestation, when the second portion of the duodenum gives rise to biliary and pancreatic buds at the junction of the foregut and midgut. The duodenum also undergoes a luminal reduction or obliteration during this time, and it is between the eighth and tenth weeks of gestation that the duodenal lumen is reestablished by recanalization. It is believed that an insult during this crucial time period leads to stenoses, atresias, and webs.

Classification

Atresia of the duodenum may result in formation of a short, fibrous cord that connects two atretic ends or may result in two completely disconnected atretic ends. Duodenal atresias, unlike more distal atresias, are often associated with other anomalies. In fact, 50% of duodenal atresias are associated with cardiac, anorectal, or genitourinary anomalies, and it is believed that up to 40% of affected patients have trisomy 21. With these factors in mind, there should be a low threshold for testing for suspected associated anomalies. It is generally the cardiac defects that are related to mortality.

Duodenal obstructions may be either preampullary or postampullary. However, a majority are considered periampullary. The degree of obstruction dictates the amount of resulting pathology. The obstruction causes dilatation of the proximal duodenum and stomach as well as hypertrophy and distension of the pylorus. The distal bowel, as with most obstructions, is collapsed. A common variation is the windsock anomaly, whereby the duodenum is dilated distal to the origin of the point of obstruction because of a prolapsing web or membrane. Heralding the lesions are maternal polyhydramnios and growth retardation of the fetus, the latter possibly due to the inability of the fetus to swallow amniotic fluid, which contains nutritive peptides.

Diagnosis

Prenatal diagnosis is possible by several methods. Ultrasound examination may indicate structural and associated abnormalities, such as a dilated stomach and proximal duodenum. Commonly associated anomalies may also be found by screening maternal serum and amniotic fluid for elevated levels of α-fetoprotein. Amniotic fluid and fetal blood are useful for karyotyping.

In the newborn, clear or bilious emesis is evident within hours of birth and as a result abdominal distension may or may not be present. As soon as intestinal obstruction is suspected, a nasogastric tube should be inserted. An output of more than 20 ml of gastric contents is indicative of possible obstruction. A delay in diagnosis is not uncommon with a stenosis or web, because these may present later as dehydration or failure to thrive. In an older child with unexplained recurrent episodes of emesis, a stenosis or web should be excluded.

Abdominal plain radiographs are helpful in making the diagnosis. Evidence of gas beyond the duodenum indicates a partial obstruction. The classic "double bubble" is associated with complete obstruction, but is not exclusive to atresia. If a partial obstruction is suspected, it is necessary to rule out malrotation and volvulus to determine if immediate operation is necessary. If surgery is not emergently indicated, a contrast study is necessary. Upright films using air or contrast medium may confirm the diagnosis.

 Copyright 2004, Elsevier (USA). All rights reserved.

Treatment

The patient diagnosed with duodenal atresia should be fully resuscitated before surgical correction is attempted, unless volvulus or malrotation is detected, in which case surgery becomes emergent. Ideally, correction of fluid status and electrolyte abnormalities should be handled preoperatively. Replacement fluid should be given and urine output should be monitored closely. Gastric decompression is essential to prevent aspiration, and thermoregulation should be monitored at all times. Screening for concurrent abnormalities should be done, if possible, prior to operation.

There are two basic options for surgical repair. The first and most common repair is a duodenoduodenostomy. The preferred approach for this option is through a supraumbilical transverse abdominal incision, which provides adequate exposure of the duodenum once the ascending and transverse colon segments are mobilized to the left. Malrotation occurs in 30% of patients with duodenal obstruction, and this diagnosis should be excluded by adequate surgical exposure. A catheter should then be inserted through the stomach to help isolate the lesion and rule out a windsock lesion, or bulging membrane. In addition, the whole small bowel should be explored carefully for other sites of obstruction or malformations, such as the associated annular pancreas. Care should be taken to examine the gallbladder to ensure that a preduodenal portal vein is not evident, and once the duodenum is open, it is important to locate the ampulla of Vater to avoid inadvertent injury. The proximal duodenum is usually dilated, compared to the thin-walled and flat distal duodenum and jejunum. A longitudinal incision is placed in the distal, collapsed end of the duodenum and it is brought up and anastomosed to the proximal, dilated portion of the duodenum, where a transverse incision is made at the inferior aspect of the bulbous blind end. This anastomosis maybe either a diamond-shaped or side-to-side duodenoduodenostomy. It is important to avoid kinking of the bowel or the anastomosis. A Ladd's procedure is warranted if malrotation is concomitantly present. The second approach, which is less common, involves a duodenotomy with excision of the web, the origin of which is identified by placing a catheter through the stomach. The ampulla of Vater should be identified to avoid injury. The web is then excised and the duodenum is closed transversely.

Gastric decompression is required until time for feeding or when the patient is able to tolerate the removal of the tube without nausea or emesis. An anastomotic leak, injury to the bile duct, and sepsis are early complications. Late or long-term complications include peptic ulceration secondary to alkaline reflux, blind-loop syndrome with duodenal stasis, abdominal pain, and diarrhea. Prognosis is generally good for patients with a repaired duodenal stenosis or atresia; however, the long-term outcome depends on coexisting diagnoses, such as Down syndrome and cardiac anomalies.

JEJUNOILEAL ATRESIA

Etiology

Many theories exist regarding the etiology of jejunoileal atresias. Most of these theories are based on experimental models trying to replicate the disease process. The most accepted theory regarding the etiology of jejunoileal atresia is that of an intrauterine vascular accident resulting in necrosis of the affected segment, with subsequent resorption or obliteration.

Classification

There are two broad categories of jejunoileal defects—stenoses and atresias. A stenosis has an intact mesentery and is a localized narrowing of the bowel. There is no loss of continuity of the lumen and the stenotic portion generally has an irregular muscularis and thickened submucosa. Four types of jejunoileal atresias were originally described and a subtype has been added more recently. The type classification generally guides both prognosis and therapy.

Type I

The mucosa and submucosa form a web, or intraluminal diaphragm, resulting in obstruction in type I atresias. A windsock effect may be evident secondary to an increase in intraluminal pressure in the proximal bowel, causing a portion of the web to protrude into the distal part of the bowel. A mesenteric defect is not present, the bowel length is not shortened, the bowel appears to be in continuity, and only the proximal and distal size disparity points to the location of the obstruction.

Type II

The mesentery is intact in type II atresias; however, the bowel is not joined. The dilated proximal portion has a bulbous blind end connected by a short, fibrous cord to the blind end of the distal flattened bowel. Small bowel length is usually not shortened.

Type IIIa

The defect in type IIIa is similar to the type II defect in that there are blind proximal and distal ends;

however, there is complete disconnection in this case, with a V-shaped mesenteric defect present. The proximal blind end is usually markedly dilated and not peristaltic. The compromised bowel undergoes intrauterine involution, and as a result the bowel in this category is shortened to varying lengths.

Type IIIb

In addition to a large defect of the mesentery, there is also significant shortening of the bowel. This lesion, also known as an "apple peel" or "Christmas tree deformity" (because the bowel has the appearance of a tinsel coil wrapped around a single perfusing vessel, or Christmas tree), is a proximal jejunal atresia near the ligament of Treitz. The distal ileum receives its blood supply from a single ileocolic or right colic artery because the better part of the superior mesenteric artery is absent. Prematurity, malrotation, and subsequent short bowel syndrome have been linked to this deformity, with an increased morbidity and mortality. There also appears to be a pattern of heredity or genetic transmission.

Type IV

Type IV involves multiple small bowel atresias of any combination of types I to III. This defect often takes on a "string of sausages" appearance because of the multiple lesions. The cause is unknown and theories range from multiple ischemic infarcts, to an early embryological defect of the gastrointestinal tract, to an inflammatory process occurring *in utero*. This type of defect may also have a hereditary component.

Diagnosis

Prenatal diagnosis is not as likely for the more distal atresias. Whereas 30% of neonates with duodenal atresias have associated anomalies, this percentage falls to 10% in jejunoileal atresias, making prenatal diagnosis a challenge. Postnatal diagnosis of jejunoileal atresia is more likely, and surely essential, because emergency treatment may be necessary for conditions such as a midgut volvulus. Neonates with a more proximal atresia may develop bilious emesis within hours of birth, whereas neonates with more distal lesions may take days to begin vomiting. The neonate with a normal or scaphoid abdomen and bilious emesis should be considered to have a proximal obstruction until proved otherwise. If there is a question of an atresia, a nasogastric aspirate should be obtained. A bilious gastric fluid may indicate an intestinal obstruction. Other factors include abdominal distension, usually more pronounced with distal lesions, and failure to pass meconium.

Abdominal radiographs are essential for diagnosis. With more proximal atresias, a few air–fluid levels are evident, with no apparent gas in the lower part of the abdomen. A more distal lesion demonstrates more gas-filled bowel loops with air–fluid levels, but the lower abdomen may remain without a gas pattern. A barium enema is best at defining the approximate location of obstruction and is also capable of ruling out other causes of lower obstruction, such as a Hirschsprung's disease, meconium plug syndrome, or a small left colon.

Neonates with intestinal stenosis are more difficult to diagnose and their obstruction may not manifest for weeks to years. These patients tend to present with a history of intermittent emesis and failure to thrive. The clinical presentation and radiographic evidence are dependent on the degree and severity of disease, but fortunately this more subtle form of disease is rare.

Treatment

Preoperative care should be the same as that described for repair of duodenal lesions, including preoperative resuscitation, unless a surgical emergency is apparent. A transverse supraumbilical incision affords adequate exposure of the abdominal contents. The full length of the bowel should be manually explored for malrotation, as well as for other atresias or stenoses. Malrotation should be corrected with a Ladd's procedure. The length of intestine that appears functional should be measured along the antimesenteric border, because bowel length affects the procedure and overall prognosis. Saline should be injected into the distal bowel and followed closely throughout to the cecum to ensure patency. The same should be done to the colon. This maneuver will assure that an unrecognized intraluminal obstructing membrane is not present.

Once patency of the entire length of the bowel has been established, the repair may proceed. The dilated proximal bowel generally does not have normal function and, as a result, if adequate intestinal length is present, it should be resected to normal size bowel to avoid problems with abnormal postoperative peristalsis. An end-to-end anastomosis can then be done. In the presence of short bowel length, the dilated proximal segment can instead be preserved by tapering. Postoperatively, antibiotics are given for a short course and gastric drainage should continue until there is evidence of return of bowel function. Total parenteral nutrition (TPN) should be started shortly after surgery and continued until enteral feeds are tolerated. A possible complication is an anastomotic leak. It is important to follow the baby closely after the operation,

monitoring abdominal girth, vital signs, temperature, fluid status, urine output, gastric drainage, and level of activity.

Prognosis for the jejunoileal atresia group is dependent on the amount of residual, functional bowel that exists after surgery. In general, 40 cm of functional small bowel is considered adequate, and if no other comorbidities exist, the prognosis is good. The potential major morbidity and mortality for this group is a short bowel syndrome ($<20-40$ cm) that requires a dependence on total parenteral nutrition, a therapy at times complicated by liver failure. Often the only solution for this latter circumstance is a small bowel and liver transplant, which involves a range of challenges.

COLONIC ATRESIA

Colonic atresia, like jejunoileal atresia, is believed to be the result of an *in utero* vascular accident resulting in ischemic injury. This occurs after the midgut has returned to the coelomic cavity. Colonic atresia is the least common type and encompasses 1.8–15% of all intestinal atresias and stenoses. Atresias may occur throughout the colon, but lesions to the right of the splenic flexure and distal to the vascular watershed area are the most common. Colonic atresia is often associated with other anomalies, especially when there is a complicated partial or complete absence of the hindgut. As a result, anterior wall defects, vesicointestinal fissures, and genitourinary defects are frequently associated (for example, the cloacal exstrophy complex).

Prenatal diagnosis is possible by ultrasound. Expected findings would include a colon that is larger than expected for gestational age or an expected bowel obstruction. Diagnosis after birth is usually timely, because the newborn demonstrates signs of distal obstruction and meconium is generally not passed. Abdominal distension is prominent in the first 24 hours and the hugely dilated proximal loop is often palpable. This can be confirmed by a radiograph showing the large loop of proximal colonic segment. A contrast enema may confirm the diagnosis.

Treatment

Treatment depends on the extent and location of the lesion and the clinical presentation of the patient. Care should be taken to avoid perforation secondary to severe distension. A staged procedure, beginning with resection of the affected portion and colostomy with mucous fistulas, is generally the initial treatment of choice. An initial resection with anastomosis may be appropriate in selected cases. For those undergoing a staged repair, ileocolic or colocolic anastomosis should be performed by 1 year of age and a close second look should be performed to rule out distal pathology. Outcome is dependent on the residual length of colon, the length of small bowel, and associated anomalies, including small bowel atresias. The prognosis is generally good.

See Also the Following Articles

Colonic Obstruction • Duodenal Obstruction • Esophagus, Development • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Malrotation • Neonatal Intestinal Obstruction • Neonatal Tracheoesophageal Anomalies • Pyloric Stenosis • Webs

Further Reading

Baglaj, S. M., Czernik, J., Kuryszko, J., and Kuropka, P. (2001). Natural history of experimental intestinal atresia: Morphologic and ultrastructural study. *J. Pediatr. Surg.* **36**, 1428–1434.

Dalla Vecchia, L. K., Grosfeld, J. L., West, K. W., Rescorla, F. J., Scherer, L. R., and Engum, S. A. (1998). Intestinal atresia and stenosis: A 25-year experience with 277 cases. *Arch. Surg.* **133**, 490–496; discussion, 496–497.

Grosfeld, J. L., and Rescorla, F. J. (1993). Duodenal atresia and stenosis: Reassessment of treatment and outcome based on antenatal diagnosis, pathologic variance, and long-term follow-up. *World J. Surg.* **17**, 301–319.

Louw, J. H. (1959). Congenital intestinal atresia and stenosis in the newborn: Observations on its pathogenesis and treatment. *Ann. R. Coll. Surg. Engl.* **25**, 209–234.

Intestinal Infarction

DAVID B. LAUTZ AND CARLOS M. MERY
Brigham and Women's Hospital and Harvard University

bowel ischemia Acute or chronic syndrome resulting from inadequate blood perfusion of any portion of the small or large bowel.
embolism Partial or complete obstruction of a vessel with a plug extraneous to the site of the obstruction.
intestinal infarction Loss of viability of any portion of the bowel due to the lack of sufficient blood supply to ensure adequate perfusion over a significant period of time.
thrombosis Partial or complete obstruction of a vessel with a clot originating at the site of the obstruction.

Intestinal infarction is the necrosis of any portion of the small or large bowel due to insufficient blood flow to supply its basic metabolic demands. It represents one of the extremes of a continuous spectrum of the bowel ischemia syndrome that ranges from completely reversible alterations of the bowel to transmural necrosis of the intestinal wall. Early diagnosis and treatment of acute bowel ischemia are mandatory to avoid its progression into necrosis. Intestinal infarction may develop within hours of the first insult and has a mortality rate as high as 70 to 100%.

ETIOLOGY

The causes of intestinal infarction are diverse and in general similar to those that cause acute bowel ischemia. Chronic mesenteric ischemia resulting from inadequate perfusion of the midgut during periods of increased oxygen demand is a chronic clinical syndrome that rarely causes bowel infarction and as such will not be considered in this article.

Acute Mesenteric Artery Embolism

Embolization of the superior mesenteric artery (SMA) accounts for approximately 50% of cases of acute mesenteric ischemia. The majority of the emboli originate from the left atrium or ventricle in patients with either atrial fibrillation or cardiac hypokinesis.

Acute Mesenteric Artery Thrombosis

Thrombosis of the SMA occurs acutely in patients with preexistent stenotic lesions and accounts for 25% of cases of acute mesenteric ischemia. It usually presents in patients with previous symptoms of chronic mesenteric ischemia such as postprandial abdominal pain, early satiety, and weight loss.

Non-occlusive Mesenteric Ischemia

Non-occlusive mesenteric ischemia (NOMI) accounts for the other 25% of cases of acute mesenteric ischemia. It occurs in the absence of any anatomical obstruction and is a consequence of low mesenteric blood flow states. It is caused by mesenteric vasospasm induced by excessive sympathetic activity during cardiogenic shock or hypovolemia or after the use of vasoactive medications such as α-adrenergic agents and digitalis.

Acute Mesenteric Venous Thrombosis

Mesenteric venous thrombosis represents a rare but potentially lethal form of mesenteric ischemia, usually associated with predisposing conditions such as hypercoagulable states, inflammation, portal hypertension, trauma, bowel obstruction, or dehydration.

Strangulated Hernias

Strangulation of a hernia occurs when the intestinal veins of the bowel contained within the hernial sac are compressed by the opening of the hernia. This compression causes edema of the bowel wall with further compression and obstruction, eventually progressing to infarction of the loops of bowel. Strangulation is more common in incarcerated hernias (hernias that are not reducible), in hernias found in patients of advanced age, and in large hernias with relatively small openings.

Strangulated Bowel Obstruction

Strangulated obstruction occurs when blood flow to an obstructed segment of bowel is compromised. It is a particularly important complication of closed-loop bowel obstruction, where both inflow and outflow of a loop of bowel are occluded. Secretion of mucus and fluid into the closed loop of bowel leads to high intralumenal pressure. When this intralumenal pressure

 Copyright 2004, Elsevier (USA). All rights reserved.

exceeds the mesenteric venous and capillary perfusion pressures, bowel viability becomes compromised.

Other Causes of Intestinal Infarction

Unusual reported causes of intestinal ischemia and infarction include arterial dissection, thrombocytosis, amyloidosis, disseminated intravascular coagulation, vasculitides, and early postoperative jejunal tube feeding.

PATHOPHYSIOLOGY

With any significant reduction of the perfusion pressure of the bowel, the intestinal vascular bed reduces its vascular resistance, arterial collateral pathways open immediately, and redistribution of flow among the different layers of the bowel occurs (with increasing flow to the mucosa, the metabolically most active layer). Furthermore, oxygen extraction increases, allowing the cells to support their metabolic demands. When oxygen extraction can no longer compensate for the diminished blood flow, ischemia ensues. Within 20 min of complete ischemia, the insult affects the villi (the most superficial area of the mucosa). The ischemic insult then progresses centrifugally with time to the outer layers of the bowel. Transmural infarction is usually seen within 8–16 h of severe ischemia.

Intestinal infarction is associated with third spacing of fluid into the lumen of the bowel. These fluid shifts induce hemoconcentration and hypovolemia. Furthermore, untreated intestinal necrosis promotes bacterial translocation into the bloodstream and production of cytokines, with a resulting sepsis-like syndrome that in association with the hypovolemia can easily lead to shock.

DIAGNOSIS

Early diagnosis of bowel ischemia and infarction is a clinical challenge and requires a high index of suspicion.

Clinical Manifestations

Bowel ischemia usually presents with poorly localized abdominal pain, out of proportion to the physical examination. Presentation tends to be more acute and severe on cases of arterial embolism and more indolent and vague on cases of venous thrombosis. Pain is usually worsened with food intake and may be associated with rapid and forceful bowel movements with guaiac-positive stools.

As ischemia progresses and infarction ensues, the abdomen becomes distended and signs of peritoneal irritation, such as increasing abdominal tenderness, rebound, and guarding, appear. Other late signs that indicate compromise of bowel viability include nausea, vomiting, hematochezia, hematemesis, massive abdominal distension, back pain, and shock.

The clinical picture can be characterized according to four progressive clinical stages:

1. Hyperactive stage. Three is intermittent severe pain with passage of loose stools, sometimes with blood and vomiting. There are hyperactive bowel sounds on auscultation, reflecting hyperperistalsis.
2. Paralytic stage. Pain usually diminishes but becomes more continuous and diffuse. Abdominal distension and tenderness appear and bowel sounds disappear.
3. Disarranged fluid balance. There is leakage of fluid, proteins, and electrolytes, associated with necrosis of the bowel and generalized peritonitis.
4. Shock.

Strangulated hernias present with extreme pain, tenderness, edema, erythema, and irreducibility. Although the diagnosis of strangulation may be relatively easy on inguinal or ventral hernias, strangulated femoral or obturator hernias may be difficult to detect.

In addition to the usual symptoms and signs of bowel obstruction, strangulated obstruction of the intestine presents with peritoneal signs. Patients manifest abdominal tenderness, involuntary guarding, decreased urine output, fever, and tachycardia. Leukocytosis, metabolic acidosis, increased serum amylase, and phosphate should raise the suspicion of intestinal infarction.

Laboratory

Laboratory abnormalities appear with advanced ischemia. Leukocytosis with a left shift, metabolic acidosis, and an elevated hematocrit as a result of hemoconcentration are manifestations of intestinal infarction. Serum amylase and phosphate may also be elevated.

Imaging

Plain Films of the Abdomen

The main utility of plain radiographs of the abdomen is to rule out other causes of abdominal pain such as bowel obstruction or perforated viscus. Bowel ischemia may present with subtle findings such as adynamic ileus and distended, air-filled loops of bowel. However, in intestinal infarction, plain films can

show "thumbprinting" of the bowel with pneumatosis of the intestinal wall and gas in the portal vein. These signs, associated with an ominous prognosis, are rarely encountered.

Computed Tomography

In current practice, abdominal computed tomography (CT) is frequently obtained in patients with abdominal pain of unclear source. In addition to helping identify the source of the ischemia in certain situations (e.g., mesenteric venous thrombosis), CT may show portions of thickened bowel wall, suggestive of ischemia. On advanced intestinal infarction, intestinal wall pneumatosis and portal vein gas may also be seen.

Angiography

Although angiography is the definitive diagnostic study for bowel ischemia, it does not provide information about the viability of the involved bowel. It allows the clinician to determine the mechanism for the ischemic insult, define the anatomy of the vessels, and on certain occasions, administer therapeutic agents such as urokinase (for thrombosis) or papaverine (for NOMI).

MANAGEMENT

Initial treatment of patients with bowel ischemia/infarction includes volume resuscitation, correction of acidosis if possible, and administration of broad-spectrum antibiotics. Intravenous heparin is indicated in cases of thromboembolism to avoid propagation of the clot and to allow restoration of flow.

Any patient with major arterial occlusion from thrombosis or embolus that does not respond completely to vasodilators on angiography, and any patient with signs of peritoneal irritation regardless of the mechanism of ischemia, should undergo immediate exploratory laparotomy. The purpose of the laparotomy is twofold: to revascularize the affected bowel (if possible) and to resect all nonviable bowel. This article focuses on the management of bowel infarction; the treatment of bowel ischemia per se is beyond its scope.

Evaluation of Intestinal Viability

Revascularization should precede any evaluation of intestinal viability because bowel that may appear nonviable initially may show complete recovery after revascularization. Once revascularization has been performed, the involved loops of bowel are covered with warm saline-soaked pads for 20–30 min and are then visually inspected. Nonviable bowel is dull gray or black in color, edematous, distended, and lacking peristaltic activity. Doppler measurements of arterial flow can also be used to assess bowel viability in the operating room by showing adequate pulsations on the anti-mesenteric border of viable bowel. Another objective method used to assess intestinal viability is the injection of fluorescein followed by the inspection of the bowel in question under Wood's lamp.

Resection of Nonviable Bowel

Minimal bowel resection with liberal use of second-look laparotomy is the hallmark of treatment of bowel infarction. Clearly nonviable bowel should be excised. If the margins are unequivocally viable, intestinal continuity may be restored. Otherwise, it is recommended to restore continuity on a second operation. Similarly, areas of questionable or marginal viability are usually left in place since they may recover completely with time and a second-look laparotomy is planned 24–48 h later. This allows a clearer delineation of viable bowel and permits supportive measures to render more of the bowel viable. Once the decision to perform a second-look laparotomy is taken (during the first operation), it is upheld regardless of the clinical course of the patient.

Since in colonic ischemia it is not unusual to have a normal serosal appearance with extensive mucosal injury, the extent of the resection should be guided by preoperative evaluation rather than by the appearance of the serosa. All specimens resected for colonic ischemia should be opened at the time of the operation to ensure that margins have normal mucosa. Similarly, as in other colonic surgeries, due to the risk of contamination of the anastomosis from an unprepared bowel, ischemic emergencies of the left colon should be treated with colostomies instead of reestablishing intestinal continuity primarily.

When obvious infarction of all or most of the bowel is found on the initial laparotomy, a decision must be made between resecting the entire bowel with the subsequent need for lifetime TPN or doing nothing and allowing the patient to die by providing comfort measures only.

Second-Look Exploration

The decision to perform a second-look laparotomy is made during the initial procedure and maintained regardless of the patient's postoperative course. At the initial laparotomy, bowel segments of marginal viability are often preserved to avoid a short bowel syndrome. These segments must be reinspected to ensure viability. Laparoscopy has been proposed as an alternative method to perform the second-look

exploration but enough experience has not been accrued to evaluate its adequacy.

CONCLUSIONS

Intestinal infarction represents the end result of severe bowel ischemia. The only potential for cure lies in prompt diagnosis and management. Once the suspicion of bowel necrosis is raised, exploratory laparotomy should not be delayed. The treatment of intestinal infarction consists of the correction of the underlying cause of ischemia, assessment of viability of the bowel, and resection of clearly nonviable segments with liberal use of second-look explorations for reassessment of marginally viable bowel left in place.

See Also the Following Articles

Circulation, Overview • Colonic Ischemia • Hernias • Intestinal Ischemia • Laparascopy • Portal Vein Thrombosis

Further Reading

Boley, S. J., and Kaleya, R. N. (1997). Mesenteric ischemic disorders. *In* "Maingot's Abdominal Operations" (M. J. Zinner, S. I. Schwartz, and H. Ellis, eds.), 20th Ed., pp. 655–689. Appleton & Lange, Stamford, CT.

Eldrup-Jorgensen, J., Hawkins, R. E., and Bredenberg, C. E. (1997). Abdominal vascular catastrophes. *Surg. Clin. North Am.* **77**, 1305–1320.

Haglund, U., and Bergqvist, D. (1999). Intestinal ischemia—The basics. *Langenbeck Arch. Surg.* **384**, 233–238.

Schwartz, L. B., and Gewertz, B. L. (eds.) (1997). Mesenteric ischemia. *Surg. Clin. North Am.* **77**, 307–502.

Intestinal Ischemia

DAVID B. LAUTZ AND SIDHU P. GANGADHARAN
Brigham and Women's Hospital, Boston

angiography Examination of blood vessels using X rays following the injection of a radio-opaque substance.
bowel resection Surgical removal of part of the large or small intestine.
bypass graft Passage created surgically to divert flow of blood or other bodily fluid or to circumvent an obstructed or diseased organ, with the transplantation or implantation of tissue to replace the damaged part.
embolectomy Surgical removal of an embolus.
embolism Obstruction or occlusion of a blood vessel caused by a mass such as an air bubble or clot.
ischemia Decrease in the blood supply to an organ, tissue, or body part caused by constriction or obstruction of a vessel.
mesentery Double layer of peritoneum attached to the abdominal wall, enclosing in its fold certain organs of the abdominal viscera.
thrombosis Formation of a fibrinous clot in a blood vessel or in a chamber of the heart.
vasodilator Something (such as a nerve or a drug) that causes dilation of a blood vessel.

Acute mesenteric ischemia is a life-threatening emergency that comprises several distinct etiologies. The overall incidence is 1 in 1000 hospital admissions. Historically, mortality rates for these conditions have been in excess of 60–70%, which underscores the importance of early diagnosis and treatment. With few exceptions, patients with acute mesenteric ischemia present with abdominal pain, often out of proportion to physical findings. The clinician should have a high index of suspicion with patients who have this finding in association with risk factors for intestinal ischemia, so that measures may be taken expeditiously to prevent the irreversible outcome of bowel infarction.

INTRODUCTION

A thorough understanding of the etiologies of intestinal ischemia is key to effective treatment. Broadly, intestinal ischemia can be separated into acute and chronic

Copyright 2004, Elsevier (USA). All rights reserved.

exploration but enough experience has not been accrued to evaluate its adequacy.

CONCLUSIONS

Intestinal infarction represents the end result of severe bowel ischemia. The only potential for cure lies in prompt diagnosis and management. Once the suspicion of bowel necrosis is raised, exploratory laparotomy should not be delayed. The treatment of intestinal infarction consists of the correction of the underlying cause of ischemia, assessment of viability of the bowel, and resection of clearly nonviable segments with liberal use of second-look explorations for reassessment of marginally viable bowel left in place.

See Also the Following Articles

Circulation, Overview • Colonic Ischemia • Hernias • Intestinal Ischemia • Laparascopy • Portal Vein Thrombosis

Further Reading

Boley, S. J., and Kaleya, R. N. (1997). Mesenteric ischemic disorders. *In* "Maingot's Abdominal Operations" (M. J. Zinner, S. I. Schwartz, and H. Ellis, eds.), 20th Ed., pp. 655–689. Appleton & Lange, Stamford, CT.

Eldrup-Jorgensen, J., Hawkins, R. E., and Bredenberg, C. E. (1997). Abdominal vascular catastrophes. *Surg. Clin. North Am.* **77**, 1305–1320.

Haglund, U., and Bergqvist, D. (1999). Intestinal ischemia—The basics. *Langenbeck Arch. Surg.* **384**, 233–238.

Schwartz, L. B., and Gewertz, B. L. (eds.) (1997). Mesenteric ischemia. *Surg. Clin. North Am.* **77**, 307–502.

Intestinal Ischemia

DAVID B. LAUTZ AND SIDHU P. GANGADHARAN
Brigham and Women's Hospital, Boston

angiography Examination of blood vessels using X rays following the injection of a radio-opaque substance.
bowel resection Surgical removal of part of the large or small intestine.
bypass graft Passage created surgically to divert flow of blood or other bodily fluid or to circumvent an obstructed or diseased organ, with the transplantation or implantation of tissue to replace the damaged part.
embolectomy Surgical removal of an embolus.
embolism Obstruction or occlusion of a blood vessel caused by a mass such as an air bubble or clot.
ischemia Decrease in the blood supply to an organ, tissue, or body part caused by constriction or obstruction of a vessel.
mesentery Double layer of peritoneum attached to the abdominal wall, enclosing in its fold certain organs of the abdominal viscera.
thrombosis Formation of a fibrinous clot in a blood vessel or in a chamber of the heart.
vasodilator Something (such as a nerve or a drug) that causes dilation of a blood vessel.

Acute mesenteric ischemia is a life-threatening emergency that comprises several distinct etiologies. The overall incidence is 1 in 1000 hospital admissions. Historically, mortality rates for these conditions have been in excess of 60–70%, which underscores the importance of early diagnosis and treatment. With few exceptions, patients with acute mesenteric ischemia present with abdominal pain, often out of proportion to physical findings. The clinician should have a high index of suspicion with patients who have this finding in association with risk factors for intestinal ischemia, so that measures may be taken expeditiously to prevent the irreversible outcome of bowel infarction.

INTRODUCTION

A thorough understanding of the etiologies of intestinal ischemia is key to effective treatment. Broadly, intestinal ischemia can be separated into acute and chronic

Copyright 2004, Elsevier (USA). All rights reserved.

syndromes. Etiologies of acute mesenteric ischemia include mesenteric arterial embolism, mesenteric arterial thrombosis, mesenteric venous thrombosis, and nonocclusive mesenteric ischemia. The hallmark of chronic mesenteric ischemia is arterial insufficiency to increased blood flow demand, most often related to atherosclerotic stenoses of the mesenteric arterial orifices, trunks, and branches. Other causes of mesenteric arterial insufficiency not covered herein include vasculopathy, arterial dissection, tumor, or radiation.

ACUTE MESENTERIC ISCHEMIA

Mesenteric Arterial Embolism

The patient with embolism to the superior mesenteric artery presents with a history of a sudden onset of acute abdominal pain. This may be accompanied by nausea, vomiting, or diarrhea. The pain is constant and persistent, and, as noted above, may be accompanied by an unremarkable physical exam in the early stages of the presentation. As the bowel progresses to infarction, bloody diarrhea and peritoneal signs may present.

The patient's history may increase the suspicion that mesenteric arterial embolism is the cause of the intestinal ischemia. The origin of the embolus is most often the heart, as a consequence of intracardiac thrombus from atrial fibrillation or other arrhythmias. Other causes of intracardiac thrombus include myocardial infarction, ventricular aneurysms, and diseased or mechanical heart valves. A thromboembolus arising from a deep venous thrombosis could enter the systemic circulation if a right-to-left intracardiac shunt exists. Tumor embolus (e.g., atrial myxoma), atheroembolus (spontaneous or iatrogenic), and septic embolus have also been reported. A thorough assessment of the patient must be made to exclude the possibility of synchronous or metachronous embolism to other vascular beds.

In general, laboratory tests are nonspecific for the diagnosis of mesenteric arterial embolism and other causes of acute mesenteric ischemia. Third-space volume losses into the bowel wall and lumen may be manifested by increased blood urea nitrogen/creatinine ratios and hemoconcentration. Leukocytosis is common. Lactic acid levels are sensitive for detecting bowel ischemia and infarct, but can be elevated in other etiologies of the acute abdomen. Serum phosphate is elevated after a significant bowel infarction has occurred.

The gold standard for diagnosis is selective mesenteric angiography. The embolus will typically lodge at a branch point several centimeters from the origin of the superior mesenteric artery, thus sparing the first few jejunal branches. However, because of the overlap of symptomatology with many other intraabdominal etiologies of pain, patients with a mesenteric arterial embolus may have already undergone several other diagnostic tests to rule out other disease processes. A plain film of the abdomen may be essentially normal in the early stages of mesenteric arterial embolism. As the ischemia worsens, small bowel dilation, air–fluid levels, and evidence of bowel wall thickening and eventual pneumatosis intestinalis may be apparent. Abdominal computed tomography (CT) scan may be more effective at diagnosing alternative causes for the patient's abdominal pain, such as diverticulitis or ureterolithiasis. Although an abdominal CT scan is able to demonstrate bowel thickening, ileus, pneumatosis, and, in some cases, occlusion of the superior mesenteric artery, it should not be considered the equivalent of mesenteric angiography.

The treatment for arterial mesenteric embolus should commence simultaneously with the completion of the workup. The patient should be aggressively volume resuscitated and broad-spectrum antibiotics should be given. Once the diagnosis of mesenteric arterial embolus has been confirmed with angiography, the patient should be brought to the operating room for surgical embolectomy. The superior mesenteric artery is approached at the root of the transverse mesocolon and a transverse arteriotomy is made. An embolectomy catheter is passed anterograde and retrograde to remove the embolus. Once vigorous flow is achieved, the arteriotomy is closed primarily or with a patch if stenosis is a concern. Assessment of distal flow should be made with direct palpation or Doppler insonation of the vessels.

Only after the blood flow is restored should any grossly nonviable intestine be resected. Questionably viable intestine should be reassessed with a "second-look" operation in 24–48 hours. This strategy is employed to minimize resection of any potentially viable bowel at the first operation, while at the same time ensuring that marginal bowel that does not recover its viability is then removed before further complications ensue. In addition to visual inspection of bowel color, peristalsis, and edema, viability may be assessed by adjunctive measures such as Doppler insonation of antimesenteric blood flow or fluorescein injection and inspection under a Wood's lamp.

Mesenteric Arterial Thrombosis

The patient with mesenteric arterial thrombosis also presents with abdominal pain out of proportion

to physical exam, although possibly of a more gradual onset. Because this acute ischemic syndrome most often represents the end stage of progressive atherosclerotic disease of the visceral arteries, it may be possible to elicit a history of chronic mesenteric ischemic symptoms, i.e., intestinal angina (abdominal pain after meals), food aversion, weight loss, and diarrhea. Similar to mesenteric arterial embolism, physical findings and laboratory studies with mesenteric arterial thrombosis may be unremarkable and nonspecific until progression of the ischemia leads to irreversible infarction of the bowel.

The plain film and abdominal CT findings are similar to those for mesenteric arterial embolism. There may be additional evidence of chronic visceral atherosclerosis on CT, such as prominent calcifications of the aorta and its branches. Angiography is the mainstay of diagnosis. Unlike embolism, thrombosis of the superior mesenteric artery usually leads to occlusion of the vessel most proximally, up to its origin of the vessel from the aorta.

Volume resuscitation, heparin anticoagulation, and broad-spectrum antibiotics should be instituted concomitantly with the diagnostic workup. The operative strategy is to revascularize first, then assess nonviable bowel for resection. The revascularization may involve transaortic endarterectomy of the orificial lesion, or retrograde or anterograde bypass grafting. Autogenous vein or prosthetic graft may be used as conduit, though it has been suggested that vein graft may be preferred in the setting of bowel infarct, with its attendant bacterial translocation or peritonitis. Multiple grafts may be necessary to adequately revascularize a diffusely diseased mesenteric arterial bed, but this more extensive procedure may be less optimal in a sick patient with infarcted bowel with marginal nutritional status. A second-look operation again may be employed to most appropriately resect nonviable bowel.

Nonocclusive Mesenteric Ischemia

Nonocclusive mesenteric ischemia results from persistent vasospasm of the visceral vessels, usually in response to diminished mesenteric blood flow. Inciting events include low-cardiac-output states, sepsis, hypotension, hypovolemia, or numerous vasoactive medications such as α-adrenergic agents, vasopressin, or digitalis. Vasospasm may also cause persistent ischemia after surgical revascularization of mesenteric embolism or thrombosis. The diagnosis of nonocclusive mesenteric ischemia should be considered if a patient has abdominal pain with these risk factors. The patient's exam may be out of proportion with complaints of pain, but with this etiology of acute mesenteric ischemia in particular, it must be recognized that the patient's comorbid condition may mask early subjective awareness of symptoms. Laboratory findings of leukocytosis, lactic acidosis, or hemoconcentration may be noted. Late findings of infarction are also similar to those in thrombotic or embolic cases.

Treatment of nonocclusive mesenteric ischemia begins with attempts to correct the underlying condition and maximize the patient's hemodynamic status. Vasoconstrictive medications should be stopped if possible. Expeditious mesenteric angiography is essential, provided the patient can tolerate it clinically, because both diagnosis and treatment may be provided with this procedure. Superior mesenteric arterial branches will show evidence of severe vasospasm, with an absence of occlusion. Vasodilators (e.g., papaverine) are then selectively administered using a catheter directed into the superior mesenteric artery.

Mesenteric vasodilator therapy should be instituted first when possible, even if the patient's initial evaluation reveals likely infarcted bowel, which would require a subsequent operative resection. Often, however, these patients cannot tolerate vasodilator therapy, in which case exploration is warranted. Conversely, during infusion of vasodilators in patients who initially appear to have threatened but not necessarily nonviable bowel, any sign of progression of bowel ischemia (e.g., worsening leukocytosis, abdominal pain, or acidosis) should trigger a trip to the operating room for abdominal exploration. Persistence of symptoms and signs of bowel ischemia after 12–24 hours of vasodilator therapy should alert the clinician that an operation will be required to resect nonviable bowel. During therapy, the patient should also undergo close hemodynamic monitoring.

Mesenteric Venous Thrombosis

The patient with mesenteric venous thrombosis may also present as an acute ischemic syndrome, but the presentation may be more protracted compared with nonocclusive mesenteric ischemia or mesenteric arterial embolism or thrombosis. Though abdominal pain is still a hallmark of the disease, several days may elapse between the onset of pain and progression of symptoms to the point where the patient seeks medical attention. Once evaluated, however, the patient with acute mesenteric venous thrombosis may display the classic finding in intestinal ischemia, i.e., pain out of proportion to physical exam. Symptoms of diarrhea, anorexia, nausea/vomiting, and bloating are common. Laboratory findings of leukocytosis or lactic acidosis may be less profound and less prevalent than what is seen with

other etiologies of acute mesenteric ischemia. Mesenteric venous thrombosis is associated with hypercoagulable states, previous deep venous or mesenteric thrombosis, as well as intraabdominal inflammation or infection and previous abdominal surgery.

Abdominal CT scan is the most useful test in diagnosing mesenteric venous thrombosis, with a sensitivity that exceeds that of mesenteric angiography. Findings on CT scan include presence of thrombus within the superior mesenteric vein or portal vein, or less specific findings of ascites, bowel wall thickening, mesenteric congestion, or stranding, or pneumatosis intestinalis. Angiography may be useful if the possibility of an arterial etiology needs to be ruled out. Mesenteric venous thrombosis is diagnosed by angiographic venous phase findings of thrombus or nonfilling of the superior mesenteric vein, or prolonged emptying of the arterial phase.

Treatment of mesenteric venous thrombosis begins with the institution of heparin anticoagulation, which will be continued as chronic oral anticoagulation after the acute episode is resolved. The decision of whether the patient should go to the operating room depends primarily on the abdominal exam. Evidence of localized or generalized peritonitis requires that the patient be explored and nonviable bowel resected.

CHRONIC MESENTERIC ISCHEMIA

As previously mentioned in regard to acute mesenteric arterial thrombosis, chronic mesenteric symptoms result most commonly from atherosclerotic lesions in the visceral arteries. This compromised blood flow manifests as intestinal angina (abdominal pain after meals) and may result in food aversion and weight loss. Diarrhea and other nonspecific abdominal complaints may be present. Though chronic mesenteric ischemia is not immediately life-threatening or bowel-threatening, its progression to an acute thrombotic state is unpredictable.

The diagnosis of chronic mesenteric ischemia is made with angiography. Multiple stenoses and occlusions may be present. Treatment is intestinal revascularization. Trap-door endarterectomy or bypass grafting are the operative means to achieve this. Depending on the anatomy of the mesenteric arteries, single or multiple grafts may be used. Recent results with percutaneous angioplasty and stenting have also demonstrated effective treatment of this condition.

It should also be noted that mesenteric venous thrombosis can have a chronic presentation as well. In these patients, venous collateral flow allows bowel viability, though symptoms of vague abdominal pain may be present. Incidental findings of partially or completely occlusive mesenteric venous thrombus may be noted during abdominal CT scanning to rule out other pathology. Long-term anticoagulation may be of benefit.

SUMMARY

Modern series of acute mesenteric ischemia still demonstrate the lethality of this set of disease entities. Early recognition of the possibility of acute mesenteric ischemia and prompt diagnostic imaging (angiography in mesenteric arterial embolism, mesenteric arterial thrombosis, and nonocclusive mesenteric ischemia; abdominal CT scan for mesenteric venous thrombosis) ensure that adequate perfusion is restored to the ischemic intestine as rapidly as possible. Even with expeditious diagnosis, however, there are futile cases in which the overwhelming extent of bowel infarct precludes any further surgical intervention.

See Also the Following Articles

Circulation, Overview • Intestinal Infarction • Malabsorption • Portal Vein Thrombosis

Further Reading

Bassiouny, H. S. (1997). Nonocclusive mesenteric ischemia. *Surg. Clin. North Am.* **77**(2), 319.

Endean, E. D., Barnes, S. L., Kwolek, C. J., *et al.* (2001). Surgical management of thrombotic acute intestinal ischemia. *Ann. Surg.* **233**(6), 801.

Matsumoto, A. H., Angle, J. F., Spinosa, D. J., *et al.* (2002). Percutaneous transluminal angioplasty and stenting in the treatment of chronic mesenteric ischemia: Results and longterm followup. *J. Am. Coll. Surg.* **194**(1), S22.

McKinsey, J. F., and Gewertz, B. L. (1997). Acute mesenteric ischemia. *Surg. Clin. North Am.* **77**(2), 307.

Ottinger, L. W., and Austen, W. G. (1967). A study of 136 patients with mesenteric infarction. *Surg. Gynecol. Obstet.* **124**, 251.

Park, W. M., Cherry, K. J., Chua, H. K., *et al.* (2002). Current results of open revascularization for chronic mesenteric ischemia: A standard for comparison. *J. Vasc. Surg.* **35**(5), 853.

Park, W. M., Gloviczki, P., Cherry, K. J., *et al.* (2002). Contemporary management of acute mesenteric ischemia: Factors associated with survival. *J. Vasc. Surg.* **35**(3), 445.

Rhee, R. Y., and Gloviczki, P. (1997). Mesenteric venous thrombosis. *Surg. Clin. North Am.* **7**(2), 327.

Rhee, R. Y., Gloviczki, P., Mendonca, C. T., *et al.* (1994). Mesenteric venous thrombosis: Still a lethal disease in the 1990's. *J. Vasc. Surg.* **20**(5), 688.

Stoney, R. J., and Cunningham, C. G. (1993). Acute mesenteric ischemia. *Surgery* **114**(3), 489.

Zelenock, G. B. (2001). Visceral occlusive disease. *In* "Surgery: Scientific Principles and Practice" (L. J. Greefield, M. W. Mulholland, K. T. Oldham *et al.,* eds.). Lippincott Williams & Wilkins, Philadelphia.

Intestinal Pseudoobstruction

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

myopathic Pertaining to diseases of the musculature.
neuropathic Pertaining to disordered function caused by disease or malfunction of the nervous system.

Pseudoobstruction is a neuromuscular disorder of motor propulsion that occurs most commonly in the small or large intestine. It is a chronic disorder for which both myopathic and neuropathic forms are recognized. Chronic intestinal pseudoobstruction is characterized by symptoms of intestinal obstruction in the absence of any form of mechanical obstruction, such as strictures or tumors. It is a condition in which the mechanisms for control of organized propulsive motility fail while the intestinal lumen is patent. The general symptoms of colicky abdominal pain, nausea, and vomiting and abdominal distension in pseudoobstruction simulate the symptoms arising from a mechanical obstruction.

MYOPATHIC FORM

Degenerative changes in the musculature underlie the myopathic form of pseudoobstruction. Familial visceral myopathies are genetic diseases characterized by degeneration and fibrosis of gastrointestinal smooth muscle. Progressive weakening of the contractile forces the musculature is able to generate compromises propulsive motility and underlies the clinical symptoms that include pseudoobstruction in the affected patients. Hypomotility characterizes this form of chronic intestinal pseudoobstruction when studied manometrically.

There are two well-recognized types of familial visceral myopathy based on the pattern of inheritance. The histopathologic appearance of degenerating muscle fibers and fibrosis is essentially the same for both types. One type is transmitted by an autosomal dominant gene and the second type by an autosomal recessive gene. Prognosis is poor for both types of myopathy because there is no effective treatment and the patients require long-term parenteral feeding to avoid severe malnutrition.

NEUROPATHIC FORM

Neuropathic forms of intestinal pseudoobstruction reflect degenerative changes in the enteric nervous system. Failure of propulsive motility in the affected length of bowel reflects loss of the neural networks that program and control the repertoire of motility patterns required for the necessary functions of the affected region of intestine. Pseudoobstruction occurs in part because contractile behavior of the circular muscle is hyperactive but disorganized in the denervated regions. Hyperactivity of the musculature is a diagnostic sign of the neuropathic form of chronic intestinal pseudoobstruction in humans. The hyperactive and disorganized contractile behavior reflects the absence of inhibitory nervous control of the muscles that are self-excitable (i.e., autogenic) when released from the braking action imposed by inhibitory motor neurons. Chronic pseudoobstruction is therefore symptomatic of the advanced stages of a progressive enteric neuropathy.

Degenerative noninflammatory and inflammatory enteric neuropathies both culminate in pseudoobstruction. Noninflammatory neuropathies can be either familial or sporadic. The mode of inheritance may be autosomal recessive or dominant. In the former, the neuropathologic findings include a marked reduction in the number of neurons in both myenteric and submucosal plexuses. Members of two families have been described with intestinal pseudoobstruction associated with the autosomal dominant form of enteric neuropathy. The numbers of enteric neurons were decreased in these patients, with no alterations being found in the central nervous system or parts of the autonomic nervous system outside the gut.

Degenerative inflammatory enteric neuropathies are characterized by a dense inflammatory infiltrate confined to the enteric nervous system. Paraneoplastic syndrome, Chagas' disease, and idiopathic enteric neuronal degenerative disease are recognizable forms of pseudoobstruction related to inflammatory neuropathies.

See Also the Following Articles

Chagas' Disease • Colonic Obstruction • Disinhibitory Motor Disorder • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Manometry • Neurogastroenterology • Paraneoplastic Syndrome

 Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Schuffler, M., and Pope, C. (1977). Studies of idiopathic intestinal pseudoobstruction. Hereditary hollow visceral myopathy: Family studies. *Gastroenterology* **73**, 344–399.

Stanghellini, V., Camilleri, M., and Malagelada, J. (1987). Chronic idiopathic intestinal pseudoobstruction. Clinical and intestinal manometric findings. *Gut* **23**, 824–828.

Summers, R., Anuras, S., and Green, J. (1983). Jejunal manometry patterns in health, partial intestinal obstruction and pseudoobstruction. *Gastroenterology* **85**, 1297–1300.

Wood, J., Alpers, D., and Andrews, P. (2000). Fundamentals of neurogastroenterology: Basic science. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 31–90. Degnon Associates, McLean, VA.

Intrinsic Factor

DAVID H. ALPERS

Washington University School of Medicine, St. Louis, Missouri

cobalamin Refers to the entire vitamin B_{12} molecule except for the cyanide moiety, which may be substituted for by other organic ligands (methyl, adenosyl) and still retain full activity.

cubilin Multifunctional protein in the apical (luminal) membrane of intestinal and renal cells; composed of multiple closely related repeating units, two of which are responsible for binding the IF–Cbl complex.

haptocorrin Cobalamin-binding glycoprotein of unknown function produced by the salivary glands, stomach, and other foregut tissues of mammals.

intrinsic factor Glycoprotein produced by the stomach or other foregut tissues of mammals; specifically binds vitamin B_{12} (cobalamin).

transcobalamin Cobalamin-binding nonglycosylated protein that is produced by all epithelial cells, circulates in the blood, and delivers cobalamin to the tissues of mammals.

vitamin B_{12} Cobalt-containing compound produced by microorganisms; consists of a planar four-pyrrole ring (corrin), with a central cobalt molecule coordinately bound to a dimethylbenzimidazole-ribose moiety below and to a cyanide above. This vitamin is essential for cell growth.

Intrinsic factor is a 399-amino-acid glycoprotein produced by mammalian stomach or other foregut tissues, primarily gastric parietal cells in humans. Intrinsic factor specifically binds cobalamin, a large water-soluble molecule that cannot be absorbed by diffusion across the intestinal wall but is essential for growth of mammalian cells.

INTRODUCTION

The intestinal epithelial cell layer does not permit absorption of peptides larger than five or six amino acids in size. The enterocytes responsible for this barrier express apical carriers and transporters that are highly efficient for the uptake of the products of digestion, such as vitamins, minerals, and amino acids. Cobalamin (Cbl), which is essential for growth of mammalian cells, is large (molecular weight, 1356) and cannot be absorbed by diffusion across the intestinal wall. Gastric intrinsic factor (IF) and salivary haptocorrin (Hc) are the main Cbl binders in the gastrointestinal lumen, and the corresponding receptor for each one is found on the enterocyte apical membrane. Transcobalamin (TC) also binds Cbl and is necessary for the transport of Cbl to peripheral tissues; its receptor is present in basolateral membranes of all somatic cells (Table I).

OCCURRENCE AND DISTRIBUTION

Intrinsic factor is produced by mammalian tissues derived from the foregut, especially in the mucosa of the gastric body (corpus), but not in cardiac or antral

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Schuffler, M., and Pope, C. (1977). Studies of idiopathic intestinal pseudoobstruction. Hereditary hollow visceral myopathy: Family studies. *Gastroenterology* **73**, 344–399.

Stanghellini, V., Camilleri, M., and Malagelada, J. (1987). Chronic idiopathic intestinal pseudoobstruction. Clinical and intestinal manometric findings. *Gut* **23**, 824–828.

Summers, R., Anuras, S., and Green, J. (1983). Jejunal manometry patterns in health, partial intestinal obstruction and pseudoobstruction. *Gastroenterology* **85**, 1297–1300.

Wood, J., Alpers, D., and Andrews, P. (2000). Fundamentals of neurogastroenterology: Basic science. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 31–90. Degnon Associates, McLean, VA.

Intrinsic Factor

DAVID H. ALPERS

Washington University School of Medicine, St. Louis, Missouri

cobalamin Refers to the entire vitamin B_{12} molecule except for the cyanide moiety, which may be substituted for by other organic ligands (methyl, adenosyl) and still retain full activity.

cubilin Multifunctional protein in the apical (luminal) membrane of intestinal and renal cells; composed of multiple closely related repeating units, two of which are responsible for binding the IF–Cbl complex.

haptocorrin Cobalamin-binding glycoprotein of unknown function produced by the salivary glands, stomach, and other foregut tissues of mammals.

intrinsic factor Glycoprotein produced by the stomach or other foregut tissues of mammals; specifically binds vitamin B_{12} (cobalamin).

transcobalamin Cobalamin-binding nonglycosylated protein that is produced by all epithelial cells, circulates in the blood, and delivers cobalamin to the tissues of mammals.

vitamin B_{12} Cobalt-containing compound produced by microorganisms; consists of a planar four-pyrrole ring (corrin), with a central cobalt molecule coordinately bound to a dimethylbenzimidazole-ribose moiety below and to a cyanide above. This vitamin is essential for cell growth.

Intrinsic factor is a 399-amino-acid glycoprotein produced by mammalian stomach or other foregut tissues, primarily gastric parietal cells in humans. Intrinsic factor specifically binds cobalamin, a large water-soluble molecule that cannot be absorbed by diffusion across the intestinal wall but is essential for growth of mammalian cells.

INTRODUCTION

The intestinal epithelial cell layer does not permit absorption of peptides larger than five or six amino acids in size. The enterocytes responsible for this barrier express apical carriers and transporters that are highly efficient for the uptake of the products of digestion, such as vitamins, minerals, and amino acids. Cobalamin (Cbl), which is essential for growth of mammalian cells, is large (molecular weight, 1356) and cannot be absorbed by diffusion across the intestinal wall. Gastric intrinsic factor (IF) and salivary haptocorrin (Hc) are the main Cbl binders in the gastrointestinal lumen, and the corresponding receptor for each one is found on the enterocyte apical membrane. Transcobalamin (TC) also binds Cbl and is necessary for the transport of Cbl to peripheral tissues; its receptor is present in basolateral membranes of all somatic cells (Table I).

OCCURRENCE AND DISTRIBUTION

Intrinsic factor is produced by mammalian tissues derived from the foregut, especially in the mucosa of the gastric body (corpus), but not in cardiac or antral

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Properties of Cbl-Binding Proteins[a]

Property	Intrinsic factor	Haptocorrin	Transcobalamin
Expression	Gastric body Medium-sized ducts (salivary, pancreatic)	Granulocytes, yolk sac, mammary glands, salivary acini, and ducts	Most tissues
Carbohydrate	15%	33–40%	0
Molecular mass	45–59 kDa	58–60 kDa	44 kDa
Number of subunits	1	1	1
Number of Cbl binding sites	1	1	1
K_a for Cbl[b]	1–4 n*M*	0.1 n*M*	0.02 n*M*

[a] Data are from Schneider and Stroinski (1987); adapted from Allen *et al*. (1978).
[b] The large range of values for K_a may represent differences in methods used, and may not reflect the true degree of relative affinity.

mucosa in most species. Both parietal cells and chief cells in the gastric body produce IF. In humans, most IF is expressed in parietal cells, but IF in rat and mouse is produced primarily in chief cells. IF is expressed in the minority of enteroendocrine cells in the gastric body of dogs and humans. The cells of medium-sized pancreatic ducts are the principal source of IF in the dog, and these cells respond to hormonal stimuli that stimulate pancreatic function (e.g., cholecystokinin).

SYNTHESIS AND SECRETION

Any increase in acid secretion from the parietal cell (by histamine, gastrin, or cholinergic stimulation) also increases IF production in those species (including humans) in which the parietal cell is the principal source of IF. The regulated step in IF expression is secretion from the cell, as the rate of IF mRNA content and rate of synthesis is constant, independent of the addition of stimulators of secretion. Drugs that directly diminish acid secretion in man can inhibit IF secretion, including histamine-2 receptor antagonists and proton pump inhibitors. In the rat and mouse, stimulators of chief cell secretion (cholinergic stimulators and dibutyryl camp), but not histamine or gastrin, increase IF secretion.

STRUCTURE AND FUNCTION

IF is a 399-amino-acid glycoprotein essential for nearly all uptake of cobalamin from the intestine of most vertebrates. IF purified by affinity chromatography using a Sepharose-coupled monocarboxylic acid derivative of Cbl has a molecular mass of 45–47 kDa. Larger apparent sizes have been reported, presumably due to the carbohydrate content of 15%. The sequence of human IF deduced from its cDNA clone is 80% identical with that of rat IF, and the gene is found on chromosome 11. The carbohydrate chains on IF are not important for binding function, but glycosylation of IF creates a protein more resistant to degradation by gastric and pancreatic proteases.

The entire Cbl molecule is important for IF binding. Binding of Cbl to IF is dependent on interactions with both the 5,6-dimethylbenzimidazole and the corrin ring, initiating a change in folding as IF closes around the Cbl molecule. This change in structure presumably accounts for the higher affinity for the intestinal IF receptor of the [Cbl : IF] complex versus the IF molecule alone. The ribazole fragment in the Cbl side chain and the Co–N coordination bonds of the lower cobalt nucleotide ligands are needed to promote conformational changes in IF to allow Cbl binding.

Although there are six regions of high sequence homology between IF, Hc, and TC, the predicted secondary structure within these proteins is not well conserved, perhaps accounting for the different binding affinities of these proteins for cobalamin (Table I) or for their individual receptors. There are two domains on IF identified by monoclonal antibody binding, one for binding Cbl and the other for IF–Cbl receptor binding. Binding of the IF–Cbl complex to its receptor, cubilin, is dependent on the N terminus of IF.

CELLULAR UPTAKE AND METABOLISM OF IF–CBL

Dietary Cbl is released from Cbl-binding enzymes by gastric proteases, and binds first to Hc in the stomach. The Hc–Cbl complex is then hydrolyzed by pancreatic proteases and the Cbl is transferred to IF in the proximal duodenum. The IF–Cbl complex survives proteolysis within the intestinal lumen due to its relative protease resistance, and is bound by the IF–Cbl receptor located on villous cells of the ileal epithelium. This receptor is now known to be the multiligand 460-kDa protein, cubilin. Transcobalamin may be involved in

TABLE II Physiological Causes of Cobalamin Malabsorption Related to Intrinsic Factor

Organ	Disorder/condition	Pathophysiology
Stomach	Pernicious anemia	Insufficient IF
	Total gastrectomy	Absence of IF
	Subtotal gastrectomy, atrophic gastritis	Decreased acid/pepsin secretion with impaired food cobalamin release
	Abnormal IF expression or structure	Genetic variation in IF
Gut lumen	Vegetarian diet, anorexia	Lack of cobalamin in foods of nonanimal origin
	Zollinger–Ellison syndrome	Low pH with impaired transfer of cobalamin from haptocorrin to IF
	Pancreatic insufficiency	Protease deficiency with impaired transfer of cobalamin from haptocorrin to IF
	Bacterial overgrowth	Competition for uptake of cobalamin by bacteria
Small bowel mucosa	Ileal resection or disease	Decreased number/function of receptors for IF–Cbl complex
	Imerslund–Grasbeck syndrome	Abnormal receptor-mediated endocytosis of IF–Cbl complex
	AIDS enteropathy	Combination of mucosal disease and decreased surface exposure time

transcytosis of Cbl across the enterocyte, because a low concentration of TC receptors has been demonstrated on the apical membrane of enterocytes.

Rapid internalization of IF–Cbl into enterocytes occurs by receptor-mediated endocytosis, and agents that inhibit lysosomal enzymes also inhibit transcytosis of Cbl. Radioautographic studies using ^{125}I-labeled IF have demonstrated localization in endosomes and early lysosomes of the apical cytoplasm. A half-life of 4 h for degradation of internalized IF has been estimated from studies in cultured epithelial cells, and this time frame is consistent with the 2–4 h needed to transfer Cbl from the luminal IF–Cbl complex to the serum TC–Cbl complex. The initial mechanism for intracellular IF degradation proceeds via cathepsin L activity. Following intracellular release of Cbl, Cbl exits the cell bound to TC, which is probably produced in the enterocyte.

Cbl deficiency occurs most commonly in pernicious anemia (loss of IF from chronic autoimmune gastritis), following resection of the gastric body (containing parietal cells the source of IF), with atrophic gastritis (a sequela of *Helicobacter pylori* infection), after resection or disease (usually Crohn's disease) of the terminal ileum (the site of IF–Cbl absorption via receptor-mediated endocytosis), or following malabsorption from any cause (especially AIDS enteropathy) (Table II). Less common causes of deficiency include congenital absence of IF or TC, Imerslund–Grasbeck syndrome (abnormal IF–Cbl receptor, or cubilin), small bowel bacterial or parasitic overgrowth (utilization of Cbl), and pancreatic insufficiency (lack of transfer of Cbl from Hc to IF).

See Also the Following Articles

Cobalamin Deficiency • Parietal Cells • Pernicious Anemia • Vitamin B12: Absorption, Metabolism, and Deficiency

Further Reading

Allen, R. H., Seetharam, B., Podell, E., and Alpers, D. H. (1978). Effect of proteolytic enzymes in the binding of cobalamin to R protein and intrinsic factor. *J. Clin. Invest.* **61**, 47–54.

Alpers, D. H., and Russell-Jones, G. J. (1999). Intrinsic factor and haptocorrin and their receptors. *In* "Chemistry and Biochemistry of B_{12}, Part II: Biochemistry of B_{12}" (R. Banerjee, ed.), pp. 411–440. John Wiley & Sons, New York.

Christensen, E. L., and Birn, H. (2001). Megalin and cubilin: Synergistic endocytic receptors in renal proximal tubule. *Am. J. Physiol. Renal Physiol.* **280**, F562–F573.

Dieckgraefe, B. K., Seetharam, B., Banaszak, L., Leykam, J. F., and Alpers, D. H. (1988). Isolation and structural characterization of a cDNA clone encoding rat gastric intrinsic factor. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 46–50.

Glass, G. B. J. (1974). "Gastric Intrinsic Factor and Other Vitamin B12 Binders." Georg Thieme Publ., Stuttgart, Germany.

Gordon, M. M., Howard, T., Becich, M. J., and Alpers, D. H. (1995). Cathepsin L mediates intracellular ileal digestion of gastric intrinsic factor. *Am. J. Physiol.* **268**, G33–G40.

Hewitt, J. E., Gordon, M. M., Taggart, T., Mohandas, T. K., and Alpers, D. H. (1991). Human gastric intrinsic factor: Characterization of cDNA and genomic clones and localization to human chromosome 11. *Genomics* **10**, 432–440.

Nicolas, J. P., and Guéant, J. L. (1995). Gastric intrinsic factor and its receptor. *Baillière's Clin. Haematol.* **8**, 515–531.

Schneider, Z., and Stroinski, A. (1987). *In* "Comprehensive B_{12}," pp. 297–330. Walter de Gruyter & Co., Berlin.

Seetharam, B. (1999). Receptor mediated endocytosis of cobalamin (vitamin B12). *Annu. Rev. Nutr.* **19**, 173–195.

Intussusception

JOEL SMITH AND ROY MCCAULEY
Tufts–New England Medical Center

intussuscepiens The intestinal loop, or sheath, into which the intussusepting bowel (the intussusceptum) invaginates.
intussusception Telescoping or invagination of one portion of the bowel into the adjacent, distal intestinal lumen.
intussusceptum Invaginating and returning bowel of the intussusception, including the adjacent mesentery.
lead point An anatomic or congenital abnormality, neoplasm, or postsurgical change in the intussusceptum that causes secondary invagination into the intussuscepiens.

Intussusception is the telescoping of one region of the small or large bowel into the adjacent intestinal lumen, secondary to focally abnormal peristalsis. The large majority of intussusceptions present as acute abdominal emergencies in children, although they may occur at any age, including *in utero*. When seen in children, the cause is almost always idiopathic, occurring near the ileocecal valve. Reduction is frequently by liquid or air enema under fluoroscopic or ultrasonic guidance, although some cases require surgery. In adults, intussusception is usually secondary to a definable cause and thus requires surgical reduction and removal of the lead point.

INTRODUCTION

Intussusception has been recognized as a cause of obstruction for nearly 300 years and has been successfully reduced by surgical, hydrostatic, and pneumatic methods for more than 100 years. The frequency is approximately 2 to 4 in 1000, but varies according to the series. It is more common in boys than in girls. Two of three cases are seen within the first year of life, with the vast majority of cases occurring between 3 months and 6 years of age. Most intussusceptions originate near the ileocecal valve (ileocolic intussusception), but may occur elsewhere in the small or large bowel. The tendency is to occur at junctions between "free" and intraperitoneal bowel. Treatment of intussusception first requires resuscitation with fluids and antibiotics. In cases of obstruction, decompression of the stomach with a nasogastric tube may be necessary.

ETIOLOGY

Idiopathic intussusception is likely secondary to kinking of the bowel secondary to inequilibrium between contraction of circular and longitudinal muscles in the bowel wall. This may be secondary to lymphoid hyperplasia of the distal ileum. This abnormal peristalsis may also occur at a lead point, such as a Meckel's diverticulum, neoplasm (lymphoma), polyp (Peutz-Jeghers syndrome, juvenile polyposis), duplication cyst, ectopic pancreas, inspissated stool (cystic fibrosis), hemangioma, or bowel wall hematoma (trauma; Henoch-Schonlein purpura). Recurrent intussusceptions are more common if a lead point is present.

PRESENTATION

Signs and Symptoms

Classically, children with intussusception are previously healthy and well nourished, with sudden onset of severe colic. Approximately 50% will have the triad of abdominal pain, "red currant jelly" stool, and a palpable abdominal mass. These symptoms are caused by compression of the intramural and mesenteric lymphatics and veins, with ultimate arterial compromise, resulting in bowel wall edema, hemorrhage, ischemia, and infarction. Symptoms may progress as continued normal peristalsis results in lengthening of the intussusception. Episodes of pallor and "doubling over" may occur at intervals of approximately 20 min, with pulling up of the knees to the chest, cringing in pain, and occasional vomiting. Between such attacks, the child may be apathetic, normal, or even playful. As the bowel mucosa becomes compromised, the stool will contain the bloody mucus (red currant jelly stool).

Physical Exam

A "sausage-shaped" mass may be palpated, most frequently in the right upper quadrant. Of course, not all cases are typical, and children may have no pain, red currant jelly stool, or mass. In such cases, Henoch-Schonlien purpura, and other forms of obstruction,

 Copyright 2004, Elsevier (USA). All rights reserved.

including malrotation with volvulus, should be considered. Older children may have pain as their only symptom. If this is the case, the differential diagnosis includes simple gastroenteritis. When bowel compromise has occurred, signs of peritonitis may be present.

DIAGNOSIS

Radiographic Evaluation

The role of abdominal radiographs in suspected intussusception is controversial, as false negatives may occur, and correct diagnosis based on radiographic findings alone occurs in less than 50% of cases (see Table I). Radiographic findings may include a lack of gas and stool and/or the presence of a soft tissue mass in the right lower quadrant. The mass may obscure the liver edge. Classic radiographic findings include the "meniscus sign" and the "target sign." A meniscus is formed if the gas-filled distal colon outlines the leading edge of the intussusceptum (Fig. 1). A target shape may be seen if the central, lucent, mesenteric fat of the intussusceptum is identifiable in the soft-tissue mass. With obstruction, distended small bowel loops may be identified. Free air, very rarely identified in children with intussusception, signifies perforation and the need for immediate surgery.

Enema Diagnosis

Until recently, barium enema was the "gold standard" diagnostic modality. Air contrast enema and ultrasound have largely replaced this technique. The diagnosis is confirmed on enema with outlining of the intussusceptum by barium or air (meniscus sign) (Fig. 1). With barium, a "coiled spring" appearance of the bowel may be present if barium coats the outer surface of the intussusceptum and the inner surface of the intussuscipiens.

Ultrasound Diagnosis

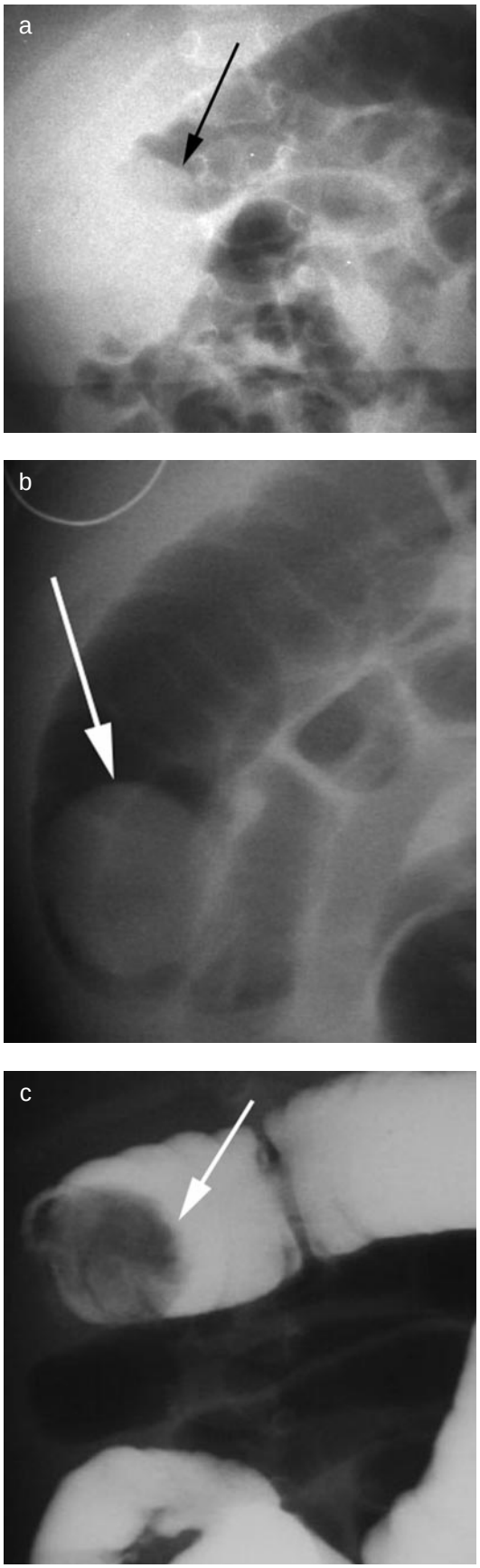
Ultrasound is very sensitive for the diagnosis, especially by an experienced radiologist. The mass is usually large (~5 cm) and displaces adjacent loops. Ultrasound diagnosis and reduction is advantageous as it precludes the need for radiation and is more comfortable for the patient. False positives are rare, but may occur with misidentification of abnormal bowel wall thickening, stool, or psoas muscles. Ultrasound is likely the best method for depicting a lead point in children. Transverse views along the axis of the bowel show central echogenic mesenteric fat and vessels and a hypoechoic outer bowel wall, sometimes referred to as the "doughnut" or target sign (Fig. 2). Longitudinal views may demonstrate a "pseudo-kidney" or "sandwich" sign when imaging through the middle of the intussusception and the "hayfork sign" at images near the apex of the intussusception. These signs are descriptive variations on the appearance of alternating hypoechoic muscle (intussusceptum), hyperechoic mesentery (intussusceptum), and hypoechoic muscle (intussuscipiens). Incidental peritoneal fluid may be identified, but does not necessarily indicate the presence of perforation or peritonitis. Lack of Doppler flow in the intussusceptum may suggest ischemia or necrosis and often indicates a lower rate of reduction.

Computed Tomography Scan Diagnosis

Computed tomography (CT) scan is sometimes used for diagnosis, most often in a symptomatic

TABLE I Comparison of Methods for Diagnosing and Treating Intussusception

	Diagnosis	Reduction rate	Advantages	Disadvantages
Radiograph	Not sensitive or specific	—	Noninvasive	Diagnostic in less than 50%; not therapeutic
Barium enema	Sensitive and specific	70–80%	High success rate; low perforation; evaluation of lead point	Barium irritates peritoneum if perforation occurs
Air enema	Sensitive and specific	Higher than barium	High success rate; low perforation; quicker than barium; less irritating	May not fully evaluate colon and lead point
Ultrasound	Sensitive and specific	Comparable to enema	High success rate; low perforation; no radiation; evaluation of lead point	Occasional false-positives
CT scan	Used for adults	Not therapeutic	Evaluates small bowel and lead points	Not therapeutic
Surgery	—	Very high	Can remove lead point	Invasive



adult. Axial images may show a complex target mass, containing eccentric mesenteric fat (Fig. 3). Mesenteric vessels are frequently identified in the intussusceptum. An appearance analogous to the "coiled spring" on barium enema may be present if the oral contrast outlines intussusceptum. A lead point is occasionally identified.

REDUCTION

Image-Guided Reduction

Introduction

Image-guided reduction is less invasive, has a lower morbidity and cost, and leads to a shorter length of hospitalization, when compared to surgery. Such reduction may be accomplished under fluoroscopic guidance with air, barium, or other contrast material or with ultrasound using saline or, occasionally, air. Ultimately, the choice of technique is often dependent on the experience of the radiologist. With any image-guided technique, reduction is approximately 90% successful. Success rates are lower in younger infants, those with longer duration of symptoms, and those with more severe obstruction.

Barium Enema Reduction

Traditionally, barium reduction follows the "rule of threes." These include raising the barium bag not more than 3–3.5 feet above the table, making not more than three attempts at reduction, and sustaining pressure for not more than 3 min during each attempt. This rule is arbitrary and often physicians attempt reduction more vigorously. Barium has demonstrated high reduction rates, 70–80%, infrequent perforation, and better evaluation of lead points and colonic pathology than air.

Air Enema Reduction

Air reduction has a higher successful reduction rate, but may result in higher perforation rates compared with barium. Reduction is often accomplished more rapidly, with a secondary decrease in radiation exposure. If perforation does occur, air in the peritoneal cavity is less irritating than barium, although tension pneumoperitoneum may rarely occur. Air insufflation

FIGURE 1 (a) Abdominal radiograph demonstrates a "meniscus sign" as the intussusceptum (arrow) is outlined by air in the proximal transverse colon. (b) Spot radiograph from air enema reduction demonstrates displacement of the filling defect to the proximal ascending colon (arrow). Complete reduction was confirmed by reflux of air into the terminal ileum. (c) Following recurrent intussusception in the same patient, reduction was accomplished by barium enema. The intussusceptum is outlined by barium (arrow).

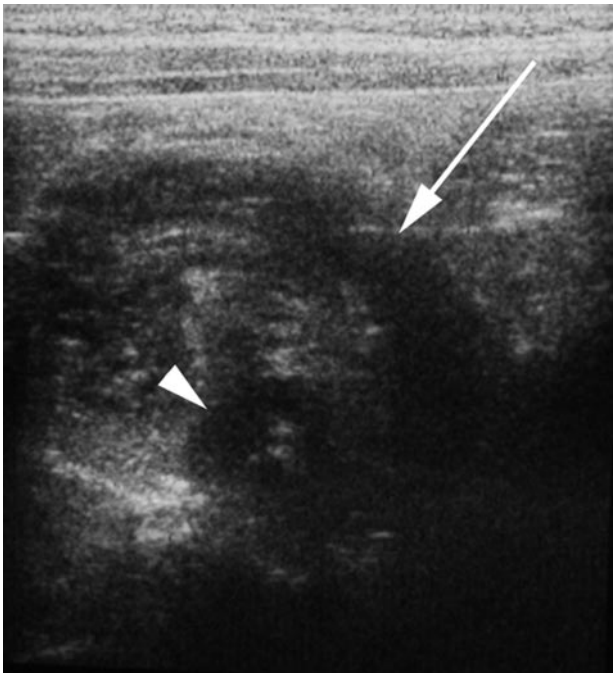


FIGURE 2 Ultrasound image of intussusception. Image obtained axially through the abnormal bowel. The hypoechoic muscle of the intussuscipiens (arrow) surrounds the hyperechoic mesenteric fat and hypoechoic wall of the intussusceptum (arrowhead).

of the colon is usually performed with a pressure release system and manometer, to avoid pressures above which perforations usually occur. The initial suspicion that air had a higher perforation rate than barium was likely related to the initial learning curve and the true rate is probably less than 1% with either method. With either technique, complete reduction is confirmed with free reflux of barium or air reflux into the distal ileum. When possible, the patient may be requested to perform the Valsalva maneuver during treatment as it may help reduction and decrease the risk of perforation.

Ultrasound Reduction

Hydrostatic (saline) reduction under ultrasound evidence is gaining in popularity. It has the advantage of having no radiation exposure. Reduction rates are high and perforation rate is comparable to that of other techniques. Another advantage is a more detailed visualization of the intussusceptum and lead point, if present. Recently, air reduction with ultrasound guidance has been advocated and also has a high success rate in initial studies. Ultrasound may have a role after enema reduction for confirming reduction and evaluating for a lead point.

Surgical Reduction

Contraindications to attempted enema reduction include the presence of peritonitis ("surgical abdomen"). When symptoms have been present for more than 48 h, careful evaluation of the abdomen is needed. Surgical reduction is also necessary if there is perforation during nonoperative reduction. Therefore, the surgical team should be informed when an enema reduction is to be attempted. At surgery, manual reduction is usually curative, but resection and primary anastomosis may be necessary, if manual reduction cannot be accomplished or if a lead point is identified.

PROGNOSIS

Delayed diagnosis or inadequate fluid and antibiotic resuscitation may lead to fatal complications. When treated promptly, however, recovery almost always occurs. There is an approximately 10% rate of recurrence. If complete reduction is accomplished by image-guided methods, each recurrence is considered a new event and alone is not an indication for surgery. To monitor for recurrence, patients are usually observed for 12–24 h after reduction. The major cause of morbidity is perforation, which tends to occur in younger patients and those with longer duration of symptoms. Perforation may occur in the intussusceptum or intussuscipiens. Ultrasound is useful in evaluating patients with recurrence of or persistent abdominal pain after air reduction, at times revealing recurrent intussusception or findings suggestive of an inflammatory process.

POSTOPERATIVE INTUSSUSCEPTION

Postoperative intussusception is an unusual entity, making up less than 1% of cases. It most often occurs after abdominal surgery, but may be seen after thoracotomy as well. Unlike idiopathic intussusception in a child, the jejunum is the most frequent site of origin. Only rarely is a lead point identified, such as a drain, tube, suture line, adhesion, or focus of bowel wall edema. When no lead point is present, the cause is likely related to bowel manipulation or an electrolyte abnormality, with secondary paralytic ileus. Diagnosis is often delayed, so this entity should be considered in cases of postoperative obstruction, especially within 2 weeks of surgery. Operative reduction is usually curative.

ADULT INTUSSUSCEPTION

Intussusception is rare in adults and, when present, rarely results in obstruction. Preoperative diagnosis is

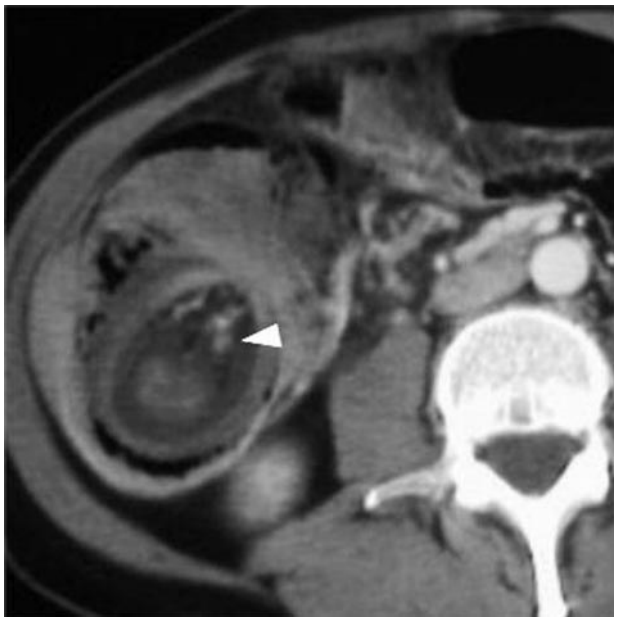


FIGURE 3 Axial CT scan with IV contrast enhancement. The intussusceptum with mesenteric vessels (arrowhead) is seen within the colonic intussuscipiens.

established in only approximately one-third of patients. Common symptoms are intermittent pain, nausea, and vomiting. These symptoms may present acutely, but are frequently chronic over weeks or months. Mild, focal tenderness may be elicited on physical exam. Unlike findings in children, bloody stool and abdominal masses are uncommon. A lead point is almost always present, necessitating surgery. The intussusception occurs most frequently within the small bowel and is almost always secondary to a benign lead point, such as adhesions. The rarer colonic intussusception is more likely to be related to a malignant lead point, predominantly metastatic disease (Fig. 3). CT may be the most accurate means of diagnosis, as the small bowel is not usually visualized by enema. Endoscopy is sometimes employed for diagnosis, and, at times, reduction via air insufflation. Intussusception of the appendix is a separate, rare entity, also seen in adults.

See Also the Following Articles

Barium Radiography • Colonic Obstruction • Computed Tomography (CT) • Radiology, Interventional • Ultrasonography

Further Reading

Allbery, S. M., Swischuk, L. E., John, S. D., and Angel, C. (1998). Post-operative intussusception: Often an elusive diagnosis. *Pediatr. Radiol.* **28**, 271.

Begos, D. G., Sandor, A., and Modlin, I. M. (1997). The diagnosis and management of adult intussusception. *Am. J. Surg.* **173**, 88–94.

Daneman, A., and Alton, D. J. (1996). Intussusception: Issues and controversies related to diagnosis and reduction. *Pediatr. Gastrointest. Radiol. RCNA* **34**, 743–756.

del-Pozo, G., Albillos, J. C., Tejedor, D., Calero, R., Rasero, M., de-la-Calle, U., and Lopez-Pacheco, U. (1999). Intussusception in children: Current concepts in diagnosis and enema reduction. *RadioGraphics* **19**, 299–319.

Dugougeat, F., Navarro, O., and Daneman, A. (2000). The role of sonography in children with abdominal pain after recent successful reduction of intussusception. *Pediatr. Radiol.* **30**, 654.

Gayer, G., Zissin, R., Apter, S., Papa, M., and Hertz, M. (2002). Adult intussusception—A CT diagnosis. *Br. J. Radiol.* **75**, 185–190.

Holcomb, G. W., Ross, A. J., and O'Neill, J. A. (1991). Postoperative intussusception: Increasing frequency or increasing awareness. *South. Med. J.* **84**, 1334–1339.

Navarro, O., Dugougeat, F., Kornecki, A., Shuckett, B., Alton, D. J., and Daneman, A. (2000). The impact of imaging in the management of intussusception owing to pathologic lead points in children. *Pediatr. Radiol.* **30**, 594–603.

Stringer, D. A., and Ein, S. H. (1990). Pneumatic reduction: Advantages, risks and indications. *Pediatr. Radiol.* **20**, 475–477.

Taraneh, A., and Berger, D. L. (1997). Adult intussusception. *Ann. Surg.* **226**, 134–138.

Yoon, C. H., Kim, H. J., and Goo, H. W. (2001). Intussusception in children: US-guided pneumatic reduction—Initial experience. *Radiology* **218**, 85–88.

Iron Absorption

DAVID F. NUTTING
University of Tennessee Health Science Center

divalent metal ion transporter-1 (DMT1) Also called divalent cation transporter-1 (DCT1) and Nramp2. DMT1 moves ferrous iron from the lumen into enterocytes.

duodenal cytochrome b (Dcytb) A ferric reductase found in the brush border membrane of proximal intestinal enterocytes. Dcytb converts the poorly absorbed ferric form of iron (Fe^{3+}) to the ferrous form (Fe^{2+}) that is transported into the cells by divalent metal ion transporter-1.

ferritin A mainly cytosolic molecule that binds excess iron, in both enterocytes and the liver. Ferritin complexes have the capacity to bind approximately 4500 iron molecules.

hereditary hemochromatosis A disorder of iron metabolism in which affected individuals absorb an excessive amount of iron that accumulates in internal organs and may eventually interfere with cellular function.

iron-responsive gene element (IRE) and iron-regulatory protein (IRP) These participate in the regulation of expression of some genes involved in iron absorption. IRPs bind to IREs and increase or decrease the expression of specific genes.

transferrin (Tf) An iron-binding molecule found in several compartments in the body. Circulating transferrin carries iron to and from body tissues and may apprise intestinal cells about body iron status. The cytosol of villus enterocytes contains a similar, but not identical, form of transferrin. Some of the cytosolic transferrin (as apotransferrin, which has no iron bound to it) is secreted into the intestinal lumen, where it promotes iron absorption by chelating two molecules of Fe^{3+} and then binding to a transferrin receptor in the brush border membrane.

Dietary iron is poorly absorbed in most circumstances, but absorption is regulated and increases when needed, such as in pregnancy and iron deficiency. In contrast, the body has few mechanisms for removing excess iron, which must therefore be stored to prevent it from catalyzing reactions that may produce undesired products, such as free radicals. Other than blood loss, as occurs during menstruation, the only mechanism to remove iron is by trapping some of the iron in enterocytes, which are then sloughed off. A typical meat-containing diet provides 10–30 mg iron each day. Of this, only approximately 1 mg is absorbed (3–10% of dietary iron). During pregnancy, iron absorption may increase to provide 5–6 mg/day. However, if a pregnant woman takes iron supplements, efficiency is down-regulated to the normal low percentage of ingested iron. Thus, the amount of iron absorbed increases linearly with the quantity ingested from all sources, but the proportion absorbed decreases. Multiple mechanisms are involved in iron absorption. Some of them are regulated; others appear not to be. Furthermore, the relative contributions of the various pathways to overall iron absorption have not yet been fully quantitated. Most iron absorption takes place in the proximal intestine, mainly the duodenum and proximal jejunum. However, a small amount of iron is absorbed by the stomach, distal small intestine, and colon.

LUMENAL IRON METABOLISM

In the stomach, acid helps to release iron from dietary nutrients and to release heme from ingested hemoglobin and myoglobin in meat. The low pH in the stomach promotes dissolution of iron in the aqueous medium; at very low pH (<3), both Fe^{2+} and Fe^{3+} are soluble. People with deficient acid production absorb iron even more poorly than do normal individuals. Most dietary iron is not in heme and most of this nonheme iron is trivalent (ferric) iron. When the pH is >3, iron (especially the less soluble Fe^{3+}) must be kept soluble; this occurs through the binding of iron to other nutrients. A variety of substances chelate iron, including organic acids (lactic acid, citric acid, ascorbic acid), bile acids, fatty acids, some amino acids, mucin, and apotransferrin secreted by enterocytes. In addition to helping solubilize iron, ascorbate reduces Fe^{3+} to Fe^{2+}. Not all dietary organic acids aid iron absorption, though, and significant amounts of some inhibitors may occur in vegetarian diets. Examples are phytate and oxalate, which are well known to interfere with calcium absorption, but they prevent iron absorption as well. In addition, iron binds to two intestinal proteins, mucin and apotransferrin. Mucin is a large glycoprotein with

 Copyright 2004, Elsevier (USA). All rights reserved.

multiple ion-binding groups. By binding many iron molecules, it reduces the formation of large, insoluble (and thus unabsorbable) iron–hydroxide complexes. Once the mucin-bound iron diffuses to the surface of the enterocyte, the iron can be released for absorption.

ENTEROCYTE IRON UPTAKE, PROCESSING, AND RELEASE

A carrier-mediated process takes up heme iron (see Fig. 1). Fe^{2+} is extracted from heme by the enzyme heme oxygenase. This iron is then taken up by cytosolic

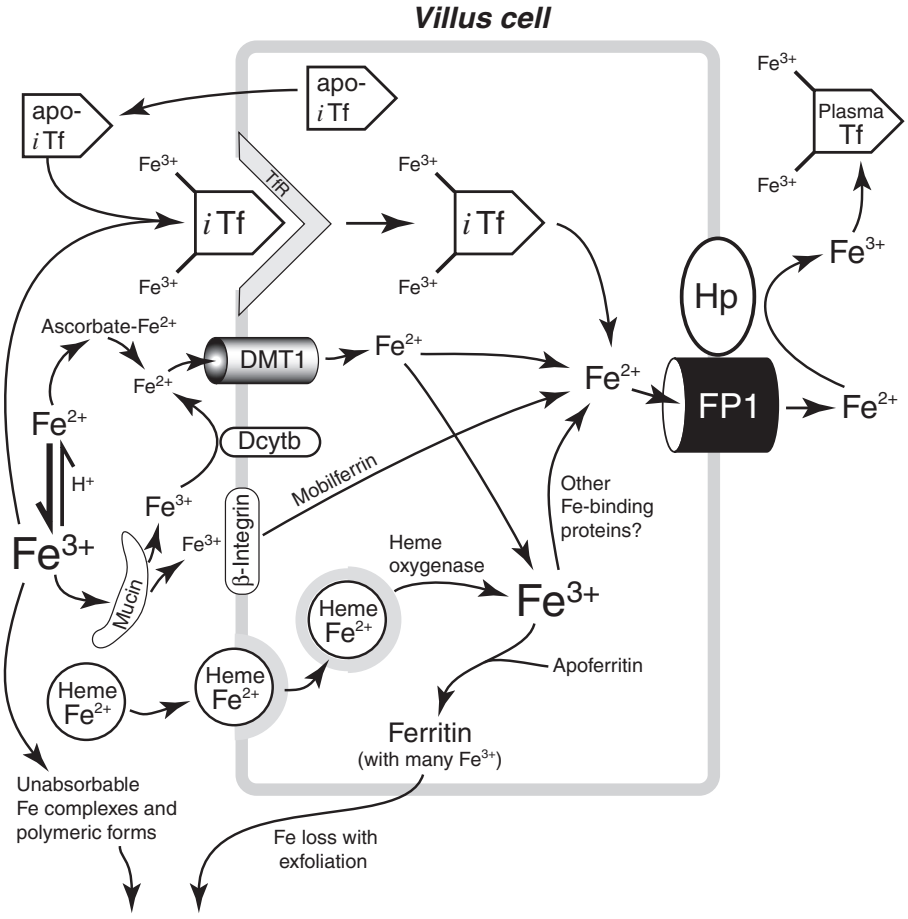


FIGURE 1 Processes involved in iron absorption in the proximal intestine. Iron is released from food by acid and enzymes. Heme is removed from hemoglobin and myoglobin. The heme is absorbed intact, probably by pinocytosis. Inside, heme iron is released and oxidized to Fe^{3+} by heme oxygenase. In the duodenal lumen, the less acidic environment promotes the oxidation of ferrous iron (Fe^{2+}) to the trivalent form. Fe^{3+} is poorly absorbed unless bound. Fe^{3+} may bind to mucin or other molecules, such as ascorbate, or to intestinal apotransferrin (apo-*i*Tf) that has been secreted by enterocytes. Tf-bound iron is carried in via a Tf receptor (TfR). Lumenal iron binding stabilizes the iron and reduces the formation of insoluble iron polymers. Ascorbate also reduces ferric iron back to Fe^{2+}. In addition, Fe^{3+} that reaches the microvilli is either transferred to β_2-integrin, which carries it into the cell, or reduced by duodenal cytochrome b (Dcytb). A well-regulated uptake pathway exists for Fe^{2+}, which is transported into the cell by DMT1 (divalent metal transporter-1). In the cytosol, iron is bound to iron-binding proteins, such as mobilferrin and apotransferrin, and carried to the basolateral side. Excess iron is captured by ferritin. A small amount of iron from ferritin may be released for export, but most stays with the cell and is lost to the lumen when the enterocyte is sloughed at the villus tip. At the basolateral membrane, iron is secreted by a transport protein, ferroportin (FP1). In order to be carried to the body via circulating Tf, it must be oxidized again. This is carried out by a protein, hephaestin (Hp), that is closely associated with FP1.

iron-binding proteins, such as mobilferrin, apotransferrin, and ferritin. The former two transport iron to the basolateral side for export. Ferritin binds Fe^{3+} in times of iron excess.

Some nonheme iron that reaches the apical surface may be absorbed after binding to β-integrin. Once inside the cell, it binds to ferritin or is carried across the enterocyte by cytosolic Fe-binding proteins. The details of the regulation of this pathway are not yet known. Another mechanism involves apotransferrin (apo-Tf). Apo-Tf secreted into the lumen binds two molecules of Fe^{3+}. The iron–transferrin complex then binds to a Tf receptor (TfR) located in the brush border membrane. This complex is absorbed, perhaps by pinocytosis. Once inside, the Fe^{3+} may be reduced, transferred to another molecule of apotransferrin or other cytosolic binding proteins for export, or bound to ferritin for storage or excretion. The physiological regulation of this pathway also has not been established.

A substantial amount of iron is taken up by carrier-mediated transport in the brush border of the villus and this is a regulated process. The transporter is divalent metal transporter-1 (DMT1), which has a marked preference for ferrous iron. Since much of the nonheme iron in the lumen is ferric iron, this carrier-mediated transport requires prior reduction. Brush border membranes have ferric reductase activity provided by the enzyme duodenal cytochrome b (Dcytb). An individual's iron status regulates both Dcytb and DMT1 expression in parallel.

Iron that is not trapped by enterocyte ferritin is carried to the basolateral side, where it is transported outward by a specific protein called ferroportin-1 (FP1). This transporter is also known as iron-regulated gene transcript and metal transport protein-1 and it carries ferrous iron. The divalent iron released at the basolateral side must be reoxidized to trivalent iron before it binds to circulating apotransferrin for transport to the rest of the body. A protein, hephaestin (Hp), is associated with FP1 and may serve this function.

REGULATION

Various stimuli increase iron absorption, for example, increased erythropoiesis following hemorrhage or ascent to high altitude, pregnancy, and iron deficiency. In most cases, the nature of the signal remains to be determined. Tissues such as bone marrow and liver may release as yet unidentified signaling molecules that regulate duodenal enterocytes. In addition, evidence supports the hypothesis that circulating levels of transferrin may play a role in some of these situations.

Intestinal iron absorption appears to operate around a "setpoint" that changes only under specific conditions, such as those listed above (see Fig. 2). When more iron is needed, absorption of both heme and nonheme iron increases, as do the rates of apical iron uptake and basolateral iron release. Expression of three apically targeted proteins, Dcytb, DMT1, and TfR, increases. Conversely, expression of apoferritin decreases, reducing

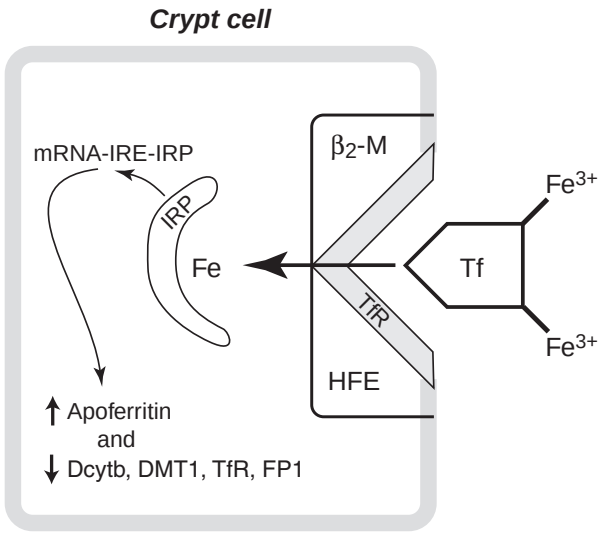


FIGURE 2 Regulation of iron absorption. One mechanism is depicted here. There may be other signals, perhaps originating from the hematopoietic system or the liver. In the mechanism shown, the setpoint for iron absorption is determined in the intestinal crypts. A high level of plasma iron feedsback negatively on the newly formed enterocytes to reduce iron absorption once these cells mature. A fraction of plasma iron-bound transferrin (Tf) escapes from the capillary lumen. This Tf binds to a Tf receptor (TfR) located on the basolateral side of cryptal enterocytes, bringing the Tf-bound iron into the cell. TfR is associated with at least two basolateral membrane proteins, β_2-macroglobulin (β_2-M) and a gene product termed HFE. Defects in HFE lead to unregulated iron absorption, because the iron overload is not sensed by the developing crypt cells. The iron is released and binds to iron-regulatory proteins (IRPs). Under conditions of iron-deficiency, the IRPs that lack iron interact with iron-responsive gene elements (IREs) of certain messenger RNAs. This interaction stabilizes some of the mRNAs, thereby increasing translation (as occurs for the proteins Dcytb, DMT1, TfR, and FP1), or it inhibits translation of other mRNAs (as occurs for apoferritin). These effects become manifest once the cells migrate to the absorptive region along the villi. Thus, iron absorption is increased and loss of absorbed iron is reduced (because less iron is bound to ferritin). The opposite occurs when plasma iron levels are normal or high: when they contain iron, the IRPs do not bind to the IREs, so the expression of proteins that promote iron absorption is diminished, whereas apoferritin expression is high. The apoferritin binds more of the iron that does cross the apical membrane and then takes it back to the intestinal lumen during exfoliation at the villus tip.

iron sequestration in the cytosol. Expression of the basolateral iron exporter FP1 increases, but expression of Hp does not. Perhaps the amount of Hp normally present is sufficient to oxidize any iron transported outward by FP1.

One system of regulation involves iron-regulatory proteins (IRPs) and iron-responsive elements (IREs), but likely there are other regulatory mechanisms. In iron deficiency, more IRPs bind to the iron-regulatory elements in the genome. Iron repletion reverses this. Binding of IRPs increases the transcription of several proteins whose IREs are at the 3′-end of the gene (by stabilizing their mRNAs), but inhibits the transcription of those proteins encoded near the 5′-end, such as apoferritin. Thus, iron absorption is facilitated, whereas iron trapping (and subsequent loss by exfoliation of villus tip cells) is reduced. The converse processes occur during iron repletion and excess, presumably as a consequence of more iron being available to bind to IRPs and thereby inhibiting IRP association with IREs.

Surprisingly, much of the regulation apparently takes place in the newly developing cells in the intestinal crypts, rather than in mature enterocytes in the villi. The strongest evidence for this comes from studies of hereditary hemochromatosis (HH), in which iron absorption increases and remains inappropriately high, due to the lack of normal down-regulating controls. Genetic defects have been identified in a gene product termed HFE. In the basolateral membranes of intestinal crypt cells, HFE associates with at least two other molecules, β_2-macroglobulin and a transferrin receptor. A working hypothesis is that by binding a fraction of transferrin that escapes into the extracellular space, this TfR monitors the circulating Tf level. When iron stores and circulating Tf are adequate, some of this Tf is taken up by crypt cells and this reprograms them so that by the time they reach the villus tip, they have low levels of key proteins such as Dcytb, DMT1, and FP1 and high levels of ferritin. This reduces iron absorption and enhances the loss of absorbed iron after sloughing of the villus tip cells. When HFE is absent or defective, as in some types of HH, the crypt cells are incapable of sensing the high levels of circulating Tf iron and thus produce enterocytes with enhanced iron absorption. Ultimately, this results in excessive deposition of iron in various tissues and disruption of cellular and organ function.

See Also the Following Articles

Dietary Reference Intakes (DRI); Concepts and Implementation • Hereditary Hemochromatosis • Neonatal Hemochromatosis

Further Reading

Cremonesi, P., Acebron, A., Raja, K. B., and Simpson, R. J. (2002). Iron absorption: Biochemical and molecular insights into the importance of iron species for intestinal uptake. *Pharmacol. Toxicol.* **91**, 97–102.

Farrell, J. J. (2002). Digestion and absorption of nutrients and vitamins. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., Vol. 2, pp. 1715–1750. W. B. Saunders, Philadelphia, PA.

Fleming, R. E., and Sly, W. S. (2002). Mechanisms of iron accumulation in hereditary hemochromatosis. *Annu. Rev. Physiol.* **64**, 663–680.

Philpott, C. C. (2002). Molecular aspects of iron absorption: Insights into the role of HFE in hemochromatosis. *Hepatology* **35**, 993–1001.

Irritable Bowel Syndrome

INES SCHWETZ AND LIN CHANG
University of California, Los Angeles

brain–gut axis Bidirectional communication between the central nervous system and the gut via nervous and endocrine pathways to ensure that digestive function is optimal for the overall state of the organism.
dysmenorrhea A condition marked by painful menstruation.
dyspareunia Difficult or painful sexual intercourse.
high-amplitude propagating contractions Colonic contractions that produce mass movements when present in the transverse and ascending colon and defecation when present in the descending or sigmoid colon.
maladaptive coping styles Beliefs and strategies on how to master a crisis or a problem that are not helpful or even counterproductive, e.g., catastrophizing, internalization.
pelvic floor dyssynergia Inappropriate contraction of failure to relax the pelvic floor and/or anal sphincter muscles during attempts to defecate, resulting in the impedance of stool passage.
Sitzmarker colon transit test Simple, validated method using radiopaque markers to calculate segmental and colon transit by counting markers on abdominal radiographs.
stress A perceived threat to homeostasis; an event or stimulus that causes an often abrupt but always a large physiologic change from what is expected.
sustained stress Chronically ongoing stressful situations in a person's life, e.g., marital problems, death of a family member, change in role in life, change in job, significant financial loss.

Irritable bowel syndrome (IBS) is a functional disorder characterized by symptoms of abdominal pain or discomfort associated with a change in bowel habit. It accounts for up to 12% of total visits to primary care providers and is one of the most common conditions diagnosed by gastroenterologists. The economic burden in terms of medical expenses, loss of productivity, and work absenteeism is considerable. The diagnosis of IBS is based on identifying characteristic symptoms and excluding organic disease. Novel treatments have broadened therapeutic options that currently include dietary recommendations, lifestyle modifications, pharmacotherapy, and behavioral and psychological therapy.

EPIDEMIOLOGY

Irritable bowel syndrome (IBS) is a very common disorder worldwide, with prevalence rates of 9–23% in the general population. In a recent report on gastrointestinal (GI) illnesses in the United States, IBS was second only to esophageal reflux disease in its prevalence (15.4 million people). Several population-based studies have found women to be at least twice as frequently affected as men, with prevalence ratios ranging from 2 : 1 to 3 : 1.

The first presentation of patients to a physician usually takes place between the ages of 30 and 50 years, with a decrease in reporting frequency among older subjects. However, not all individuals with IBS seek medical care for their symptoms. Based on epidemiological studies conducted in different countries, 20–75% of persons that meet diagnostic criteria for IBS will seek medical attention at some point in their lives. Cultural factors as well as differences in healthcare systems and access to medical care may influence the wide range of these rates. Although most persons with IBS do not consult a physician, there are still between 2.4 and 3.5 million physician visits annually for IBS in the United States. The burden of illness of the disease was recently estimated to be $1.6 to $10 billion in direct costs and $19.2 billion in indirect costs.

PATHOPHYSIOLOGY

In the past, IBS has been considered to be primarily a disorder of altered gut motility or a purely psychological disorder. Currently, a dysregulated brain–gut axis resulting in altered gut motility and visceral hypersensitivity is felt to be the principal pathophysiologic mechanism underlying IBS. Many factors (both central and peripheral) may contribute to this dysregulation in IBS and include genetic predisposition, chronic stress, inflammation/infection, and environmental factors.

Central dysregulation can result in alterations of autonomic outflow, pain modulation pathways, and

 Copyright 2004, Elsevier (USA). All rights reserved.

neuroendocrine systems. Downstream, these alterations subsequently may lead to disturbances in intestinal motility, visceral sensitivity, and mucosal immune response and permeability. Peripheral mechanisms that can contribute to a dysregulated brain–gut axis include enteric infections or inflammation leading to peripheral and spinal sensitization and/or the release of inflammatory mediators (e.g., cytokines) that can further perpetuate alterations in the brain–gut axis.

Altered Intestinal Motility

In the ileum, colon, and rectum, IBS patients show an enhanced response to a variety of stimuli including meals, distension, and stress. Altered motility patterns have been reported in IBS patients but are not consistently found. Patients with diarrhea and abdominal pain experience significantly more high-amplitude propagating contractions (HAPCs), which are of higher amplitude than those found in healthy controls. These HAPCs are more likely to be associated with a sensation of pain. Motility abnormalities can interact with altered perception of visceral events to produce typical symptoms in IBS. Delayed transit of gas with enhanced perception of gas causes patients to feel bloated. It should be noted, however, that altered motility is present in a subset of patients and cannot be used as a diagnostic marker.

Altered Visceral Perception

The initial clinical observations that led to the hypothesis that IBS patients have visceral hypersensitivity include recurrent abdominal pain, tenderness during palpitation of the sigmoid colon during physical examination, and excessive pain during endoscopic evaluation of the sigmoid colon. Experimental evidence suggests that a variety of perceptual alterations exist in IBS patients: visceral hypersensitivity involving the upper and lower GI tract, as well as with a heightened perception of physiologic intestinal contractions. Multiple studies using various balloon distension paradigms have reported lowered colorectal perceptual thresholds and increased sensory ratings and viscerosomatic referral areas in IBS patients than in healthy individuals. In contrast, most studies have demonstrated that IBS patients do not exhibit generalized hypersensitivity to noxious somatic stimulation. At least two underlying distinct mechanisms contribute to the visceral hypersensitivity in IBS: a hypervigilance toward expected aversive events arising from the viscera and a hyperalgesia, which is inducible by sustained noxious visceral stimulation.

Alteration of Gut Immune Function

A subset of patients associate the development of IBS symptoms with the onset of gastroenteritis. In recent studies, IBS-like symptoms were found in 7–30% of patients who had recovered from a proven bacterial gastroenteritis. Risk factors associated with the development of postinfective IBS include female gender, duration of acute diarrheal illness, and the presence of significance life stressors occurring around the time of the infection. Furthermore, increased cellularity of colonic mucosa and lamina propria has been described in IBS patients. These findings suggest that in a subgroup of patients there is an up-regulation of gut immune function. Further studies are needed to explore these findings and their possible association with clinical factors (e.g., symptom severity, bowel habit, and gender) in unselected IBS patients.

Central Nervous System Modulation

In general, brain–gut interactions play a key role in the modulation of gastrointestinal functioning in health and disease. Signals from the brain to the gut ensure that digestive function is optimal for the overall state of the organism. Signals from the gut to the brain play an important role in reflex regulation of the GI tract and in modulation of mood states.

A unifying hypothesis to explain functional GI disorders is that they result from a dysregulation of the brain–gut neuroenteric system. These proposed alterations in the brain–gut axis are best supported by recent findings in functional neuroimaging studies. Using distal colonic stimulation, several studies have demonstrated alterations in regional brain activation in IBS patients compared to healthy control subjects. These brain regions include the anterior and midcingulate cortices, insula, and dorsal pons (in the region of the periaqueductal gray), which are some of the most consistently activated brain areas in response to visceral as well as to somatic nociceptive stimuli. One area that is consistently activated to a greater degree in IBS patients than control subjects is the anterior midcingulate cortex, a brain region concerned with cognitive processing of sensory input including attentional processes and response selection. Furthermore, midcingulate activation correlates with the subjective unpleasantness of visceral and somatic pain. These observations suggest that IBS patients may fail to use CNS down-regulating mechanisms in response to incoming or anticipated visceral pain. They show altered activation or deactivation of brain areas involved in the emotional or cognitive processing of visceral stimuli, ultimately resulting in the amplification of pain perception.

Role of Stress and Psychosocial Factors

Stress is widely believed to play a major role in the pathophysiology and clinical presentation of IBS. It has been postulated that in the predisposed individual, sustained stress can result in permanent increased stress responsiveness of central stress circuits and vulnerability to develop functional and affective disorders. Stress may be central (e.g., psychological distress) or peripheral (e.g., infection, surgery). Numerous studies indicate that IBS patients report more lifetime and daily stressful events, including abuse, than medical comparison groups or healthy controls. In addition, in IBS patients stress is strongly associated with symptom onset, exacerbation, and severity. Even though the effects of stress on gut function are universal, patients with IBS appear to have greater reactivity to stress than healthy individuals.

A large proportion of patients with IBS or other functional bowel disorders have concurrent psychological disturbances, particularly those with severe symptoms or those seen in tertiary care referral centers. In tertiary care centers, the prevalence of psychiatric disorders in IBS patients ranges from 40 to 90%. Psychosocial factors have been recognized to modify the illness experience and influence healthcare utilization and treatment outcome. Factors that adversely affect health status and clinical outcome include a history of emotional, sexual or physical abuse, stressful life events, chronic social stress, anxiety disorders, or maladaptive coping styles.

A current conceptual model of the role of psychosocial factors in IBS suggests that early adverse life experiences influence later physiologic functioning, stress responsiveness, and susceptibility to developing a functional disorder. A stressor may produce symptoms through changes in gastrointestinal function, through central amplification of normal gut signals, or through a combination of both. The combined effect of altered physiology and psychosocial status via the brain–gut axis will modulate symptom experience, illness behaviors, and ultimately clinical outcome. Thus, although psychosocial factors are not etiologic to IBS, they are important for the understanding and treatment of the disorder.

Gender Differences in IBS

It is well established that a greater number of women seek healthcare services for symptoms of IBS than men. Reports of nausea, bloating, constipation, and extraintestinal symptoms seem more prevalent in women with IBS than in men. Furthermore, recent clinical trials suggest that gender differences in response to pharmacological treatments also occur. Numerous factors need to be considered in the exploration of these differences. These factors include biological, hormonal, behavioral, psychological, and sociocultural differences between men and women. Several studies indicate that the menstrual cycle influences GI symptoms, which are reportedly increased immediately before and during menses. In addition, brain imaging studies have reported gender differences in the central processing of aversive information originating from pelvic viscera in IBS. Further studies are needed to provide additional insight into gender differences in the epidemiology, pathophysiology, symptom expression, and response to treatment in IBS.

In summary, IBS is best viewed as a biopsychosocial disorder in which dysregulation of neuroenteric systems is heavily influenced by central and/or peripheral stress. Modulation of brain–gut pathways, including autonomic and pain modulation systems, by psychologic and biologic processes results in physiologic alterations affecting gut motility and visceral perception.

CLINICAL FEATURES

Gastrointestinal Symptoms

Abdominal Pain/Discomfort

The hallmark symptoms of IBS are chronic abdominal pain and/or discomfort. The pain is usually described as crampy or as a generalized ache with superimposed periods of abdominal cramps. The intensity and location of abdominal pain is highly variable even within a single patient. Several factors can exacerbate or reduce pain in IBS. Patients often report worsening of symptoms in periods of stress or emotional distress such as problems at work or in a relationship. Ingestion of certain foods may exacerbate symptoms, usually up to 60 to 90 min after the meal.

Altered Bowel Habits

Based on bowel habits, patients are usually subclassified into those having diarrhea-predominant IBS (IBS-D) or constipation-predominant IBS (IBS-C). In many patients, the predominant bowel habit is not stable over time but may alternate between diarrhea and constipation. If this occurs in a short time period, the pattern is referred to as alternating bowel habit (IBS-A). As many as 30–50% of patients alternate between diarrhea and constipation.

IBS-D accounts for approximately one-third of cases. Stools in these patients are typically loose and watery but of normal daily volume. Usually, diarrhea occurs only during the waking hours. Many patients

experience fecal urgency and loose stools associated with stressful events and an urgent desire to defecate is often noted postprandially.

In IBS-C (30% of IBS cases), stools are usually hard and pellet-like and long periods of straining on the toilet are often reported. A sensation of incomplete evacuation may prompt patients to make multiple attempts at stool passage over a short time period.

Other Gastrointestinal Symptoms

Symptoms of upper GI tract dysfunction are common in IBS patients. These symptoms include early satiety, heartburn, nausea, abdominal fullness, bloating, and vomiting. It is possible that upper GI symptoms are exclusively related to IBS. However, there is evidence for an overlap in patient populations between IBS and another functional gastrointestinal disorder called functional dyspepsia (FD). FD is defined as the presence of persistent or recurrent pain or discomfort centered in the upper abdomen, without identifiable cause by conventional diagnostic means. Epidemiological data suggest that the overlap of FD and IBS may range from 13 to 87% of patients. The wide range of reported overlap can be explained by the different diagnostic criteria applied across studies, in particular in regards to FD.

Extraintestinal Symptoms

IBS patients commonly report extraintestinal complaints. Prevalent symptoms include fatigue, musculoskeletal pain, urinary complaints, sleep, and sexual disturbances. IBS patients have a high incidence of genitourinary dysfunction (e.g., dysmenorrhea, dyspareunia, impotence, nocturia). Sleep-related disturbances such as difficulty falling asleep or repeated awakening also occur frequently in IBS patients. Furthermore, a high prevalence of psychiatric diagnoses, most commonly anxiety disorders and depression, has been reported. Epidemiologic studies have confirmed the clinical impression that IBS and other functional disorders often overlap in the same patients. In particular, there seems to be a significant overlap with fibromyalgia, a chronic somatic pain disorder, and interstitial cystitis, a chronic pelvic pain syndrome with symptoms referable to the bladder.

DIAGNOSIS

Symptom-Based Criteria

The clinical diagnosis of IBS is based on identifying symptom criteria with a "positive diagnosis" and excluding, in a cost-effective manner, other conditions with similar clinical presentation, which may include organic (e.g., inflammatory bowel disease, GI infections, and malabsorption syndromes) or other functional disorders (e.g., functional diarrhea or bloating, pelvic floor disorders, or slow-transit constipation). Because IBS is a functional disorder and biologic markers do not exist, it is diagnosed using the recently revised symptom-based ROME II criteria (Table I). Additional symptoms, which support the diagnosis of IBS, include bloating, urgency, sensation of incomplete evacuation, and mucus in the stool. When "alarm signs" such as weight loss, fever, anemia, GI bleeding, family history of colon cancer, or onset of disease late in life are excluded, the ROME II criteria reach a sensitivity of 65%, a specificity of 100%, and a positive predictive value of 100%. In a recent study, 93% of patients diagnosed based on ROME II criteria and the absence of red flags were found to be true positive cases after 2 years. Hence, if diagnosis is made properly, the risk of missing organic disease is low.

Evaluation

Initial Evaluation

During the first visit, a detailed history should be followed by a physical examination. This is useful to identify findings not consistent with IBS (e.g., enlarged liver, abdominal mass, signs of bowel obstruction) and to meet the patient's expectations of a thorough examination. In general, physical examination in IBS patients is unremarkable with the exception of mild–moderate abdominal tenderness consistent with visceral hypersensitivity in these patients. Care should be taken to avoid unnecessary investigations that are costly and

TABLE I Diagnostic Criteria for Irritable Bowel Syndrome as Developed by International Consensus

At least 12 weeks, which need not be consecutive, in the preceding 12 months of abdominal discomfort or pain that has two out of three features:
1. Relieved with defecation; and/or
2. Onset associated with a change in frequency of stool; and/or
3. Onset associated with a change in form (appearance) of stool.

Symptoms that cumulatively support the diagnosis of IBS:
- Abnormal stool frequency (>3 bowel movements per day or <3 bowel movements per week)
- Abnormal stool form (lumpy/hard or loose/watery stool)
- Abnormal stool passage (straining, urgency, or feeling of incomplete evacuation)
- Passage of mucus
- Bloating or feeling of abdominal distension

Adapted from ROME II (2000), with permission.

harmful. A recent systematic review of studies investigating the utility of diagnostic tests in IBS concluded that there is insufficient evidence to recommend the routine performance of a standardized battery of tests in patients who meet symptom-based criteria.

Minimal diagnostic procedures to exclude organic disease include complete blood count, erythrocyte sedimentation rate, blood chemistries, thyroid panel, and examination of stool for occult blood and ova, and parasites (latter in diarrhea-predominant patients). A colon examination consisting of a colonoscopy or barium enema and flexible sigmoidoscopy should also be performed in patients 50 years or older. Once the initial therapeutic strategy is found to be unrevealing, the physician should initiate symptomatic treatment for IBS and reassess in 3 to 6 weeks. When red flag signs are present and for those patients that present with short symptom duration or worsening severity, onset late in life, or a family history of colon cancer, a more detailed diagnostic evaluation should be considered (Fig. 1).

Further Diagnostic Approach Based on Predominant Symptoms

Depending on the quality and severity of a patient's symptoms and his or her response to treatment, further diagnostic studies may be performed and are based on whether diarrhea, constipation, or abdominal pain is the predominant symptom.

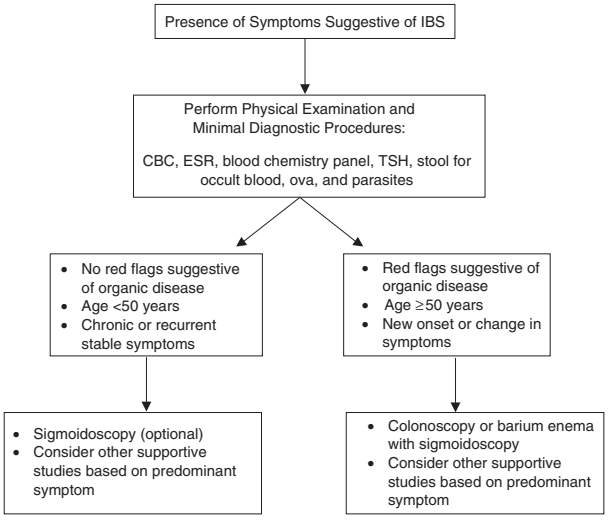


FIGURE 1 The diagnosis of IBS should be based on the history and physical examination as well as the presence or absence of red flag signs suggestive of organic disease. Stool should be examined for ova and parasites in patients who are diarrhea predominant and not in those who are constipation predominant. CBC, complete blood count; ESR, erythrocyte sedimentation rate; TSH, thyroid-stimulating hormone.

In patients with constipation-predominant IBS with infrequent bowel movements, a colon transit study by Sitzmark technique or scintigraphy is recommended. When symptoms, such as dyschezia or incomplete evacuation, suggest obstruction to defecation, further anorectal testing should be considered. This includes anorectal motility testing with balloon expulsion, defecography, and possibly dynamic pelvic magnetic resonance imaging. These studies may be useful in the diagnosis of pelvic floor dyssynergia, rectal prolapse, rectocele, and anismus.

If diarrhea is persistent, colonoscopy or sigmoidoscopy with mucosal biopsies should be performed to rule out collagenous or lymphocytic colitis even in the absence of macroscopic mucosal findings. In patients with refractory diarrhea, especially in those with associated weight loss, further testing should include 24 h stool volume and 72 h fecal fat. If an increase in stool volume is determined (> 400 cc/day), screening for electrolytes and laxatives is useful. Upper endoscopy with biopsy specimen from distal duodenum or proximal jejunum is helpful in excluding celiac sprue, Whipple's disease, intestinal lymphoma, and infections. A jejunal aspirate can be obtained to rule out *Giardia lamblia* or bacterial overgrowth. Finally, when postprandial symptoms of bloating and gaseousness accompany diarrhea and suggest bacterial overgrowth, a breath hydrogen test can be considered.

In patients with pain predominance, an abdominal and pelvic CT scan may help exclude intra-abdominal masses.

TREATMENT OPTIONS

The treatment of IBS can be challenging to both the clinician and the patient. It is therefore important for the physician to strive to gain the patient's confidence with a concise, appropriate work-up and by offering reassurance that IBS is a common, functional disorder that can impact quality of life but without long-term health risks. The therapeutic management of IBS patients requires a multidisciplinary approach including education, reassurance, dietary recommendations, lifestyle modifications, pharmacotherapy, and psychosocial intervention in more severe cases.

Severity of Symptoms

Patients can be categorized into groups based on symptom severity. Many IBS patients have mild intermittent symptoms that correlate closely with

gut-related physiologic events (e.g., pain/diarrhea worse after meal, stress, or menses). Primary care physicians, rather than specialists, usually treat these patients. Treatment is directed toward education, reassurance, and dietary/lifestyle changes. Some patients suffer from moderate symptoms that are intermittent but can at times be disabling. Symptoms can be associated with considerable symptom-related distress. These patients are often evaluated by specialists and seem to benefit from a combination of pharmacotherapy and psychological intervention. Finally, a relatively smaller subgroup of IBS patients present with severe symptoms characterized as refractory and severely disabling. They predominate among patients seen in tertiary referral centers and frequently have psychiatric co-morbidity and psychosocial difficulties (history of adverse life events or maladaptive coping styles) associated with high healthcare utilization. In addition to pharmacological treatment, often with antidepressant medication, and a referral to mental health professionals, an ongoing and effective relationship with a healthcare provider through brief and regularly scheduled visits is beneficial.

Establishing an Effective Physician–Patient Relationship

An effective physician–patient relationship is the cornerstone of successful management of an IBS patient. The physician is encouraged to listen actively to identify the patient's concerns and his or her understanding of the disorder. It is important to set realistic and shared goals and to involve the patient in treatment decisions rather than issuing directives.

After the physician assesses the patient's level of knowledge about the disorder, information should be provided verbally and/or in print to enhance the patient's understanding. Patients commonly want to understand the basis of their symptoms and sometimes seek validation that their symptoms are "real." The clinician can explain that the intestine is overly sensitive to a variety of stimuli such as food, hormonal changes, or stress. These stimuli can produce enhanced visceral motility or visceral perception or both. These alterations in physiologic function are experienced as pain or discomfort, diarrhea or constipation, bloating, or any combination anywhere in the abdomen.

Reassurance that IBS is a benign disorder even though it can drastically affect quality of life should be given after the physician has appreciated the patient's concerns and worries and after an adequate diagnostic workup. If reassurance is given prior to evaluation or in a perfunctory manner, the patient is likely to reject it.

This kind of treatment approach is associated with improved patient satisfaction and reduced physician visits. Patients who feel adequately informed tend to have less healthcare visits.

Dietary Recommendations and Lifestyle Modifications

Symptom Monitoring

Monitoring symptoms by asking the patient to use a diary for 2–3 weeks can assist the physician in identifying the presence of possible aggravating factors. The diary may help to reveal dietary indiscretions or specific stressors that were not previously considered and can give the patient a sense of participation in the planning of care. Physician and patient together can review diet, lifestyle, or behavioral modifications.

Dietary Measures

Patients often attribute their symptoms to specific food substances. There may be some patients that have intolerance to certain foods and may benefit from elimination of foods that trigger their symptoms. It is more likely, however, that most patients experience symptoms as a generalized effect of eating without a specific food trigger. Foods known to be associated with increased flatulence include beans, onions, celery, carrots, raisins, bananas, apricots, prunes, brussels sprouts, wheat germ, pretzels, and bagels. Patients suffering from excess gas may benefit from avoiding these foods.

Lactose intolerance can mimic symptoms of IBS and may co-exist with IBS. A dietary history for carbohydrate intolerance should be assessed, particularly in patients with excess gas with and without diarrhea. Intolerance to other carbohydrates (fructose and sorbitol) may also contribute to GI symptoms. Therefore, a history of carbohydrates should be obtained and managed accordingly if found. However, care should be taken to avoid an unnecessarily restrictive diet.

Role of Fiber in IBS

The most widely recommended agents for treatment of IBS are dietary fiber supplements. These agents serve to enhance water-holding properties of the stool, form gels to provide stool lubrication, provide bulk for the stool, and bind agents such as bile acids that may be responsible for some symptoms in IBS. Soluble fiber, such as pectin, psyllium, or oat bran, is thought to act by enhancing water-retentive properties of stool, whereas insoluble fiber, such as cellulose or lignin, is likely to be a more effective bulking agent. Increased dietary fiber or psyllium products are frequently

recommended for IBS-C patients, even though, as a group, these patients do not consume less dietary fiber than controls. In IBS-C patients, fiber accelerates oro-anal transit time and significant improvement of constipation was demonstrated when 20 to 30 g of fiber was consumed. In summary, whereas fiber has an established role in relieving constipation symptoms, its value in IBS for the relief of diarrhea is controversial and it is not helpful for pain.

Pharmacological Treatment

At present, no pharmacological agent has been found to improve all symptoms of IBS. Therefore, the pharmacotherapeutic approach should be tailored to remedy the patient's most bothersome symptom. However, novel agents have been shown to improve the multiple symptoms of IBS.

The Constipation-Predominant Patient

The principles of management of IBS-C patients are to improve stool frequency and consistency and reduce the effort needed for defecation. Bulking agents such as psyllium, bran, and methylcellulose, in concert with a high-fiber diet, are recommended. Patients should be informed that fiber supplementation may take several weeks to produce a satisfactory result and that fiber should be introduced slowly to prevent excess distension and gas. Patients with an inadequate response should be given osmotic agents such as milk of magnesia or polyethylene glycol. If the patient is unresponsive or intolerant of osmotic laxatives, stimulant laxatives can be prescribed on an intermittent basis, e.g., three times per week. Tegaserod is a recently approved 5-HT_4 (serotonin type 4) agonist that is recommended for treating the multiple symptoms of IBS with constipation in women (see below).

The Diarrhea-Predominant Patient

IBS-D is associated with acceleration of small bowel and proximal colonic transit. The management of these patients focuses on the reduction of defecation frequency and urgency and improvement in stool consistency. The agents most commonly employed are the synthetic opiate derivatives loperamide (Imodium) and diphenoxylate (Lomotil). Loperamide decreases intestinal transit, enhances intestinal water and ion absorption, and increases anal sphincter tone at rest. These physiologic actions are thought to explain the improvement in diarrhea, urgency, and fecal soiling observed in patients with IBS. The effect on resting anal sphincter tone possibly helps to reduce fecal soiling at nighttime when the internal sphincter is the predominant mechanism of continence. Systematic review of randomized clinical trials has demonstrated placebo superiority of loperamide in controlling IBS-associated diarrhea but not in relieving abdominal pain. Loperamide is prescribed at a dose of 2 mg orally after each loose stool with a maximum of 16 mg/kg but can also be used to reduce postprandial urgency. It can also be used prophylactically as a means of improving control at times of anticipated stress or other adverse stimuli (e.g., before exercise, social gatherings, or eating in a restaurant). Common side effects include drowsiness, dizziness, dry mouth, constipation, nausea, and abdominal pain/distension. Loperamide is generally favored over diphenoxylate because the latter contains atropine and may cause adverse anticholinergic side effects that can be worrisome in the elderly (e.g., urinary retention, glaucoma, tachycardia, and glaucoma).

Cholestyramine may be considered for a subgroup of patients with cholecystectomy or who may have bile acid malabsorption. However, evidence for its efficacy in IBS is limited. The starting dose is 4 g orally before meals and can be increased up to 4 g three times daily.

The Pain-Predominant IBS Patient

Severe abdominal pain represents one of the most difficult symptoms to treat in IBS. Antispasmodic drugs, especially anticholinergics, are most commonly prescribed and some patients, especially those with postprandial symptom exacerbation, may benefit from such therapy. The rationale behind the use of antispasmodic agents is that they relax intestinal smooth muscle and thereby decrease cramps and painful sensations. In one meta-analysis of smooth muscle relaxants in IBS, five drugs showed efficacy over placebo in improving abdominal pain: cimetropium bromide, pinaverium bromide, octylonium bromide, mebeverine, and trimebutine. However, none of these drugs underwent extensive trials in North America and they have not been approved by the Food and Drug Administration (FDA) for use in the United States. Although the available agents have not been shown to be superior over placebo, they are often used on an as-needed basis up to four times a day for acute attacks of pain, distension, or bloating. Agents such as dicyclomine or mebeverine seem to retain efficacy when used on a *pro re nata* basis, but become less effective with chronic use.

Tricyclic antidepressants (TCAs; e.g., amitriptyline, imipramine, doxepin) are now frequently used in the treatment of IBS, particularly in patients with more severe or refractory symptoms, impaired daily function, and associated depression and panic attacks. A recent systematic review uncovered only one high-quality trial

of amitriptyline, which showed a trend for global symptom improvement. The efficacy of TCAs as a visceral analgesic agent at lower doses than those used to treat depression suggests neuromodulatory analgesic properties. Treatment with these agents should start with low doses (e.g., 10–25 mg nightly) and titrated up as needed to the lowest most effective therapeutic dose, e.g., amitriptyline (Elavil) or doxepin (Sinequam) at a dose of 10–50 mg at bedtime. These agents may have strong sedative effects and can be used to promote sleep with a single nightly dose.

Selective serotonin reuptake inhibitors are often used to treat IBS patients although they have not yet been shown to reduce IBS symptoms in well-designed published studies. They have been shown to be effective in treating fibromyalgia alone or in combination with a TCA. They may be most useful in treating the co-morbid psychological distress symptoms in IBS patients.

Emerging Therapies

Serotonin type 3 and 4 receptors are involved in mediating reflexes that control gastrointestinal motility and secretion, bowel function, and perception of pain. They have recently been studied extensively as novel targets for pharmacological therapy of IBS.

Tegaserod (Zelnorm), a partial 5-HT_4 agonist, appears to be a promising agent for treatment of constipation-predominant IBS. Tegaserod was associated with an acceleration of orocecal transit time and superiority to placebo in global relief of IBS symptoms in female IBS patients. The effective dose is a total of 12 mg per day in two divided doses prior to meals. The drug appears to be safe with no serious adverse events reported in clinical trial programs. Only small but significant increases in the prevalence of headache and diarrhea, often transient, were associated with tegaserod compared to placebo. The FDA approved this medication in July 2002 for the short-term treatment of the multiple symptoms of IBS with constipation in females.

Prucalopride, a full 5-HT_4 agonist, has recently been demonstrated to accelerate colonic transit time in functional constipation. The effect of the agent on abdominal pain has not been adequately investigated. Clinical trials have been discontinued because of carcinogenicity in animals.

5-HT_3 antagonists reduce visceral pain, rate of colonic transit, and small intestinal secretion. Alosetron (Lotronex), a selective 5-HT_3 antagonist, is effective in relieving pain, normalizing bowel frequency, and reducing urgency in diarrhea-predominant, female patients with IBS. The most common side effect is constipation. Acute ischemic colitis was observed in 0.1–1% of patients receiving the medication, which led to the voluntary withdrawal from the market in November 2000. In June 2002, the FDA approved the restricted use of alosetron to treat women with severe diarrhea-predominant IBS. This medication became available in November 2002. Other drugs in this class (i.e., cilansetron) are currently undergoing multicenter phase III clinical trials.

Other novel drugs that are being investigated as therapeutic agents for IBS include neurokinin (NK)-1 and NK-3 antagonists, cholecystokinin-1 (CCK_1) antagonist, and κ-opioid agonist. It is hoped that continued research in the pathophysiologic mechanisms of IBS will result in the identification of potential targets of action from which more effective therapies for IBS can be developed.

Psychological Treatment

Psychological treatment can be considered when IBS symptoms are moderate to severe, when symptoms have failed to respond to medical treatments, or when there is evidence that stress or psychological factors are contributing to GI symptom exacerbation. The symptoms associated with a favorable response are diarrhea, abdominal pain, and psychological distress. The patient's understanding of the rationale for such a treatment and his or her motivation to engage in it are critical to a successful outcome. A recent review demonstrated the superiority of psychological treatment over conventional medical therapy in 8 of 13 studies. To date, no psychological intervention technique is known to be superior to another. Therefore, the choice of the specific treatment option depends on patient requirements and resource availability.

Cognitive-behavioral therapy consists of a wide range of strategies and procedures designed to bring about alterations in the patient's perception of his or her situation and ability to control GI symptoms. Relaxation training includes a variety of different methods to teach patients to counteract the psychological effects of stress and anxiety. Among the most widely used techniques are progressive muscle relaxation training, biofeedback, and autogenic training.

The efficacy of gut-directed hypnotherapy in IBS has been demonstrated in several studies. The hypnotic state is a state of heightened suggestibility. Following induction, the therapist couples progressive muscle relaxation with "gut-directed" therapeutic suggestions. Recent results suggest that hypnotic relaxation increases the distension volume required to induce discomfort, whereas anger induced in the hypnotic state reduces this threshold.

SUMMARY

IBS is one of the most common functional bowel disorders and is characterized by the presence of chronic or recurrent abdominal pain and/or discomfort associated with altered bowel habits. It is prevalent in 9–23% of the general population and accounts for 12% of primary care visits and 28–37% of visits to gastroenterologists. It appears to be more common in women, particularly those with constipation-predominant symptoms. The healthcare and economic burden of IBS is considerable and has been estimated to be up to $30 billion per year. A unifying hypothesis to explain functional GI disorders is that they result from a dysregulation of the brain–gut neuroenteric system, which may result in altered GI motility, visceral perception, autonomic responses, and central nervous system modulation of visceral sensory information. The diagnostic approach in IBS is based on making a positive symptom-based diagnosis of IBS and excluding organic disease. The management of IBS includes nonpharmacologic (e.g., education, reassurance, and psychological treatment) and pharmacologic approaches and should be based on predominant symptoms, symptom severity, and the presence of co-morbid psychological features.

See Also the Following Articles

Anti-Diarrheal Drugs • Brain-Gut Axis • Constipation • Diarrhea • Dietary Fiber • Food Intolerance • Gastric Motility • Nausea • Psychosociology of Irritable Bowel Syndrome • Stress

Further Reading

Bouin, M., Plourde, V., Boivin, M., Riberdy, M., Lupien, F., Laganiere, M., Verrier, P., and Poitras, P. (2002). Rectal distention testing in patients with irritable bowel syndrome: Sensitivity, specificity, and predictive values of pain sensory thresholds. *Gastroenterology* **122**, 1771–1777.

Camilleri, M., Northcutt, A. R., Kong, S., Dukes, G. E., McSorley, D., and Mangel, A. W. (2000). Efficacy and safety of alosetron in women with irritable bowel syndrome: A randomised, placebo-controlled trial. *Lancet* **25**, 1035–1040.

Cash, B. D., Schoenfeld, P., and Chey, W. D. (2002). The utility of diagnostic tests in irritable bowel syndrome patients: A systematic review. *Am. J. Gastroenterol.* **97**, 2812–2819. [Review]

Chang, L., and Heitkemper, M. M. (2002). Gender differences in irritable bowel syndrome. *Gastroenterology* **123**, 1686–1701.

Drossman, D. A., Camilleri, M., Mayer, E. A., and Whitehead, W. E. (2002). AGA technical review on irritable bowel syndrome. *Gastroenterology* **123**, 2108–2131.

Drossman, D. A., Creed, F. H., Olden, K. W., Svedlund, J., Toner, B. B., and Whitehead, W. E. (1999). Psychosocial aspects of the functional gastrointestinal disorders. *Gut* **45**(Suppl. 2), 1125–1130.

Houghton, L. A., Calvert, E. L., Jackson, N. A., Cooper, P., and Whorwell, P. J. (2002). Visceral sensation and emotion: A study using hypnosis. *Gut* **51**, 701–704.

Jackson, J. L., O'Malley, P. G., Tomkins, G., Balden, E., Santoro, J., and Kroenke, K. (2000). Treatment of functional gastrointestinal disorders with antidepressant medications: A meta-analysis. *Am. J. Med.* **108**, 65–72.

Jailwala, J., Imperiale, T. F., and Kroenke, K. (2000). Pharmacologic treatment of the irritable bowel syndrome: A systematic review of randomized, controlled trials. *Ann. Intern. Med.* **18**, 136–147.

Mayer, E. A., Naliboff, B. D., Chang, L., and Coutinho, S. V. (2001). Stress and irritable bowel syndrome. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G519–G524.

Mertz, H., Morgan, V., Tanner, G., Pickens, D., Price, R., Shyr, Y., and Kessler, R. (2000). Regional cerebral activation in irritable bowel syndrome and control subjects with painful and nonpainful rectal distention. *Gastroenterology* **118**, 842–848.

Muller-Lissner, S. A., Fumagalli, I., Bardhan, K. D., Pace, F., Pecher, E., Nault, B., and Ruegg, P. (2001). Tegaserod, a 5-HT(4) receptor partial agonist, relieves symptoms in irritable bowel syndrome patients with abdominal pain, bloating and constipation. *Aliment Pharmacol. Ther.* **15**, 1655–1666.

ROME II (2000). "The Functional Gastrointestinal Disorders. Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. A. Drossman, E. Corazziari, N. J. Talley, W. G. Thompson, and W. E. Whitehead, eds.), 2nd Ed. Degnon Associates, McLean, VA.

Sandler, R. S., Everhart, J. E., Donowitz, M., Adams, E., Cronin, K., Goodman, C., Gemmen, E., Shah, S., Avdic, A., and Rubin, R. (2002). The burden of selected digestive diseases in the United States. *Gastroenterology* **122**, 1500–1511.

Vanner, S. J., Depew, W. T., Paterson, W. G., DaCosta, L. R., Groll, A. G., Simon, J. B., and Djurfeldt, M. (1999). Predictive value of the ROME criteria for diagnosing the irritable bowel syndrome. *Am. J. Gastroenterol.* **94**, 2912–2917.

Kaposi's Sarcoma

SONALI S. MASTER AND JOHN M. CARETHERS
University of California, San Diego

hemorrhage Loss of blood from blood vessels.
intussusception Telescoping or invagination of one portion of the bowel into the adjacent, distal intestinal lumen.

Kaposi's sarcoma, a multicentric malignant vascular neoplasm, was first described in 1872. In its classic form, it was originally seen in elderly men of Eastern European or Mediterranean background. With the advent of acquired immunodeficiency syndrome (AIDS), epidemic Kaposi's sarcoma became more widely recognized in its cutaneous form and is one of the most common AIDS-associated cancers in the United States.

EPIDEMIOLOGY

Extracutaneous Kaposi's sarcoma (KS), including visceral KS, has been well described in the setting of cutaneous disease. Reported sites of visceral involvement include the esophagus, stomach, small and large bowels, liver, spleen, pancreas, mesentery, and biliary tree. Although KS can develop anywhere in the gastrointestinal (GI) tract, the stomach and proximal small bowel are most often involved. Less commonly, KS may involve the gastrointestinal tract in the absence of cutaneous disease.

At one time, gastrointestinal KS was estimated to occur in up to 40% of patients with AIDS, but more recently its incidence has declined. In addition to occurring in patients with AIDS, visceral KS has also been confirmed in organ transplant recipients as well as in patients receiving immunosuppressants for other diseases. In these patients, KS occurs several months to a few years after initiation of immunosuppression. Although skin lesions may regress with a reduction in immunosuppression, patients with visceral involvement may succumb to their disease.

DIAGNOSIS

Early gastrointestinal KS consists of asymptomatic macular, plaquelike lesions with a red–blue appearance. Endoscopically, these lesions are described as being multiple, flat, violaceous lesions ranging in size from a few millimeters to 1–2 cm (see Fig. 1). The lesions may become progressively bulky and nodular, with central umbilication or ulceration. Tumors may also be sessile or polypoid. Clinically, nausea, abdominal pain, intussusception, obstruction, or perforation may occur. Gastrointestinal hemorrhage from KS lesions has also been described.

Endoscopic biopsies, using standard biopsy forceps, are low yield in the diagnosis of gastrointestinal KS due to the submucosal location of these tumors. In one study, biopsy specimens were positive for KS in only 7 of the 17 patients with macroscopic gastrointestinal lesions. Diagnosis of gastrointestinal KS is, therefore, largely based on gross endoscopic findings in the appropriate clinical setting.

Early gastrointestinal KS is not detectable on air-contrast barium radiography due to its macular nature. As a lesion progresses and develops central umbilication, it may be seen as a target, or bull's eye lesion. Small intestinal KS may be seen as focal nodular

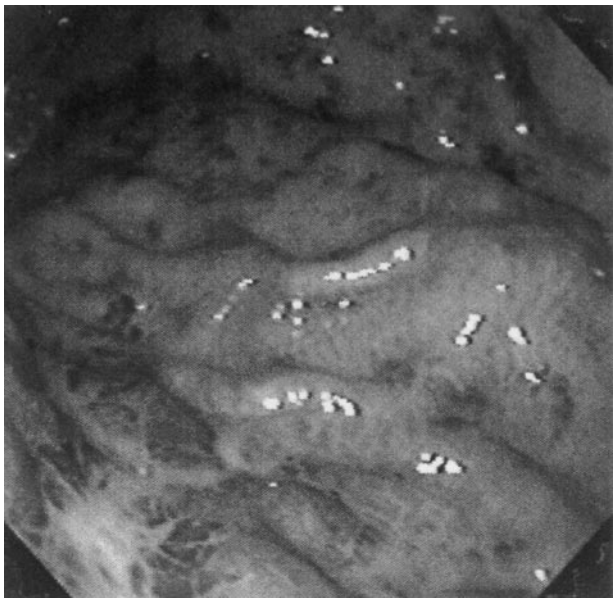


FIGURE 1 Gastric body. Endoscopic appearance of gastric Kaposi's sarcoma in an AIDS patient.

 Copyright 2004, Elsevier (USA). All rights reserved.

thickening. However, because the findings are nonspecific, barium studies may be of limited use in diagnosing gastrointestinal KS.

PATHOLOGY/PATHOGENESIS

Microscopic examination of KS lesions reveals a proliferation of spindle-shaped cells in the lamina propria and submucosa. The spindle cells form slits and vascular clefts in which there are extravasated erythrocytes. Hemorrhage and central necrosis are also often seen within the tumor. KS is thought to arise from a malignant proliferation of lymphatic endothelial cells, which may explain its multicentric nature. Evidence suggests that KS results from an infectious agent. In 1994, a specific type of human herpesvirus was found to be present in KS lesions. The genomic sequences of this herpesvirus, also called KS-associated herpesvirus (KSHV), are found in virtually all KS lesions. KSHV is also referred to as human herpesvirus-8 (HHV-8) and is part of the same gammaherpesvirus subfamily that includes Epstein–Barr virus.

PROGNOSIS

Diagnosis of gastrointestinal KS is important, because prognosis worsens with visceral involvement. The 2-year survival of AIDS patients with endoscopic evidence of KS was found to be 11% compared to an 88% 2-year survival in those without intestinal KS. This survival difference reinforces the observation that KS is yet another manifestation of the compromised immune state seen in AIDS.

Visceral KS lesions require treatment only if they are symptomatic. Treatment with immunomodulation, radiation, or chemotherapy can reduce tumor bulk and improve symptoms but does not change overall survival in AIDS patients.

SUMMARY

Gastrointestinal Kaposi's sarcoma is an uncommon neoplasm that should be considered in patients with AIDS-related or iatrogenic compromise of their immune system.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • Vascular Abnormalities • Vascular Abnormalities, Pediatric

Further Reading

Chang, Y., Cesarman, E., Pessin, M. S., *et al.* (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865–1869.

Friedman, S. L., Wright, T. L., and Altman, D. F. (1985). Gastrointestinal Kaposi's sarcoma in patients with acquired immunodeficiency syndrome. *Gastroenterology* **89**, 102–108.

Kadakia, S. C., Kadakia, A. S., and Westphal, K. W. (1992). Gastrointestinal Kaposi's sarcoma as the first manifestation of acquired immunodeficiency syndrome. *South. Med. J.* **85**(1), 37–39.

Parente, F., Cernuschi, M., Orlando, G., Rizzardini, G., Lazzarin, A., and Bianchi Porro, G. (1991). Kaposi's sarcoma and AIDS: Frequency of gastrointestinal involvement and its effect on survival. *Scand. J. Gastroenterol.* **26**, 1007–1012.

Kawasaki Syndrome

H. CODY MEISSNER
Tufts University School of Medicine

atypical or incomplete Kawasaki syndrome Variation of classical Kawasaki syndrome in which fewer than five criteria are present but coronary artery aneurysms still develop.

giant aneurysm Most severe complication of Kawasaki syndrome in which a coronary artery aneurysm develops and measures more than 8 mm in diameter.

Kawasaki syndrome Vasculitis; has replaced rheumatic fever as the most common cause of acquired heart disease in children and is diagnosed on the basis of fever plus five criteria.

superantigen Protein antigen that differs from a conventional antigen in several aspects, but one of the most important is its ability to elicit a massive cytokine response, typical of what occurs in Kawasaki syndrome.

Kawasaki syndrome is an acute, multisystem vasculitis that primarily affects infants and young children under 5 years of age. The illness was first described by Tomisaku Kawasaki in 1967 in Japanese in the *Journal of Allergology*. Although Kawasaki syndrome was originally described as a benign illness of early childhood, it is now recognized that this disease is a leading cause of acquired heart disease in children in the United States and Western Europe. Children with Kawasaki syndrome who do not receive treatment with intravenous immunoglobulin and aspirin within the first 10 days of onset of fever have a 20–25% risk of developing coronary artery abnormalities. Early recognition and prompt treatment reduce the prevalence of coronary artery abnormalities to less than 5%, placing emphasis on the need for early diagnosis and treatment.

INTRODUCTION

A diagnosis of Kawasaki syndrome (KS) is based on the presence of fever plus at least four of five criteria, just as a diagnosis of rheumatic fever is based on the Jones criteria. Although the etiology of Kawasaki syndrome is not fully defined, features of this syndrome are similar to those found in certain illnesses that are known to be caused by toxin- or superantigen-producing bacteria, including toxic shock syndrome and scarlet fever.

Kawasaki syndrome has been reported from countries throughout the world, although the highest rates of disease are found in Japan and children of Japanese ancestry living outside of Japan. In the United States the prevalence of Kawasaki syndrome is highest in Asians, intermediate in blacks, and lowest in Caucasians. The hospitalization rate for Kawasaki syndrome among American Indians and Alaskan natives appears to be lower than that among Caucasians. Disease occurs in males about 1.5 times more often than in females. The peak age at which disease occurs is 12–24 months. More than 80% of cases occur in patients before 5 years of age and most cases occur before 2 years. The onset of disease after 8 years is rare, representing less than 10% of cases reported to the Centers for Disease Control and Prevention (CDC). Recurrent disease develops in fewer than 2% of patients. It is not clear whether older patients are predisposed to more severe disease or whether older patients are more likely to have a delay in diagnosis and initiation of treatment later in the course of their illness, with correspondingly greater morbidity. Infants less than 12 months of age and particularly those less than 6 months appear to be at increased risk of developing coronary artery abnormalities.

CLINICAL DISEASE

The possibility of Kawasaki syndrome should be considered in any patient with a fever lasting 5 or more days without an alternative explanation and the presence of at least four of the five clinical criteria listed in Table I. Kawasaki syndrome typically begins with an acute onset, with daily fevers to 40°C or greater in a toxic-appearing child. Fever may last 2 weeks or longer without treatment. The mucocutaneous manifestations of Kawasaki syndrome are varied and not all patients will exhibit each feature. In the first days of acute febrile illness, approximately 90% of children with Kawasaki syndrome develop a polymorphous exanthem, which may demonstrate a variety of forms. The eruption

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Diagnostic Criteria for Kawasaki Syndrome[a]

Fever for ≥ 5 days without other explanation, and at least four of the following five criteria:
Nonexudative bulbar conjunctival injection
Oropharyngeal changes, including injected or fissured lips, injected pharynx, or strawberry tongue
Extremity changes, including erythema of the palms or soles, edema of the hands or feet, or periungal desquamation
Polymorphous rash
Acute nonsuppurative cervical lymphadenopathy

[a] A diagnosis of atypical Kawasaki syndrome can be made when more than four criteria are present when coronary artery aneurysms are present.

tends to be most prominent on the trunk and extremities. One early sign may be accentuation of a perineal rash.

In most patients a nonexudative, bilateral conjunctival injection begins shortly after the onset of fever and generally involves the bulbar conjunctiva to a greater extent than the palpebral conjunctiva. Conjunctival vessels become engorged and dilated. Purulent discharge is generally not present. Conjunctival injection is associated with anterior uveitis in about 80% of patients.

Oral mucosal findings occur in almost all typical cases. The lips become red, dry, and often cracked, producing small hemorrhagic fissures. There are no punctate ulcerations such as those seen in herpes gingivostomatitis and no erosions suggestive of Stevens–Johnson syndrome. The tongue is often "strawberry" in appearance with hypertrophied papillae and hyperemia, similar to that seen with streptococcal infections. A generalized erythema of the oropharynx is common although ulceration of the mucosal surface is not characteristic.

Changes in the extremities may be the most distinctive finding among the five criteria. Erythema and edema of the hands and feet with fusiform swelling of the fingers and toes are often observed. Swelling usually begins within a few days of the onset of illness. The hyperemic areas desquamate during the second or third week. The desquamation characteristically begins at the tips of the fingers and toes (the periungal region) and may extend to involve the palms and soles in a manner similar to that seen in scarlet fever. Lymphadenopathy is the least frequent finding, seen in 50–75% of patients. It is most often unilateral and the nodes are firm and nontender.

Other associated clinical features of Kawasaki syndrome are listed in Table II. Polyarticular arthralgia and arthritis may occur soon after onset of fever and involve the small joints. Pauciarticular arthritis involving the large weight-bearing joints (hips, knees, and ankles) may occur in the second or third week of illness. Urethritis associated with sterile pyuria is common. A mononuclear cell pleocytosis of the cerebrospinal fluid, with normal glucose and protein levels, occurs in approximately 1/3 of patients who undergo lumbar puncture. It should be remembered that a small number of patients will develop aseptic meningitis (fever, headache, vomiting, and nuchal rigidity) within 48 hours following intravenous immunoglobulin infusion.

Cardiac abnormalities are the major complication of Kawasaki syndrome and constitute the most serious complication of this disease. Myocarditis may develop during the first few days following onset of fever and is manifest by tachycardia out of proportion to fever elevation, conduction irregularities, a gallop rhythm, or electrocardiogram changes. Pericardial effusion may develop as a manifestation of carditis. Congestive heart failure may develop as a complication of myocardial dysfunction secondary to ischemia or infarct. Coronary artery abnormalities occur in 20–25% of untreated patients. Dilatation of the coronary arteries can be detected by echocardiography soon after the onset of fever. Aneurysms of the coronary arteries may be demonstrable by echocardiography as soon as a few days after onset of illness, but more typically occur between 1 and 4 weeks of illness. Rarely, patients experience aortic regurgitation or mitral regurgitation due to valvulitis, transient papillary muscle dysfunction, or myocardial infarction. Occasionally, patients develop aneurysm of the brachial, renal, or iliac arteries. This usually occurs in association with coronary artery abnormalities.

The likelihood of resolution of the aneurysm is determined by the initial size of the aneurysm, with smaller aneurysms having a greater likelihood of regression. Patients with giant aneurysms (diameter >8 mm) have the worst prognosis. Nakano *et al.* have reported that 71% of patients with giant aneurysms progress to stenosis or obstruction over an 11-month followup

TABLE II Associated Features of Kawasaki Syndrome

Cardiovascular abnormalities, including myocarditis, arterial aneurysms, pericarditis, aortic or mitral regurgitation, and ventricular arrhythmias
Arthralgia and arthritis
Hepatic dysfunction
Urethritis with sterile pyuria
Aseptic meningitis
Hydrops of the gallbladder
Diarrhea, vomiting, or abdominal pain
Peripheral gangrene
Uveitis
Sensorineural hearing loss

period. At a mean followup of 32 months, 30% of giant aneurysms develop obstruction. Nearly all late deaths from Kawasaki syndrome occur in patients with this complication.

Kawasaki syndrome is based on diagnostic criteria using clinical signs and symptoms that overlap with other illnesses. This can result in diagnostic dilemmas, particularly in atypical cases that do not completely fulfill the diagnostic criteria but are associated with the development of coronary artery abnormalities. Reports of atypical or incomplete Kawasaki syndrome have increased in recent years. In certain cases, the decision to treat with intravenous immunoglobulin (IG) and acetylsalicylic acid (ASA) can be a difficult diagnostic dilemma. The risk of coronary artery abnormalities increases in direct proportion to the interval of time between the onset of fever and intravenous administration of IG. Instances of atypical Kawasaki syndrome are most common in infants, the age group that is at greatest risk of coronary artery abnormalities. The decision to initiate therapy in children who do not satisfy the American Heart Association criteria can be supported by laboratory results showing acute-phase reactants (elevated white blood cell count, elevated sedimentation rate), an ultrasound showing a pericardial effusion, or a slit lamp exam showing anterior uveitis.

GASTROINTESTINAL ABNORMALITIES

Kawasaki syndrome is a multisystem inflammatory disorder and a number of organ systems may be involved, as shown in Table II. Early in the acute stage, gastrointestinal complaints include nausea, abdominal pain, and diarrhea. Mild hepatic dysfunction in association with modest elevation of serum transaminases and elevated bilirubin values may occur during the acute febrile phase of disease. Hypoalbuminemia is associated with more severe disease. Hydrops of the gallbladder is well described in patients with Kawasaki syndrome and typically presents with a right upper quadrant mass in the first 10–14 days of illness and may occur with or without obstructive jaundice. Gallbladder hydrops can be detected and monitored by ultrasonography. Cholecystectomy is not indicated. Involvement of the liver and gallbladder resolves soon after administration of intravenous IG and aspirin and is not associated with any long-term abnormality. Pancreatitis has also been reported in association with Kawasaki syndrome.

The laboratory features of Kawasaki syndrome include leukocytosis with an increased band count and an elevated platelet count by the second week of illness. During the acute early febrile stage, "acute-phase reactants" are generally elevated, including the erythrocyte sedimentation rate, C-reactive protein, α2-globulin, and α1-antitrypsin. An increase in fibrinogen and a prolongation of prothrombin time (PT) and partial thromboplastin time (PTT) values have been reported.

DIFFERENTIAL DIAGNOSIS

The clinical manifestations of Kawasaki syndrome are not specific and they occur in a number of other infectious and rheumatologic diseases. Other diseases that may resemble Kawasaki syndrome are shown in Table III. A firm diagnosis of KS is often difficult because the symptoms are not unique and because not all symptoms are seen in all patients. Because therapy within the first 10 days of onset of fever is clearly desirable in terms of reducing the risk for cardiac complications, rapid diagnosis and treatment are important. In many patients, differentiation of KS from other illnesses that resemble this syndrome may be difficult. Viral illnesses that may enter the differential diagnosis include measles, Epstein-Barr virus, adenovirus, and influenza virus infections. Nasal secretions should be cultured in a patient whose symptoms include evidence of an upper respiratory tract infection, especially to rule out adenovirus infection. Group A β-hemolytic streptococcus and *Staphylococcus aureus* infections may also mimic KS. Noninfectious diseases that frequently appear in the differential diagnosis include drug reactions and juvenile rheumatoid arthritis. Findings at physical examination that are not typical of KS include discrete intraoral ulcerations, exudative conjunctivitis, and generalized lymphadenopathy.

THERAPY

Management of patients with Kawasaki syndrome is directed at reducing inflammation in the myocardium and coronary artery wall during the acute phase. Once

TABLE III Differential Diagnosis of Kawasaki Syndrome

Measles
Scarlet fever
Drug reactions
Toxic shock syndrome
Staphylococcal scalded skin syndrome
Stevens–Johnson syndrome
Rocky Mountain spotted fever
Viral exanthems
Leptospirosis
Juvenile rheumatoid arthritis
Adenovirus infection
Epstein–Barr virus infection

TABLE IV Treatment of Kawasaki Syndrome

Acute phase
Intravenous immunoglobulin 2 g/kg over 10–12 hours
Aspirin, 80–100 mg/kg/day in four divided doses until afebrile
Convalescent phase in patients with uncomplicated Kawasaki syndrome
Aspirin, 3–5 mg/kg/day once daily for 6–8 weeks
For patients with coronary artery disease
Aspirin, 3–5 mg/kg/day once daily
Dipyridamole, 1 mg/kg/day in selected patients
Anticoagulant therapy as needed in patients with arterial thrombosis

the acute stage has passed, therapy is directed at prevention of coronary artery thrombosis. ASA in combination with high-dose intravenous IG forms the basis of current therapy (Table IV). ASA is used for both antiinflammatory and antithrombotic actions, although convincing data that ASA reduces coronary artery abnormalities are not available. Aspirin is administered at a dose of 80–100 mg/kg/day in four divided doses to achieve a serum salicylate level of 20–25 mg/dl during the acute phase of the illness. This is the only dose of ASA that has been carefully studied in the United States. Although gastrointestinal hemorrhage is a known complication of aspirin therapy, this adverse event is a rare complication of high-dose salicylate therapy in patients with Kawasaki syndrome. Efficacy from intravenous IG therapy has only been demonstrated when administered within the first 10 days of illness. Patients who present beyond the tenth day of fever should still be treated with intravenous IG and ASA, although supporting data are not available. It is not clear whether different preparations of intravenous IG have similar efficacy in the prevention of coronary artery abnormalities. In afebrile children, the ASA dose is reduced to 3–5 mg/kg/day to continue antithrombosis activity. ASA is discontinued if no coronary abnormalities have been detected by 6–8 weeks after onset of the illness. ASA therapy is continued indefinitely if coronary artery aneurysms develop.

At present, the treatment of choice for acute Kawasaki syndrome is a single intravenous dose of IG at 2 g/kg administered over 10–12 hours in combination with ASA. Using this regimen, the prevalence of coronary artery abnormalities falls to less than 5%. Peak adjusted serum globulin levels are lower among patients who subsequently develop coronary artery abnormalities and are inversely related to fever duration and laboratory indices of acute inflammation.

All children diagnosed with Kawasaki syndrome should have two-dimensional echocardiography at the time of diagnosis. Repeat studies are recommended at 4–6 weeks and again at 6–12 months. The major long-term morbidity in Kawasaki syndrome is related to cardiovascular complications. Approximately 50% of children with arterial aneurysms will show angiographic regression within 6 months to 2 years after onset of their disease.

MECHANISM OF ACTION OF INTRAVENOUS IMMUNOGLOBULIN

The mechanism by which high-dose intravenous IG works to reduce the vasculitis associated with Kawasaki syndrome is unknown. The observation that intravenous IG works rapidly in reducing the laboratory parameters of the acute-phase response associated with KS suggests a generalized antiinflammatory effect. In this regard, it has been reported that prior to intravenous IG therapy, peripheral blood mononuclear cells from patients with acute Kawasaki syndrome secrete high levels of interleukin-1 (IL-1), an endogenous pyrogen, and tissue vascular endothelial cells express IL-1-inducible endothelial activation antigens. IL-1 secretion remains elevated in intravenous IG-treated patients in whom coronary artery abnormalities develop. However, IL-1 secretion levels fall to normal in-patients who respond to intravenous IG therapy. These data support the notion that intravenous IG may work in Kawasaki syndrome by reducing cytokine-inducible endothelial activation.

Intravenous IG has been shown to contain high concentrations of neutralizing antibodies that inhibit the T cell response to staphylococcal superantigens. Using affinity absorption techniques, it has been shown that this T cell inhibitory effect is mediated by antitoxin-specific antibodies in intravenous IG. Thus the beneficial effect of intravenous IG may be due in part to antibodies that inhibit bacterial toxin-induced stimulation of the immune response.

The following hypothesis has been proposed to explain the pathogenesis of this illness (although this has not been uniformly accepted as the universal pathogenetic mechanism): an organism colonizes the mucous membranes of the gastrointestinal tract of a genetically susceptible host and produces a toxin that behaves as a superantigen. Toxin is absorbed through the inflamed mucosal surface and stimulates local or circulating mononuclear cells to produce proinflammatory cytokines, which in turn cause fever and the clinical picture of Kawasaki syndrome. In response to cytokine-induced stimulation, antigens are expressed on the surface of vascular endothelial cells, rendering them susceptible to attack by cytotoxic antibodies and activated T cells.

Neoantigens on endothelial cells render the vessels more thrombogenic.

Understanding the etiology of Kawasaki syndrome remains a major unresolved issue of pediatrics. It is important that the etiology of this illness be resolved so that a definitive test can be developed to identify children with typical Kawasaki syndrome as well as those who present with atypical disease and who do not satisfy the diagnostic criteria, but are still at risk for coronary artery disease. In addition, it is unlikely that a more specific form of therapy than intravenous immune globulin will be found without understanding the etiology.

See Also the Following Article

Vasculitis

Further Reading

Dajani, A. S., Taubert, K. A., Gerber, M. A., Shulman, S. T., Ferrieri, P., Freed, M., *et al.* (1993). Diagnosis and therapy of Kawasaki disease in children. *Circulation* **87**, 1776–1780.

Dajani, A. S., Taubert, K. A., Takahashi, M., Bierman, F. Z., Freed, M. D., Ferrieri, P., Gerber, M., Shulman, S. T., Karchmer, A. W., Wilson, W., Peter, G., Durack, D. T., and Rahimtoola, S. H. (1994). Guidelines for long-term management of patients with Kawasaki disease. *Circulation* **89**, 916–922.

Dale, R., Saleem, M. A., Daw, S., and Dillon, M. J. (2000). Treatment of severe complicated Kawasaki disease with oral prednisolone and aspirin. *J. Pediatr.* **137**, 723–726.

Durongpisitkul, K., Gururaj, V. J., Park, J. M., and Martin, C. F. (1995). Prevention of coronary artery aneurysms in Kawasaki disease: A meta-analysis on the efficacy of aspirin and immunoglobulin treatment. *Pediatrics* **96**, 1057–1061.

Kawasaki, T. (1967). Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children: Clinical observations of 50 cases. *Jpn. J. Allergol.* **16**, 178–222.

Leung, D. Y. M., Meissner, H. C., Fulton, D. R., Murray, D. L., Kotzin, B. L., and Schlievert, P. M. (1993). Toxic shock syndrome toxin-secreting *Staphylococcus aureus* in Kawasaki syndrome. *Lancet* **342**, 1385–1388.

Leung, D. Y. M., Meissner, H. C., Shulman, S. T., Mason, W. H., Gerber, M. A., Glode, M. P., Myones, B. L., Wheeler, J. G., Ruthazer, R., and Schlievert, P. M. (2002). Prevalence of superantigen-secreting bacteria in patients with Kawasaki disease. *J. Pediatr.* **140**, 742–746.

Mark, T., Witt, M. D., Minich, L., Bohnsack, J. F., and Young, P. C. (1999). Kawasaki Disease: More Patients are being diagnosed who do not meet America Heart Association criteria. *Pediatrics* 104, e10.

Meissner, H. C., and Leung, D. Y. M. (2000). Superantigen, conventional antigens and the etiology of Kawasaki syndrome. *Pediatr. Infect. Dis. J.* **19**, 91–94.

Newburger, J. W., Takahashi, M., Beiser, A. S., Burns, J. C., Bastian, J., Chung, K. J., *et al.* (1991). A single intravenous infusion of gamma globulin as compared with four infusions in the treatment of acute Kawasaki syndrome. *N. Engl. J. Med.* **324**, 1633–1639.

Newburger, J. W., Takahashi, M., Burns, J. C., Beiser, A. S., Chung, K. J., Duffy, C. E., *et al.* (1986). The treatment of Kawasaki syndrome with intravenous gammaglobulin. *N. Engl. J. Med.* **315**, 341–347.

Rosenfield, E. A., Corydon, K. E., and Shulman, S. T. (1995). Kawasaki disease in infants less than one year of age. *J. Pediatr.* **126**, 524–529.

Shinohara, M., Sone, K., Tomomasa, T., and Morikawa, A. (1999). Corticosteroids in the treatment of the acute phase of Kawasaki disease. *J. Pediatr.* **135**, 465–469.

Terai, M., and Shulman, S. T. (1997). Prevalence of coronary artery abnormalities in Kawasaki disease is highly dependent on gamma globulin dose but independent of salicylate dose. *J. Pediatr.* **131**, 888–893.

Kernicterus

RENE DAVILA AND CLAUDIO R. TOMBAZZI
University of Tennessee Health Science Center

bilirubin A bile pigment formed during the catabolism of heme-containing compounds, primarily hemoglobin.

Crigler-Najjar Syndrome of severe unconjugated hyperbilirubinemia that presents as congenital familial nonhemolytic jaundice and is related to a lack of ability to conjugate bilirubin, as well as a blockage in bilirubin excretion from the hepatocyte.

Kernicterus (German: *kern*, nucleus; *icterus*, yellow) is the condition associated with severe unconjugated hyperbilirubinemia that causes yellow staining of certain regions of the brain and presents with neurological manifestations. Bilirubin-related brain damage occurs most commonly in the neonatal period, but it has also been reported in adults with Crigler-Najjar syndrome.

ETIOLOGY

All conditions related to increases in serum unconjugated bilirubin levels are potentially capable of inducing kernicterus. Some of the most common causes are fetal–maternal blood group incompatibilities, accumulation of extravascular blood in body tissues due to hemorrhage, polycythemia, and red blood cell abnormalities.

PATHOPHYSIOLOGY

Kernicterus develops when the rate of bilirubin deposition becomes overwhelming as a result of high serum bilirubin concentration, low albumin-binding capacity, or low serum pH. Low serum albumin levels or the use of drugs that displace bilirubin from albumin can increase the risk for kernicterus. Kernicterus is likely to occur when serum levels of unconjugated bilirubin are greater than 30 mg/dl and is unlikely to occur when levels are lower than 20 mg/dl. The areas of the brain damaged by bilirubin toxicity are the basal ganglia, hippocampus, cerebellum, and nuclei of the floor of the fourth ventricle. Early evidence of bilirubin toxicity in the brainstem has been detected by auditory evoked potentials.

CLINICAL FINDINGS

Acute clinical manifestations include sluggish Moro's reflex, opisthotonos, hypotonia, vomiting, high-pitched cry, hyperpyrexia, seizures, paresis of gaze, oculogyric crisis, and death. Long-term clinical manifestations include hearing loss, spasticity, and choreoathetosis. Mild forms can present as cognitive dysfunction and learning disabilities.

MANAGEMENT

Preventive measures such as avoidance of administration of pharmacologic agents that bind to albumin and displace bilirubin are reasonable. Therapeutic options include phototherapy, exchange transfusion, pharmacologic interruption of the enterohepatic circulation, enzyme induction, and alteration of breast-feeding.

Phototherapy consists of irradiation of the jaundiced child with light. Photon energy changes the structure of bilirubin, allowing its excretion without hepatic glucuronidation. Special blue light appears to be better than white or green light. In general, phototherapy is used to prevent the serum bilirubin concentration from rising to a level that necessitates exchange transfusion. Intensive phototherapy should induce a decline in total serum bilirubin of 1 to 2 mg/dl within 4 to 6 h. The bilirubin level should continue to decline until it remains below the level that necessitates exchange transfusion (see below). Failure of phototherapy is evident when the newborn continues to experience dangerously high levels of bilirubin.

Exchange transfusion is the best method to acutely lower the serum bilirubin concentration. Mortality associated with this procedure is 0.6%. The indications for exchange transfusion in otherwise healthy term newborns are shown in Table I.

Interruption of the enterohepatic circulation is achieved by the use of agents that prevent the reabsorption of bilirubin from the intestines, such as cholestyramine, activated charcoal, and calcium phosphate.

Conjugation of bilirubin takes place because of the action of the enzyme bilirubin uridine diphosphate

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Indication for Exchange Transfusion to Prevent Kernicterus

Bilirubin levels >20 mg/dl in newborns between the ages of 25 and 48 h who fail to respond to phototherapy
Bilirubin levels >25 mg/dl in newborns between the ages of 25 and 48 h
Bilirubin levels >25 mg/dl in newborns age 49 h or older that fail to respond to phototherapy
Bilirubin levels >30 mg/dl in newborns age 49 h or older
Bilirubin levels >15 mg/dl for more than 48 h
Ratio of total serum bilirubin level to total serum protein level >3.7

glucuronosyltransferase (BUGT). The activity of this enzymatic system is reduced in neonates. The use of phenobarbital in the neonate has proven to increase the activity of BUGT, thus inducing an improvement in the level of unconjugated bilirubin.

Increases in the frequency of breast-feeding to an average interval between feeds of 2 h and no feeding supplements may increase intestinal peristalsis with a subsequent decrease in the absorption of bilirubin.

See Also the Following Articles

Bilirubin and Jaundice • Neonatal Hyperbilirubinemia

Further Reading

Brodersen, R., and Stern, L. (1990). Deposition of bilirubin acid in the central nervous system—A hypothesis for the development of kernicterus. *Acta Paedriatr. Scand.* **79**, 12–19.

Gourley G. R. (2001). Neonatal jaundice and disorders of bilirubin metabolism. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), 2nd Ed., pp. 275–314. Lippincott Williams & Wilkins, Philadelphia, PA.

Newman, T. B., and Maisels, M. J. (1990). Does hyperbilirubinemia damage the brain of healthy full-term infants? *Clin. Perinatol.* **17**, 331–358.

Laparoscopy

EDWARD G. SOLTESZ AND DAVID C. BROOKS
Brigham and Women's Hospital and Harvard University

adnexal disease Diseases of the uterine appendages (ovaries, fallopian tubes, and uterine ligaments).

pneumoperitoneum Insufflation of carbon dioxide gas into the abdominal cavity performed in all laparoscopic procedures to increase the work space available inside the abdomen.

Laparoscopy refers to the minimally invasive technique of assessing the abdominal cavity through small incisions for the purpose of diagnosis and treatment of various intra-abdominal conditions. Although simple laparoscopy has been in use since the turn of the 20th century as a tool for intra-abdominal diagnosis, its popularity has only recently grown as a result of the introduction of many therapeutic laparoscopic procedures. The success of laparoscopic cholecystectomy, confirmed in multiple randomized trials over the past 10 years, has reawakened surgeons' interest in laparoscopy and fueled the development of this new field of modern general surgery. Although the laparoscopic approach to cholecystectomy and fundoplication has proven superior to the open method, whether these same advantages exist for other procedures, such as hernia repair, colectomy, and appendectomy, awaits confirmation of randomized trials. This article will summarize the current status of diagnostic and therapeutic laparoscopy and provide an overview of "minimally invasive" gastrointestinal procedures for the gastroenterologist.

HISTORY

Modern laparoscopy traces its origins to the Italian physician Bozzini, who in the 19th century viewed the peritoneal cavity using reflected candlelight. In 1901, Kelling performed the first reported "minimally invasive" examination of the abdomen after creating pneumoperitoneum and inserting a cystoscope through a trocar in a dog. Soon thereafter, Jacobeus used laparoscopy to diagnose syphilis, tuberculosis, cancer, and cirrhosis. The technique of diagnostic laparoscopy was advanced with the subsequent introduction of the Veress spring-loaded needle in 1938 as well as CO_2 pneumoperitoneum, which improved visualization and increased the intra-abdominal working environment. Unfortunately, the early years of laparoscopy required the operator to peer into the small lens of the scope, thus limiting visualization and concomitant bimanual instrumentation. The advent of computerized video monitors in the 1980s revolutionized the field and liberated the surgeon from the role of "scope holder." The surgeon could now operate with both hands, using two separate instruments, while another operator held the camera. Until a general surgeon named Semm performed the first laparoscopic appendectomy in 1983, laparoscopy was used almost exclusively by gynecologists. Shortly thereafter, the French surgeon Mouret performed the first recorded laparoscopic cholecystectomy in 1987 followed by Reddick and Olsen and also McKernan and Saye in the United States a year later. The ensuing decade saw an explosion of interest in laparoscopic techniques, partly fueled by the overwhelming zeal of the lay public for more minimally invasive surgical procedures as well as the growing interest among general surgeons in these new, intriguing techniques.

TECHNIQUE OF LAPAROSCOPY

Laparoscopic surgery is defined by its equipment that can be grouped into three categories: (1) image production, (2) intra-abdominal access, and (3) instruments. The basic set-up for either diagnostic or therapeutic laparoscopy includes a camera attached to a video monitor for viewing, trocars that traverse the abdominal wall and provide airtight access to the peritoneal cavity, and instruments used to manipulate abdominal contents.

Every laparoscopic procedure begins with safe access to the peritoneal cavity. The open technique of intra-peritoneal access, popularized by Hasson, involves making a small, cutaneous incision, usually periumbilical, which is then extended under direct visualization through the midline abdominal fascia into the peritoneal cavity in a technique similar to diagnostic peritoneal lavage (DPL). After the absence of abdominal adhesions in the area of access is confirmed, the Hasson trocar with a blunt-tipped obturator is advanced directly into the peritoneal cavity and secured in place. The closed technique for access is

 Copyright 2004, Elsevier (USA). All rights reserved.

slightly different and calls for the blunt-tipped Verres needle. Here, the surgeon advances the Verres needle through a small cutaneous periumbilical incision into the pelvis at a 45° angle while maintaining upward traction on the abdominal wall with clips. Two or three "clicks" of the obturator sheath may be heard as the needle passes through the fascia and peritoneum into the peritoneal cavity. Absence of blood, urine, or stool on aspiration and rapid flow of saline through the needle by gravity corroborate correct needle location. Although the open technique is claimed to be safer, both methods can cause serious complications. The open technique is advantageous in patients with a "hostile" abdomen, including those who have undergone previous laparotomies, as well as in pregnancy, where the exact position of the tip of the Verres needle is in doubt.

After safe access to the peritoneal cavity is achieved, a working space is created either by insufflating gas into the peritoneum or by setting up devices that lift the abdominal wall. Most centers use CO_2 as the agent of choice to create a pneumoperitoneum (PNP) since it is noncombustible (allowing the use of electrocautery) and highly soluble in blood (preventing air embolism). However, the absorbed CO_2 is converted into H_2CO_3 and can lower the plasma pH as well as irritate the peritoneal surface. Thus, other gases with varying properties have been used to create PNP. Nitrous oxide (NO) causes less peritoneal irritation than CO_2, but is combustible, obviating its use in any case requiring electrocautery. Helium (He) is inert and noncombustible, yet poorly soluble in plasma, thus increasing the risk of air embolism. Mechanical lifters have been compared to CO_2 and He PNP in animal studies and have been found to create less hemodynamic embarrassment than either technique of PNP; however, other animal studies have confirmed that the effects of open or laparoscopic surgery with gasless lifters, He, or CO_2 PNP on hemodynamic parameters are minimal and no method is superior to another. Typically, the abdominal cavity is insufflated to an intra-abdominal pressure (IAP) of 12–15 mm Hg. The video laparoscope, connected to an external monitor, is introduced through the previously placed trocar and the abdominal cavity is inspected. The standard 0° Hopkins rod-lens laparoscope ranges in size between 5 and 10 mm in diameter with options for oblique viewing scopes (30° and 45°).

Various specially designed laparoscopic instruments are available for diagnostic as well as therapeutic maneuvers. Basic equipment includes graspers, dissectors, scissors, and a needle holder, along with a clip applier, stapling device, and suction/irrigator. More advanced procedures require the use of specialized equipment. Tissue coagulation can be accomplished with simple electrocoagulation as well as with more advanced tools such as the harmonic scalpel, a tool that tamponades a blood vessel between the jaws of the instrument and seals the vessel with denatured protein, achieving precise cutting with minimal lateral thermal tissue damage. Traditional laparoscopic instruments are 5 to 12 mm in diameter, but a new generation of fine-caliber instruments with outer diameters ranging from 1.7 to 2.5 mm have been developed. These second-generation instruments produce minimal, if any, postoperative pain, obviate the need for trocar-site closure, prevent trocar-site hernia, and achieve excellent cosmetic results, yet are much more fragile to work with. Reports have confirmed the feasibility of several operations (including cholecystectomy, appendectomy, and fundoplication) with fine-caliber instrumentation using traditional sized laparoscopes. Objective clinical data must demonstrate a clinical benefit before widespread use of fine-caliber instrumentation can be adopted.

PHYSIOLOGY OF LAPAROSCOPY

A wide variety of physiologic changes occur with the administration of PNP during laparoscopic surgery. As with any surgical procedure, there is a certain upregulation of the acute-phase and postoperative inflammatory response. This response occurs in proportion to the generalized physiologic insult. Laparoscopic approaches to traditional procedures involve limited abdominal access and less systemic stress and subsequently produce a blunted acute-phase response compared to their open counterparts. At the same time, however, the additional insult of PNP creates a whole other set of altered physiologic parameters of which the surgeon and anesthesiologist must be cognizant for each individual patient.

During laparoscopy with CO_2 PNP, there is a measurable increase in central venous pressure and pulmonary capillary wedge pressure, yet a paradoxical decrease in cardiac preload and a resultant slight decrease in stroke volume. Furthermore, there is an overall increase in mean arterial pressure and systemic vascular resistance, contributing to an increase in afterload, likely a result of both increased IAP and a systemic catecholamine surge. Most studies have shown no appreciable decline in cardiac output (CO), in healthy individuals with IAP between 10 and 15 mm Hg, the typical insufflation pressures used in most diagnostic and therapeutic laparoscopies. In canine studies, though, there is a noticeable decrease in CO at an IAP of 25 mm Hg or greater. While most would

agree that these subtle changes in CO during PNP are unlikely to adversely affect cardiac function in healthy subjects, patients with cardiopulmonary disease may face risks. These concerns were evaluated in multiple prospective trials that concluded that the effects of PNP on CO are clinically insignificant in patients with compromised cardiopulmonary function and that laparoscopy is safe in this population if appropriately managed intraoperatively. In addition, this specific population benefits the most from the diminished postoperative pulmonary embarrassment and faster return to function with laparoscopic approaches. In addition to a decline in CO with PNP, there is a measurable decrease in mesenteric blood flow, which is clinically most evident in a decline in renal perfusion. Again, in healthy subjects, this is well tolerated; yet there are no data documenting the response of patients with preoperatively impaired renal function to this transient, albeit real decrease in renal perfusion.

Intraoperative pulmonary function is affected by laparoscopy with CO_2 PNP. The increase in IAP forces the diaphragm cephalad, compromising ventilation and increasing the work of breathing, resulting in alveoli collapse. In plasma, the increased CO_2 absorbed intraperitoneally causes a metabolic acidosis that typically leads to an increased respiratory rate. Ventilated patients under anesthesia require higher minute ventilations to maintain the same pH. Postoperatively, many studies have clearly documented that there is less pulmonary dysfunction and quicker return of spirometry values in patients undergoing laparoscopic procedures than in those undergoing open procedures.

A decrease in systemic immune response following surgery or other major stresses is well documented. The acute-phase response generated is proportional to the inciting stress and creates a state of postoperative immunosuppression. Systemic cell-mediated immunity is better preserved following laparoscopic procedures, owing to a blunted acute-phase response, yet intraperitoneal cell-mediated immunity appears to be impaired by PNP. Macrophages incubated in CO_2 produce significantly less tumor necrosis factor and interleukin-1 in response to stimulus than those incubated in He or air, thus implicating CO_2 as a mechanism behind the impaired intraperitoneal immunity. Though this may explain initial reports of increased trocar site recurrences following laparoscopic colon resections, more recent studies have found no recurrences at a 5-year follow-up. Compared to open procedures, laparoscopy produces less of an inflammatory response and thus fewer intra-abdominal adhesions.

Intestinal function is affected by surgical procedures, and the rapidity of return of bowel function as evidenced by flatus and bowel movement is directly correlated with the degree of abdominal stress. Bowel function returns faster after laparoscopic cholecystectomy and colon resection than after similar open procedures. Unfortunately, studies of myoelectric activity do not completely corroborate these clinical findings.

As with all surgical procedures, laparoscopy produces a transient state of hypercoaguability. Increased IAP with PNP and certain positions required during laparoscopic procedures produce a decrease in antegrade femoral venous flow, further increasing the risk of deep venous thrombosis (DVT) and pulmonary embolism (PE). Since these factors were recognized early and aggressive DVT prophylaxis has been the rule at most all centers, there have been relatively few reported cases of PEs.

ADVANTAGES OF LAPAROSCOPY

There are multiple advantages of laparoscopy over traditional open procedures. The lessened physiologic stress encountered with laparoscopic approaches translates into easily identifiable clinical parameters that the lay press has promulgated. The most frequently reported benefits include decreased postoperative pain, improved cosmetic result, and a shorter hospital stay. Detailed prospective randomized studies are currently under way to discover the advantages of laparoscopy compared to traditional open surgery with respect to individual procedures. Currently, only laparoscopic cholecystectomy and fundoplication have become the "gold standards" of their respective procedures.

COMPLICATIONS OF LAPAROSCOPY

Serious complications following laparoscopy are rare, occurring in less than 1% of all cases. However, there are specific complications of laparoscopic access that must be considered in addition to those that occur during traditional open surgery. The initial intra-abdominal access and trocar placement, the creation of PNP, and laparoscopic instrumentation are all unique to laparoscopy and can result in complications. The technique of initial intra-abdominal access, by either Hasson or Veress needle technique, has been associated with injury to internal organs, of which the small bowel is the most commonly injured (52%), followed by colon, duodenum, and stomach. Though damage to the small bowel is easily and safely repaired by externalization and direct visual repair, injury to other internal structures can result in more devastating consequences. During initial intra-abdominal access, major vascular

structures can be violated, resulting in sudden hemoperitoneum or more delayed retroperitoneal hematoma formation, both of which must be recognized and addressed with immediate conversion to open laparotomy. The overall morbidity rate for trocar complications is between 0.2 and 0.5%, with mortality rates of 0.0033 to 0.1%. In addition to immediate complications following trocar placement, delayed complications identical to any abdominal incision can occur, namely, bleeding, wound infection, or hernia formation. Bleeding from a trocar site is easily managed with electrocautery or direct suture ligation. Wound infection can also occur and is more common at the umbilical port, especially after laparoscopic cholecystectomy and removal of the infected gallbladder through that port. Finally, the creation of PNP can cause complications, the most devastating of which is an insufflation gas embolization, which can cause sudden cardiovascular collapse by obstructing the pulmonary outflow tract. Overall, though, these complications are exceedingly rare, and with appropriately trained surgeons, laparoscopy is very safe.

DIAGNOSTIC LAPAROSCOPY

Laparoscopy has recently become a very useful modality in the evaluation of an acute or chronic abdominal process, obviating the need for a laparotomy and providing addition data to traditional imaging. Laparoscopy is being increasingly applied in the diagnosis and workup of acute and chronic abdominal pain and masses, triage of trauma, and staging of malignancy. Compared to traditional noninvasive imaging techniques, diagnostic laparoscopy can more accurately characterize an intra-abdominal process and detect lesions less than 1 mm, which is below the resolution of magnetic resonance imaging or computed tomography (CT) scans. Furthermore, compared to exploratory laparotomy, diagnostic laparoscopy results in a smaller incision, less postoperative pain, and faster return of bowel function. Clearly, diagnostic laparoscopy has a place in the management of both acute and chronic abdominal processes.

The evaluation of the acute abdomen includes an extensive list of differential diagnoses that must be quickly and accurately considered. For example, the premenopausal female who presents with lower abdominal pain may have appendicitis, pelvis inflammatory disease (PID), or a ruptured ectopic or ovarian cyst. Laparoscopy has been shown to provide both a diagnosis and therapeutic options in such situations. There are several settings where diagnostic laparoscopy is favored over traditional exploratory laparotomy:

1. in patients who have a suspected intra-abdominal process that could be easily managed laparoscopically; such patients may be those in whom appendicitis or adnexal pathology is suspected;
2. in patients in whom the suspected diagnosis would not typically be managed operatively, especially those with suspected PID, pancreatitis, or low-flow mesenteric ischemia; in these situations, a definitive diagnosis is imperative and usually cannot be accurately provided by conventional imaging modalities;
3. in trauma patients in whom the inclusion or exclusion of major intra-abdominal pathology is critical to the management of other life-threatening injuries;
4. in critically ill ICU patients in whom bedside laparoscopy may avoid the need for hazardous transport to the operating room.

Similarly, there are situations where diagnostic laparoscopy should not be performed:

1. in trauma patients who are suspected of having a myocardial injury, significant hypovolemia, or closed head injury; and
2. in ICU patients who are suspected of having a highly correctable lesion that would necessitate an operative procedure (whether open or laparoscopic).

Suspected Acute Appendicitis or Gynecologic Abdominal Disease

Diagnostic laparoscopy has become an accepted technique for the assessment and management of lower abdominal pain. Today, appendicitis is often detected through a combination of CT scan, history, and physical examination. However, there are many cases where the radiographic findings are inconclusive, thus necessitating abdominal exploration. Traditionally, appendicitis was solely a clinical diagnosis and a negative appendectomy rate of 20% was accepted since the morbidity and mortality associated with a perforated appendix was so much higher. Several studies have documented a decline in the negative appendectomy rate with the use of diagnostic laparoscopy for cases of acute lower abdominal pain. Adnexal disease in young women can mimic acute appendicitis and it is in this population that diagnostic laparoscopy is especially useful, decreasing "unnecessary" laparotomies by one-third. Ruptured ovarian cysts, ovarian torsion, PID, and tubal pregnancy can all present similarly to appendicitis and can be conclusively diagnosed at laparoscopy. Though the treatment of PID and ruptured ovarian cysts is usually supportive, management of ovarian torsion and tubal

pregnancy can easily be accomplished laparoscopically. In fact, laparoscopic treatment of tubal pregnancy is as safe as open laparotomy and results in a decreased length of stay, lower cost, and earlier return to full activity. Clearly, laparoscopic abdominal exploration is an excellent method of conclusively determining the etiology of lower abdominal pain and in most instances can be used to treat the offending condition with no added morbidity or mortality.

Suspected Mesenteric Ischemia and Abdominal Sepsis

Abdominal sepsis of unknown cause has been successfully diagnosed by a systematic method of laparoscopic abdominal inspection. In a recent series, laparoscopic exploration has discovered multiple causes of abdominal sepsis, including perforated viscus (colon, appendix), gangrenous cholecystitis, ischemic bowel disease, and closed-loop small bowel obstruction, most of which can be successfully treated laparoscopically. If evaluation reveals no abdominal source, the patient is spared the physiological insult of a traditional exploratory laparotomy. Mesenteric ischemia can often be a difficult diagnosis to confirm, especially in the elderly patient with abdominal pain, multiple medical problems, and nonspecific laboratory findings. Laparoscopy can provide a means of examining the adventitial side of the bowel, whereas concurrent colonoscopy can evaluate the mucosa for evidence of intestinal ischemia. Diagnosis can be aided with the laparoscopic use of adjunctive tools for assessing blood flow, including Doppler ultrasound, laser Doppler, pulse oximetry, and spectrophotometry. When mesenteric ischemia is confirmed and a bowel resection is performed, laparoscopy can be used to provide a "second look" 24–28 h later, sparing a reoperative laparotomy.

Trauma

The use of laparoscopy in the evaluation of trauma remains controversial and review of the literature reveals multiple varying opinions as to appropriate candidates for diagnostic laparoscopy. Most importantly, trauma patients must be hemodynamically stable and not have significant hypovolemia, closed head injury, or a definitive diagnosis that demands immediate laparotomy. Previously, all penetrating wounds of the abdomen underwent exploration to exclude fascial and/or peritoneal penetration, whereas significant blunt abdominal injury was evaluated with CT scan or DPL. Diagnostic laparoscopy has been used in the acute evaluation of both blunt and penetrating abdominal trauma; however, certain limitations exist. Laparoscopy frequently misses injuries to the flank and epigastric regions as well as the retroperitoneum; for injuries to hollow viscera, it carries 100% specificity but only 18% sensitivity. Many authors would suggest that the patients most likely to benefit from laparoscopy are those with stab wounds to the flank or back or tangential gunshot wounds to the abdomen. These patients are typically stable and need abdominal or wound exploration to confirm that the peritoneum has not been violated. Laparoscopy could easily accomplish this task in this limited trauma subgroup. Thus, while the role of laparoscopy in trauma patients awaits the results of multiple randomized trials, diagnostic laparoscopy offers an effective means to timely diagnosis with little co-morbidity in certain carefully selected patients.

Small Bowel Obstruction

The use of laparoscopy in the management of small bowel obstruction (SBO) is novel, but has recently been championed by Metzger and colleagues. In a series of 40 patients with symptoms of complete or partial SBO, laparoscopy was successful in identifying and treating the cause (usually adhesions) in a majority of patients. Furthermore, this method led to a decreased length of stay by avoiding exploratory laparotomy. Unfortunately, many patients with SBO already have significant abdominal distension, which precludes a thorough diagnostic or therapeutic laparoscopy.

Chronic Abdominal Pain

Laparoscopy is being used with increasing frequency in the evaluation of patients with chronic abdominal complaints. Compared with exploratory laparotomy, laparoscopy allows a thorough abdominal evaluation and results in fewer postoperative adhesions. Recent studies have suggested that laparoscopy used in patients with chronic abdominal complaints usually uncovers adhesions, hernias, or endometriosis, all of which can be easily treated laparoscopically. Interestingly, it has also been noted that even in the absence of definitive findings, many patients report an improvement in their symptoms after diagnostic laparoscopy.

Staging of Malignancy

Diagnostic laparoscopy has found its place in conjunction with conventional imaging techniques in determining the nature of intra-abdominal malignancies. Minimally invasive methods of viewing intra-abdominal contents and obtaining visually guided biopsies can help differentiate benign from malignant processes as well as assess for metastatic spread within the abdomen and pelvis. With the aid of laparoscopic ultrasound,

diagnostic laparoscopy can determine the resectibility of a lesion before proceeding to formal laparotomy and has been used in the management of benign and malignant hepatic tumors, pancreatic cancer, gastric cancer, and distal esophageal cancer. Laparoscopy can also be used to formally stage Hodgkin's lymphoma. Furthermore, if a lesion is found to be unresectable, laparoscopy can be used to perform palliative procedures such as loop colostomies, gastrojejunostomies, cholecystojejunostomies, and feeding tube placement.

THERAPEUTIC LAPAROSCOPY

The scope of laparoscopic surgery is constantly expanding and with it the need for formal studies comparing minimally invasive methods with traditional open techniques. After the introduction of a new procedure, refinements in technique occur with an eye toward increasing safety and efficacy. Eventually, investigators focus on more wide-ranging measures of patient outcome in an attempt to objectify the overall benefits and disadvantages of a particular laparoscopic procedure in comparison to the current standard of care. This section will review the current status of laparoscopy surgery.

Cholecystectomy

Laparoscopic cholecystectomy has replaced the conventional open method as the gold standard for the treatment of symptomatic cholelithiasis, biliary dyskinesia, acalculous cholecystitis, and select patients with asymptomatic gallstones after several well-designed multicenter trials clearly demonstrated that laparoscopic cholecystectomy results in shorter length of hospital stay, less postoperative pain, less postoperative pulmonary embarrassment, quicker return of bowel function, and overall improved patient satisfaction. Following Soper's retrospective and Barkum's randomized prospective studies in 1992 demonstrating the safety and efficacy of laparoscopic over open cholecystectomy, the National Institutes of Health Consensus Conference in 1993 recommended that laparoscopic cholecystectomy replace open cholecystectomy as the preferred method of management of patients with symptomatic cholelithiasis. MacFayden *et al*. showed in a multicenter retrospective study of 114,005 patients that mortality from laparoscopic cholecystectomy was 0.06% with a conversion to open procedure rate of 2.1% and a common duct injury rate of 0.50%. Though the original operation was designed for young, thin patients without previous abdominal procedures, now even elderly and obese patients, as well as those having had past abdominal surgery, have been shown to benefit from laparoscopic cholecystectomy.

Appendectomy

Although multiple studies have demonstrated the feasibility and safety of laparoscopic appendectomy, most surgeons find it difficult to improve on the success of open appendectomy, a procedure that enjoys a short length of stay, minimal pain, and a small incision, especially in young, thin male patients. The laparoscopic approach allows for a comparable, or shorter, length of stay and similar complication rates when compared to open appendectomy. In a meta-analysis of 17 randomized, controlled trials, Chung *et al*. found that laparoscopic appendectomy resulted in less postoperative pain, faster recovery, and lower infection rate than open appendectomy. However, most surgeons reserve the laparoscopic approach for a select subgroup of patients. Obese patients or those who have a retrocecally located appendix typically require a larger incision for adequate mobilization of the appendix. These patients may very well benefit from a laparoscopic approach as the larger incision with the open approach increases postoperative pain and the risk of wound infection. Finally, patients with lower abdominal pain and an unclear diagnosis of appendicitis are appropriate candidates for laparoscopy, as a thorough evaluation of the abdomen and pelvis to exclude other diagnoses is possible with the laparoscope but not through a limited lower abdominal incision. Future outcomes research will need to determine the clear benefits of laparoscopic appendectomy and define its position in the routine treatment of uncomplicated appendicitis.

Splenectomy

Laparoscopic splenectomy has become a recognized modality for the treatment of idiopathic thrombocytopenic purpura (ITP) in patients who are surgical candidates. Studies have shown that laparoscopic splenectomy is safe and feasible, resulting in shorter hospital stays with costs and complication rates comparable to its open counterpart. Patients typically leave the hospital 2–4 days earlier and recover much quicker if they have a laparoscopic splenectomy. Laparoscopy does not allow as thorough a search for accessory spleens as does open splenectomy, but the long-term complications of missing accessory spleens have not been realized. Clearly, in the hands of skilled laparoscopists, laparoscopic splenectomy has become the gold standard for treatment of ITP.

Inguinal Hernia Repair

Traditional inguinal hernia repair has evolved significantly over the past two decades. The original

time-honored primary repairs (e.g., Bassini, McVay) have recently been supplanted by tension-free repairs (e.g., Lichtenstein) that involve the placement of a prosthetic mesh to either reinforce existing tissue or plug a defect. The open Lichtenstein repair can be accomplished under local anesthesia, with a relatively minimal groin incision, short length of stay, and minimal level of postoperative pain, and it has enjoyed excellent short-term results in various randomized trials. Currently, there are two approaches to laparoscopic inguinal hernia repair, the transabdominal preperitoneal approach (TAPP) and the totally extraperitoneal approach (TEPP). In each of these repairs, a mesh is placed over the myopectineal defect and held in place by the pressure created from the intra-abdominal contents. TAPP requires intraperitoneal access and PNP, and thus general anesthesia, whereas TEPP can be performed with regional anesthesia since it entails access to the properitoneal space and maintenance of a working environment with CO_2 insufflation, avoiding any intraperitoneal instrumentation. In randomized controlled trials comparing open Lichtenstein with laparoscopic hernia repairs, there was no significant difference in lengths of hospital stay and morbidity, but patients undergoing laparoscopic repair recovered quicker and had less postoperative pain and narcotic requirement. Laparoscopic repairs, however, take longer and have higher in-hospital costs, thus raising questions as to the actual benefits of laparoscopic hernia repair. Most would agree that laparoscopic repair is clearly superior in cases of bilateral inguinal hernias (obviating bilateral groin incisions) and recurrent repairs (thus avoiding the scar tissue and distorted tissue planes in the previously used anterior open approach). The largest and most complete randomized controlled trial thus far was undertaken by Liem *et al.* in 1997. They treated 487 patients with the TEPP technique and 507 patients with a traditional open technique (suture or mesh repair) and found that those patients who had the laparoscopic repair had fewer hernia recurrences than those who had the open repair (3% versus 6%). As with laparoscopic appendectomy, randomized controlled trials quantifying outcomes measures will be needed to conclusively demonstrate benefits of laparoscopic inguinal hernia repair in the typical patient population.

Fundoplication

Gastroesophageal reflux disease (GERD) is a very common disorder in the United States, with one survey reporting approximately 20% of the population with reflux symptoms occurring on a monthly basis, 14% weekly, and 7% daily. With its introduction in 1991 by Dallemagne of Belgium, laparoscopic Nissen fundoplication has replaced open fundoplication as the gold standard for control of severe GERD. The Nissen repair entails mobilization of the gastric fundus and gastroesophageal junction, then wrapping of the distal esophagus full-circle with the gastric fundus. The Nissen method of GERD surgical correction has become the accepted method for the majority of patients, with long-term results showing excellent control of symptoms and minimal complications at a mean follow-up of 10 years in patients who had undergone the traditional open Nissen fundoplication. The laparoscopic Nissen repair is the same fundamental operation, just performed in a less invasive manner. For this reason, many authors believe that the long-term results of the laparoscopic repair will parallel those of the traditional open procedure. Multiple large clinical series have demonstrated that laparoscopic Nissen fundoplication affords patients excellent control of reflux symptoms with a significantly shorter length of hospital stay, less postoperative pain, and faster return to normal function. Patients are typically discharged within 3 days and more than 90% report complete relief of reflux symptoms at a mean follow-up of 5 years.

Small Intestinal Surgery

Laparoscopic small bowel resection for benign and malignant disease is both safe and feasible and the recent advances in endoscopic anastomotic devices have rekindled interest in laparoscopic bowel surgery. Studies have shown that laparoscopy can be used to reliably explore the small bowel, resect various segments, and complete an anastomosis in a variety of conditions, including leiomyomas, carcinoids, diverticula, ischemic segments of bowel, and localized complications of IBD. An excellent application of laparoscopy is in patients with isolated small bowel Crohn's disease who require limited resection or strictureplasty. This can usually be easily accomplished through externalization of the involved segment of bowel through a limited incision and does not necessarily require totally laparoscopic repair. Also, patients requiring a loop ileostomy to provide temporary fecal diversion for anorectal Crohn's disease are excellent candidates for laparoscopy. Laparoscopy can also be used to create a feeding jejunostomy in patients who require enteral access for feeding but have significant gastric reflux or gastric outlet obstruction, which would prevent the use of a percutaneously placed gastrostomy tube.

Colon Resection

Laparoscopic colon resection has been successfully performed for both benign and malignant colon disease and initial studies showed that it is safe and effective, resulting in less postoperative pain, a shorter hospital stay, and faster return to normal function. Initial enthusiasm, however, was quelled by reports of an alarmingly high rate of port site recurrences compared to open procedures (21% versus 3.3%). In a recent randomized, prospective clinical study, though, Milsom and colleagues documented no trocar site recurrences after 5 years of follow-up. When performed by adequately trained surgeons, laparoscopic colon resection achieves equivalent margins of resection, lymph node dissection, and survival compared to open procedures. Tumor localization, however, must be accomplished without the aid of direct palpation and sensory feedback and therefore relies on preoperative barium enema, colonoscopy (with the use of India ink or colored marker), or other radiographic study. Enthusiasm for laparoscopic or laparoscopic-assisted colon resection for cancer continues to rise, but long-term follow-up data will be required to confirm the excellent results thus far. Laparoscopic colectomy has also been successfully applied to a variety of nonmalignant lesions, including complications from inflammatory bowel disease (IBD), benign polyps, and diverticular disease. Multiple authors have reported favorable results for laparoscopic colon resections for Crohn's disease. Laparoscopy is especially successful in managing isolated ileal or ileocolic Crohn's disease, for which a limited resection or strictureplasty is required. Patients typically have faster return of bowel function, less postoperative narcotic requirement, shorter length of hospital stay, and no more complications than comparable open procedures. When severe inflammation, abscess, or fistula is present, the laparoscopic approach becomes more difficult and conversion rates are higher; however, the overall benefit to the patient is significantly higher if successfully completed laparoscopically. With colonic Crohn's disease, in which a total abdominal colectomy or total proctocolectomy is required, laparoscopy has not yet proven superior to open procedures.

Laparoscopy has also become a more commonly accepted method for the treatment of all forms of diverticular disease, including acute and chronic forms. Sher *et al*. compared laparoscopic and open colectomies for diverticular disease of varying severity. The 5-year series found that laparoscopic resection of diverticulitis can be successfully completed with no additional morbidity or mortality in patients with pericolic, distant, or complex abscesses with or without a fistula. Patients in whom there was generalized purulent or fecal peritonitis, however, should not undergo laparoscopic resection.

Laparoscopic Surgery in Pregnancy

Laparoscopy during pregnancy does not pose any threat to the developing fetus. Diseases requiring surgical correction that are normally encountered during pregnancy, including appendicitis, cholecystitis, and bowel obstruction, can be managed laparoscopically, thus saving the expectant female from significant postoperative pain and fetal narcotic exposure. The same advantages of laparoscopy exist for the pregnant patients, namely, decreased postoperative pain, quicker return of bowel function, and faster recovery. The second trimester is the safest time to intervene surgically in pregnancy and thus surgical procedures should be appropriately scheduled as best as possible.

Other Procedures

Many common gastrointestinal surgical procedures have been performed laparoscopically. There has been recent growth in laparoscopic bariatric surgery for morbid obesity. The traditional vertical banded gastroplasty and Roux-en-Y gastric bypass have both been performed laparoscopically with slight modifications from their open counterparts. Initial reports from multiple centers document excellent results. Living donor nephrectomy for kidney transplantation has also become a popular means of procuring a kidney for living-related renal transplants. Limited hepatic resections, adrenalectomies, gastrectomies, and biliary bypasses have been performed laparoscopically. Randomized prospective trials will need to be performed to determine the benefits of these procedures.

SUMMARY

Great advances have been made in laparoscopic surgery over the past decade. Outcomes data have led to new consensus statements that shifted the gold standard for cholecystectomy and fundoplication away from convention open techniques to laparoscopic techniques. Sophisticated databases have been created to record and analyze details of patient follow-up in an effort to define the benefits of laparoscopic approaches to various procedures. Advances in instrumentation have led to the miniaturization of laparoscopic equipment and prospective trials evaluating the efficacy of these instruments will soon become available.

See Also the Following Articles

Appendicitis • Cholecystectomy • Colectomy • Fast-Track Surgery • Hernias • Minimally Invasive Surgery • Splenectomy • Trauma, Overview

Further Reading

Brooks, D. C. (1998). "Current Review of Minimally Invasive Surgery." Springer-Verlag, Philadelphia, PA.

Chekan, E. G., and Pappas, T. N. (1999). General principles of minimally invasive surgery. *In* "Surgery: Scientific Basis and Current Practice" (J. A. Norton, R. Bollinger, and A. E. Chang, eds.), pp. 429–453. Springer-Verlag, New York.

Gupta, A., and Watson, D. I. (2001). Effect of laparoscopy on immune function. *Br. J. Surg.* **88**, 1296–1306.

MacFayden, B. V., Jr., Wolfe, B. M., and McKernan, J. B. (1992). Laparoscopic management of the acute abdomen, appendix, and small and large bowel. *Surg. Clin. North Am.* **72**, 1169–1183.

Meyers, W. C., Branum, G. D., and Farouk, M. (1991). A prospective analysis of 1518 laparoscopic cholecystectomies. The Southern Surgeons Club. *N. Engl. J. Med.* **324**, 1073–1078.

Park, A., Targarona, E. M., and Trias, M. (2001). Laparoscopic surgery of the spleen: State of the art. *Langenbecks Arch. Surg.* **386**, 230–239.

Pejovic, T., and Nezhat, F. (2001). Laparoscopic management of adnexal masses: The opportunities and the risks. *Ann. N. Y. Acad. Sci.* **943**, 255–268.

Rosen, M., and Ponsky, J. (2001). Minimally invasive surgery. *Endoscopy* **33**, 358–366.

Rosenthal, R. L., Freidman, R. L., and Phillips, E. H. (1998). "The Pathophysiology of Pneumoperitoneum." Springer-Verlag, New York.

Schauer, P. R., and Ikramuddin, S. (2001). Laparoscopic surgery for morbid obesity. *Surg. Clin. North Am.* **81**, 1145–1179.

Zucker, K. A. (2001). "Surgical Laparoscopy," 2nd Ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Large Intestine, Development

ROBERT K. MONTGOMERY
Children's Hospital Boston

Hirschsprung's disease A disease that is characterized by the absence of parasympathetic intrinsic ganglion cells in the hindgut, caused by the failure of neural crest cells to migrate into the distal end of the intestine.

neural crest cells Cells originating from the dorsal region of the neural tube, a subpopulation of which migrate into the wall of the developing intestine and give rise to enteric neurons.

The primitive gut tube consists of three regions, the foregut, midgut, and hindgut, demarcated by the origins of their blood supply. The large intestine is derived from part of the midgut and from the hindgut. As development proceeds, length increases and the intestines protrude into the umbilical cord and then are retracted, and the ascending colon, transverse colon, and descending colon become identifiable. Less is known of the mechanisms regulating the development of the large intestine than of other portions of the gastrointestinal tract. The most intensively studied aspect of large intestine development is that of neural crest cell migration and development into the enteric nervous system, because of its role in Hirschsprung's disease. Because of their application as *in vitro* model systems, a number of cell lines derived from tumors of the large intestine have been widely studied.

MORPHOGENESIS

As the gut tube lengthens and the vascular system develops, the foregut, midgut, and hindgut regions may be distinguished as regions supplied by the celiac, superior mesenteric, and inferior mesenteric arteries, respectively. The cecum, appendix, ascending colon, and right and middle thirds of the transverse colon arise from the distal portion of the midgut. The hindgut will give rise to the left third of the transverse colon, the descending colon, the rectum, and the upper third of the anal canal. The gut tube rapidly increases in length as development proceeds, protruding into the umbilical cord and subsequently being retracted into

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Appendicitis • Cholecystectomy • Colectomy • Fast-Track Surgery • Hernias • Minimally Invasive Surgery • Splenectomy • Trauma, Overview

Further Reading

Brooks, D. C. (1998). "Current Review of Minimally Invasive Surgery." Springer-Verlag, Philadelphia, PA.

Chekan, E. G., and Pappas, T. N. (1999). General principles of minimally invasive surgery. *In* "Surgery: Scientific Basis and Current Practice" (J. A. Norton, R. Bollinger, and A. E. Chang, eds.), pp. 429–453. Springer-Verlag, New York.

Gupta, A., and Watson, D. I. (2001). Effect of laparoscopy on immune function. *Br. J. Surg.* **88**, 1296–1306.

MacFayden, B. V., Jr., Wolfe, B. M., and McKernan, J. B. (1992). Laparoscopic management of the acute abdomen, appendix, and small and large bowel. *Surg. Clin. North Am.* **72**, 1169–1183.

Meyers, W. C., Branum, G. D., and Farouk, M. (1991). A prospective analysis of 1518 laparoscopic cholecystectomies. The Southern Surgeons Club. *N. Engl. J. Med.* **324**, 1073–1078.

Park, A., Targarona, E. M., and Trias, M. (2001). Laparoscopic surgery of the spleen: State of the art. *Langenbecks Arch. Surg.* **386**, 230–239.

Pejovic, T., and Nezhat, F. (2001). Laparoscopic management of adnexal masses: The opportunities and the risks. *Ann. N. Y. Acad. Sci.* **943**, 255–268.

Rosen, M., and Ponsky, J. (2001). Minimally invasive surgery. *Endoscopy* **33**, 358–366.

Rosenthal, R. L., Freidman, R. L., and Phillips, E. H. (1998). "The Pathophysiology of Pneumoperitoneum." Springer-Verlag, New York.

Schauer, P. R., and Ikramuddin, S. (2001). Laparoscopic surgery for morbid obesity. *Surg. Clin. North Am.* **81**, 1145–1179.

Zucker, K. A. (2001). "Surgical Laparoscopy," 2nd Ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Large Intestine, Development

ROBERT K. MONTGOMERY
Children's Hospital Boston

Hirschsprung's disease A disease that is characterized by the absence of parasympathetic intrinsic ganglion cells in the hindgut, caused by the failure of neural crest cells to migrate into the distal end of the intestine.

neural crest cells Cells originating from the dorsal region of the neural tube, a subpopulation of which migrate into the wall of the developing intestine and give rise to enteric neurons.

The primitive gut tube consists of three regions, the foregut, midgut, and hindgut, demarcated by the origins of their blood supply. The large intestine is derived from part of the midgut and from the hindgut. As development proceeds, length increases and the intestines protrude into the umbilical cord and then are retracted, and the ascending colon, transverse colon, and descending colon become identifiable. Less is known of the mechanisms regulating the development of the large intestine than of other portions of the gastrointestinal tract. The most intensively studied aspect of large intestine development is that of neural crest cell migration and development into the enteric nervous system, because of its role in Hirschsprung's disease. Because of their application as *in vitro* model systems, a number of cell lines derived from tumors of the large intestine have been widely studied.

MORPHOGENESIS

As the gut tube lengthens and the vascular system develops, the foregut, midgut, and hindgut regions may be distinguished as regions supplied by the celiac, superior mesenteric, and inferior mesenteric arteries, respectively. The cecum, appendix, ascending colon, and right and middle thirds of the transverse colon arise from the distal portion of the midgut. The hindgut will give rise to the left third of the transverse colon, the descending colon, the rectum, and the upper third of the anal canal. The gut tube rapidly increases in length as development proceeds, protruding into the umbilical cord and subsequently being retracted into

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Development of the Human Large Intestine

4 weeks	Intestine present as simple tube
5 weeks	Cecum identifiable
7 weeks	Intestine protrudes into umbilical cord
8 weeks	Circular muscle layer identifiable; Meissner's plexuses present; peristaltic waves and motility
10 weeks	Intestine reenters abdominal cavity; enteroendocrine cells identifiable; longitudinal muscle layer
12 weeks	Auberbach's plexuses present; goblet cells identifiable
14 weeks	Villi; sucrase–isomaltase and other enzymes detectable
20 weeks	Mature anatomic position
23 weeks	Disappearance of colonic villi
30 weeks	Circular folds of colon complete

the abdominal cavity (see Table I). During this process, the growing intestines rotate around the superior mesenteric artery, leaving the large intestine ventral to the duodenum. As the intestines rotate and return to the abdominal cavity, the cecum and appendix, initially in the upper right, come to lie in the right lower quadrant. The dorsal peritoneum of the ascending and descending colon fuses with the posterior abdominal wall, making these segments retroperitoneal. Failures in the rotational process or fixation can lead to a number of malformations. Initially, the developing gut tube is contiguous with the urogenital sinus. The anal canal normally separates from the urogenital tract, but developmental anomalies sometimes result in fistulas or an imperforate anus. The upper and lower portions of the anorectal canal have distinct embryonic origins that are reflected in their respective blood supplies and enervation. The pectinate line marks the boundary between the region derived from the distal hindgut and that derived from the anal pit, which is ectodermal in origin.

A striking characteristic of the developing fetal colon is its initial similarity to the small intestine. The development of the colon is marked by three important cytodifferentiative stages: the appearance of a primitive stratified epithelium, similar to that found in the early development of the small intestine, the conversion of this epithelium to a villus architecture with developing crypts, and the remodeling of the epithelium when villi disappear and the adult-type crypt epithelium is established. Stem cells are located at the base of the colonic crypts and proliferation occurs in the lower third of the crypt. As they migrate toward the surface, the epithelial cells differentiate into colonocytes, goblet cells, and enteroendocrine cells. The three cell lineages become evident around the time of villus formation. Emerging evidence indicates that regulation of colonic goblet cell differentiation is distinct from that of small intestinal goblet cells. The mRNAs for mucins are detectable in the early intervillus epithelium, with strong expression in the crypts later in development.

The apical surface of columnar colonocytes displays enterocyte-like microvilli and glycogen stores are abundant. The quantity of intracellular glycogen, as well as the number of glycogen-positive cells, decreases from the end of the first trimester. By the end of gestation, very few glycogen-positive cells are found in the fetal colon. During malignant transformation, glycogen expression reappears and is a prominent feature of cell lines derived from human colonic adenocarcinomas.

Concurrent with the presence of villus morphology, the colonic epithelial cells express differentiation markers similar to those in small intestinal enterocytes. Sucrase–isomaltase becomes detectable early, increases as the villus architecture emerges, and then decreases rapidly to barely detectable levels at term. Alkaline phosphatase and aminopeptidase follow a pattern generally similar to that of sucrase–isomaltase. Analysis of isolated fetal colonic epithelial cells demonstrated low levels of the disaccharidases typical of the small intestine, including lactase. In addition, the fetal human colon during the second trimester can synthesize and process lipids, phospholipids, and apolipoproteins. The villi disappear by the end of the second trimester. Data on the functional development of the human colon remain sparse. Examination of premature infants suggests that sodium transport develops in the last trimester, whereas anion-exchange mechanisms remain immature until the end of the first year.

Analysis of *hox* gene expression patterns in experimental animals suggests that these genes are responsible for patterning of the large intestine and may delineate the boundaries between different parts. Mice with null alleles of several distally expressed *hox* genes display defects in formation of the anal region. Null alleles for members of the sonic hedgehog signaling pathway also result in anal region defects. These data suggest that, these gene families regulate, at least in part, overall patterning. In addition, some of the genes regulating specification of the different colonic cell lineages, for example, goblet cells, are now being identified, but much of the regulatory network remains to be elucidated.

HIRSCHSPRUNG'S DISEASE

The most extensively studied aspect of the development of the large intestine is the mechanism of Hirschsprung's disease, which is characterized by the absence of

parasympathetic intrinsic ganglion cells in both the submucosal (Meissner's) and the myenteric (Auerbach's) plexuses of the hindgut. The disorder is caused by failure of vagal neural crest cells to migrate into the distal end of the intestine and represents the most common form of congenital bowel obstruction.

Analysis of mutant mouse models has elucidated the causes of the disease. In the lethal spotted mutant mouse, the terminal segment of the gut is congenitally aganglionic. The aganglionic colon results from the failure of migrating neural crest cells to penetrate an abnormally thickened basal lamina and colonize the bowel normally. Laminin in this basal lamina interacts with a receptor on the neural crest cells and causes them to stop migration prematurely. In a different mouse model, the congenital megacolon of transgenic mice overexpressing *hoxa-4* appears to arise from a similar defective interaction between enteric neuron precursors and smooth muscle.

The genetic and molecular mechanisms underlying the disease are the subject of intense investigation and have generated an extensive and growing literature. Mutations in five genes affecting neural crest cell development that may give rise to Hirschsprung's disease have been identified thus far. The most extensively studied are mutations in the *ret* receptor tyrosine kinase gene and in the endothelin B receptor gene. The *ret* proto-oncogene encodes a transmembrane receptor tyrosine kinase. Its ligand has been identified as glial cell-line-derived neurotrophic factor (GDNF), which is expressed in the developing gut. Current evidence indicates that GDNF binds to another receptor, GDNFRα, and together they form a signaling complex with Ret. Mice homozygous for a truncated Ret protein lacking the kinase domain failed to form enteric ganglia. Numerous mutations in the *ret* gene have been identified in patients with Hirschsprung's disease. Similarly, a targeted disruption of the endothelin-B receptor gene produces mice with aganglionic megacolon, consistent with a critical role for this receptor as well in migration or differentiation of the neural crest cells that give rise to the enteric ganglia. In addition to Ret and endothelin B receptor defects, deficits in the ligands GDNF and endothelin 3, as well as in endothelin-converting enzyme, all result in Hirschsprung's-like phenotypes in mouse models. Unrelated to either of these receptor–ligand systems, a mutation in *sox-10*, a transcription factor expressed early in neural crest cell development, has also been identified in Hirschsprung's patients.

ONCOFETAL RELATIONSHIPS IN CULTURED CELLS

The expression of differentiation markers, for example, sucrase–isomaltase, and of small intestinal enterocyte-like morphology similar to that transiently seen in fetal colon, by several cell lines derived from human colonic adenocarcinomas has provided model systems for *in vitro* studies of enterocyte development. Of the available cell lines, the Caco-2, HT-29, and T-84 cell lines are the most widely used, as under suitable conditions they display some characteristics of differentiated small intestinal epithelial cells. Caco-2 cells have been shown to express a mixture of colonic and enterocytic markers, which varied over time in culture. Although useful models for many purposes, differentiation patterns in these cell lines may be quite heterogeneous and experimental data must be interpreted with caution.

See Also the Following Articles

Colonic Motility • Colonic Obstruction • Development, Overview • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Small Intestine, Development

Further Reading

Manie, S., Santoro, M., Fusco, A., and Billaud, M. (2001). The RET receptor: Function in development and dysfunction in congenital malformation. *Trends Genet.* **17**, 580–589.

Montgomery, R. K., Mulberg, A. E., and Grand, R. J. (1999). Development of the human gastrointestinal tract: Twenty years of progress. *Gastroenterology* **116**, 702–731.

Robertson, K., Mason, I., and Hall, S. (1997). Hirschsprung's disease: Genetic mutations in mice and men. *Gut* **41**, 436–441.

Laxatives

LAWRENCE R. SCHILLER
Baylor University Medical Center and University of Texas Southwestern Medical Center, Dallas

constipation A symptom complex that includes infrequency of defecation, straining to have a bowel movement, or increased hardness of stool consistency.
dyschezia Difficulty in evacuation of stool, including straining, painful defecation, and incomplete evacuation.
laxatives Agents that induce defecation and thereby increase stool frequency, soften stools, and ease the passage of stools.

Laxatives are drugs that induce defecation. Types of laxatives include purgatives and cathartics, which produce voluminous stools, stool softeners and emollients, which have only a mild effect, and lavage solutions, which are ingested in large volumes and flush out the gastrointestinal tract in preparation for diagnostic procedures such as colonoscopy. Laxatives can be usefully categorized as bulking or hydrophilic agents, which add additional hydrophilic solids to the stools, osmotic laxatives, which reduce fluid absorption by retaining water intralumenally, secretagogues, and agents with direct effects on epithelium, nerves, or smooth muscles, and the lubricating agent, mineral oil.

PHARMACOLOGY OF LAXATIVES

Bulking or Hydrophilic Agents

Organic polymers, such as dietary and medicinal fiber preparations, complex with substantial amounts of water when they are in aqueous solution. This physiochemical property allows such molecules to hold extra amounts of water intraluminally, thereby increasing intraluminal volume, speeding transit through the colon and increasing fecal weight. Stool consistency may be softened and defecation eased.

The effectiveness of these agents depends on the amount ingested, the water-holding capacity of the agent, the amount of the agent surviving bacterial fermentation in the colon, and the net effect of fermentation products on water absorption by the colon. Water-holding capacity varies by an order of magnitude among the various agents. For example, *in vitro* bran complexes with 4.2 g of water per gram of bran, whereas pectin holds 56.2 g of water per gram. Fermentation of fiber by colonic bacteria destroys the polymer and results in the production of short-chain fatty acids and gas (mainly carbon dioxide and hydrogen), accounting for the bloating and flatulence that complicate the use of these agents. To the extent that short-chain fatty acids are absorbed by the colonic mucosa, the laxative effectiveness of fiber is reduced. To the extent that they remain within the colonic lumen, they obligate water retention osmotically and increase the laxative effect. The interactions of these factors are largely unpredictable. Thus, bran is more effective at increasing stool weight than is pectin despite the large difference in water-holding capacity. Supplementing a general diet with increasing amounts of cereal rich in bran increases stool weight linearly by 2.7 g for every gram of increased fiber intake. Constipated patients should take 20–40 g of fiber daily.

Instead of dietary fiber, "medicinal" bulking agents, such as psyllium, polycarbophil, or altered cellulose, can be used. Psyllium is a natural product that consists of hydrophilic polysaccharides that form a gel when mixed with water. Ingestion of 10 g of psyllium daily increases stool weight by an average of 45 g/24 h, largely due to excretion of an additional 40 g of water daily. Polycarbophil is a derivative of polyacrylic acid and forms a gel that is not subject to fermentation when it interacts with divalent cations and water. Methylcellulose and carboxymethylcellulose are relatively resistant to fermentation by gut bacteria and tend to cause less bloating than dietary fiber. Medicinal fiber should always be taken with extra water to avoid intestinal obstruction.

Osmotic Laxatives

Osmotic agents are osmotically active ions or molecules that are poorly absorbed by the intestine and thereby obligate water retention intralumenally to maintain isotonicity with plasma. Magnesium salts, phosphate and sulfate salts, poorly absorbed disaccharides, such as lactulose, sugar alcohols, such as mannitol and sorbitol, and polyethylene glycol are classified as osmotic agents. These agents work because gut contents are in osmotic equilibrium with plasma after equilibration in the upper intestine and so any water-soluble

 Copyright 2004, Elsevier (USA). All rights reserved.

material remaining intraluminally must retain sufficient fluid to maintain intralumenal osmolality at 290 mosm/kg. For example, stool weight increases linearly by 7.3 g for each additional millimole of soluble magnesium in stool water. Quite large stool outputs can be achieved; for example, a standard 10 oz. bottle of citrate of magnesia contains 116 mmol of magnesium and therefore could increase stool weight by as much as 850 ml.

The side effects of saline laxatives relate to excess absorption of the poorly absorbed ion in the face of large intakes. Thus, hypermagnesemia, hyperphosphatemia, and hypernatremia have been reported with ingestion of magnesium and phosphate laxatives, especially when given in high doses or to patients with renal insufficiency. Complications, such as seizures or congestive heart failure, have also been observed with these agents. Mannitol, sorbitol, and lactulose can be fermented by colonic bacteria; this can produce excess flatus and bloating.

Polyethylene glycol is a large polymer (molecular weight ~3350) that should not exert much osmotic activity. However, polymers that interact strongly with water, such as polyethylene glycol, exert an anomalous osmotic activity that is proportional to the number of monomer units in the polymer. Thus, ingestion of 17 g of polyethylene glycol obligates intraluminal retention of approximately 80 ml of water. Larger doses of polyethylene glycol and supplemental electrolytes are used in gastrointestinal lavage solutions that are designed to be consumed in large volumes and to produce no net water or electrolyte absorption or secretion. Because polyethylene glycol is not absorbed appreciably and is not fermented by bacteria, it typically does not produce electrolyte disturbances or result in excess flatus.

Secretagogues and Agents with Direct Effects on Epithelium, Nerves, or Smooth Muscle

Many different categories of drugs produce laxation by various effects on the intestine and its regulatory elements (Table I). These agents have been called "stimulant" laxatives because of presumed stimulation of motility, but they also may have independent effects on mucosal transport.

The least active of these agents are the surface-active agents, including docusates and bile acids. These substances are detergents that inhibit fluid and electrolyte absorption or stimulate secretion. Although these chemicals may have measurable (if small) effects in increasing stool water and thus "softening" stools, clinical studies suggest that docusates are of little help in the prophylaxis of constipation in the bed-bound elderly.

TABLE I Classification of Laxatives

I. Bulking or hydrophilic agents
  A. Dietary fiber
  B. Psyllium (*Plantago*)
  C. Polycarbophil
  D. Methylcellulose, carboxymethylcellulose
II. Osmotic agents
  A. Poorly absorbed ions
    1. Magnesium sulfate (epsom salt)
    2. Magnesium hydroxide (milk of magnesia)
    3. Magnesium citrate
    4. Sodium phosphate
    5. Sodium sulfate (Glauber's salt)
    6. Potassium sodium tartrate (Rochelle salt)
  B. Poorly absorbed disaccharides, sugar alcohols
    1. Lactulose
    2. Sorbitol, mannitol
  C. Polyethylene glycol
III. Secretagogues and agents with direct effects on epithelium, nerves, or smooth muscle
  A. Surface-active agents
    1. Docusates (dioctyl sulfosuccinate)
    2. Bile acids
  B. Diphenylmethane derivatives
    1. Bisacodyl
    2. Sodium picosulfate
    3. (Phenolphthalein)[a]
  C. Ricinoleic acid (castor oil)
  D. Anthraquinones
    1. Senna
    2. Cascara
    3. Aloe
IV. Lubricating agent
  A. Mineral oil

[a] No longer available in the United States.

Diphenylmethane derivatives are frequently used stimulant laxatives. Bisacodyl is a common component of over-the-counter laxatives, especially since the withdrawal of phenolphthalein because of animal studies suggesting carcinogenicity. Bisacodyl is hydrolyzed in the small intestine and colon by endogenous esterases to its free, active form, which inhibits water absorption by effects on prostaglandins, kinins, and perhaps mucosal Na^+,K^+-ATPase. Bisocodyl can be given as 10 mg tablets, which produce an effect in 6–8 h, or as 5 and 10 mg suppositories, which work within 15–30 min.

Ricinoleic acid is the active component of castor oil (ricinoleic acid triglyceride). Castor oil is hydrolyzed by lipase, releasing free ricinoleic acid, which seems to stimulate both intestinal secretion and motility. Doses of 30–60 ml of castor oil produce catharsis within a few hours. Cramping frequently complicates its use and limits long-term use. Castor oil is most frequently used for bowel preparation.

Anthraquinones are a group of chemicals that contain the tricyclic anthracene nucleus. Substitutions in the ring structure define various monoanthrones, including aloe-emodin, rhein, and frangula. The monoanthrones are usually coupled together to form dianthrones, which are conjugated with sugars to produce glycosides. Plants that produce these agents include senna, aloe, cascara, and rhubarb. Available preparations are derived from plant material and are mixtures of various chemicals. The conjugated anthraquinones pass into the colon, where bacterial metabolism converts them to active forms. The drugs seem to work by both stimulating secretion and speeding transit through the colon. Defecation usually occurs 6–8 h after ingestion. Side effects include allergic reactions, electrolyte depletion, and melanosis coli, a benign discoloration of the mucosa due to apoptosis of the colonic epithelial cells. Anthraquinones have been blamed for the development of "cathartic colon," a neurogenic atony, but this does not seem to occur frequently with currently available anthraquinone laxatives.

Lubricating Agent

Mineral oil (petrolatum) is a lubricating agent that is not chemically active within the body, but instead alters the physical characteristics of stool, thereby easing defecation. It can be taken by mouth or as an enema. Aspiration is a hazard with oral intake and may produce lipoid pneumonia. Depletion of fat-soluble vitamins with chronic use and seepage through the anus are additional side effects.

THERAPEUTIC USE

Laxatives are used to treat the symptom of constipation, to prepare the colon for procedures, such as colonoscopy or surgery, and to empty the gastrointestinal tract after ingestion of poisons.

Use in Constipation

The symptom of constipation may mean infrequency of defecation, straining to have bowel movements (dyschezia), or excessively hard stools. Constipation may be due to systemic diseases, such as diabetes mellitus or hypothyroidism, or to disorders of the colon, such as diverticulitis or colon cancer, or may be idiopathic. Idiopathic constipation is attributed to the slow transit of material through the colon or functional outlet obstruction due to dysfunction of the anal sphincters or pelvic floor muscles or to transient anatomical obstruction.

For mild symptoms, increased dietary fiber or the use of medicinal fiber may be all that is needed. Docusates are frequently utilized by patients with chronic constipation, but are usually insufficient for use in patients consulting their physicians with constipation. For severe, acute constipation, more potent laxatives are needed; stimulant laxatives or large doses of osmotic laxatives work well in this setting. For more chronic symptoms, osmotic agents are preferred. When straining (dyschezia) is the main problem, modification of stool consistency with osmotic agents or mineral oil can be useful. Fiber supplementation should be avoided in patients with straining, since ingestion of more fiber results in the need to pass bulkier stools.

Use for Colon Preparation before Procedures

Adequate preparation of the colon for colonoscopy, radiography, or surgical procedures is essential for ideal viewing conditions and patient safety. Gastrointestinal lavage solutions are the standard against which other preparation regimens are compared. These solutions are designed to be ingested rapidly in large volumes to "wash out" colon contents. Unfortunately, even with flavorings they taste badly and it sometimes becomes difficult for patients to ingest the full dose. They remain the safest option for thorough colon cleansing, however.

Ionic osmotic laxatives, such as sodium phosphate or magnesium citrate, are popular alternatives. Laxatives that contain lactulose, mannitol, or sorbitol should not be used for colon preparation because of their potential to produce potentially explosive hydrogen gas. The ionic osmotic agents can be absorbed and may produce hyperphosphatemia or hypermagnesemia. They should not be used in patients with renal insufficiency. Sufficient water should be ingested with the ionic laxatives to avoid dehydration.

Traditional cathartics, such as castor oil and bisacodyl, can be used in various combinations or with enemas to prepare the colon. Their use may be complicated by cramping and dehydration.

Use in Toxic Ingestions

The role of laxatives in reducing absorption of orally ingested poisons is limited. Most cathartics work too slowly to have much effect and studies with gastrointestinal lavage solution suggest little effect on the kinetics of absorption of potential toxins.

See Also the Following Articles

Colonoscopy • Constipation • Defecation • Dietary Fiber • Over-the-Counter Drugs

Further Reading

Schiller, L. R. (2001). Review article: The therapy of constipation. *Aliment. Pharmacol. Ther.* **15**, 749–763.

Schiller, L. R. (1997). Cathartics, laxatives, and lavage solutions. *In* "Gastrointestinal Pharmacology and Therapeutics" (G. Friedman, E. D. Jacobson, and R. W. McCallum, eds.), pp. 159–174. Lippincott-Raven, Philadelphia, PA.

Xing, J. H., and Soffer, E. E. (2001). Adverse effects of laxatives. *Dis. Colon Rectum* **44**, 1201–1209.

Lipoproteins

NICHOLAS O. DAVIDSON

Washington University School of Medicine, St. Louis, Missouri

ABC-type transporter Members of a large multigene family of transport proteins expressed in the membranes of epithelial cells. ABC-type transporters contain an ATP-binding cassette, signifying that transport of the particular substrate is energy-dependent. Members of this family play an important role in lipoprotein metabolism by regulating lipid transport (particularly cholesterol and other sterols) both into and out of cells.

apolipoprotein Members of a large multigene family of proteins with the ability to bind lipids (cholesterol and triglyceride) in order to facilitate their transport in plasma.

brush border membrane Apical membrane of the enterocyte, facing the intestinal lumen. This is the site from which uptake of dietary lipid occurs.

chylomicron The largest lipoprotein particle. This is the particle secreted by the intestine following lipid absorption and allows the transport of triglyceride and cholesterol out of the enterocyte, into the lymphatic circulation, and ultimately into plasma. The major protein components are apolipoprotein B48, apolipoprotein A-I, and apolipoprotein A-IV.

chylomicron remnant Chylomicron particles enter the plasma and undergo metabolism of the triglyceride cargo through the actions of an enzyme called lipoprotein lipase. This results in liberation of fatty acids to adipose tissue and muscle. Removal of the core triglyceride leads to shrinkage of the surface and production of a smaller particle (chylomicron remnant) that is targeted to the liver for receptor-dependent uptake, a process facilitated by apolipoprotein E.

complex lipid The major lipid products of intestinal and hepatic lipid assembly. The major classes of neutral lipid include triglyceride and cholesterol ester, whereas complex polar lipids are represented by phospholipids.

endoplasmic reticulum Compartment of the cell responsible for the synthesis of complex lipids and apolipoproteins; it is the site where initiation of lipoprotein synthesis occurs.

fatty acid-binding protein Two classes of fatty acid-binding proteins are discussed. The first class comprises members of a large multigene family of cytoplasmic transporter proteins participating in shuttling fatty acids in the liver and small intestine. The second consists of a family of membrane-associated fatty acid transport proteins presumed to be involved in the uptake of fatty acids from plasma.

Golgi complex Compartment of the cell responsible for maturation of lipoprotein particles prior to their secretion into the plasma.

high-density lipoprotein The smallest and densest lipoprotein particle. Sites of formation include the small intestine, liver, and the plasma compartment following catabolism of large triglyceride-rich lipoproteins.

lipoprotein receptor Several classes of receptor demonstrate affinities for different lipoproteins. Some, such as the low-density lipoprotein receptor, cluster in coated pits and participate in the regulation of cholesterol homeostasis. Others, such as scavenger receptor type B1 and the ABC-type transporters ABC-A1, ABC-G5, and ABC-G8, play a role in cholesterol secretion and/or uptake that is yet to be fully understood.

low-density lipoprotein (LDL) The major carrier of cholesterol in plasma of humans. LDL is formed predominantly from very-low-density lipoprotein as a result of catabolism of the core triglyceride, analagous to the formation of chylomicron remnants. The major protein component is apolipoprotein B100.

microsomal triglyceride transfer protein (MTTP) A resident endoplasmic reticulum protein that is absolutely

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Schiller, L. R. (2001). Review article: The therapy of constipation. *Aliment. Pharmacol. Ther.* **15**, 749–763.

Schiller, L. R. (1997). Cathartics, laxatives, and lavage solutions. *In* "Gastrointestinal Pharmacology and Therapeutics" (G. Friedman, E. D. Jacobson, and R. W. McCallum, eds.), pp. 159–174. Lippincott-Raven, Philadelphia, PA.

Xing, J. H., and Soffer, E. E. (2001). Adverse effects of laxatives. *Dis. Colon Rectum* **44**, 1201–1209.

Lipoproteins

NICHOLAS O. DAVIDSON

Washington University School of Medicine, St. Louis, Missouri

ABC-type transporter Members of a large multigene family of transport proteins expressed in the membranes of epithelial cells. ABC-type transporters contain an ATP-binding cassette, signifying that transport of the particular substrate is energy-dependent. Members of this family play an important role in lipoprotein metabolism by regulating lipid transport (particularly cholesterol and other sterols) both into and out of cells.

apolipoprotein Members of a large multigene family of proteins with the ability to bind lipids (cholesterol and triglyceride) in order to facilitate their transport in plasma.

brush border membrane Apical membrane of the enterocyte, facing the intestinal lumen. This is the site from which uptake of dietary lipid occurs.

chylomicron The largest lipoprotein particle. This is the particle secreted by the intestine following lipid absorption and allows the transport of triglyceride and cholesterol out of the enterocyte, into the lymphatic circulation, and ultimately into plasma. The major protein components are apolipoprotein B48, apolipoprotein A-I, and apolipoprotein A-IV.

chylomicron remnant Chylomicron particles enter the plasma and undergo metabolism of the triglyceride cargo through the actions of an enzyme called lipoprotein lipase. This results in liberation of fatty acids to adipose tissue and muscle. Removal of the core triglyceride leads to shrinkage of the surface and production of a smaller particle (chylomicron remnant) that is targeted to the liver for receptor-dependent uptake, a process facilitated by apolipoprotein E.

complex lipid The major lipid products of intestinal and hepatic lipid assembly. The major classes of neutral lipid include triglyceride and cholesterol ester, whereas complex polar lipids are represented by phospholipids.

endoplasmic reticulum Compartment of the cell responsible for the synthesis of complex lipids and apolipoproteins; it is the site where initiation of lipoprotein synthesis occurs.

fatty acid-binding protein Two classes of fatty acid-binding proteins are discussed. The first class comprises members of a large multigene family of cytoplasmic transporter proteins participating in shuttling fatty acids in the liver and small intestine. The second consists of a family of membrane-associated fatty acid transport proteins presumed to be involved in the uptake of fatty acids from plasma.

Golgi complex Compartment of the cell responsible for maturation of lipoprotein particles prior to their secretion into the plasma.

high-density lipoprotein The smallest and densest lipoprotein particle. Sites of formation include the small intestine, liver, and the plasma compartment following catabolism of large triglyceride-rich lipoproteins.

lipoprotein receptor Several classes of receptor demonstrate affinities for different lipoproteins. Some, such as the low-density lipoprotein receptor, cluster in coated pits and participate in the regulation of cholesterol homeostasis. Others, such as scavenger receptor type B1 and the ABC-type transporters ABC-A1, ABC-G5, and ABC-G8, play a role in cholesterol secretion and/or uptake that is yet to be fully understood.

low-density lipoprotein (LDL) The major carrier of cholesterol in plasma of humans. LDL is formed predominantly from very-low-density lipoprotein as a result of catabolism of the core triglyceride, analagous to the formation of chylomicron remnants. The major protein component is apolipoprotein B100.

microsomal triglyceride transfer protein (MTTP) A resident endoplasmic reticulum protein that is absolutely

Copyright 2004, Elsevier (USA). All rights reserved.

required for the transfer of complex lipid (above) to the growing polypeptide chain of apolipoprotein B. Defects in the gene encoding MTTP result in the autosomal recessive disorder abetalipoproteinemia, a condition that is associated with lipid accumulation in enterocytes and hepatocytes.

reverse cholesterol transport The process through which high-density lipoprotein acquires cholesterol from peripheral cells and transports and delivers it to the liver for secretion into bile or for use in the synthesis of bile salts.

RNA editing A form of posttranscriptional modification in which the nucleotide sequence of the transcript is modified from that encoded in the genome.

transcript messenger RNA The required intermediate between the genetically encoded sequence and protein product.

very-low-density lipoprotein (VLDL) A large lipoprotein particle synthesized in the liver and small intestine containing complex lipid for export. VLDLs are the major lipoproteins secreted by the liver and contain apolipoprotein B100 as the predominant protein.

Lipoproteins are huge multicomponent complexes containing lipids and proteins. These complexes are essential for the transport of cholesterol and triglyceride throughout the body. Lipoproteins allow the transport of dietary lipids, mostly in the chemical form of cholesterol and triglyceride, following their absorption by the small intestine. In addition, the liver synthesizes cholesterol and triglyceride to supply tissues that may lack the ability to make sufficient quantities of these forms of fat. The ability of the small intestine to absorb dietary lipid and of the liver to package newly synthesized lipid is absolutely dependent on the process of lipoprotein assembly and secretion. The small intestine and liver produce a range of different lipoproteins. These different lipoproteins have distinct importance in lipid delivery as a result of the protein components that accompany each particular lipoprotein class. The different protein components play a central role in targeting each lipoprotein to a specific receptor on the cell surface. This process in turn facilitates the regulation of cellular lipid metabolism and provides a mechanism for the cell to regulate its content of cholesterol. Genetic defects in the expression of certain transport proteins or of receptors for certain lipoprotein classes are associated with abnormalities in plasma lipid levels and in some circumstances accumulation of lipid within intestinal and hepatic cells.

INTRODUCTION

The challenge of transforming water-insoluble lipids such as cholesterol and triglyceride into a form that can be transported through the aqueous medium of plasma is achieved by the formation of large macromolecular particles containing both complex lipid and apolipoproteins into structures called lipoproteins. Different classes of lipoproteins are produced by the liver and small intestine, each with distinct functions in lipid transport and delivery. In view of the importance of dietary fat as an energy source (~30–40% of the daily energy requirement of the average U.S. diet), these transport vehicles assume considerable importance in maintaining nutritional balance. In addition, lipoproteins are required for cholesterol delivery to all cells in connection with their growth, renewal, and proliferation. The need for cells to acquire cholesterol from circulating lipoproteins is particularly noteworthy since all cells have the innate capacity to synthesize their own cholesterol from two-carbon precursors. This latter observation indicates that the demands for cholesterol may exceed the capacity of cellular production in certain circumstances, particularly during growth or for specialized metabolic purposes. Certain specialized cells, for example, hepatocytes and adrenal cortical cells, require cholesterol for the synthesis of bile acids and steroid hormones, respectively, and cholesterol destined for these purposes is derived largely from circulating lipoproteins. In addition, the developing brain and central nervous system require prodigious amounts of cholesterol and certain fat-soluble vitamins, notably vitamins A and E, for appropriate development. Taken together, lipoproteins represent a vital adaptation for organ-specific delivery and control of nutrient metabolism in addition to playing a role in the regulation of cholesterol homeostasis. These features are summarized in Fig. 1, elements of which are discussed in the following paragraphs.

INTESTINAL LIPOPROTEIN ASSEMBLY

Complex Lipid Synthesis

Complex lipid synthesis and the formation of lipoproteins within the intestinal enterocyte require several coordinated steps. Dietary lipid is presented to the brush border in the form of a mixed micelle containing fatty acid, monoglyceride, and bile salts, from which fatty acid and monoglyceride are transported into the enterocyte. It is likely that one of several fatty acid transporter proteins (FATPs) participates in this process, the most reasonable candidate being FATP4. Cholesterol and other sterols are transported across the intestinal brush border membrane, but the mechanisms and transporters involved are not fully defined. Potential candidates include ABC-type transporters and scavenger receptor type B1 (SR-B1) (Fig. 1). Once across the enterocyte brush border membrane, fatty acids are

activated to their coenzyme A derivative and transferred to the endoplasmic reticulum (ER) for assembly into complex lipid. The role of fatty acid-binding proteins in this transfer process has been proposed but has not yet been formally proven. Two fatty acid-binding proteins, known as intestinal fatty acid-binding protein and liver fatty acid-binding protein, are expressed at high levels in the enterocyte. Once delivered to the ER, fatty acids and monoglycerides are used for the assembly of complex lipids, including phospholipids, triglycerides, and cholesterol esters. Enzymes that participate in this process include monoacylglycerol acyltransferase (MGAT), diacylglycerol acyltransferase I and II, and acylcholesterol acyltransyltransferase I and II (ACAT I and II). Some of these enzymes have a more or less important role in the intestine (MGAT and ACAT II) than in the liver and other tissues. Once complex lipid assembly has been initiated in the ER, the process of intestinal lipoprotein assembly begins almost simultaneously (Fig. 1).

Role of Microsomal Triglyceride Transfer Protein and Apolipoprotein B

The efficient assembly of lipoproteins in the intestine requires the coordinated interaction of many genes, among which two gene products in particular are indispensable. These include microsomal triglyceride transfer protein (MTTP) and apolipoprotein B (apoB). Neutral lipid, including cholesterol ester and triglyceride, is synthesized in the ER and accumulates as small droplets. These lipid droplets are transferred to and appear to fuse with the nascent peptide chains of apoB within the lumen of the ER (Fig. 1). This neutral lipid transfer step allows the appropriate folding and conformation of the apoB protein such that it is transferred through the secretory pathway. In the absence of lipid transfer, for example, in the setting of genetic MTTP deficiency accompanying abetalipoproteinemia, apoB is degraded and neutral lipid accumulates within the ER. Similarly, in the absence of apoB or in the setting of truncating mutations that eliminate the lipid-binding domains of apoB, neutral lipid accumulates within the ER. The prototype for genetic deficiency of apoB is homozygous hypobetalipoproteinemia.

ApoB RNA Editing

ApoB is required as an integral structural component of lipoproteins. The protein is synthesized in the liver and in the small intestine, but each organ produces a distinct isoform. The liver produces a large protein referred to on a centile scale as apoB100. The intestine produces a protein approximately half the size of the liver form, referred to as apoB48. A specialized form of apoB mRNA processing, which results in the formation of these two protein products from a single structural *APOB* gene, occurs in the small intestine. In this process, referred to as RNA editing, a cytidine to uridine deamination of the nuclear apoB mRNA results in the introduction of a UAA stop codon and leads to translation of a truncated apoB protein (apoB48). In humans, the liver expresses apoB mRNA but synthesizes only the full-length protein (apoB100) since apoB mRNA is not edited in the liver. ApoB100 is the major protein component of low-density lipoproteins (LDL), the principal transport vehicle for cholesterol in humans (Fig. 1). ApoB48 is the major protein component of chylomicrons (Fig. 1). These two isoforms of apoB have distinct metabolic fates, as detailed below. ApoB mRNA editing is postulated to represent a metabolic advantage by way of targeting dietary lipid from the small intestine to the liver for rapid delivery as an energy substrate and also for the delivery of fat-soluble vitamins that are targeted to liver stores.

INTESTINAL LIPOPROTEIN SECRETION

Enterocytes secrete two major classes of lipoprotein. Bulk triglyceride transport is achieved through the assembly and secretion of chylomicrons, as detailed above. These are large (75−1200 nm) particles that are well adapted for the transport of triglyceride into the circulation and ultimately to the liver. By virtue of their size, a limited amount of apoB is capable of transporting a large amount of lipid (Fig. 1). ApoB is proposed to wrap itself around the lipid droplet, belt-like, and, together with the addition of other apolipoproteins, notably apoA-I, apoA-IV is secreted into the lymphatic circulation, where it acquires apoE (Fig. 1). Chylomicrons undergo hydrolysis of their core triglyceride in the peripheral circulation through the action of an enzyme, lipoprotein lipase, which is tethered within the capillary endothelium of muscle and adipose tissue beds. Following hydrolysis of the core triglyceride, the surface of the chylomicron shrinks and the particle becomes smaller, allowing more apoE to bind. This step is required for the receptor-dependent uptake of chylomicron remnants by the liver, through the chylomicron remnant receptor (Fig. 1). Enterocytes also secrete high-density lipoprotein (HDL). These are small (8−12 nm) particles composed mostly of phospholipid, apoA-I, and apoA-IV, together with a small amount of neutral lipid (Fig. 1). The importance of HDL secretion from the intestine is unknown although estimates suggest that ∼50% of the daily HDL input arises from this source. HDL plays a central role in "reverse

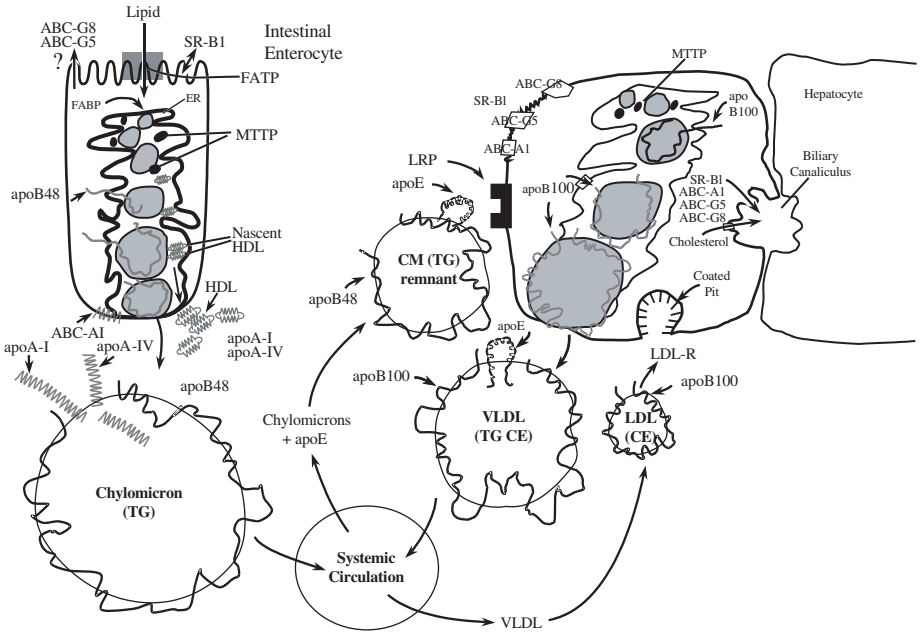


FIGURE 1 Intestinal and hepatic lipoprotein assembly and secretion and an integrated view of lipoprotein metabolism. Dietary lipid is absorbed into the intestinal enterocyte with the participation of members of the fatty acid transport protein (FATP) family. Fatty acids are bound by fatty acid-binding proteins (FABP) and are transferred to the endoplasmic reticulum (ER) for complex lipid assembly. Neutral complex lipids (cholesterol ester and triglyceride) are transferred by the microsomal triglyceride transfer protein (MTTP) to the growing chain of apolipoprotein B (apoB48) and this process initiates lipoprotein formation. The lipoprotein particles enlarge through the addition of progressively more lipid to the core and are secreted into the lymph as a chylomicron. The core lipid in chylomicrons is largely triglyceride (TG). Additional proteins are added to the chylomicron in the later stages of lipoprotein assembly, including apoA-I and apoA-IV. Intestinal enterocytes also secrete small high-density lipoprotein (HDL) particles containing apoA-I and apoA-IV. HDL metabolism in the intestine is also influenced by the expression of various HDL receptors, including ABC-type transporters (ABC-A1, ABC-G5, and ABC-G8) as well as the scavenger receptor, SR-B1. These HDL receptors and transporters likely also play a role in regulating cholesterol absorption from the diet. The assembly and secretion of lipoproteins from the liver follow many of the same patterns as in the intestinal cell. Fatty acids and cholesterol are assembled into complex lipid within the ER and are transferred to the growing chain of apoB100 through the actions of MTTP. Hepatocytes assemble very-low-density lipoproteins (VLDLs), which are smaller than chylomicrons and contain apoB100 rather than apoB48. The core lipid in VLDL is both TG and cholesterol ester (CE). Both VLDL and chylomicrons enter the systemic circulation where the core lipids undergo hydrolysis, releasing fatty acids to muscle and adipose tissue. Chylomicrons are converted into chylomicron remnants and are directed to the chylomicron remnant receptor (also known as low-density lipoprotein receptor-related protein, or LRP). This targeting is dependent on the transfer of apoE to the remnant particle. VLDLs are converted into low-density lipoproteins (LDLs), which are the major transport carrier of cholesterol in humans. LDLs are directed to the LDL receptor, which localizes in coated pits on the surface of all cells. This interaction is dependent on a domain in the carboxyl-terminus of apoB100 that is absent from apoB48. HDL transporters and HDL receptors are also expressed on hepatocytes and may participate in the regulation of cholesterol secretion into bile.

cholesterol transport," the process by which cholesterol is removed from the membranes of peripheral cells and delivered to the liver for excretion. This pathway represents an important mechanism of the body to rid itself of excess cholesterol and perhaps is the reason that HDL (the so-called good cholesterol) is protective against atherosclerosis. Intestinal cells express the transporter proteins ABC-A1, ABC-G5, and ABC-G8, recently identified as cholesterol and sterol transport proteins (Fig. 1). In addition, enterocytes express the macrophage scavenger receptor protein SR-B1, presumed to function as an HDL receptor. The precise role of these

receptors and transporters in reverse cholesterol transport and their importance in the recognition and transport of cholesterol—both into and out of the enterocyte—are important questions for future investigation. Increased HDL levels in plasma are protective against the development of atherosclerosis in humans and factors that regulate intestinal secretion of HDL are under investigation. A further aspect of the ability of intestinal epithelial cells to secrete lipoproteins is that expression of the ABC-type transporters may allow these cells to discriminate between cholesterol and other, more toxic sterols. Rare genetic diseases are associated with the indiscriminate absorption of plant sterols, which are normally not absorbed by humans. Some of these defects have been linked to mutations in ABC-type transporters.

HEPATIC LIPOPROTEIN ASSEMBLY

Assembly of lipoproteins within the liver follows much of the same basic principles outlined above for the enterocyte. Fatty acids are transported into hepatocytes through fatty acid transport proteins and are transported to the ER by the hepatic fatty acid-binding protein. Once inside the ER, complex lipid assembly occurs and transfer of the neutral lipid to a nascent apoB100 protein again requires binding by MTTP (Fig. 1). Neutral lipid is fused to the growing apoB protein and protects the peptide from degradation within the ER. During the later stages of lipoprotein assembly, apoE is added to the particle, which is then secreted into the circulation. Lipoprotein assembly in the hepatocyte and enterocyte differs in several respects (Fig. 1). First, hepatic lipoprotein assembly and secretion involve the delivery of complex lipid into the space of Disse in the form of very-low-density lipoprotein (VLDL) rather than chylomicrons (Fig. 1). VLDLs differ from chylomicrons in two important respects, namely, their smaller size (30−80 nm) and the presence of apoB100 and apoE in place of apoB48. In addition, there are two distinct routes for lipid secretion from hepatocytes since these cells secrete cholesterol directly into bile in addition to secreting lipoproteins into plasma (Fig. 1). The metabolic regulation of this partitioning process (i.e., VLDL versus biliary cholesterol) is poorly understood. A further major distinction between hepatic and intestinal lipoprotein assembly is that VLDL production is driven by both substrate (e.g., fatty acid delivery) and metabolic demands (e.g., insulin and other hormones) and is subject to feedback and feed-forward regulation by numerous genes involved in the bile acid biosynthetic pathway. Hepatic VLDLs are the major source of circulating LDL and thus factors that regulate hepatic secretion rates of VLDL in turn regulate the levels of LDL.

HEPATIC LIPOPROTEIN SECRETION

The liver plays a central role in the secretion of lipoproteins into the circulation as a mechanism for delivering newly synthesized complex lipid and cholesterol to peripheral tissues and organs. In addition to its role in lipoprotein secretion, however, the liver plays an essential role in regulating circulating lipid levels through lipoprotein uptake and catabolism. In the process of receptor-dependent LDL uptake, the liver is quantitatively the most important organ. This property of hepatocytes is relevant to an understanding of how cholesterol levels in plasma are maintained in humans and the accompanying increase in atherosclerotic heart disease and stroke that accompanies elevated LDL cholesterol levels. Most of the plasma cholesterol in humans is transported in LDL, and since elevated levels are strongly predictive of susceptibility to atherosclerosis, LDL is often referred to as the "bad cholesterol." Two major mechanisms contribute to the regulation of LDL levels. These include VLDL synthesis and LDL clearance from the circulation. Most of the circulating LDL is cleared from the circulation through its interaction with a receptor, the LDL receptor (Fig. 1).

LDL receptors are clustered in coated pits throughout the basal-lateral membranes of hepatocytes and permit binding and internalization of LDL, which in turn reduces LDL levels in plasma. The clearance of LDL through its interaction with the LDL receptor is quantitatively the dominant mechanism for regulating plasma cholesterol levels in humans. Genetic defects in the LDL receptor result in familial hypercholesterolemia, which is associated with extremely high levels of LDL in plasma and an increased risk of atherosclerotic heart disease.

The physical interaction of LDL with the LDL receptor is mediated through two regions that are present in the carboxyl-terminus of apoB100. The regions or domains in apoB100 responsible for the high-affinity binding of LDL to the LDL receptor are absent from apoB48, which cannot bind to the LDL receptor as efficiently as apoB100. As a result, intestinal apoB48-containing lipoproteins (chylomicrons and remnant particles) are directed to a different receptor (chylomicron remnant receptor or LDL receptor-related protein) that recognizes apoE (Fig. 1). ApoB RNA editing thus permits rapid delivery of intestinal lipoproteins to the liver (the half-life of chylomicrons and remnants is <60 min) versus the fine regulation of plasma

cholesterol levels achieved through their association with LDL and LDL receptor-dependent transport (a half-life of approximately 2 days).

Hepatocytes express the transporter proteins ABC-A1, ABC-G5, and ABC-G8, as well as the HDL receptor SR-B1. The ABC-type transporters have an important role in allowing cholesterol to be exported from cell membranes. This process is separate from the lipoprotein secretion pathways outlined above. Cell membranes accumulate cholesterol and other sterols within the bilayer and this may be transferred to an HDL acceptor particle from the plasma in the presence of an ABC-type transporter. The classical example of such a transporter that functions in this manner is ABC-A1. This transporter is widely expressed and is proposed to act as a cholesterol pump, transferring cholesterol to an acceptor HDL particle. Genetic defects in ABC-A1 have been linked to Tangier disease, in which circulating HDL levels are extremely low. The role of these ABC-type transporters in cholesterol secretion into bile and their participation in regulating levels of HDL in plasma are under investigation.

See Also the Following Articles

Apoproteins • Barrier Function in Lipid Absorption • Hepatocytes • Hyperlipidemia

Further Reading

Anant, S., and Davidson, N. O. (2001). Molecular mechanisms of apolipoprotein B mRNA editing. *Curr. Opin. Lipidol.* **12**, 159–165.

Boord, J. B., Fazio, S., and Linton, M. F. (2002). Cytoplasmic fatty acid-binding proteins: Emerging roles in metabolism and atherosclerosis. *Curr. Opin. Lipidol.* **13**, 141–147.

Davidson, N. O. (1999). Intestinal lipid absorption. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 1, pp. 428–456. Lippincott Williams & Wilkins, Philadelphia, PA.

Davidson, N. O., and Shelness, G. S. (2000). Apolipoprotein B: mRNA editing, lipoprotein assembly and presecretory degradation. *Annu. Rev. Nutr.* **20**, 169–193.

Kim, E., and Young, S. G. (1998). Genetically modified mice for the study of apolipoprotein B. *J. Lipid Res.* **39**, 703–723.

Leung, G. K., Veniant, M. M., Kim, S. K., Zlot, C. H., Raabe, M., Bjorkegren, J., Neese, R. A., Hellerstein, M. K., and Young, S. G. (2000). A deficiency of microsomal triglyceride transfer protein reduced apolipoprotein B secretion. *J. Biol. Chem.* **275**, 1501–1510.

Raabe, M., Flynn, L. M., Zlot, C. H., Wong, J. S., Veniant, M. M., Hamilton, R. L., and Young, S. G. (1998). Knockout of the abetalipoproteinemia gene in mice: Reduced lipoprotein secretion in hererozygotes and embryonic lethality in homozygotes. *Proc. Natl. Acad. Sci. USA* **95**, 8686–8691.

Schaffer, J. E. (2002). Fatty acid transport: The roads taken. *Am. J. Physiol. Endocrinol. Metab.* **282**, E239–E246.

Schmitz, G., Langmann, T., and Heimerl, S. (2001). Role of ABCG1 and other ABCG family members in lipid metabolism. *J. Lipid Res.* **42**, 1513–1520.

Liver Abscess

NATALIE J. TÖRÖK AND GREGORY J. GORES
Mayo Clinic and Foundation

hematogenous spread Dissemination via the circulation.
hydatid cyst Vesicular structure formed as a consequence of *Echinococcus* infection.
pyogenic Pus forming.
scoleces Anterior parts of the *Echinococcus*; they are formed within the hydatid cyst.
trophozoite Vegetative form of a parasite.

Liver abscesses do not represent a specific liver disease, but rather a final common pathway of many pathologic processes. In the preantibiotic era, liver abscesses were typically sequelae of intraabdominal infectious processes (e.g., appendicitis) or of tuberculosis. Currently, most liver abscesses are secondary to biliary tract diseases or amebiasis. Many advances, including the development of antibiotics, recognition of the role of the anaerobic bacteria, the advent of noninvasive imaging, and the use of nonsurgical drainage for pyogenic abscesses, have improved outcome.

EPIDEMIOLOGY

The incidence of liver abscess shows a wide range of variation, depending on the etiology. The incidence of pyogenic liver abscess ranges from 8 to 20 cases per million people. With the availability of effective antimicrobials, pyogenic liver abscess has become a disease of middle-aged persons, probably because of the prevalence of biliary disease as a major cause is higher in this population. No significant sex, ethnic, or geographic differences seem to exist in disease prevalence, in contrast to the epidemiology of amebic liver abscess. Amebic liver abscess is marked by a significant male preponderance and is a disease seen most commonly in patients who reside in or have emigrated from an endemic area, or who have travel history to an endemic area. Countries with the highest *Entamoeba histolytica* endemic activity include Mexico, India, East and South Africa, and Central and South America. As estimated by the World Health Organization (WHO) in 1995, approximately 40 million to 50 million people worldwide are symptomatic with amebic colitis or liver abscess, resulting in 40,000–100,000 deaths each year. In the United States, immunosuppression seems to be an important risk factor. Patients with HIV infection, malnourishment, and alcoholism are at risk of developing amebiasis.Echinococcosis may result in hydatid liver cyst formation; the most common pathogen is *Echinococcus granulosus*. The resulting cystic hydatid disease has worldwide distribution. In tuberculosis, the liver is commonly involved, with a granulomatous reaction, but the formation of tuberculous liver abscess is rare worldwide and usually is not considered in the differential diagnosis of liver abscesses. Worldwide, only 43 cases have been reported.

PATHOGENESIS

Pyogenic liver abscesses are classified by the presumed route of hepatic invasion: (1) biliary tree, (2) portal vein, (3) hepatic artery, (4) direct extension from contiguous focus of infection, and (5) penetrating trauma. Cholangitis is now the major cause of pyogenic liver abscess. Patients with Caroli's disease and primary sclerosing cholangitis (PSC) have a particularly high rate of pyogenic abscess formation. In Third World countries, infection of the biliary tree by *Cryptococcus neoformans* or direct invasion by *Ascaris lumbricoides* may occur. Mechanical manipulation of the biliary tree, such as after endoscopic retrograde cholecystopancreatography (ERCP), or percutaneous biliary tube placement can be complicated by abscess formation. The portal venous system drains almost all of the abdominal viscera. Pylephlebitis from diverticulitis, inflammatory bowel disease, pancreatitis, and prostatitis can result in pyogenic abscess formation (Table I). Historically, untreated appendicitis was a major cause of liver abscess in the past, but its occurrence has greatly diminished with the use of antibiotics. Systemic bacteremia may also result in microabscess formation in the liver because bacteria are disseminated via the hepatic artery. Direct extension from a contiguous source of infection can occur with cholecystitis and subphrenic abscesses.

In terms of pathogens, *Escherichia coli* and *Klebsiella* species are by the far the most common isolates.

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Occult Causes of Pyogenic Liver Abscesses

Diverticulitis
Inflammatory bowel disease
Prostatitis
Pelvic inflammatory disease
Pancreatitis
Appendicitis

Klebsiella abscesses are frequently associated with gas formation. Enterococci and the *viridans* streptococci are also common, especially in polymicrobial abscesses, whereas *Staphylococcus aureus* is more commonly associated with monomicrobial abscesses. In hemochromatosis, abscesses are often caused by *Yersinia enterocolitica*.

Amebiasis is caused by ingestion of infective *E. histolytica* cysts through a fecal–oral route of exposure. The cysts are resistant to degradation in the stomach and pass to the small intestine, where excystation and liberation of the trophozoite form of the parasite occur. The trophozoites then pass to the colon, where they cause mucosal invasion and subsequent spread to the liver. The pathology in amebic abscess involves three consecutive stages: acute inflammation, granuloma formation, and progressively advancing necrosis with abscess formation and periportal fibrosis. The abscess contains necrotic tissue surrounded by a rim of amebic trophozoites.

Echinococcosis causes human disease when humans become hosts for a cystic intermediate stage of canine tapeworms. *Echinococcus granulosus* most often infects humans in contact with sheep-herding dogs. The hydatid liver cysts are fluid-filled structures lined by a parasite-derived membrane. The cysts caused by *Echinococcus multilocularis* are less defined; they tend to invade the liver parenchyma and also seed the adjacent organs.

CLINICAL PRESENTATION

Only 10% of patients with pyogenic liver abscess present with the classic triad of fever, jaundice, and right upper quadrant pain. Fever and constitutional symptoms including malaise, fatigue, and anorexia are common. In the presence of cholangitis, pruritus and jaundice may also be additional features. There seems to be an association between the cause of the abscess and the duration of the symptoms: hematogenous liver abscesses present most acutely (3 days), whereas those secondary to pylephlebitis (42 days) have the longest duration of symptoms.

In the case of amebic liver abscesses, 80% of patients present with symptoms that develop over 2–4 weeks. The symptoms are similar to those in patients with pyogenic abscesses; gastrointestinal symptoms also occur in 10–35% of patients and may include nausea, vomiting, diarrhea, abdominal cramping, diarrhea, or constipation.

Echinococcal liver cysts are often asymptomatic and present as a hepatic mass with a typical appearance on computer tomography (CT) or magnetic resonance imaging (MRI).

DIAGNOSIS

Laboratory tests in cases of pyogenic liver abscesses reveal leukocytosis in many cases. Liver biochemistries are also abnormal in most of the patients; however, normal results do not exclude the diagnosis. Serum alkaline phosphatase level is elevated in two-thirds of patients and tends to be more markedly elevated than transaminase levels. Blood cultures are positive in about half of the patients. Multiple samples should be obtained, because this may be the only clue to the pathogenic agent prior to starting antimicrobial therapy.

Leukocytosis is moderate in cases of amebic liver abscesses, and eosinophilia is rare. Anemia is a typical finding. Overall, the incidence of liver test abnormalities is the same in patients with amebic liver abscess or with pyogenic liver abscess. Most patients ($>70\%$) with amebic abscess do not have detectable parasites in the stool, therefore serologic testing for antibodies to *E. histolytica* is the most useful diagnostic test. The most commonly used serologic test currently is the enzyme immunoassay (EIA), because it is rapid, stable, and more specific and sensitive than other diagnostic tests.

In hydatid liver disease, the *Echinococcus* cysts are rarely found in the stool. Eosinophilia is usually present. Enzyme-linked immunosorbent assay (ELISA) is the test of choice and is positive in about 90% of cases with the disease.

Radiographic imaging studies are essential in making the diagnosis. Ultrasonography (US) and CT have proved particularly useful for abscess visualization and subsequent drainage (see Fig. 1). US is the study of choice in patients with suspected biliary disease. Intravenous contrast-enhanced CT offers improved sensitivity and is superior for guiding complex draining procedures. MRI studies are now also recognized to be useful alternatives to CT imaging, especially if the abscesses arise as a result of cholangitis, allowing better definition of disease extent. Magnetic resonance cholangiography may further add to the sensitivity, especially if there is a suspicion of a cholangiocarcinoma. Chest radiographs are abnormal about half of the time,

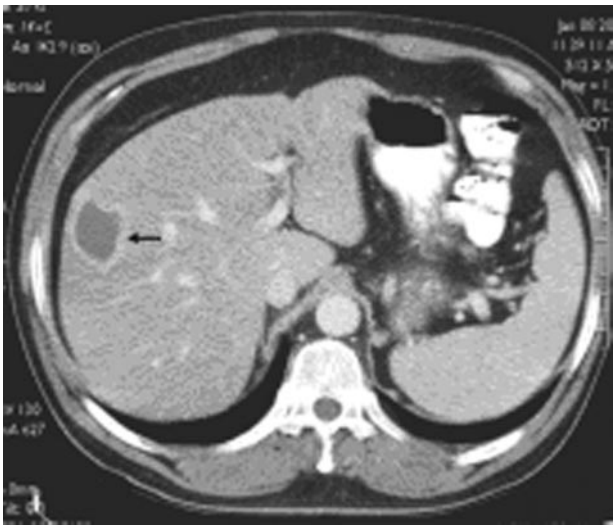


FIGURE 1 CT image of a pyogenic liver abscess (arrow); note the peripheral enhancement.

but are of no real value in making the diagnosis. A ^{99m}Tc nuclear hepatic scan is useful in differentiating an amebic abscess from a pyogenic abscess. Because amebic liver abscesses do not contain leukocytes, they appear as "cold" lesions on a nuclear scan, with a typical "hot" halo of radioactivity surrounding the abscess due to increased focal activity of leukocytes.

TREATMENT

The treatment of pyogenic abscesses consists of antibiotic administration and drainage of the abscesses. Radiographically guided percutaneous drainage is now pursued in most patients. Over the past 20 years, several studies have shown that percutaneous catheter drainage with antibiotic therapy has a success rate between 70 and 90%. Observations by Giorgio and co-workers suggest that repeated needle aspiration in selected patients with low viscous material within the abscess cavity may be a reasonable alternative.

Antibiotics should be started as soon as the diagnosis is suspected, after blood cultures are obtained. Pyogenic abscesses that arise in patients with biliary disease often include enterococci and enteric gram-negative bacilli. Biliary drainage is also essential if the abscess is secondary to an obstructed bile duct. If the source of the infection is the colon or pelvic organs, antibiotic coverage for coliforms and anaerobes is needed. The usual course of treatment is 2–3 weeks of parenteral antibiotic therapy with adequate abscess cavity drainage. The clinical response is followed by serial interval studies.

Treatment options for uncomplicated amebic liver abscess include amebicidal drugs and, if indicated, percutaneous or open aspiration of the abscesses. Therapeutic aspiration of the amebic liver abscess should be considered in patients with (1) high risk of abscess rupture as defined by a cavity size of 5 cm or greater, (2) left lobe abscess (because of the associated higher mortality rate), and (3) failure to respond to drug therapy. Metronidazole remains the drug of choice for treating amebic liver disease. Metronidazole, 750 mg orally three times a day for 10 days, has been reported to be curative in more than 90% of patients with amebic liver abscess. If a patient fails to respond to metronidazole in 5 days, then chloroquine (base) may be substituted or added to the therapy.

Albendazole has become the current standard for medical therapy of hydatid disease. It is generally administered two to three times daily at doses ranging between 10 and 50 mg/kg/day, for 12 weeks. Frequent monitoring for leukopenia and elevation of aminotransferase levels is required. The standard and final therapy is surgery, with the careful aspiration the of cyst fluid to avoid spillage of viable scoleces or anaphylaxis. Laparoscopic drainage and ultrasound-guided drainage have also been reported.

See Also the Following Articles

Amebiasis • Computed Tomography (CT) • Liver Cysts • Magnetic Resonance Imaging (MRI) • Percutaneous Drainage • Ultrasonography

Further Reading

Akgun, Y., Tacyildiz, I. H., *et al.* (1999). Amebic liver abscess: Changing trends over 20 years. *World J. Surg.* **23**(1), 102–106.

Babba, H., Messedi, A., *et al.* (1994). Diagnosis of human hydatidosis: Comparison between imagery and six serologic techniques. *Am. J. Trop. Med. Hyg.* **50**(1), 64–68.

Balci, N. C., Semelka, R. C., *et al.* (1999). Pyogenic hepatic abscesses: MRI findings on T1- and T2-weighted and serial gadolinium-enhanced gradient-echo images. *J. Magn. Reson. Imaging* **9**(2), 285–290.

Chou, F. F., Sheen-Chen, S. M., *et al.* (1997). Single and multiple pyogenic liver abscesses: Clinical course, etiology, and results of treatment. *World J. Surg.* **21**(4), 384–388; discussion, 388–389.

Giorgio, A., Tarantino, L., *et al.* (1995). Pyogenic liver abscesses: 13 years of experience in percutaneous needle aspiration with US guidance. *Radiology* **195**(1), 122–124.

Gottstein, B. (1992). Molecular and immunological diagnosis of echinococcosis. *Clin. Microbiol. Rev.* **5**(3), 248–261.

Hansen, P. S., and Schonheyder, H. C. (1998). Pyogenic hepatic abscess. A 10-year population-based retrospective study. *Apmis* **106**(3), 396–402.

Liver Biopsy

CHEN LIU AND JAMES M. CRAWFORD
University of Florida

alcoholic liver disease Liver abnormality caused by excessive ingestion of alcohol.
hematoxylin and eosin stain A tissue stain that is routinely used for histology evaluation.
immunostain A laboratory histochemical procedure that detects a specific antigen using a specific antibody.
Masson trichrome A histochemical stain that highlights the connective tissues.
needle biopsy A surgical procedure to obtain liver tissue using a needle.
periodic acid-Schiff stain A histochemical stain that highlights glycogen and glycoproteins.
reticulin stain A histochemical stain that highlights the delicate collagen fibers.
viral hepatitis Liver inflammation caused by viral infection.
wedge biopsy A surgical procedure to obtain liver tissue by excision.

Liver biopsy is a medical procedure to remove a small piece of liver tissue to diagnose liver disease and assist in disease management. Liver biopsy may be percutaneous, i.e., through the abdominal skin, or open, as during laparoscopy or laparotomy. It is the required technique for visualizing disease processes in the liver at the microscopic level. Liver biopsy may be carried out by "cutting," whereby intact pieces of tissue are obtained either by cutting needle or by scalpel for histological processing. Liver biopsy may also be carried out by "aspiration," in which a thin needle is inserted into the liver substance and cellular material is withdrawn under suction; this variant yields dispersed specimens for cytologic analysis and/or clumps of tissue for histological examination. The first cutting-needle biopsy device was introduced by Vim and Silverman in 1938 and was used in procedures requiring several minutes for percutaneous placement of the needle and withdrawal of tissue specimens. A key refinement was the introduction of Menghini cutting needle in 1958 and the use of this needle in percutaneous needle biopsy procedures requiring only a second or two of penetration and withdrawal. The resultant substantial decrease in bleeding complications enabled percutaneous liver biopsy to become a routine procedure in the evaluation and management of patients with suspected liver disease. More recently introduced cutting-needle biopsy devices, such as the Tru-cut biopsy and biopsy guns, provide for semiautomation of the percutaneous procedure.

TYPES OF LIVER BIOPSY

Liver biopsy is an invasive procedure. It requires a skilled clinician and special monitoring of the patient for postbiopsy complications, of which intra-abdominal bleeding is the chief concern. The mortality rate from percutaneous liver biopsy is on the order of 0.1% for cutting needle biopsies and much less for aspiration biopsies that utilize a finer needle. Percutaneous liver biopsy is by far the most commonly performed procedure. Blind needle biopsy uses abdominal percussion only as the guide for needle placement. Guided biopsy uses computed tomography (CT) or ultrasound to visualize liver anatomy and guide the placement and direction of the percutaneous needle biopsy thrust. In patients with clotting abnormalities, a transvenous route through jugular vein permits specimens to be obtained adjacent to the root of the hepatic vein, without risk of extravascular bleeding. Cutting or aspiration biopsies may be obtained at the time of laparoscopy or laparotomy. With a scalpel, a triangular segment of liver approximately 0.5 cm in maximal dimension is cut out from the anterior convexity of the liver surface or from the lower edge of a liver lobe; the defect can be sutured closed and hemostasis verified before exiting the abdomen.

The typical cutting-needle biopsy specimen is a 2 to 3 cm cylindrical core of tissue, between 0.4 and 0.8 mm in diameter, depending on the diameter of the cutting needle (Fig. 1). This sample represents approximately 1/50,000 of the organ. Liver biopsy thus has an inherent problem of sampling error in a variety of hepatic disorders. Liver biopsy is well suited to assessment of diffuse disorders, since the entire liver corpus is affected by many injurious events, particularly inflammatory and toxic process. Liver biopsy to diagnose disorders that exhibit irregular distribution of focal lesions within the liver is more problematic. The latter disorders

 Copyright 2004, Elsevier (USA). All rights reserved.

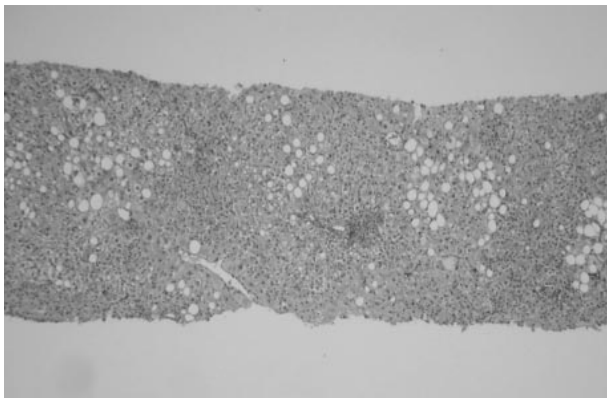


FIGURE 1 Middle portion of a liver needle core biopsy tissue from a patient with hepatitis C viral infection. Original magnification, ×40.

include primary biliary cirrhosis, primary sclerosing cholangitis, graft-versus-host disease, and nonhepatotropic viral infections such as cytomegalovirus. The needle may harvest tissue exhibiting definitive diagnostic histology or it may miss such tissue.

Liver wedge biopsy is excellent for the evaluation of focal lesions that are immediately below the capsule of the liver. It is not as well suited to general diagnosis of liver disease, since the liver tissue obtained from the subcapsular region is not representative of the whole liver. In particular, the normal penetration of fibrous septa from Glisson's capsule into the superficial 0.5 cm of the liver substance makes assessment of liver fibrosis in the immediate subcapsular region unreliable. If a wedge biopsy is to be obtained for assessment of diffuse liver disease, a cutting-needle biopsy penetrating deep into the liver should also be obtained. Despite the smaller size of the sample obtained from the cutting-needle biopsy, it frequently is the more reliable source of histologic information.

Fine-needle aspiration biopsy is particularly useful for evaluating hepatic mass lesions. It is usually performed with ultrasound or CT imaging. Considering that neoplasms such as liver cell adenomas and hepatocellular carcinomas have a rich arterial circulation, the lower risk of postbiopsy bleeding often makes fine-needle aspiration biopsy greatly preferable to cutting-needle biopsy. If the fine needle is properly placed, the samples obtained often (but not always) provide the needed information for diagnosis of the mass lesion. Liver aspiration is not suitable for evaluation of medical hepatic diseases because the samples do not preserve much if any of the architecture of the tissue. It should also be noted that no form of needle biopsy is recommended for diagnosis of hemangiomas. Noninvasive imaging techniques are usually sufficient to identify hemangiomas without incurring the strong likelihood of bleeding following biopsy of such lesions.

INDICATIONS FOR LIVER BIOPSY

Serum biochemical tests that remain of consistent value for screening for liver disease include aminotransferases as measures of hepatocellular injury; alkaline phosphatase and γ-glutamyl transpeptidase as measures of cholestatic injury to either hepatocytes or bile ducts; serum bilirubin (direct and total), albumin, and clotting times as measures of hepatocellular function; and serological markers for infections and autoimmune events. With the continued evolution of laboratory testing for liver disease, a clinical differential diagnosis is usually well established at the time of liver biopsy. The indications for liver biopsy—percutaneous or open—include the following: *de novo* histologic diagnosis of liver disease; assessment of liver disease severity; guidance in dealing with a difficult differential diagnosis; monitoring of multiple causes of liver disease simultaneously affecting the liver; and assessment of focal lesions. Liver biopsy is now frequently used to provide prognostic as well as diagnostic information. For these reasons, liver biopsy plays a critical role in clinical hepatology practice.

In the community care setting, the most common indication for an initial liver biopsy is an abnormal liver laboratory test. Many insidious chronic liver diseases cause mild liver abnormalities and liver biopsy plays an indispensable role in determining the underlying etiology. Sustained fever of unknown origin or unexplained hepatic or splenomegaly also may prompt a liver biopsy. Identification of a liver mass lesion, either incidentally or as part of evaluating right upper quadrant symptomatology, also frequently leads to liver biopsy. When diagnoses have been established, especially for viral hepatitis, liver biopsy is used to monitor the progression of disease and/or efficacy of treatment. Liver biopsy is a critical part of managing liver transplant patients, as early histological identification of graft injury, allograft rejection, recurrent disease, or new causes of liver damage greatly assists effective management of this patient population.

LIVER BIOPSY SPECIMEN PREPARATION IN THE LABORATORY

In order to accurately interpret a liver biopsy, it is essential to handle and prepare the tissue specimen

correctly. Depending on the nature of the suspected diseases, a variety of laboratory techniques can be used to enhance the diagnostic sensitivity and specificity. Therefore, there must be effective communication between the clinician and the pathologist about the patient's status before the biopsy is taken, so that the correct processing is chosen. In most instances, all liver biopsy tissue is routinely fixed in buffered formalin and processed for microscopy, but this is not always the best choice. In the setting of possible hematological disorders affecting the liver, it is essential to have some fresh (not fixed) tissue submitted for flow cytometric analysis. When inherited metabolic disorders are suspected, it is necessary to preserve a portion of the liver tissue by immediate immersion in liquid nitrogen for biochemical, enzymatic, and/or molecular analysis; a small portion also should be placed in electron microscopy fixative for potential ultrastructural analysis. If an infectious etiology other than hepatotropic viral infection is suspected, a portion of fresh tissue may be submitted for microorganism cultures.

Liver tissue fixed in buffered formalin can be stained with a variety of techniques. The most common tissue stain is hematoxylin and eosin. The commonly used stains and their utility are listed in Table I.

PRINCIPLES FOR LIVER BIOPSY INTERPRETATION

A systematic approach is best for evaluating liver biopsy. Individual histopathologic features should be integrated into a clinically relevant differential diagnosis. It is true that definitive diagnosis based on a liver biopsy cannot be rendered all the time. However, a thorough and systematic description of the histologic findings forms an important basis for ongoing clinical management of patients. Conversely, morphological findings must be considered in the context of clinical information.

Microscopic examination must first determine the hepatic architecture. This is usually performed under light microscopy with low magnification, e.g., $\times 40$. This examination includes noting the presence of adequate portal tracts, the presence of focal lesions, the presence of distortion of hepatic architecture, the distribution of portal tracts and terminal hepatic veins, sinusoidal and liver plate architecture, the presence of inflammatory cell infiltrates in the portal tracts or in the lobules, and the presence of zonal injury (i.e., periportal, midzonal, or centrilobular).

Next, attention should be directed to specific anatomic compartments. Portal tracts should be examined for the status of hepatic arteries and portal veins, the presence or absence of bile ducts, evidence of bile duct epithelial damage, bile ductular proliferation, the composition of inflammatory cells if present, the integrity of the hepatocyte-limiting plate at the margin of the portal tract, the presence of granulomas, and the presence of significant portal tract or periportal fibrosis. Examination of the lobular parenchyma should be focused on the presence or absence of an inflammatory cell infiltrate; the extent of hepatocellular apoptosis, necrosis, atrophy, ballooning degeneration (swollen hepatocytes), and steatosis (microvesicular or macrovesicular); and the presence or absence of retained bile pigment (cholestasis) and Mallory bodies (clumps of intermediate filaments in hepatocytes). In addition, sinusoids should be examined for dilation and congestion, increased Kupffer cell numbers, and sinusoidal fibrosis (the result of stellate cell activation and collagen deposition in the space of Disse). Terminal

TABLE I Selected Histological Stains in Liver Biopsy Interpretation

Stain	Comment
Hematoxylin and eosin	All-purpose stain for histological assessment
Masson trichrome	Connective tissue in portal tracts, parenchyma, and around terminal hepatic veins
Reticulin	Delicate collagen fibers of the subendothelial space, parenchymal architecture
Iron	Hepatocellular pigmented granules, Kupffer cell granules
Periodic acid-Schiff	Hepatocellular glycogen, basement membrane around bile ducts
Periodic acid-Schiff with diastase	Hepatocellular glycoproteins (especially α-1-antitrypsin), cellular debris in macrophages
Shikata	Elastin fiber deposition in long-standing fibrosis
Orcein	Copper-binding protein in Wilson disease, elastin fibers, hepatitis B surface antigen
Rubeanic acid	Copper granules in Wilson disease
Hall	Bilirubin stains green, connective tissue stains red
Oil red O	Lipid droplets stain red
Congo red	Amyloid

hepatic veins are examined for vascular injury and occlusion.

Certain patterns of hepatocellular damage are particularly significant for diagnosis. Apoptosis of individual hepatocytes is common in many liver diseases. It is pronounced in the setting of viral hepatitis, especially when it is associated with the inflammatory disruption of limiting plates (interface hepatitis) and a portal tract and parenchymal infiltrate that is predominantly lymphocytic. The term confluent necrosis is used when there are contiguous areas of hepatocyte death. It occurs in the setting of viral or drug-induced hepatitis. The term "bridging necrosis" refers to confluent hepatocyte death that links portal tracts one to another or to terminal hepatic veins. It is commonly associated with acute hepatitis or fulminant hepatic failure. Ischemic necrosis of hepatocytes, as results from cardiovascular shock, leads to ghosts of poorly staining hepatocytes in the pericentral area. Microvesicular steatosis of hepatocytes may result from the metabolic overload of acute alcohol exposure or may signify mitochondrial injury from drugs, acute fatty liver of pregnancy, or Reye's syndrome. Macrovesicular steatosis is featured in alcoholic liver disease and nonalcoholic fatty liver disease.

Evaluation of cholestasis is an essential part of the liver biopsy examination. Cholestasis is defined as retention of bile substances in hepatocytes, biliary channels, or both. Cholestasis may be nonobstructive, the result of hepatocellular dysfunction from many forms of injury. Hepatocytes become discolored by the brown bilirubin pigments and bile canaliculi may be expanded by retained bile. Sepsis may lead to dilation of bile ductules with massive retention of bile. Obstruction of bile flow in major bile ducts leads to edema and neutrophilic inflammation of portal tracts, inspissation of bile in bile ducts, and bile ductular proliferation, as well as hepatocellular and canalicular cholestasis.

In addition to the above important findings from the routine hematoxylin and eosin stain, special tissue stains are an important adjunct in the assessment of a liver biopsy. A Masson trichrome stain is helpful for determination of the degree of the fibrosis, up to and including identification of the liver disruption of cirrhosis. A reticulin stain defines the architecture of the parenchyma, in particular the width of hepatocellular plates and whether they are expanded, atrophied, or collapsed from hepatocyte destruction. Periodic acid-Schiff stains—without and with diastase digestion of glycogen—are useful for evaluating parenchymal injury at the portal tract interface, accumulation of debris-containing macrophages, bile duct architecture, and whether metabolic storage diseases are present (especially α-1-antitrypsin storage disorder and type IV glycogen storage disease). An iron stain helps to characterize the nature of hepatocellular and Kupffer cell pigment granules.

Additional information can be gained from immunohistochemical staining, which allows the pathologist to identify particular molecular markers in the liver tissue (Table II). These include immunostains for the antigens of hepatitis B virus, cytomegalovirus, adenovirus, or herpesvirus. Immunostain for hepatitis C virus remains of questionable clinical utility, particularly given the excellent serological and virological markers that are available. Immunostaining has also been widely used to characterize the nature of a neoplasm based on the differential molecular expression of tumor cells. It should be noted that immunohistochemistry permits evaluation of metastatic disease to the liver

TABLE II Selected Immunohistochemical Stains in Liver Biopsy Interpretation

Stain	Comment
Hepatitis B surface antigen	Active hepatitis B infection or the carrier state with retained HBsAg in hepatocytes
Hepatitis B core antigen	Active hepatitis B infection
Hepatitis D antigen	Active hepatitis D infection (with hepatitis B)
Pancytokeratin	All hepatocytes and bile duct epithelial cells
Cytokeratin 7	Bile duct epithelial cells, hepatocytes to variable degrees in cholestatic injury
Cytokeratin 19	Bile duct epithelial cells
Cytokeratin 20	Metastatic adenocarcinoma (as from colon), but not bile duct malignancy
HepPar 1	Hepatocytes, variable staining of hepatocellular carcinoma
α-Fetoprotein	Variable staining of hepatocellular carcinoma
Polyclonal carcinoembryonic antigen	Bile canaliculi of hepatocytes (including hepatocellular carcinoma), apical membrane of bile duct epithelial cells
Monoclonoal carcinoembryonic antigen	Metastatic adenocarcinoma, cytoplasmic staining
CD34	Endothelium, including sinusoidal endothelium

and determination of whether a tumor is hepatocellular carcinoma, cholangiocarcinoma, or a mixture of both. However, immunohistochemistry is not the basis for determining whether a hepatocellular nodule is malignant. Rigorous assessment of tissue architecture and cytology on hematoxylin and eosin stain, with the assistance of connective tissue stains such as the reticulin stain, remain the gold standard for diagnosis of hepatocellular carcinoma. Finally, on occasion *in situ* RNA hybridization is of value, particularly for identifying Epstein-Barr virus early mRNA in the diagnosis of posttransplantation lymphoproliferative disorder.

All the above-mentioned histologic findings must be considered in the context of the clinical presentation, laboratory tests, and other physical findings in order to render an accurate and helpful clinical diagnosis.

LIVER DISEASES WITH CHARACTERISTIC LIVER BIOPSY FINDINGS

Comments are made about some of the more common liver diseases.

Viral Hepatitis

Biopsies from patients who are infected with hepatitis B or hepatitis C virus usually exhibit a portal tract lymphocytic infiltrate, interface hepatitis, lymphocytosis in the lobules with associated hepatocyte apoptosis, and a variable degree of portal fibrosis. Hepatitis B liver can exhibit hepatocytes with a smudgy "ground glass" cytoplasm, the result of accumulated hepatitis B surface antigen (HBsAg). Hepatitis C liver tends to have more lymphoid aggregates in the portal tracts and focal hepatocyte macrovesicular steatosis.

Alcoholic Liver Disease

The characteristic features of alcohol-related liver histopathology are hepatocellular steatosis (microvesicular and macrovesicular—fatty liver) and, with severe alcohol exposure, a parenchymal neutrophilic infiltrate, hepatocyte ballooning degeneration and apoptosis, Mallory bodies (clumps of intermediate filaments in the hepatocyte cytoplasm), and sinusoidal and perivenular fibrosis (alcoholic hepatitis).

Nonalcoholic Fatty Liver Disease

Bland macrovesicular steatosis may occur in patients who do not consume alcohol but have mild elevations in serum aminotransferase levels. This may signify nonalcoholic fatty liver disease. When the macrovesicular steatosis is accompanied by a parenchymal neutrophilic infiltrate, ballooning degeneration, and apoptosis, nonalcoholic steatohepatitis merits consideration.

Drug-Induced Hepatitis

Drug-induced liver injury can produce almost any form of histologic outcome, ranging from a frank hepatocyte necrosis with direct hepatotoxins, an inflammatory hepatitis with hepatocyte destruction, or variable hepatocyte injury with drugs that induce immunological injury, ranging from moderate to severe cholestasis to outright cirrhosis. The possibility of drug-induced liver injury should be in the differential diagnosis of any histological assessment of liver pathology.

Hemochromatosis

The most significant finding for a diagnosis of hemochromatosis is the marked increase in iron storage, especially in the hepatocytes. The more reliable criterion is the quantification of iron by chemical analysis of liver tissue.

α-1-Antitrypsin Storage Disorder

The failure of the liver to secrete α-1-antitrypsin leads to α-1-antitrypsin deficiency, with its attendant lung injury. Some, but not all, patients with α-1-antitrypsin deficiency accumulate mutated protein within hepatocytes and develop severe liver injury, constituting α-1-antitrypsin storage disorder. The critical finding for this disease is the presence of periodic acid-Schiff-diastase-resistant globules in the cytoplasm of hepatocytes. Most of the affected hepatocytes surround the portal tracts.

Primary Biliary Cirrhosis

The key histology of primary biliary cirrhosis is a dense infiltrate of lymphocytes and plasma cells in portal tracts, with lymphocytic infiltration of bile ducts and bile duct destruction. Portal tract granulomas centered upon degenerating bile ducts is even more pathognomonic. Secondary changes include hepatocellular cholestasis, bile ductular proliferation, and progressive portal and periportal fibrosis.

Primary Sclerosing Cholangitis

Progressive fibrosis around portal tract bile ducts leads to their obliteration in this disease; secondary

changes of hepatocellular cholestasis and progressive periportal fibrosis develop.

Cirrhosis

Cirrhosis is defined as diffuse subdivision of the liver by interconnecting fibrous tissue, isolating individual islands of hepatocytes. The islands may undergo regenerative proliferation, generating spherical nodules embedded in constricting fibrous tissue. The nodules may be consistently less than 3 mm in diameter (micronodular cirrhosis) or variable in size and greater than 3 mm in diameter (macronodular cirrhosis). Liver biopsy of the cirrhotic liver frequently yields a fragmented specimen consisting of spherical tissue fragments composed of hepatocyte nodules surrounded by a rim of fibrous tissue. If the needle successfully cuts through the cirrhotic liver tissue, the biopsy shows tracts of fibrous septa subdividing a disorganized hepatocellular parenchyma.

Hepatocellular Carcinoma

Hepatocellular carcinoma is diagnosed on the basis of both architecture and cytology. Markedly thickened liver cell plates consistently greater than three cells thick is the single most important architectural feature. Unlike essentially any other form of solid malignancy, hepatocellular carcinoma tissue consists of thick tongues of hepatocytes (trabeculae) coated by a thin external endothelial layer. Unlike normal liver cell plates, there is diminished to absent reticular staining of collagen fibers in the subendothelial space of malignant trabeculae in hepatocellular carcinoma. Other architectural configurations of hepatocellular carcinoma include circular "pseudoacinar" configurations of hepatocytes and thick cords of hepatocytes invading fibrous tissue of an otherwise cirrhotic liver. The cytologic features of hepatocellular carcinoma are a modest increase in hepatocellular diameter, with an increased nuclear: cytoplasmic ratio and variable degrees of nuclear atypia. Well-differentiated hepatocellular carcinomas may have little cytologic atypia, hence the importance of assessing architecture.

Metastatic Malignancy

By far the most common malignancy in the liver is metastatic disease. Metastases in the liver frequently originate from colon carcinoma, pancreatobiliary carcinoma, breast carcinoma, lung carcinoma, hematologic malignancy, and malignant melanoma.

CONCLUSION

The key to liver biopsy evaluation is systematic examination of the parenchyma, portal tracts, and component cellular populations. Rigorous identification of patterns of injury and their distribution among cellular elements permits alignment of histologic findings with the clinical differential diagnosis and possible definitive diagnosis of the cause of liver injury. Even without definitive understanding of the cause of liver disease, the liver biopsy findings provide important information for patient management and prognostication.

See Also the Following Articles

Alcoholic Liver Injury, Hepatic Manifestations of • Alpha-1-Antitrypsin (α1AT) Deficiency • Cholestatic Diseases, Chronic • Cirrhosis • Hepatitis B • Hepatitis C • Hepatocellular Carcinoma (HCC) • Hepatocytes • Hereditary Hemochromatosis • Non-Alcoholic Fatty Liver Disease

Further Reading

Buscarini, L., Fornari, F., Bolondi, L., Colombo, P., Livraghi, T., Magnolfi, F., Rapaccini, G. L., and Salmi, A. (1990). Ultrasound-guided fine-needle biopsy of focal liver lesions: Techniques, diagnostic accuracy and complications. A retrospective study on 2091 biopsies. *J. Hepatol.* **11**, 344–348.

Crawford, A. R., Lin, X. Z., and Crawford, J. M. (1998). The normal adult human liver biopsy: A quantitative reference standard. *Hepatology* **28**, 323–331.

Dienstag, J. L. (2002). The role of liver biopsy in chronic hepatitis C. *Hepatology* **36**, S152–S160.

Geller, S. and Pitman, M. (2001). Morphological diagnostic procedures (liver biopsy). *In* "Pathology of the Liver" (R. N. M. MacSween, A. D. Burt, B. C. Portmann, *et al.,* eds.), 4th Ed., pp. 1–66. Churchill Livingston, London.

Gerber, M. A., and Thung, S. N. (1987). Histology of the liver. *Am. J. Surg. Pathol.* **11**, 709–722.

Goodman, Z. D. (2002). Drug hepatotoxicity. *Clin. Liver Dis.* **6**, 381–397.

Ishak, K. G. (2000). Pathologic features of chronic hepatitis: A review and update. *Am. J. Clin. Pathol.* **113**, 40–55.

Karamshi, M. (2001). Transjugular liver biopsy: Method of choice for taking a biopsy. *Br. J. Nurs.* **10**, 727–730.

Lebrec, D., Goldfarb, G., Degott, C., Rueff, B., and Benhamou, J. P. (1982). Transvenous liver biopsy: An experience based on 1000 hepatic tissue samplings with this procedure. *Gastroenterology* **83**, 338–340.

Maeda, T., Kajiyama, K., Adachi, E., Takenaka, K., Sugimachi, K., and Tsuneyoshi, M. (1996). The expression of cytokeratins 7, 19, and 20 in primary and metastatic carcinomas of the liver. *Mod. Pathol.* **9**, 901–909.

McAfee, J. H., Keeffe, E. B., Lee, R. G., and Rosch, J. (1992). Transjugular liver biopsy. *Hepatology* **15**, 726–732.

Pitman, M., and Szyfelbein, W. (1994). "Final Needle Aspiration Biopsy of the Liver." Butterworth-Heinemann, Boston.
Reddy, K., and Jeffers, L. (1999). Liver biopsy and laparoscopy. *In* "Diseases of the Liver" (E. R. Schiff, M. F. Sorrel, and W. C. Maddrey, eds.), 8th Ed., pp. 245–266. Lippincott-Raven, Philadelphia, PA.
Scheuer, P., and Lefkowitch, J. (2000). "Liver Biopsy Interpretation," 6th Ed. W. B. Saunders, Philadelphia, PA.
Snover, D. (1992). Technical aspects of the evaluation of liver biopsies. *In* "Biopsy Diagnosis of Liver Diseases" (D. C. Snover, ed.), pp. 2–23. Williams & Wilkins, Baltimore, MD.

Liver Cysts

STANLEY MARTIN COHEN AND PHILLIP Y. CHUNG
University of Chicago

autosomal dominant polycystic kidney disease A disease characterized by the presence of multiple kidney, and often liver, cysts. The condition is inherited in an autosomal dominant pattern.
Caroli's disease Congenital cystic dilatation of the intrahepatic bile ducts.
cavernous hemangioma A benign vascular malformation presenting as a vascular mass in the liver. These may show cystic characteristics.
choledochal cyst A congenital, cystic structure arising from the biliary tree, often in association with abdominal pain and jaundice.
ciliated hepatic foregut cyst A solitary cystic lesion lined with ciliated, pseudo-stratified columnar epithelium.
cystadenocarcinoma A malignant cystic lesion derived from glandular epithelium.
cystadenoma A benign cystic lesion derived from glandular epithelium.
Echinococcus A genus of tapeworms, capable of causing hydatid cysts in humans.
hydatid cyst A cyst formed by the larval form of *Echinococcus*. Often appears as a "mother" cyst with several smaller "daughter" cysts.
microhamartoma A rare, benign lesion of hepatocytes and biliary epithelium encased by a fibrous stroma.
peliosis hepatis A condition associated with blood-filled cavities in the liver. This disease is found mostly in patients with acquired immune deficiency syndrome and anabolic steroid users.
polycystic liver disease A disease characterized by the presence of multiple liver cysts. This condition often occurs in association with autosomal dominant polycystic kidney disease. The condition is inherited in an autosomal dominant pattern.

Hepatic cysts are a common group of disorders that differ in etiology, physical manifestations, and clinical significance. Most cysts occur without symptoms and may be found incidentally by radiological imaging. The majority of cystic lesions are benign and only a small minority will be associated with serious morbidity or mortality. A universally accepted, standardized classification of hepatic cysts is lacking and there is also no consensus regarding the treatment of symptomatic lesions. This article reviews the various etiologies and diagnostic and treatment modalities of hepatic cysts.

BENIGN CYSTIC HEPATIC LESIONS

Simple Cysts

Simple cysts are cystic lesions that do not communicate with the intrahepatic biliary tree. Although generally occurring as solitary lesions, patients may present with multiple cysts. Simple cysts are common lesions with an autopsy prevalence of approximately 1%. The majority of the cysts are less than 1 cm in diameter, but rarely can be large enough to occupy the entire upper abdomen. Simple cysts are principally located in the right lobe of the liver. They are more common in women, with a male to

Copyright 2004, Elsevier (USA). All rights reserved.

Pitman, M., and Szyfelbein, W. (1994). "Final Needle Aspiration Biopsy of the Liver." Butterworth-Heinemann, Boston.
Reddy, K., and Jeffers, L. (1999). Liver biopsy and laparoscopy. *In* "Diseases of the Liver" (E. R. Schiff, M. F. Sorrel, and W. C. Maddrey, eds.), 8th Ed., pp. 245–266. Lippincott-Raven, Philadelphia, PA.
Scheuer, P., and Lefkowitch, J. (2000). "Liver Biopsy Interpretation," 6th Ed. W. B. Saunders, Philadelphia, PA.
Snover, D. (1992). Technical aspects of the evaluation of liver biopsies. *In* "Biopsy Diagnosis of Liver Diseases" (D. C. Snover, ed.), pp. 2–23. Williams & Wilkins, Baltimore, MD.

Liver Cysts

STANLEY MARTIN COHEN AND PHILLIP Y. CHUNG
University of Chicago

autosomal dominant polycystic kidney disease A disease characterized by the presence of multiple kidney, and often liver, cysts. The condition is inherited in an autosomal dominant pattern.
Caroli's disease Congenital cystic dilatation of the intrahepatic bile ducts.
cavernous hemangioma A benign vascular malformation presenting as a vascular mass in the liver. These may show cystic characteristics.
choledochal cyst A congenital, cystic structure arising from the biliary tree, often in association with abdominal pain and jaundice.
ciliated hepatic foregut cyst A solitary cystic lesion lined with ciliated, pseudo-stratified columnar epithelium.
cystadenocarcinoma A malignant cystic lesion derived from glandular epithelium.
cystadenoma A benign cystic lesion derived from glandular epithelium.
Echinococcus A genus of tapeworms, capable of causing hydatid cysts in humans.
hydatid cyst A cyst formed by the larval form of *Echinococcus*. Often appears as a "mother" cyst with several smaller "daughter" cysts.
microhamartoma A rare, benign lesion of hepatocytes and biliary epithelium encased by a fibrous stroma.
peliosis hepatis A condition associated with blood-filled cavities in the liver. This disease is found mostly in patients with acquired immune deficiency syndrome and anabolic steroid users.
polycystic liver disease A disease characterized by the presence of multiple liver cysts. This condition often occurs in association with autosomal dominant polycystic kidney disease. The condition is inherited in an autosomal dominant pattern.

Hepatic cysts are a common group of disorders that differ in etiology, physical manifestations, and clinical significance. Most cysts occur without symptoms and may be found incidentally by radiological imaging. The majority of cystic lesions are benign and only a small minority will be associated with serious morbidity or mortality. A universally accepted, standardized classification of hepatic cysts is lacking and there is also no consensus regarding the treatment of symptomatic lesions. This article reviews the various etiologies and diagnostic and treatment modalities of hepatic cysts.

BENIGN CYSTIC HEPATIC LESIONS

Simple Cysts

Simple cysts are cystic lesions that do not communicate with the intrahepatic biliary tree. Although generally occurring as solitary lesions, patients may present with multiple cysts. Simple cysts are common lesions with an autopsy prevalence of approximately 1%. The majority of the cysts are less than 1 cm in diameter, but rarely can be large enough to occupy the entire upper abdomen. Simple cysts are principally located in the right lobe of the liver. They are more common in women, with a male to

Copyright 2004, Elsevier (USA). All rights reserved.

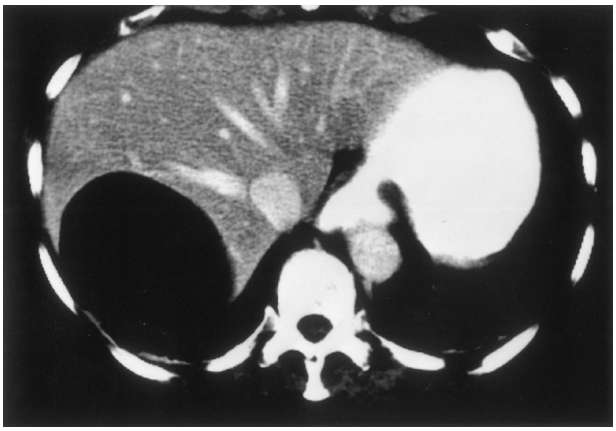


FIGURE 1 CT scan showing a large simple liver cyst. Courtesy of Abraham H. Dachman, M.D.

female ratio of 1 : 5 for asymptomatic cysts and 1 : 9 for symptomatic or complex cysts. Symptomatic cysts tend to present in the fourth and fifth decades of life.

Small cysts are usually asymptomatic and found incidentally on radiologic studies. Larger cysts tend to cause symptoms such as abdominal fullness or pain and nausea, due to compression of adjacent intra-abdominal structures. Physical examination of these patients generally reveals a palpable right upper quadrant mass. Laboratory examination is almost always normal unless the cyst impinges on the biliary tree and gives evidence of obstructive jaundice. Larger lesions may cause atrophy of the adjacent hepatic tissue with compensatory hypertrophy of the other lobe. Hemorrhage into the cyst, bacterial superinfection, torsion, or biliary obstruction can be seen with larger cysts. There is no convincing evidence of neoplastic transformation from simple cysts.

The most useful radiologic modality may be ultrasound, with simple cysts appearing as unilocular, thin-walled (less than 1 cm), anechoic, fluid-filled lesions with posterior acoustic enhancement. Computerized tomography (CT) scan shows well-demarcated, non-enhancing lesions of fluid density (see Fig. 1). If intracystic hemorrhage has occurred, septations may be visible, making diagnosis more complicated. Magnetic resonance imaging (MRI) of simple cysts reveals well-defined, water-attenuation lesions that do not enhance with gadolinium. T2-weighted images reveal a high level of enhancement, whereas the level of enhancement in T1 images is low.

Histologic assessment is rarely necessary for diagnosis. Needle aspiration reveals sterile fluid ranging from clear to brown in color. Cytologic examination of the fluid does not reveal any cellular atypia or dysplasia. On pathologic sections, the cysts demonstrate a single layer of cuboidal or columnar epithelium without mesenchymal stroma or cellular atypia. They are usually surrounded by three layers of dense fibrous tissue.

Asymptomatic cysts do not require any treatment, but periodic ultrasonic monitoring of larger lesions may be prudent to ensure stability of size. If the cyst remains stable for 2 to 3 years, further monitoring is probably unnecessary. If the lesions increase in size, alternate diagnoses such as cystadenoma or cystadenocarcinoma should be considered and surgical intervention may be indicated.

Patients with symptomatic cysts should be considered for treatment. Percutaneous needle drainage with or without injection of sclerosants may provide immediate relief, but has been associated with a high failure rate and rapid recurrence. Internal drainage via cystgastrostomy or cystjejunostomy can be performed via upper endoscopy with or without the aid of endoscopic ultrasound. Surgical unroofing of the cyst has been associated with low recurrence rates and few complications. A laparoscopic approach has been shown to be as effective as the open approach and obviates the need for a large abdominal incision. Recurrence and morbidity rates are less than 15% and this has become the intervention of choice for accessible lesions. Hepatic resection, including lobectomy, is generally reserved for large symptomatic lesions refractory to other treatments.

Cystadenomas

Cystadenomas are rare cystic lesions that occur within the liver parenchyma, extrahepatic bile ducts, or gallbladder. Cystadenomas are more prevalent in women. More than 60% present in patients over 40 years of age. These lesions can reach up to 15 cm in diameter.

Most cystadenomas are asymptomatic and are incidental findings on abdominal imaging studies. When present, the most common clinical manifestations are abdominal fullness, a palpable upper abdominal mass, abdominal pain or discomfort, and anorexia. Biliary obstruction is an uncommon complication. The most significant complication is malignant transformation, occurring in up to 15% of patients.

Ultrasound imaging reveals a well-defined cystic lesion with or without septations, often with slight irregularities of the cyst wall. CT scanning reveals a low-attenuation, uni- or multilocular cystic mass, often with septations, and a thickened and/or irregular cyst wall. MRI reveals a fluid-containing, multiloculated mass with homogenous low signal

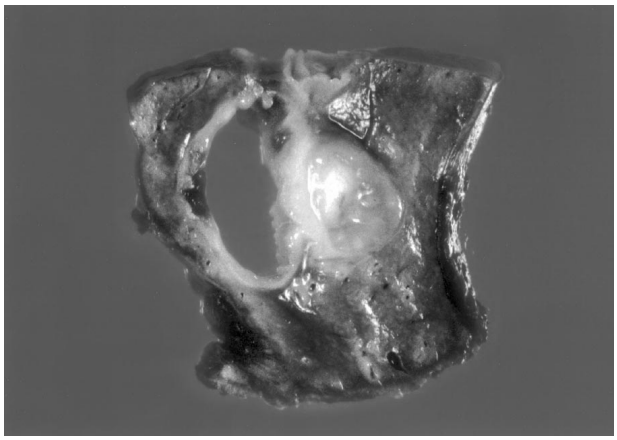


FIGURE 2 Gross liver specimen showing a cross-sectional view of a hepatic cystadenoma. The white rim of tissue lining the cyst cavity is composed of mesenchymal stroma. Courtesy of John Hart, M.D.

intensity on T1-weighted images and high signal intensity on T2-weighted images.

Radiologic studies cannot always differentiate cystadenomas from cystadenocarcinomas; thus, the definitive diagnosis of a cystadenoma requires histology. Gross pathologic examination typically reveals a multilocular cystic lesion with a smooth external surface. The wall is usually thick with a smooth internal lining (see Fig. 2) and the lesions often contain blood or dark brown liquid. Histologically, the cysts are lined by biliary-appearing mucous-producing cuboidal or columnar epithelium. The lining of the cyst is often surrounded by a dense mesenchymal fibrous stroma and a loose collagenous layer. Cystadenomas surrounded by a more cellular (and less fibrous) stroma may be a subtype with less associated risk of malignancy.

Due to the high rate of malignant degeneration and the high rates of recurrence with incomplete resection or needle aspiration, therapy almost exclusively involves complete surgical resection. The excision of the cystadenoma is accomplished by enucleating the lesion from the surrounding liver parenchyma.

Polycystic Liver Disease

Liver cysts can occur in association with autosomal dominant polycystic kidney disease (ADPKD). The incidence increases with age and worsening renal dysfunction. The incidence of polycystic liver disease is similar between the sexes, but the presence of massive cysts is almost exclusively observed in multiparous females. This may be due to sensitivity of the cysts to female hormones, a fact supported by the observation of cyst enlargement in certain patients receiving oral contraceptives or postmenopausal hormone replacement therapy.

The pathogenesis of ADPKD appears to involve a generalized defect in epithelial cell differentiation and/or extracellular matrix function. The liver lesions are derived from biliary epithelium. In addition to renal, hepatic, and pancreatic cysts, afflicted patients may have cerebral aneurysms, cardiac valvular disease, colonic diverticuli, and abdominal or inguinal hernias. Most patients are asymptomatic with preserved hepatic function, but some patients may present with abdominal pain from distension of the liver capsule or compression of adjacent structures. Acute pain suggests the possibility of cyst infection, hemorrhage, rupture, or torsion.

Radiologic (see Fig. 3) and histologic characteristics of the lesions are similar to those of simple cysts, with much larger numbers of cysts visualized.

Treatment should be reserved for symptomatic patients. Varying degrees of success have been reported with partial hepatectomy, but surgical intervention should be performed only at centers with experience in hepatic resection. Orthotopic or living-related liver transplantation has been performed as definitive therapy in patients with disease not amenable to local resection. The limited availability of donor organs and the preserved liver function in patients with ADPKD have made this intervention somewhat controversial.

Cavernous Hemangiomas

Cavernous hemangiomas are benign blood vessel tumors of the liver that may appear cystic in nature.

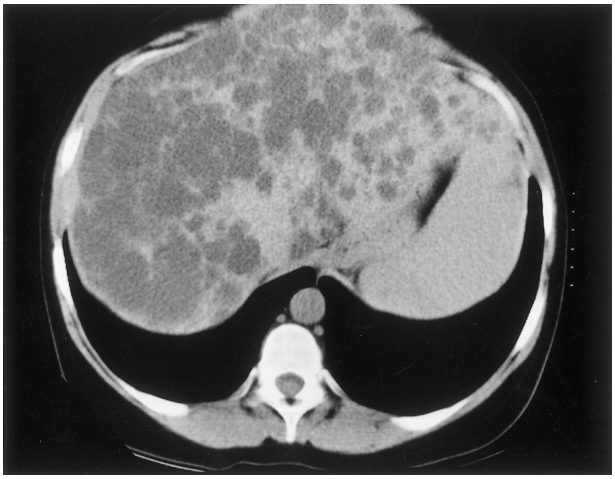


FIGURE 3 CT scan showing polycystic liver disease. Note the multiple simple cysts of varying sizes throughout both hepatic lobes.

They represent the most common benign hepatic tumors, with an autopsy incidence of 5–7%. The lesions appear to be equally divided among the sexes, but larger, symptomatic lesions tend to occur in women. The cysts are usually solitary and tend to involve the right lobe. However, 20% may involve both lobes. Size ranges from several millimeters to 45 cm in diameter.

Symptoms are more common in elderly patients with larger (greater than 4 cm) lesions. Symptoms include abdominal fullness, pain, nausea, and vomiting. Examination may reveal abdominal distension or a mass. Rare complications can include rupture, hemorrhage, or Kasabach-Merritt syndrome (consumptive coagulopathy within the hemangioma presenting as disseminated intravascular coagulopathy). There is no evidence of malignant transformation of hemangiomas.

Ultrasound demonstrates a well-defined hyperechoic mass. CT shows a mass with contrast accumulation spreading from the periphery to the center of the lesion. MRI is extremely useful, demonstrating a hyperintense signal on T2-weighted images (light bulb sign). Tagged red cell scanning shows initial hypoperfusion followed by gradual increase of the tracer uptake.

Pathologic examination should come from surgical specimens as biopsy is contraindicated due to the extremely vascular nature of these tumors. Grossly, the lesions are blue to purple in color and very vascular. Histology reveals large, dilated channels lined by vascular endothelium. A variable amount of surrounding fibrous tissue is present.

Treatment is necessary only for symptomatic patients. Wedge resection and lobectomy are the treatments of choice. Radiation therapy or arterial embolization may be useful as temporizing measures in patients with acute bleeding or Kasabach-Merritt syndrome.

Miscellaneous Benign Cysts

Ciliated Hepatic Foregut Cysts

This rare, benign condition presents as a solitary hepatic cyst lined with ciliated, pseudo-stratified columnar epithelium surrounded by subepithelial connective tissue, a smooth muscle layer, and an outer fibrous capsule. Only 60 cases have been reported, with cyst sizes ranging from 0.4 to 9.0 cm. These cysts are found more commonly in men and occur more frequently in the left lobe. Malignant transformation has not been reported and the clinical significance is restricted to differentiating ciliated hepatic foregut cysts from other lesions with malignant potential.

Microhamartomas (von Meyenberg Complexes)

Microhamartomas are rare lesions presenting in young children. They may occasionally be cystic in nature. They tend to occur in males. The presentation is that of a painless right upper quadrant mass. Imaging techniques demonstrate a multiloculated cystic lesion, predominantly in the right hepatic lobe. Histology reveals hepatocytes and biliary channels trapped in a variable cellular and fibrous stroma. Treatment is generally not necessary due to the benign nature of these lesions. For symptomatic patients, surgical excision is curative.

Peliosis Hepatis

Peliosis hepatis is a rare condition characterized by blood-filled cavities in the hepatic parenchyma. This condition is most commonly associated with anabolic steroid use. Peliosis also occurs in human immunodeficiency virus (HIV)-positive patients infected with *Bartonella henselae*. Peliosis is generally asymptomatic, with no evidence of malignant degeneration. Treatment, although inconsistently effective, targets the likely cause. In cases associated with anabolic steroid use, discontinuation of these agents has been shown to cause disease regression. In HIV-positive patients, peliosis may respond to prolonged courses of erythromycin or doxycycline.

Miscellaneous

Glands adjacent to the biliary tree can expand and present as periductal cysts. In patients with prolonged cholestasis, biliary cysts can form. Blood-filled cysts can develop as a result of trauma, most commonly after a liver biopsy. Duodenal duplication cysts, extremely rare embryologic abnormalities involving duplication of the duodenal lumen, can present as cystic lesions in the hepatic parenchyma.

MALIGNANT CYSTIC HEPATIC LESIONS

Cystadenocarcinoma

Cystadenocarcinomas likely arise from the malignant transformation of cystadenomas (see above). These lesions occur in the hepatic parenchyma, but have not been found arising from the extrahepatic biliary tree. Although they have been reported in patients as young as 30 years of age, this tumor is generally found in the elderly.

The presenting symptoms are similar to those seen with a cystadenoma (i.e., abdominal pain, mass). However, acute worsening of the symptoms associated with a rapid increase in size of a previously stable lesion is suggestive of malignant transformation.

Radiologically, cystadenomas and cystadenocarcinomas appear similar on ultrasound and CT. Both reveal multilocular, septated cystic lesions.

The differentiation between cystadenoma and cystadenocarcinoma requires histologic examination. Preoperative diagnosis is difficult, as needle aspiration carries the risk of peritoneal seeding of tumor. On pathologic examination, cystadenocarcinomas have a thicker cyst wall than cystadenomas and may have tissue extruding from the internal cyst lining. Histologically, cellular atypia, increased cellular proliferation, and malignant changes are typically found in the inner epithelial lining.

If cystadenocarcinoma is suspected, complete excision by segmental liver resection is the treatment of choice. Enucleation of the cyst, as performed for cystadenoma, is not acceptable, as there is an increased risk of recurrence. Nonsurgical treatments such as radiation or chemotherapy are unproven and thus not recommended.

Cystadenocarcinomas may invade locally and metastasize distantly, but may be curable with complete excision. The prognosis is generally better than that associated with cholangiocarcinoma.

Liver Metastases

The liver is a common site of metastasis for many abdominal and pelvic neoplasms. Rarely, metastatic lesions may outgrow their blood supply, resulting in central necrosis, which may appear as cystic lesions on radiologic studies. The primary lesions include colorectal, ovarian, pancreatic, kidney, and neuroendocrine tumors. Segmental resection of accessible metastases is the only curative intervention, but there must be no involvement of the hepatic artery, portal vein, or major bile ducts, and there must be sufficient postresection functional hepatic reserve. Although laparoscopy may identify occult metastases obviating the need for laparotomy, a clinical risk score involving five factors (lymph node positivity, relapse-free interval <12 months, >1 hepatic metastases on preoperative imaging, serum CEA (carcinoembryonic antigen, a serologic colon cancer marker) >200 ng/ml within 1 month of surgery, and any metastasis >5 cm) has been suggested to predict nonresectability in metastatic colon cancer. Only 12% of patients with a score ≤ 2 had occult metastases on laparoscopy versus 42% of those with a score >2. Five-year survival after resection of hepatic metastasis is 24–38% in most studies. However, fewer than 10% of patients with hepatic metastasis are candidates for curative resection at the time of presentation. Other alternatives for liver metastases, regardless of the primary tumor, include local tumor ablation with injection of ethanol or acetic acid, intraoperative cryotherapy, percutaneous hyperthermic coagulation, radiofrequency or microwave ablation, and hepatic intra-arterial and systemic chemotherapy.

Primary Squamous Cell Carcinoma

Several reports have been published describing the development of primary squamous cell carcinoma within solitary congenital hepatic cysts lined with stratified squamous epithelium. The prognosis appears poor despite surgical resection due to local recurrence. The number of published cases remains small and no formal trials of therapy have been conducted.

CHOLEDOCHAL CYSTS AND CAROLI'S DISEASE

Choledochal cysts are congenital cystic dilatations of the intra- and extrahepatic biliary tree. A system of classification based on morphology has been well accepted (see Fig. 4). Class I choledochal cysts, defined as segmental or diffuse fusiform dilatations of the extrahepatic biliary tree, account for 80 to 90% of cases. Class II cysts are true choledochal diverticuli, class III consists

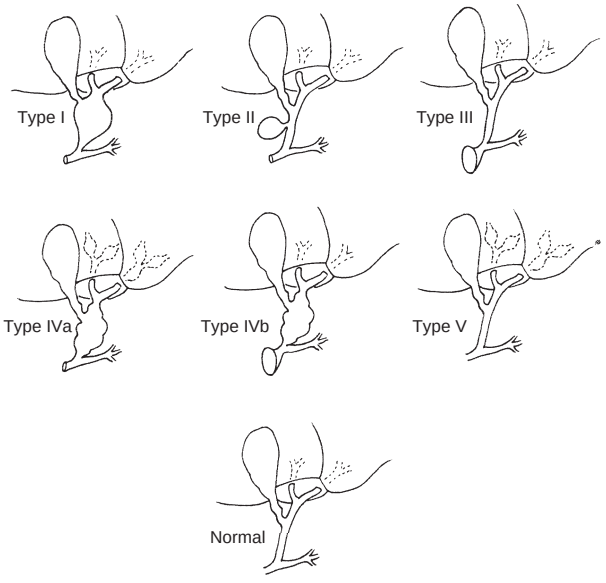


FIGURE 4 Subtypes of choledochal cysts including Caroli's disease. Courtesy of Kenneth D. Chi, M.D.

of dilatations of the intraduodenal common bile duct, and class IV is made up of either multiple intra- and extrahepatic cysts (subtype IVa) or just multiple extrahepatic cysts (subtype IVb). Caroli's disease (Class V) consists of cystic dilatations of the intrahepatic ductal system.

The incidence of choledochal cysts varies geographically, with approximately 1 : 100,000–150,000 in Western countries and up to 1 : 10,000 in Japan. Although choledochal cysts have been documented at all ages, 60–80% of cases present before the age of 10 years. Females are more commonly affected with choledochal cysts, whereas Caroli's disease occurs more commonly in males. Neither of these conditions is familial.

Patients often present in infancy with the classic triad of pain, jaundice, and a palpable abdominal mass. Other symptoms include cholestatic jaundice, acholic stools, hepatomegaly, vomiting, irritability, and failure to thrive. Portal hypertension and ascites may also be evident on presentation, especially in Caroli's disease patients with associated congenital hepatic fibrosis. Adult patients tend to present with epigastric pain or intermittent fever and jaundice. Only approximately 20% of adult patients will present with the classic symptom triad. Complications of choledochal cysts include biliary obstruction, stone formation, cholangitis, hepatic abscess, biliary perforation, pancreatitis, and malignant degeneration. The risk of cholangiocarcinoma correlates with duration of disease. The risk is approximately 1% in the first decade of life, 6–7% in the second decade, and 14% in successive decades. The prognosis of these malignancies is extremely poor.

The diagnosis of choledochal cysts is made by ultrasonography, which reveals the cystic dilations of the bile ducts. Percutaneous cholangiography transhepatic (PTC) or endoscopic retrograde cholangiopancreatography (ERCP) can define the site of biliary origin and the extent of both extra- and intrahepatic disease. PTC and ERCP can also reveal an anomalous arrangement of the pancreaticobiliary junction, which occurs in up to 40% of cases.

On pathologic examination, the cysts are composed of a fibrous wall, with or without a lining of columnar epithelium. Calculi frequently develop within the dilated bile ducts. Liver biopsy may reveal portal tract edema and fibrosis in early disease or biliary cirrhosis with or without carcinoma of the cyst wall in older patients.

Due to a 20-fold increased risk of malignancy in the cystic lesions, the preferred method of treatment is surgical excision of the cyst and reconstruction of the extrahepatic biliary tree, usually with a Roux-en-Y hepaticojejunostomy or choledochojejunostomy for biliary drainage. Cholecystectomy should also be performed due to the subsequent risk of gallbladder cancer. Recurrent cholangitis, stone formation, pancreatitis, and malignancy anywhere along the remaining pancreaticobiliary tree may develop many years after reconstruction, necessitating long-term follow-up. Hepatic resection is indicated for disease limited to a single lobe. Simple decompression and internal drainage are reserved for inoperable cases or as temporizing therapy prior to surgery. Ursodeoxycholic acid therapy may be effective in dissolving intrahepatic stones. Liver transplantation is a therapeutic option in patients with frequent complications, in patients with extensive disease that precludes other surgical alternatives, and in patients with Caroli's disease.

INFECTIOUS/PARASTIC HEPATIC CYSTIC LESIONS

Echinococcal/Hyatid Cysts

Echinococcal disease is caused by the metacestode stage of the tapeworm, *Echinococcus*, family Taeniidae. There are four species that cause disease in human. *E. granulosus* and *E. multilocularis* are the most common, causing cystic and alveolar disease, respectively. *E. vogeli* and *E. oligarthrus* are rarely associated with disease in humans.

Humans are an incidental host and do not play a role in the biologic cycle of the organism (see Fig. 5). Therefore, human to human infection does not occur. Inter- and intraspecies differences exist in regards to geographical distributions and hosts. The adult tapeworm is usually found in the intestinal tract of the canine host acquired by ingestion of sheep viscera infected with hydatid cysts. Scoleces adhere to the dog's intestine and become adult taenia, which shed eggs into the intestinal lumen. The eggs are expelled in the feces and are ingested by sheep, cattle, or rodents or directly contaminate the environment. Humans are usually infected by ingestion of contaminated vegetables or by direct contact with an infected pet through fecal–oral contact. The pathophysiology in humans is similar to that of the typical intermediate host, with formation of hyatid cysts, typically in the liver.

Approximately 50% of hydatid cysts occur in asymptomatic patients and many more are found incidentally at autopsy. The average age at diagnosis is 36 years. The right lobe is affected in 60–85% of cases. The cysts typically grow at a rate of 1 to 5 cm per year. Symptoms

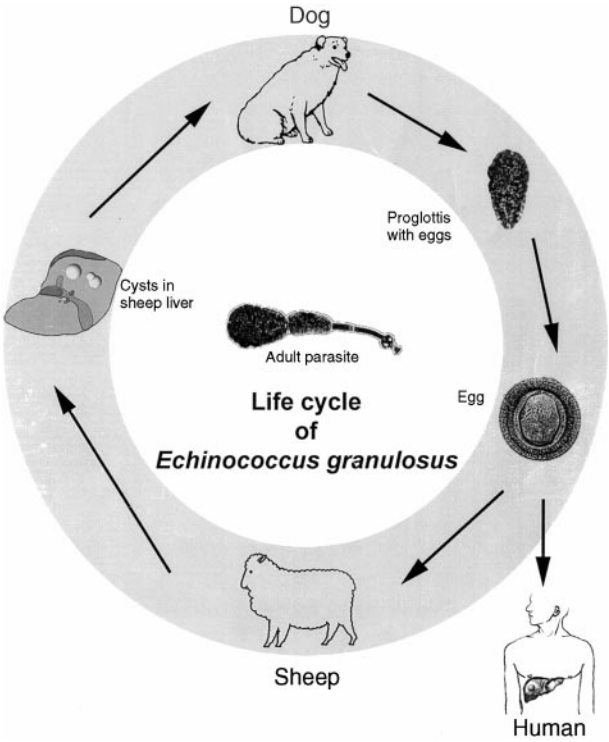


FIGURE 5 The life cycle of *Echinococcus*. Courtesy of Kenneth D. Chi, M.D.

of abdominal fullness, right upper quadrant pain, nausea, and vomiting typically occur when cysts reach 10 cm or more in diameter. Clinical features are due to mass effect on adjacent organs, obstruction of blood or lymphatic flow, or complications such as rupture or bacterial superinfection.

Diagnosis is made by a combination of radiologic and serologic studies. The hyatid cysts may appear as calcified cystic lesions on plain radiograph, whereas the ultrasound appearance is that of an anechoic, smooth, round cyst, similar in appearance to simple cysts. A flattened or elliptical shape, detachment of the germinal layer from the cyst wall (the "water lily" sign), or calcifications within the cyst wall suggest an inactive lesion. When daughter cysts are present, characteristic internal septations are also visible on ultrasound or CT imaging. "Hyatid sand," consisting of scoleces and hooklets of protoscoleces, may also be visualized, particularly if the patient shifts position during imaging. CT has been reported to be more sensitive than ultrasound in diagnosis (95–100% versus 90–95%, respectively).

Many serologic tests have been used to aid in the diagnosis of echinococcal infection. Immunoglobulin G enzyme-linked immunosorbent assay and hemagglutination assays have sensitivities of 80–90%. Immunoblot or antigen gel diffusion assays are more specific and may be used for confirmation.

Diagnostic aspiration or biopsy carries the risk of cyst rupture and anaphylaxis, or secondary spread of infection, and should be reserved for situations when other diagnostic methods are inconclusive. Aspiration should be performed only under CT or ultrasound guidance. Grossly, the cysts are thick-walled lesions often containing smaller, thin-walled daughter cysts (see Fig. 6). Microscopically, a cyst with possible calcifications in the wall is seen. The cyst contains cellular debris and occasionally viable scoleces.

Surgery is the treatment of choice in patients who can tolerate the operation and who have accessible cysts (approximately 90% of patients). Techniques include removal of the intact cyst, aspiration and cyst excision or external drainage, and partial hepatectomy. The instillation of protoscolicidal agents has not been definitively shown to reduce recurrence risk and may cause sclerosing cholangitis or pancreatitis if communication exists between the cyst and the biliary tree. Mebendazole or albendazole should be administered at least 4 days before surgery and for 1 (albendazole) to 3 (mebendazole) months afterward.

In poor surgical candidates, patients with multiple cysts, or patients with cyst recurrence, percutaneous aspiration/instillation of a protoscolicidal agent/reaspiration (PAIR) may be used as a therapeutic option. The procedure involves ultrasound-guided aspiration of the cyst fluid, followed by instillation of a scolicidal agent. The remaining cystic contents are then removed. PAIR has been shown to be effective in 70–99% of cases. Medical therapy alone can be considered in patients who cannot tolerate surgical or PAIR interventions. Although mebendazole or albendazole may be associated

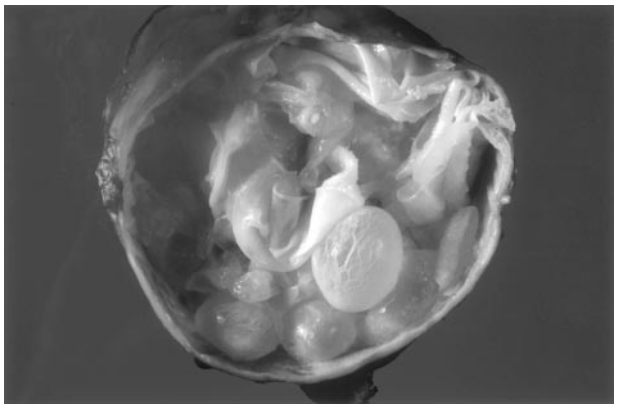


FIGURE 6 Gross specimen of a hydatid cyst filled with multiple daughter cysts. Courtesy of John Hart, M.D.

with shrinkage of the lesions in 50% of cases, cure occurs in less than 30% of patients.

Hepatic Abscesses

Liver abscesses may cavitate and form cystic lesions due to infection with gas-forming organisms or ameba.

SUMMARY

Cystic lesions of the liver are increasingly being diagnosed. Clinical and radiologic characteristics can help differentiate benign from malignant conditions. Needle aspiration and histology are generally not required for diagnosis of simple cysts. Complex, symptomatic, or enlarging lesions require further evaluation. The treatment for such lesions is generally surgical; however, percutaneous and endoscopic approaches are available.

See Also the Following Articles

Biliary Tract, Developmental Anomalies of the • Computed Tomography (CT) • Magnetic Resonance Imaging (MRI) • Percutaneous Transhepatic Choliangiography (PTC) • Ultrasonography

Further Reading

Ammori, B., Jenkins, B., Lim, P., Prasad, K., Pollard, S., and Lodge, J. (2002). Surgical strategy for cystic diseases of the liver in a Western hepatobiliary center. *World J. Surg.* **26**, 462–469.

Balik, A., Basoglu, M., Celebi, F., Oren, D., Polat, K., Atamanalp, S., and Akcay, M. (1999). Surgical treatment of hydatid disease of the liver: Review of 304 cases. *Arch. Surg.* **134**, 166–169.

Beckingham, I., and Krige, J. (2001). Liver tumors. *Br. Med. J.* **322**, 477–480.

Fulcher, A., and Sterling, R. (2002). Hepatic neoplasms: Computed tomography and magnetic resonance features. *J. Clin. Gastroenterol.* **34**, 463–471.

Gaia, S., Alessandria, C., Marzano, A., and Rizzetto, M. (2001). Polycystic liver disease. *Liver Transplant.* **10**, 912.

Gigot, J., Metairie, S., Etienne, J., Horsmans, Y., van Beers, B., Sempoux, C., Deprez, P., Materne, R., Geubel, A., Glineur, D., and Gianello, P. (2001). The surgical management of congenital liver cysts. *Surg. Endosc.* **15**, 357–363.

Gitlin, N., and Ros, P. R. (eds.) (2002). "Clinics in Liver Disease: Hepatic Imaging and Intervention," Vol. 6, No. 1. W. B. Saunders, Philadelphia, PA.

Hansman, M., Ryan, J., Holmes, J., Hogan, S., Lee, F., Kramer, D., and Biehl, T. (2001). Management and long-term follow-up of hepatic cysts. *Am. J. Surg.* **181**, 404–410.

Kew, M. C. (1998). Hepatic tumors and cysts. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., Vol. 2, pp. 1364–1387. W. B. Saunders, Philadelphia, PA.

Martin, I., McKinley, A., Currie, E., Holmes, P., and Garden, O. (1998). Tailoring the management of nonparasitic liver cysts. *Ann. Surg.* **228**, 167–172.

Regev, A., Reddy, K., Berho, M., Sleeman, D., Levi, J., Livingstone, A., Levi, D., Ali, U., Molina, E., and Schiff, E. (2001). Large cystic lesions of the liver in adults: A 15-year experience in a tertiary center. *J. Am. College Surg.* **193**, 36–45.

Regev, A., and Reddy, K. R. (2002). Diagnosis and management of large cystic lesions of the liver. Available at http://www.uptodate.com.

Sherlock, S. (1999). Cystic diseases of the liver. *In* "Schiff's Diseases of the Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 2, pp. 1083–1090. Lippincott Williams and Wilkins, Philadelphia, PA.

Weyant, M. J., Maluccio, M. A., Bertagnolli, M. M., and Daly, J. M. (1998). Choledochal cysts in adults. *Am. J. Gastroenterol.* **93**, 2580–2583.

tree. However, studies in model systems are now making the long list of developmental defects seem more rational (Fig. 2).

See Also the Following Articles

Biliary Tract, Development • Biliary Tract, Developmental Anomalies of the • Development, Overview • Gallbladder, Pediatric • Neonatal Cholestasis and Biliary Atresia

Further Reading

Crawford, J. M. (2002) Development of the intrahepatic biliary tree. *Semin. Liver Dis.* **22**, 213–226.

Matsumoto, K., Yoshitomi, H., Rossant, J., and Zaret, K. S. (2001). Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563.

Onuchic, L. F., Furu, L., Nagasawa, Y., Hou, X., Eggermann, T., Ren, Z., Bergmann, C., Sanderek, J., Esquivel, E., Zeltner, R., Rudnik-Schoneborn, S., Mrug, M., Sweeney, W., Avner, E. D., Zerres, K., Guay-Woodford, L. M., Somlo, S., and Germino, G. G. (2002). PKHD1, the polycystic kidney and hepatic disease gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix repeats. *Am. J. Hum. Genet.* **70**, 1305–1317.

Sherlock, S., and Dooley, J. (eds.) (2002). "Diseases of the Liver and the Biliary System," 11th Ed., pp. 579–591. Blackwell Science, Oxford.

Zakim, D., and Boyer, T. (eds.) (2003). "Hepatology: A Textbook of Liver Disease," 4th Ed., pp. 1649–1698. W. B. Saunders, Philadelphia.

Zaret, K. S. (2001). Hepatocyte differentiation: From the endoderm and beyond. *Curr. Opin. Genet. Dev.* **11**, 568–574.

Zaret, K. S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nat. Rev. Genet.* **3**, 499–512.

Liver Disease, Pregnancy and

J. EILEEN HAY
Mayo Clinic and Foundation

acute fatty liver of pregnancy Acute liver failure due to microvesicular fatty infiltration of hepatocytes in the third trimester of pregnancy.

HELLP syndrome Triad of hemolysis, elevated liver tests, and low platelet count in the third trimester of pregnancy.

hyperemesis gravidarum Intractable vomiting in the first trimester of pregnancy.

intrahepatic cholestasis of pregnancy Pruritus and liver dysfunction in the second half of pregnancy.

preeclampsia triad of hypertension, edema, and proteinuria in the third trimester of pregnancy.

Pregnancy is an altered physiologic state associated with many hemodynamic and metabolic adaptations to support the placenta and fetus. Although most pregnant women remain healthy, pathophysiologic changes can occur, including five unique liver diseases seen only in the pregnant or postpartum patient. These pregnancy-associated liver diseases have characteristic clinical features and timing of onset in relation to pregnancy. Hyperemesis gravidarum is intractable nausea and vomiting in the first trimester; intrahepatic cholestasis of pregnancy, a disease of pruritus and abnormal liver tests, occurs in the second half of pregnancy; and preeclampsia and its complications of HELLP (hemolysis, elevated liver tests, and low platelets) syndrome and acute fatty liver of pregnancy are third-trimester diseases.

INTRODUCTION

Any liver disease, acute or chronic, occurring in young adult women may also occur coincidentally in pregnancy. Viral hepatitis is a common cause of jaundice in pregnancy and hepatitis A, hepatitis B, and hepatitis C occur with the same frequency and with the same clinical features as in nonpregnant individuals, with pregnancy and the viral hepatitis having little effect on each other. Gallstones are common, particularly in multiparous patients, but fortunately they are rarely symptomatic with biliary colic, pancreatitis,

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Incidence and Clinical Features of Liver Diseases Unique to Pregnancy

Disease	Incidence (per 1000 pregnancies)	Trimester	Risk factors
Hyperemesis gravidarum	2–10	1st	Multiparity, adolescent mother, twins, hydatiform mole
Intrahepatic cholestasis of pregnancy	1 (USA)	2nd and 3rd	Family history
Preeclampsia with liver dysfunction	1–6	3rd	
HELLP syndrome	1–6	3rd	Older mother, multiparity, preeclampsia
Acute fatty liver of pregnancy	0.05–1	3rd	Primigravida, twins, preeclampsia

or cholecystitis. Pregnancy in the patient with cirrhosis and/or portal hypertension is rare and guidelines for management are sparse; the main maternal risk is variceal bleeding.

This article will be limited to discussion of the five liver diseases unique to pregnancy. The incidence and etiology of the pregnancy-associated liver diseases will be discussed, followed by a description of the clinical features and management of each of these disorders (Tables I and II).

INCIDENCE

A recent prospective population study in the United Kingdom found that, of 4377 deliveries, 3% of patients had liver dysfunction, with elevations of transaminases (84%), bilirubin (17%), and bile acids (14%). Surprisingly, 80% of these patients had pregnancy-associated disorders and only 12% (0.4% of the total population) had coincidental or preexisting hepatobiliary problems; sepsis and postpartum problems contributed in approximately one-third of affected patients.

Nausea and vomiting are very common in pregnancy but the overall reported incidence of hyperemesis gravidarum is 2–10 per 1000 pregnancies with high transaminase levels in 25–50% of patients. Intrahepatic cholestasis of pregnancy (ICP) is a worldwide disease with striking geographical variations. In the United States, it occurs in 0.1% of pregnancies, with jaundice in 20% of cases.

The preeclampsia-associated liver diseases are preeclampsia itself, the HELLP (hemolysis, elevated liver tests, and low platelets) syndrome, and acute fatty liver of pregnancy (AFLP). Preeclampsia occurs in 5–10% of pregnancies and, although liver involvement is considered uncommon, it was the commonest cause of liver dysfunction in pregnancy (1.4% of all deliveries) in a recent prospective study. In 2–12% of cases of severe preeclampsia (0.1–0.6% of all pregnancies), this is further complicated by the HELLP syndrome. The most severe of the pregnancy-associated liver diseases, AFLP, is the least common in most studies, reportedly occurring in 0.005% of pregnancies but its incidence was considerably higher in the recent prospective study.

TABLE II Management of Liver Diseases Unique to Pregnancy

Disease	Hospitalization	Fetal management	Maternal management
Hyperemesis gravidarum	Yes	Routine	Intravenous hydration Antiemetics ? High-dose steroids
Intrahepatic cholestasis	No	Fetal monitoring, delivery <35 weeks	UDCA ? Cholestyramine ? Steroids
Preeclampsia with liver dysfunction	Yes	Immediate delivery	Antihypertensives Seizure prophylaxis
HELLP syndrome	Yes	Immediate delivery	Treatment for preeclampsia CT abdomen Blood/blood products ? Plasmapheresis/OLT
Acute fatty liver of pregnancy	Yes	Immediate delivery	Supportive care for liver failure ? OLT

ETIOLOGY

The etiologies of the liver diseases unique to pregnancy remain obscure but new knowledge is emerging, particularly about ICP and AFLP. The pathophysiology of hyperemesis gravidarum is complex, with evidence for hormonal (elevated estrogen levels, transient hyperthyroidism, and elevated chorionic gonadotropin) and immunologic abnormalities; the role of psychological factors remains poorly defined. The etiology of ICP is probably multifactorial. The geographic and seasonal variability, the potential role of dietary factors such as selenium, and the 45–70% recurrence rate in multiparous patients clearly suggest the importance of exogenous factors. Female sex hormones are of pathogenic importance, perhaps by a genetically abnormal or exaggerated hepatic metabolic response to the physiologic increase in estrogens during pregnancy. Patients with ICP have abnormalities in progesterone metabolism and exogenous progesterone is the only proven exogenous precipitant of ICP. The familial cases and high ethnic variations strongly suggest a genetic predisposition to ICP. Mothers with ICP who are heterozygous, with homozygous babies, for genetic abnormalities in biliary (canalicular) transport proteins have been identified; these abnormalities are responsible for the rare group of diseases known as the progressive familial intrahepatic cholestasis syndromes. Abnormalities in placental bile acid transport systems and/or high circulating bile acid levels with vasospasm of the placental vessels may contribute to fetal loss in ICP.

Preeclampsia is associated with generalized endothelial dysfunction. In the HELLP syndrome, microangiopathic hemolytic anemia causes periportal hemorrhage, necrosis, and fibrin deposition in the liver. With increasing severity, areas of infarction enlarge and hemorrhage dissects from zone 1 diffusely to involve the whole lobule, leading to widespread necrosis, large hematomas, capsular tears, and hepatic rupture.

In AFLP, acute hepatic failure is due to microvesicular fatty infiltration of hepatocytes, perhaps due to abnormalities in intramitochondrial fatty acid oxidation (FAO). FAO helps to meet the increased metabolic demands of pregnancy but is reduced in later pregnancy even as metabolic demands are increasing, especially with preeclampsia. Some babies of AFLP mothers are homozygous for deficiencies in enzymes necessary for normal FAO, the best characterized being LCHAD (long-chain 3-hydroxyacyl-coenzyme A dehydrogenase) deficiency. Maternal heterozygosity, with fetal homozygosity, for deficiency of an enzyme essential for the normal metabolism of fatty acids may overwhelm the increased demands of FAO in pregnancy and cause AFLP.

CLINICAL FEATURES AND MANAGEMENT

Hyperemesis Gravidarum

Clinical Features and Diagnosis

Hyperemesis gravidarum (HG) is intractable vomiting of such severity as to necessitate intravenous hydration. HG occurs in the first trimester of pregnancy, typically between 4 and 10 weeks of gestation, with risk factors being adolescent pregnancy, multiparity, twins, and hydatiform mole. Patients become severely dehydrated and may lose $>5\%$ of prepregnancy weight. High transaminase levels (up to a 20-fold increase) occur in 25–50% of cases and occasionally patients have mild jaundice. A clinical diagnosis is made based on the severity and timing of vomiting; mild to moderate vomiting in pregnancy does not cause liver dysfunction. Viral hepatitis must be considered but liver biopsy is necessary only to exclude other serious diagnoses.

Management

Hospitalization is necessary for hydration and, in severe cases, parenteral nutrition. Symptomatic therapy with antihistamines is safe and successful in many patients; oral ginger root extract or pyridoxine (vitamin B6) may occasionally give relief. Short-term, high-dose steroids may be effective in refractory cases.

Intrahepatic Cholestasis of Pregnancy

Clinical Features and Diagnosis

ICP is characterized by severe, generalized pruritus and liver dysfunction in the second half of pregnancy, typically at 25–32 weeks of gestation, in a patient with no other signs of liver disease; it disappears after delivery, to recur in subsequent pregnancies. Maternal distress can be great and skin excoriations are obvious. Jaundice will follow the pruritus in 10–25% of patients but rarely occurs without pruritus. Severe cholestasis may be complicated by steatorrhea and malabsorption of fat-soluble vitamins.

Mild to 20-fold elevations of transaminases occur but bilirubin remains <5 mg/dl. With placental production, alkaline phosphatase is of little diagnostic value in pregnancy and the most specific and sensitive diagnostic marker of ICP is a serum bile acid level that

is always elevated, up to 100 times above normal, and which may correlate with fetal risk. Generally, in a primigravida, a presumptive clinical diagnosis is made based on typical clinical features, consistent biochemical changes, and no evidence of obstructive gallstone disease, viral hepatitis, or other skin condition; the diagnosis is confirmed by rapid postpartum resolution. A liver biopsy is needed only to exclude more serious liver disease.

Management

Maternal management is symptomatic relief of pruritus. Three small trials have shown that ursodeoxycholic acid (UDCA) is safe and effective at doses of 600–1200 mg/day with symptomatic relief, biochemical improvement, and perhaps improved fetal outcome. Higher doses of UDCA (up to 2 g/day) have been shown to improve the bile acid profile in both mothers and babies—hopefully a benefit to improve fetal outcome. Cholestyramine may be used for symptomatic relief but does not affect the cholestasis. Dexamethasone (12 mg/day for 7 days), by suppressing fetoplacental estrogen production, may give symptomatic relief and it promotes fetal lung maturity to allow early delivery of the baby. No other agents are of proven efficacy for maternal relief in ICP. Therapy with vitamin K may be required at the time of delivery.

ICP results in a high-risk pregnancy for the fetus with premature deliveries, perinatal deaths, and fetal distress. Fetal monitoring for chronic placental insufficiency is necessary but will not prevent intrauterine deaths from acute anoxic injury. Careful obstetric intervention will improve fetal prognosis and early delivery is recommended by 35 weeks or earlier, as soon as there is fetal lung maturity. Whether maternal UDCA therapy will improve fetal outcome remains to be proven.

Symptoms of ICP will resolve by 4 weeks postpartum but will recur in 45–70% of subsequent pregnancies. Rare familial cases of apparent ICP have persisted after delivery with eventual progression to fibrosis and cirrhosis.

Preeclampsia

Preeclampsia is the triad of hypertension, edema, and proteinuria in the third trimester of pregnancy and, if severe, right upper abdominal pain and tenderness and mild jaundice (bilirubin < 5 mg/dl) may develop. Transaminases are elevated, from mild to 20-fold elevations. Hepatic involvement in preeclampsia signifies severe disease but requires no therapy other than that for the preeclampsia; the patient must be stabilized and then immediate delivery effected to avoid eclampsia, hepatic rupture, or necrosis.

HELLP Syndrome

Clinical Features and Diagnostic Criteria

The typical patient with HELLP syndrome is the older, multiparous patient who presents between 27 and 36 weeks of gestation. Most, but not all, have preeclampsia with hypertension, edema, and weight gain. The most common presenting symptoms are epigastric or right upper quadrant pain and tenderness, nausea and vomiting, a flu-like illness, and headache. Jaundice is uncommon (5%). Early diagnosis and management are essential to reduce maternal and fetal mortality.

Diagnosis requires the presence of all three of the following criteria: (1) hemolysis with an abnormal blood smear, an elevated lactate dehydrogenase (> 600 U/liter), and an increase in indirect bilirubin; (2) an aspartate aminotransferase level of > 70 U/liter (up to 20-fold elevation); and (3) a platelet count of $< 100{,}000 \times 10^6$/liter or $< 100 \times 10^9$/liter and, in severe cases, $< 50{,}000$. Coagulation is usually normal but occasionally disseminated intravascular coagulation (DIC) occurs. Bilirubin is usually < 5 mg/dl. Computed tomography of the liver is indicated to detect hepatic rupture, subcapsular hematomas, intraparenchymal hemorrhage, or infarction.

Management

Expeditious management is required for optimal maternal and fetal outcomes. Treatment of hypertension and seizure prophylaxis is implemented. Severe cases will require full obstetric, surgical, and intensive care facilities and transfer to a tertiary referral center is essential. Delivery is the only definitive therapy and should be effected immediately near term or with fetal lung maturity. Uncomplicated vaginal delivery can be allowed to proceed but many patients require a cesarean section. Blood or blood products are required for hypovolemia, anemia, or coagulopathy, and broad-spectrum antibiotics are given. Management remote from term is more controversial. In severe cases, delivery must proceed despite the fetal risk but occasionally in milder cases at < 34 weeks, high-dose glucocorticoids are given to improve fetal lung maturity and to prolong the pregnancy.

Large subcapsular hematomas and diffuse intrahepatic hemorrhage without rupture are managed with close hemodynamic monitoring in an intensive care unit, correction of coagulopathy, availability of

large volumes of blood/blood products, and immediate intervention for hemodynamic instability or rupture. Exogenous trauma such as abdominal palpation, convulsions, emesis, and unnecessary transportation should be avoided. Liver rupture is rare but life-threatening, and rapid, aggressive medical management and immediate surgery are essential for survival.

In most cases, HELLP syndrome will rapidly resolve within 5 days of delivery with normalization of platelet counts but some patients have persisting thrombocytopenia, hemolysis, worsening hyperbilirubinemia, and impaired renal function. Life-threatening complications persisting for >72 h after delivery are usually taken as an indication for plasmapheresis or plasma exchange but if there is no response, liver transplantation must be considered. Unfortunately, life-threatening maternal complications are common and include DIC, abruptio placentae, acute renal failure, pulmonary edema, and rarely acute respiratory distress syndrome, hepatic failure, or hepatic rupture. The risk of recurrence of HELLP in subsequent pregnancies is poorly established but with subsequent pregnancies, these patients have an increased risk of preeclampsia, preterm delivery, intrauterine growth retardation, and abruptio placentae.

Acute Fatty Liver of Pregnancy

Clinical Features and Diagnosis

Many patients with AFLP are primigravidae, especially with twin pregnancies. The commonest time of presentation is 36 weeks, but it can occur at any time in the third trimester and in the postpartum period. The typical patient has 1–2 weeks of anorexia, nausea, vomiting, and right upper quadrant pain and is ill-looking with jaundice, hypertension (50% of patients have preeclampsia), edema, ascites, a small liver, and hepatic encephalopathy.

Transaminases are often 10-fold elevated but can vary from near normal to 1000 U/liter. Other typical abnormalities are anemia, a high white blood cell count, a normal to low platelet count, abnormal protime, activated partial thromboplastin and decreased fibrinogen with or without DIC, metabolic acidosis, renal dysfunction (often progressing to oliguric renal failure), hypoglycemia, high ammonia, and often biochemical pancreatitis. Liver biopsy is rarely indicated for management but is essential for a definitive diagnosis of AFLP; it shows microvesicular, and infrequently macrovesicular, fatty infiltration, most prominent in zone 3. Occasionally, in the severely ill third-trimester patient, it is difficult to differentiate between HELLP syndrome and AFLP and in some patients these conditions may overlap. Fortunately, delivery of the baby and supportive care for the mother are required in both conditions.

Management

AFLP must be diagnosed quickly with immediate termination of pregnancy and intensive supportive care for the mother until liver function recovers. Transfer to a center with intensive care and liver transplant facilities should ideally be effected. If conditions are favorable, rapid controlled vaginal delivery with fetal monitoring is probably safest but often a cesarean section is performed with blood products to maintain an international normalized ratio of <1.5 and a platelet count of $>50{,}000 \times 10^6$/liter during and after delivery and with prophylactic antibiotics. With correction of the coagulopathy, epidural anesthesia is probably the best choice and will allow better ongoing assessment of the patient's level of consciousness.

Liver tests and encephalopathy will start to improve after 2–3 days but intensive supportive care is needed to manage the many complications, particularly infections and bleeding complications, of fulminant hepatic failure until this recovery occurs over days or many weeks. There is no specific effective therapy for AFLP. Liver transplantation has a very limited role here because of the great potential for recovery with delivery but should be considered in patients whose clinical course continues to deteriorate with advancing fulminant hepatic failure after the first 1–2 days postpartum without signs of hepatic regeneration.

Many patients have no further pregnancies after the devastating illness of AFLP, either by choice or from complications of the illness, but AFLP rarely recurs in subsequent pregnancies. Because LCHAD deficiency is present in many of these infants, later neonatal problems may occur.

See Also the Following Articles

Emesis • Hyperemesis Gravidarum • Nausea • Pregnancy and Gastrointestinal Disease

Further Reading

Bacq, Y., Sapey, T., Brechot, M.-C., Pierre, F., Fignon, A., and Dubois, F. (1997). Intrahepatic cholestasis of pregnancy: A French prospective study. *Hepatology* **26**, 358–364.

Barton, J. R., and Sibai, B. M. (1999). HELLP and the liver diseases of preeclampsia. *Clin. Liver Dis.* **3**, 31–48.

Ch'ng, C. L., Morgan, M., Hainsworth, I., and Kingham, J. G. (2002). Prospective study of liver dysfunction in pregnancy in Southwest Wales. *Gut* **51**, 876–880.

de Vree, J. M. L., Jacquemin, E., Sturm, E., Cresteil, D., Bosma, P. J., Aten, J., Deleuze, J.-F., Desrochers, M., Burdelski, M., Bernard, O., Oude Elferink, R. P. J., and Hadchouel, M. (1998).

Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. *Proc. Natl. Acad. Sci. USA* **95**, 282–287.
Hay, J. E. (2000). Viral hepatitis in pregnancy. *Viral Hepatitis Rev.* **6**, 205–215.
Ibdah, J. A., Bennett, M. J., Rinaldo, P., Zhao, Y., Gibson, B., Sims, H. F., and Strauss, A. W. (1999). A fetal fatty-acid oxidation disorder as a cause of liver disease in pregnant women. *N. Engl. J. Med.* **340**, 1723–1731.
Koch, K. L., and Frissora, C. L. (2003). Nausea and vomiting during pregnancy. *Gastroenterol. Clin. North Am.* **32**, 201–234.
Kuscu, N. K., and Koyuncu, F. (2002). Hyperemesis gravidarum: Current concepts and management. *Postgrad. Med. J.* **78**, 76–79.
Mazzella, G., Rizza, N., Azzaroli, F., Simoni, P., Bovicelli, L., Miracolo, A., Simonazzi, G., Colecchia, A., Nigro, G., Mwangemi, C., Festi, D., and Roda, E. (2001). Ursodeoxycholic acid administration in patients with cholestasis of pregnancy: Effects on primary bile acids in babies and mothers. *Hepatology* **33**, 504–508.
Meng, L.-J., Reyes, H., Palma, J., Hernandez, I., Ribalta, J., and Sjovall, J. (1997). Profiles of bile acids and progesterone metabolites in the urine and serum of women with intrahepatic cholestasis of pregnancy. *J. Hepatol.* **27**, 346–357.
Mullally, B. A., and Hansen, W. F. (2002). Intrahepatic cholestasis of pregnancy: Review of the literature. *Obstet. Gynecol. Surv.* **57**, 47–52.
Reyes, H. (1999). Acute fatty liver of pregnancy: A cryptic disease threatening mother and child. *Clinics in Liver Disease* **3**, 69–81.
Sibai, B. M., Ramadan, M. K., Usta, I., *et al.* (1993). Maternal morbidity and mortality in 442 pregnancies with hemolysis, elevated liver enzymes and low platelets (HELLP syndrome). *Am. J. Obstet. Gynecol.* **169**, 1000–1006.

Liver Failure, Pediatric

PHILIP ROSENTHAL
University of California, San Francisco

acute-on-chronic liver failure Refers to an individual with chronic liver disease who decompensates as a result of an intercurrent infection or as the result of progressive liver disease.

fulminant liver failure Acute hepatitis occurring within 2–8 weeks of the onset of illness in the absence of preexisting liver disease, complicated by hepatic encephalopathy, massive hepatic necrosis, and a prolonged prothrombin time.

subfulminant liver failure Acute hepatitis accompanied by encephalopathy that begins after 2 weeks, but within 8–12 weeks from the presentation of jaundice.

Fulminant liver failure is a rare but devastating sequel of hepatic injury. It is defined as an acute hepatitis occurring within 2–8 weeks of the onset of illness in the absence of preexisting liver disease, complicated by hepatic encephalopathy, massive hepatic necrosis on histologic examination of the liver, and a prolonged prothrombin time. In spite of advances in medical therapy and intensive care, the mortality rate in the literature ranges between 40 and 80% and is often related to complications of cerebral edema, hypoglycemia, sepsis, gastrointestinal bleeding, and acute renal failure. Liver transplantation currently is the only definitive treatment for those who do not spontaneously recover, and early recognition may allow for stabilization of the patient prior to transfer to a liver transplant center.

INTRODUCTION

In assessing fulminant liver failure, the time of onset of encephalopathy is an important prognostic factor. Severe acute hepatitis consists of hepatic injury accompanied by jaundice and coagulopathy without encephalopathy and typically has an excellent prognosis. Fulminant liver failure consists of severe acute hepatitis with hepatic encephalopathy occurring within 2 weeks of the onset of jaundice. Survival is only about 40% without liver transplantation. Subfulminant liver failure consists of acute hepatitis accompanied by encephalopathy that begins after 2 weeks, but within 8–12 weeks from the presentation of jaundice. Subfulminant hepatitis has a poor prognosis.

Often, the term "liver failure" is used to define the status of an individual with chronic liver disease who

Copyright 2004, Elsevier (USA). All rights reserved.

Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. *Proc. Natl. Acad. Sci. USA* **95**, 282–287.
Hay, J. E. (2000). Viral hepatitis in pregnancy. *Viral Hepatitis Rev.* **6**, 205–215.
Ibdah, J. A., Bennett, M. J., Rinaldo, P., Zhao, Y., Gibson, B., Sims, H. F., and Strauss, A. W. (1999). A fetal fatty-acid oxidation disorder as a cause of liver disease in pregnant women. *N. Engl. J. Med.* **340**, 1723–1731.
Koch, K. L., and Frissora, C. L. (2003). Nausea and vomiting during pregnancy. *Gastroenterol. Clin. North Am.* **32**, 201–234.
Kuscu, N. K., and Koyuncu, F. (2002). Hyperemesis gravidarum: Current concepts and management. *Postgrad. Med. J.* **78**, 76–79.
Mazzella, G., Rizza, N., Azzaroli, F., Simoni, P., Bovicelli, L., Miracolo, A., Simonazzi, G., Colecchia, A., Nigro, G., Mwangemi, C., Festi, D., and Roda, E. (2001). Ursodeoxycholic acid administration in patients with cholestasis of pregnancy: Effects on primary bile acids in babies and mothers. *Hepatology* **33**, 504–508.
Meng, L.-J., Reyes, H., Palma, J., Hernandez, I., Ribalta, J., and Sjovall, J. (1997). Profiles of bile acids and progesterone metabolites in the urine and serum of women with intrahepatic cholestasis of pregnancy. *J. Hepatol.* **27**, 346–357.
Mullally, B. A., and Hansen, W. F. (2002). Intrahepatic cholestasis of pregnancy: Review of the literature. *Obstet. Gynecol. Surv.* **57**, 47–52.
Reyes, H. (1999). Acute fatty liver of pregnancy: A cryptic disease threatening mother and child. *Clinics in Liver Disease* **3**, 69–81.
Sibai, B. M., Ramadan, M. K., Usta, I., *et al.* (1993). Maternal morbidity and mortality in 442 pregnancies with hemolysis, elevated liver enzymes and low platelets (HELLP syndrome). *Am. J. Obstet. Gynecol.* **169**, 1000–1006.

Liver Failure, Pediatric

PHILIP ROSENTHAL
University of California, San Francisco

acute-on-chronic liver failure Refers to an individual with chronic liver disease who decompensates as a result of an intercurrent infection or as the result of progressive liver disease.

fulminant liver failure Acute hepatitis occurring within 2–8 weeks of the onset of illness in the absence of preexisting liver disease, complicated by hepatic encephalopathy, massive hepatic necrosis, and a prolonged prothrombin time.

subfulminant liver failure Acute hepatitis accompanied by encephalopathy that begins after 2 weeks, but within 8–12 weeks from the presentation of jaundice.

Fulminant liver failure is a rare but devastating sequel of hepatic injury. It is defined as an acute hepatitis occurring within 2–8 weeks of the onset of illness in the absence of preexisting liver disease, complicated by hepatic encephalopathy, massive hepatic necrosis on histologic examination of the liver, and a prolonged prothrombin time. In spite of advances in medical therapy and intensive care, the mortality rate in the literature ranges between 40 and 80% and is often related to complications of cerebral edema, hypoglycemia, sepsis, gastrointestinal bleeding, and acute renal failure. Liver transplantation currently is the only definitive treatment for those who do not spontaneously recover, and early recognition may allow for stabilization of the patient prior to transfer to a liver transplant center.

INTRODUCTION

In assessing fulminant liver failure, the time of onset of encephalopathy is an important prognostic factor. Severe acute hepatitis consists of hepatic injury accompanied by jaundice and coagulopathy without encephalopathy and typically has an excellent prognosis. Fulminant liver failure consists of severe acute hepatitis with hepatic encephalopathy occurring within 2 weeks of the onset of jaundice. Survival is only about 40% without liver transplantation. Subfulminant liver failure consists of acute hepatitis accompanied by encephalopathy that begins after 2 weeks, but within 8–12 weeks from the presentation of jaundice. Subfulminant hepatitis has a poor prognosis.

Often, the term "liver failure" is used to define the status of an individual with chronic liver disease who

Copyright 2004, Elsevier (USA). All rights reserved.

decompensates as a result of an intercurrent infection or as the result of progressive liver disease. This is better defined as acute-on-chronic liver failure. It is critically important that definitions are clear so that there is no misunderstanding regarding the disorder under discussion here, i.e., fulminant or acute liver failure.

ETIOLOGY

A list of potential etiologies of acute liver failure is presented in Table I. Etiologies are often age dependent. Although the cause of acute liver failure is unknown in over half of pediatric cases; in known cases, drug-induced hepatic injury and hepatotrophic viral infections are the most commonly recognized causes in the United States. Hepatitis A and hepatitis B viral infections are the most common infectious causes in all age groups beyond the neonatal period.

In the neonate, the differential diagnosis of liver failure includes infectious, metabolic, ischemic, and perfusion disorders. Infectious etiologies include herpesvirus, echovirus, adenovirus, and hepatitis B. Inborn errors of metabolism presenting as acute liver failure in the neonate include hereditary fructose intolerance, galactosemia, tyrosinemia type I, urea cycle defects, inborn errors of bile acid synthesis, and neonatal hemochromatosis. Ischemic and perfusion defects may be the result of congenital heart disease, cardiac surgery, myocarditis, or severe asphyxia. In infants or toddlers, drug and toxin ingestion, including seizure medications (valproate), antituberculin drugs (isoniazid), acetaminophen, and poison mushrooms (*Amanita*), must be considered. Additionally, infants and children may be exposed to hepatitis A from siblings or day care. Hereditary fructose intolerance may present with liver failure after the introduction of sucrose- and/or fructose-containing formulas or solid foods (fruits/ vegetables) into the diet.

In older children and adolescents, besides infection, drugs, toxins, and perfusion defects, Wilson's disease must be considered in the differential diagnosis. In the appropriate adolescent female, acute fatty liver of pregnancy and fatty acid oxidation defects can be seen. Autoimmune hepatitis with or without a hemolytic anemia may be present. Rarely, malignancy (hepatoblastoma, leukemia, non-Hodgkin lymphoma), radiation, and hyperthermia may result in acute liver failure.

Reye's syndrome is associated with microvesicular fat on liver biopsy, encephalopathy, elevated serum aminotransferase levels, and prolonged prothrombin time, usually arising during convalescence from a viral syndrome such as influenza or varicella. The incidence of Reye's syndrome in recent years has significantly diminished with the reduced use of aspirin for fever control in children and improved recognition and diagnosis of fatty acid oxidation defects that may exhibit similar findings.

It is becoming increasingly apparent that some herbal preparations may also result in acute liver failure and a careful history that includes herbal remedy ingestion must be sought.

PRESENTATION AND DIFFERENTIAL DIAGNOSIS

The hallmarks of acute liver failure are the acute onset of jaundice, biochemical evidence of elevated serum

TABLE I Potential Etiologies of Acute Liver Failure in Children[a]

Cause	Examples
Infection	Hepatitis A or B, hepatitis B with hepatitis D super- or coinfection, hepatitis E, non-A–G hepatitis, cytomegalovirus, varicella-zoster virus, adenovirus, yellow fever, Q-fever
Drugs	
Dose-dependent	Acetaminophen, haloalkanes
Idiosyncratic	Isoniazid, dilantin, valproate, phenytoin, propylthiouracil, penicillin, tetracyclines, sulfonamides, amiodarone, ketoconazole, nonsteroidal antiinflammatory drugs
Synergistic drug reactions	Alcohol and acetaminophen, isoniazid + rifampicin, bactrim, barbiturates + acetaminophen, amoxacillin + clauvulinic acid
Metabolic disorders	Wilson's disease, acute fatty liver of pregnancy, galactosemia, tyrosinemia, errors of fatty acid metabolism
Toxins	*Amanita phalloides*, industrial solvents, carbon tetrachloride, herbal medicines
Systemic disorders	Autoimmune hepatitis, malignancy, hyperthermia, sepsis
Miscellaneous	Budd–Chiari syndrome, portal vein thrombosis, indeterminate

[a] Adapted with permission from Perkin (2003), Acute liver failure *In* "Pediatric Hospital Medicine: Textbook of Inpatient Management." Copyright Lippincott Williams & Wilkins.

aminotransferase levels, prolonged prothrombin time, and encephalopathy in a previously healthy patient. Initial symptoms may be nonspecific, such as jaundice, nausea, vomiting, and abdominal pain. A careful medical history focusing on a history of ingestion (accidental or intentional), medications, mushroom picking, recreational drugs, blood transfusions, immunizations (particularly hepatitis A and B vaccines), and travel must be sought. An accurate acetaminophen history should include the amount and interval of doses, carefully identifying all acetaminophen derivatives because many combination medications include acetaminophen as a component. On occasion, parents have mistakenly substituted acetaminophen infant drops for the children's elixir, resulting in a significant overdose. In infants, a previous hypoglycemic event, developmental delay, or history of seizures may suggest a metabolic disorder. Rarely, in adolescents, acute liver failure secondary to Wilson's disease may be preceded by neuropsychiatric manifestations.

On physical examination, the child is usually jaundiced and icteric. There may or may not be hepatomegaly. The remainder of the physical exam may be noncontributory. There may be occasional petechiae or bruises attributable to the coagulopathy. A careful assessment of neurological status is imperative but is often a challenge in the young child. The grading scale utilized by the Pediatric Acute Liver Failure Study Group is presented in Table II as a means of evaluating and scoring the degree of encephalopathy.

Biochemical features of acute liver failure include marked elevations in serum aminotransferases (aspartate transaminase, alanine transaminase), total and direct-reacting bilirubins, and prolongation of the prothrombin time. Although the serum aminotransferase levels may rise into the thousands (units/liter), the peak value and rate of decline are not predictive of outcome. The most frequently observed electrolyte disturbances are hyponatremia, hypokalemia, hypocalcemia, and hypomagnesemia. Due to a combination of liver glycogen dysregulation, and central nervous system and multiorgan dysfunction, hypoglycemia is frequently observed. More commonly accompanying non-A—non-G hepatitis is an aplastic anemia. Wilson's disease may present with a constellation of laboratory findings, including a Coombs-negative hemolytic anemia, low serum alkaline phosphatase level, and hyperuricemia. Serum copper values are often markedly elevated. Serum ceruloplasmin levels classically are decreased in Wilson's disease. However, a normal serum ceruloplasmin does not rule out Wilson's disease, and analysis of 24-hour urinary copper excretion a may

TABLE II Encephalopathy Staging Criteria (Pediatric Acute Liver Failure Study Group)[a]

Stage	Clinical	Asterixis/reflexes	Neurologic signs	EEG changes
0	None	None/normal	Psychological testing only	Normal
I	Confused, mood changes, altered sleep habits, loss of spatial orientation, forgetful	None/normal	Tremor, apraxia, impaired handwriting	Normal or diffuse slowing to theta rhythm, triphasic waves
II	Drowsy, inappropriate behavior, decreased inhibitions	None/hyperreflexic	Dysarthria, ataxia	Abnormal generalized slowing, triphasic waves
III	Stuporous obeys simple commands	None/hyperreflexia, upgoing toes (+ Babinski)	Rigidity	Abnormal generalized slowing, triphasic waves
IV	Comatose, arouses with painful stimuli (IVa), or no response (IVb)	Absent	Decerebrate of decorticate	Abnormal, very slow delta activity
For younger children				
Early (I and II)	Inconsolable crying, sleep reversal, inattention to task	Unreliable/normal or hyperreflexic	Untestable	
Middle (III)	Somnolence, stupor, combativeness	Unreliable/ hyperreflexic	Most likely untestable	
Late (IV)	Comatose, arouses with painful stimuli (IVa), or no response (IVb)	Absent	Decerebrate or decorticate	

[a] Adapted with permission from Perkin (2003), Acute liver failure *In* "Pediatric Hospital Medicine: Textbook of Inpatient Management." Copyright Lippincott Williams & Wilkins.

be necessary. If autoimmune hepatitis is suspected, autoimmune markers including antinuclear antibody, anti-smooth-muscle antibody, anticardiolipin, anti-liver–kidney microsomal antibody, and antimitochondrial antibody should be obtained. A reversed albumin/globulin serum protein ratio may also be suggestive of autoimmune disease.

THERAPY

Initial management should focus on stabilization and support to allow for possible recovery or liver transplantation if necessary. All patients with severe acute hepatitis who progress to any stage of encephalopathy should be managed in an intensive care unit. Early referral to a liver transplant center is imperative for all children whose condition is not improving. Room noise should be kept to a minimum and room lighting should be used sparingly to avoid agitating the child. Significant morbidity and mortality may result from the complications of acute liver failure even if the liver spontaneously recovers. Venous access is mandatory. An arterial line should be considered if the child's condition is deteriorating and blood gas analysis is necessary prior to intubation and monitoring of ventilatory support. A nasogastric tube may be required to monitor for gastric bleeding and for administration of medications. A urinary catheter may be placed to monitor fluid status and renal function. If available, the patient's bed should have a built-in scale that enables weighing the patient with minimal disturbance. Frequent scheduled repositioning and involvement of a specialized skin care team especially in the obtunded or comatose patient should avoid bedsores and decubitus ulcers.

Medical management of acute liver failure must be multidisciplinary because of the complexity of the disease and its potential progression to multiorgan failure, characterized by increased cardiac output, peripheral vasodilatation, respiratory distress, and renal failure. Useful laboratory tests include a complete blood count with differential and platelet count, and levels of serum electrolytes, blood urea nitrogen, creatinine, calcium, magnesium, phosphorus, blood sugar, alanine transaminase, aspartate transaminase, alkaline phosphatase, γ-glutamyl transferase, total and direct-reacting bilirubins, prothrombin and partial thomboplastin times, international normalized ratio (INR), amylase, lipase, and serum ammonia. If the history is suggestive, a toxicology screen, acetaminophen, or aspirin levels should be obtained. Hepatitis serologies, including antihepatitis A virus immunoglobulin M (IgM), hepatitis B surface antigen, antihepatitis B core antibody (IgM), and antihepatitis C virus, should be acquired. Culture specimens (blood, urine, stool) should be sent to the lab. Prothrombin times (a measure of vitamin K-dependent clotting factors synthesized by the liver) or INRs are often the most useful laboratory parameters to follow serial progression of the patient's condition. Obviously, administration of fresh frozen plasma or clotting factors will negate the usefulness of serial prothrombin time values. Thus, the judicious use of fresh frozen plasma and clotting factors is mandatory in these patients and is reserved for the presence of active bleeding. If the prothrombin time is prolonged, a trial of intravenous vitamin K daily for 3 consecutive days should be undertaken. Giving the dose intravenously eliminates the possibility of poor intestinal absorption of this fat-soluble vitamin in a patient who is cholestatic. Improvement in the prothrombin time suggests viable hepatocytes capable of synthesizing clotting factors and having the potential for regeneration and recovery, whereas no improvement or a continued prolongation of the prothrombin time is an indication of a poor prognosis.

Because electrolyte and acid–base disturbances frequently occur in acute liver failure, fluid management and electrolyte replacement and supplementation must be meticulously monitored. Fluid management requires a balance to avoid fluid overload and dehydration as multiorgan failure develops. Renal failure can occur in the face of pulmonary edema and peripheral vascular collapse. Cerebral edema and intracranial hypertension necessitate careful fluid administration to avoid intravascular space expansion and exacerbation of cerebral edema.

Nutritional support in acute liver failure is difficult. Hypoglycemia occurs frequently, necessitating intravenous dextrose infusion. Amino acids should be restricted in the initial management of acute liver failure to avoid an excess nitrogen load. However, to prevent a catabolic state, limited protein supplementation (0.5–1 g/kg/day) by day 3 of illness may be considered. The type of proteins and amino acids may influence hepatic encephalopathy. α-Keto analogues of amino acids are transaminated in the intestinal mucosa into amino acids without supplying additional nitrogen. Limited studies suggest that ketoacids are efficiently converted to amino acids and converted into protein while maintaining nitrogen balance. Branched-chain amino acid solutions administered parenterally have also been studied with inconclusive results. Their high expense with limited effectiveness suggests they be restricted at present to investigational protocols.

An abdominal ultrasound with Doppler interrogation is useful to evaluate anatomic or vascular etiologies.

Unless the patient has clinical evidence suggesting cerebral edema, computed tomographic examination of the head is not useful in the evaluation of hepatic encephalopathy.

N-Acetylcysteine should be administered to any patient suspected of toxic acetaminophen ingestion to replenish glutathione stores. Most effective within 12 hours of ingestion, it has demonstrated effects up to 36 hours after the initial overdose. The published acetaminophen level nomogram is useful only in guiding the treatment for a one-time single overdose and does not predict outcome in chronic ingestion cases. Adolescent medicine/psychiatric consultation is imperative if the overdose was intentional or represented a suicidal threat and should be obtained prior to the patient progressing into coma, because this assessment may be crucial in determining the appropriateness of liver transplant.

The cause of hepatic encephalopathy is multifactorial. Current theory includes the accumulation of neurotoxic and neuroactive substances. Although most are poorly defined, hyperammonemia has been associated with acute liver failure-induced encephalopathy. Limiting protein intake to $\leq$1 g/kg/day may decrease colonic bacterial ammonia production and allow the compromised liver to metabolize the protein load. Lactulose, a nonabsorbable carbohydrate, acts as a cathartic to induce diarrhea and ammonia removal via the intestinal tract, and it acidifies the colon to trap ammonia (NH_3) by converting it to NH_4^+ in the gut, promoting ammonia excretion. Oral neomycin has been advocated as a relatively nonabsorbable antibiotic to rid the colon of its ammonia-producing bacterial flora. Unfortunately, chronic oral administration of neomycin may result in its absorption, allowing oto- and nephrotoxicity, which may outweigh any possible benefit. All sedating drugs, even if the child is combative, are to be avoided because most are predominantly metabolized via the liver and their administration may confound the mental status examination. Restraints may be necessary to protect the comatose child from self-injury. Acid suppression with histamine-2 (H2) blockers or proton pump inhibitors is indicated as prophylaxis for gastrointestinal bleeding.

Cerebral edema and brain herniation are the most frequent complications leading to death in acute liver failure. The pathophysiology is unclear and treatment is controversial. Intensive care unit admission, transfer to a liver transplant center, and listing for liver transplantation should be instituted early in the course to optimize outcomes. Cerebral edema must be managed aggressively. Neurologic physical examination may not be sensitive enough for detecting increased intracranial pressure, so clinical suspicion must be maintained. Early signs and symptoms of cerebral edema include systolic hypertension and increased muscle tone. Headache, vomiting, and papilledema are present inconsistently. If progressive, cerebral edema will be associated with decerebrate posturing, hyperventilation, myoclonus, seizures, trismus, and opisthotonos. Although computed axial tomography may not be sensitive for detecting early cerebral edema, its use may be warranted to rule out intracranial bleeding. Intracranial pressure monitoring is required in comatose patients and in those with neurologic deterioration. Several types of devices are available. Intraventricular monitors are accurate and precise, but suffer from complications, especially bleeding. Epidural monitors are the safest but are often inaccurate. Subdural placement of monitors has intermediate risk and relatively accurate readings. All intracranial pressure monitors increase the risk for infection.

Increased intracranial pressure management includes use of hyperventilation, minimal stimulation, head elevation to 30°, maintenance of hemodynamic stability, and mannitol infusion. Steroid administration has not been found beneficial and should be avoided. Attention to avoidance of rapid fluid boluses or aggressive correction of electrolyte imbalances that can aggravate the situation is important. The goal is to maintain cerebral perfusion pressure (the difference between mean arterial pressure and intracranial pressure) above 50 torr. This usually requires maintaining the intracranial pressure in children in the normal range (10–15 torr).

Mechanical hyperventilation lowers intracranial pressure but does not improve patient morbidity. Between 20 and 60 torr, cerebral blood flow is linearly related to the partial pressure of blood carbon dioxide. Lowering carbon dioxide content to 25–30 torr maximizes cerebral vascular constriction and reduces blood flow. This vascular effect progressively diminishes after 6 hours of therapy, although the clinical response may last for a few days. Hyperventilating the patient by increasing tidal volume may raise positive end expiratory pressure and restrict venous return. This may occur especially if there is pulmonary edema. Increased positive end expiratory pressure also reduces diastolic filling volume and may reduce cerebral perfusion pressure.

In patients without renal failure, mannitol infusions have an 80% response rate and may improve survival. Serum osmolality must be monitored frequently and maintained in the range of 300–320 mOsm. Mannitol should not be given if serum osmolality is over 320 mOsm, if renal failure ensues, or if oliguria and a rising serum osmolality develop simultaneously.

Repeated mannitol infusions may reverse the osmotic gradient. Mannitol should be discontinued if the intracranial pressure does not respond after the first few boluses.

Renal failure occurs in up to half of patients with acute liver failure, often due to hepatorenal syndrome. Hepatorenal syndrome is the unexplained development of reversible renal failure in both acute and chronic liver disease, characterized by azotemia, oliguria, low urinary sodium excretion (< 10 mEq/day), and an increased urine:serum osmolality ratio in the absence of active urinary sediment. Attention should be paid to diminished cardiac output and blood pooling from peripheral vasodilatation as possible etiologic factors for poor renal perfusion. Adequate urine output can be maintained with judicious use of loop diuretics and low-dose dopamine infusion. Depleted intravascular volume can be corrected with blood components or volume expanders. Because plasma albumin is invariably low, salt-poor albumin solutions may be preferred to carbohydrate-based volume expanders. If oliguria develops, hemodialysis/hemofiltration may be required to maintain optimal fluid volume and serum osmolality. Fluid restriction must be used if renal failure is established. Sodium and potassium losses should be estimated and replaced appropriately. Hyponatremia should not be replaced by use of hypertonic saline unless a hyponatremic seizure occurs. Excess free water may be removed by dialysis and will resolve the hyponatremia problem without altering the total body sodium content.

Pulmonary complications occur frequently in acute liver failure patients. These patients have a predisposition to low pressure pulmonary edema and development of adult respiratory distress syndrome. Pulmonary edema is observed in about 40% of acute liver failure patients. It is believed to result from pulmonary capillary changes that cause increased permeability. Management is supportive. Continuous pulse oximetry, frequent arterial blood gas measurements, administration of supplemental oxygen, and use of mechanical ventilation are all appropriate. Sedative and paralytic agents may be required to ensure tolerance of mechanical ventilation. These need to be used sparingly if they will alter neurologic evaluation, and they may have prolonged effects due to longer half-lives from decreased hepatic metabolism and renal excretion. Further, the neurologically depressed patient is at risk for the development of aspiration pneumonia.

Infection plays a serious threat to the patient with acute liver failure. It places the patient at risk for sepsis and is a contraindication to liver transplantation. Bacterial infections of either the respiratory or the urinary tract are prevalent. Organisms frequently seen include *Staphylococcus* spp., *Streptococcus* spp., and gram-negative rods. Iatrogenic sources of infection must also be considered because these patients often have indwelling lines and urinary catheters. Despite the high rates of infection, prophylactic antibiotic administration is not advocated. However, a low threshold for starting antibiotics should be used because the usual presentation of infection with fever and leukocytosis may be absent. Surveillance cultures for bacteria and fungi from blood, urine, sputum, and open wounds should be routine. Ascitic fluid should be cultured if present. High-dose broad-spectrum antibiotics should be started at the first sign of infection, with coverage narrowed as soon as a specific organism is identified.

OUTCOME

In general, acute liver failure is associated with significant morbidity and mortality. Fortunately, it is a rare disorder in children. Spontaneous recovery is unusual except for prompt treatment of acetaminophen overdose in young children. Intensive care support and liver transplantation are crucial for the management of these critically ill children, necessitating early referral to transplant centers. The shortage of donor organs continues to influence survival for children with fulminant hepatic failure. Improved outcomes may be obtained by the use of living-related transplants, split-liver transplants, and auxiliary transplants in selected patients. Experimental therapies, including the use of bioartificial liver support and hepatocyte transplantation, hold promise in providing temporary hepatic support, allowing the native liver to repair, recover, and regenerate, thereby reducing complications and avoiding transplantation and lifelong immunosuppression.

Unfortunately, there are several major limitations to attempts to better define and develop therapeutic regimens for severe, acute liver failure. Acute liver failure has varying etiologies (viral, drug, toxin, etc.) and thus may actually be several different conditions. There is no clear understanding of the pathophysiology of the disease, making it difficult to develop specific treatment strategies. There is no single marker of liver dysfunction allowing for stratification of hepatic injury. Relatively few individuals with acute liver failure are seen at a single institution. Progression of the disease is rapid and little time is available to initiate therapy. Therefore, multicenter collaborative trials, such as those of the Pediatric Acute Liver Failure Study Group, are imperative so that standardized data collection and management protocols can be utilized to define and

improve therapies for children with this devastating diagnosis.

LIVER-ASSIST DEVICES

The development of an "artificial" liver system capable of supporting liver functions in patients with liver failure has been in preparation for over four decades. Recent work has focused on the use of mammalian hepatocytes to provide metabolic assistance to the failing liver. Several extracorporeal hepatocyte liver-assist systems have been reported in a small number of pilot clinical trials. One such device was utilized successfully in the care of a 10-year-old child with fulminant liver failure as a bridge to liver transplant. A large, randomized controlled trial in adults utilizing porcine hepatocytes in a hollow-fiber cartridge (Circe Biomedical, HepatAssist 2000) showed encouraging results, but the design of the study was such that it failed to demonstrate statistical significance. A recent systematic review of artificial and bioartificial support systems for acute and acute-on-chronic liver failure has suggested that artificial support systems reduce mortality in acute-on-chronic liver failure compared with standard medical therapy. Artificial and bioartificial support systems do not appear to affect mortality in acute liver failure.

SUMMARY

Acute liver failure is a dramatic syndrome in which previously healthy children rapidly lose hepatic function, develop encephalopathy, and become critically ill. In spite of liver transplantation, many patients die due to complications of the disease or because of a lack of timely donors. Because acute liver failure occurs infrequently, no single center has a large enough experience to conduct important clinical studies to investigate natural history or treatment. Investigations by the Pediatric Acute Liver Failure Study Group should define patient demographics, presenting symptoms, clinical course, suspected etiology, and outcome of children with acute liver failure. This important data collection registry should eventually aid in refining our understanding of the etiology, management, and treatment of this devastating disorder in children.

See Also the Following Articles

Fulminant Hepatic Failure • Hepatic Encephalopathy • Hepatitis A • Hepatitis B • Liver Transplantation

Further Reading

Kjaergard, L. L., Liu, J., Als-Nielsen, B., and Gluud, C. (2003). Artificial and bioartificial support systems for acute and acute-on-chronic liver failure. *JAMA* **289**, 217–222.

Lee, W. M. (1993). Acute liver failure. *N. Engl. J. Med.* **329**, 1862–1872.

Rivera-Penera, T., Gugig, R., Davis, J., *et al.* (1997). Outcome of acetaminophen overdose in pediatric patients and factors contributing to hepatotoxicity. *J. Pediatr.* **130**, 300–304.

Watanabe, F., and Rosenthal, P. (1994). Medical therapy. *In* "Support of the Acutely Failing Liver" (A. A. Demetriou ed.), pp. 22–32. R. G. Landes, Austin.

Whitington, P. F., Soriano, H. E., and Alonso, E. M. (2001). Fulminant hepatic failure in children. *In* "Liver Disease in Children," 2nd Ed. (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), pp. 63–88. Lippincott Williams & Wilkins, Philadelphia.

Liver Ischemia

HANNES A. RÜDIGER AND PIERRE A. CLAVIEN
University Hospital of Zurich, Switzerland

cold ischemia Technique to slow the metabolism of an organ in order to reduce oxygen consumption; used in the transplant setting and during organ harvesting and preservation.

intermittent clamping Strategy to prevent ischemic injury by interrupting long ischemic insults with multiple short intervals of reperfusion.

ischemic preconditioning Regimen used to prime an organ against subsequent long ischemic insults; consists of a short period of ischemia followed by a brief interval of reperfusion prior to the prolonged ischemic insult.

warm ischemia Interruption of inflow of blood and oxygen during liver surgery, shock, and trauma; occurs at normothermic (body) temperature.

Ischemia reperfusion injuries in the liver have important implications in a number of clinical settings, including transplantation, liver resection, trauma, and shock. Ischemia reperfusion injury occurs in two main forms, cold ischemic injury (occurring during organ preservation) and warm ischemic injury (occurring during liver surgery, shock, and trauma). The mechanisms of injury are complex, and only very few protective strategies are currently applied in routine clinical practice.

INTRODUCTION

Injuries to the liver caused by prolonged periods of interrupted blood flow or low-level oxygen delivery are important in many clinical situations. Lack of insight into the mechanisms associated with these types of injuries has significantly hampered the development of hepatic surgery and transplantation. Over the past two decades, efforts have been invested in the search for protective strategies, and important discoveries have been made over the past few years. In this article, the current literature on the mechanisms of ischemic injury in the liver is summarized and novel protective approaches are discussed.

Hepatic ischemia/reperfusion (I/Rp) injury can be divided into two types: (1) warm (normothermic) I/Rp, which occurs during liver resection, trauma, various forms of shock, and transplantation, and (2) cold I/Rp injury, which occurs in the transplant setting during the period of organ preservation at 1°−4°C. Although several mechanisms of injury appear to share common pathways, recent evidence suggests important differences between warm and cold I/Rp.

WARM (NORMOTHERMIC) ISCHEMIA

Warm ischemia injury is common during liver surgery, trauma, and transplantation. For example, ischemia is intentionally achieved during major liver resections by occluding hepatic arterial inflow (Pringle maneuver) during the time of transsection of the liver parenchyma, to prevent bleeding. Ischemia is tolerated only for a period of about 1 hour in normal livers and about 30 minutes in cirrhotic livers, and the duration of ischemia correlates with the degree of reperfusion injury.

Many studies have shown that death of endothelial cells and hepatocytes does not occur during the period of ischemia but rather during subsequent reperfusion. Furthermore, accumulating evidence suggests that cell injury occurs by apoptosis (programmed cell death), which, unlike necrosis, is an active and energy-requiring form of cell death. Some investigators, however, suggest that necrosis, rather than apoptosis, is the major cause of cell death in the ischemic liver. The reason for this ongoing controversy might be related to the fact that the distinction between apoptosis and necrosis is not always clear. There are shared pathways, common features, and transitions from apoptotic pathways to necrosis. In addition, differences in the energy status and ATP levels (e.g., starved vs. fed condition) might be critical to induce a switch from apoptotic to necrotic forms of cell death. Nevertheless, active apoptotic mediators have been well documented in the ischemic liver, mainly after reperfusion.

The balance between pro and antiapoptotic mediators prevents and regulates death in healthy cells. In the setting of I/Rp, this tightly regulated "survival" balance is severely disturbed in favor of cell death. Recent advances in our understanding of several mediators of

 Copyright 2004, Elsevier (USA). All rights reserved.

apoptosis have enabled new insights into the mechanisms of injury in the ischemic liver. For example, apoptotic cell death is initiated by binding of tumor necrosis factor α (TNFα) to the specific receptor, TNF-R1, on the cell membrane of hepatocytes in a mouse model of normothermic I/Rp injury (Fig. 1). The apoptotic signal is then transferred into the cell, where the specific protease caspase-8 is activated. This initial step is followed by the activation of a cascade of various proteases and other cellular events. First, a protein that is a BH3-interacting death domain (BID) agonist is activated and translocates to the mitochondria, inducing the release of cytochrome C, which binds to the apotosis protease-activating factor-1 (APAF-1). This compound activates caspase-9 with subsequent activation of caspase-3, the "executioner" caspase. This process leads to DNA fragmentation and cell death. Due to the presence of inactive proforms of the specific apoptotic proteases, signal transduction does not depend on protein synthesis and can occur rather quickly. In this process, the mitochondria act as amplifiers of the deadly signal, allowing even weak signals to have deleterious effects.

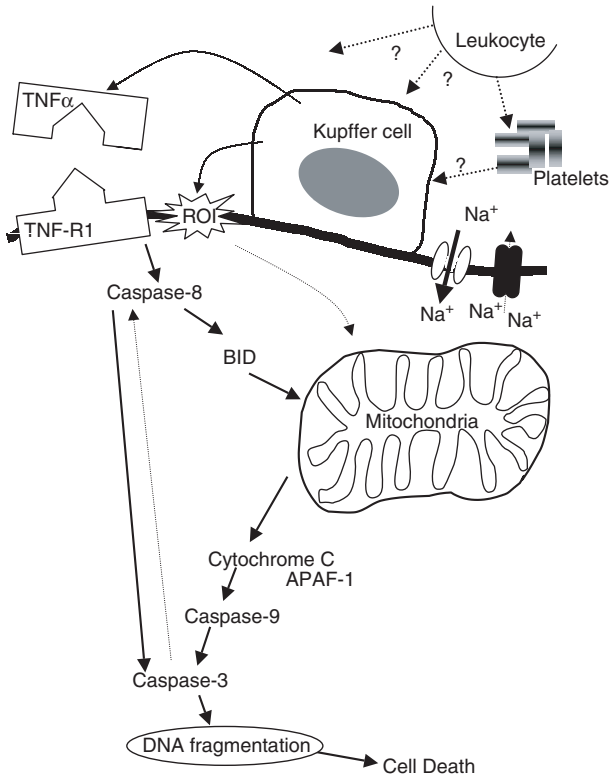


FIGURE 1 Several mechanisms of normothermic ischemic injury have been postulated to date, including tumor necrosis factor α (TNFα)-mediated apoptosis, dysregulation of ion distribution, and generation of reactive oxygen intermediates (ROI). The mechanisms of injury appear to involve multiple cell systems and pathways. APAF-1, Apotosis protease-activating factor-1; BID, BH3-interacting death domain agonist.

The activation of Kupffer cells also contributes to I/Rp injury in the liver. Activated Kupffer cells release a variety of potentially harmful mediators, including TNFα and reactive oxygen species. In high concentrations, oxygen radicals induce peroxidation of lipids, an effect that leads to membranous dysfunction and favors cell death. The release of highly reactive radicals might be critical for initiation of apoptotic cell death after ischemia, inasmuch as several cell types, including hepatocytes, are resistant to TNFα alone. From this perspective, oxidative stress might act as a sensitizer for the deleterious effect of TNF signaling.

Prolonged hypoxic conditions are associated with an impaired energy status within cells, which has severe consequences for the regulation of cellular pH and cell volume. The ATP-dependent efflux of Na^+ through Na^+,K^+-ATPase as well as the impaired function of Na^+/H^+ exchangers and Na^+/HCO_3^- cotransporters lead to acidification of the cytosol and cell swelling. However, the link between these disturbances of ion distribution, cell swelling, and simultaneously ongoing apoptotic processes remains unclear.

Ischemia in the Diseased Liver

Chronic liver diseases such as steatosis and cirrhosis are associated with increased sensitivity to ischemia and hypoxia. Although mild steatosis (< 30%) has little impact on the results of liver resection and transplantation, severe steatosis (> 60%) is associated with an increased risk of liver failure after surgery. We recently hypothesized that the devastating effects of ischemia on the fatty liver are due to a dysfunction in the apoptotic pathway and to a predominant necrotic form of cell death. We used a model of Zucker rats, which develop severe steatosis at the age of 8 weeks due to overeating. After 1 hour of hepatic ischemia, there was an associated dramatic increase in liver injury (e.g., serum transaminases) and impaired survival. However, markers of apoptosis [caspase-8 and caspase-3 activity, cytochrome C release, and terminal deoxynucleotidyl transferase-mediated dUTP nick end-labeling (TUNEL) staining] were decreased compared to lean animals, suggesting a low apoptotic activity level. Morphological analysis of a hematoxylin/eosin-stained biopsy revealed massive coagulative necrosis in fatty animals, whereas apoptosis was predominant in lean animals.

Although the increased sensitivity of the cirrhotic liver to warm or cold ischemia has been well described, the underlying mechanisms of cell injury are still

unknown. Future studies in animal models are needed to further identify mechanisms and to develop effective protective strategies in cirrhosis.

Protective Strategies

Many strategies to protect against warm ischemic injury have been proposed in various animal models, but only two, intermittent clamping and ischemic preconditioning, have finally made the transition into clinical practice. Intermittent clamping consists of multiple cycles of short intervals of ischemia (10–30 minutes) and reperfusion (5–15 minutes) (Fig. 2b). The protocol for ischemic preconditioning consists of a single brief ischemic interval (10 minutes) followed by a short period of reperfusion (10–15 minutes) prior to the prolonged ischemic interval (Fig. 2c). We have recently evaluated protection conferred by these two approaches using a mouse model of prolonged hepatic ischemia. We found that protection was similar for ischemic insults of up to 75 minutes, whereas intermittent clamping was more effective for ischemic intervals of 120 minutes. Ischemic preconditioning appears preferable in most cases because it avoids the blood loss that occurs during the multiple periods of reperfusion inherent to intermittent clamping. In addition, a great majority of cases require an ischemic interval shorter than 60 minutes, and protection appears comparable in ischemic preconditioning and intermittent clamping.

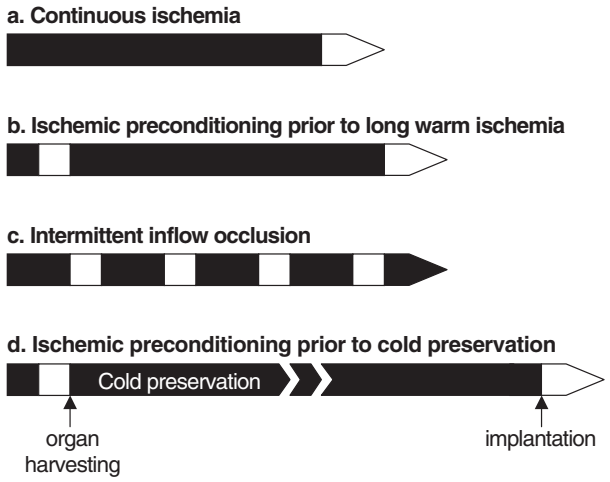


FIGURE 2 Several surgical strategies exist to overcome the negative effects of (a) prolonged ischemia (■) and reperfusion (□). To overcome warm ischemic injuries, options include ischemic preconditioning (b) and intermittent inflow occlusion (c). Experimental studies suggest that normothermic ischemic preconditioning prior to hypothermic preservation also protects against cold ischemic injuries (d).

The mechanisms of the biologic adaptation to ischemia initiated by ischemic preconditioning are still poorly understood. Several mediators have been suggested to be involved in these protective pathways, including adenosine, nitric oxide, oxygen radicals, protein kinases, ATPases, and prostaglandins. We have recently shown that preconditioning causes a downregulation of the apoptotic pathway following ischemia. Better understanding of these protective pathways may lead to the design of specific protective interventions by pharmacological means. Such approaches are appealing because the negative effects of preconditioning due to the short period of ischemia would be avoided (e.g., a possible decrease in ATP levels).

COLD ISCHEMIA

Liver transplantation has experienced a dramatic success over the past 15 years and has become the treatment of choice for patients with end-stage liver disease. Although major progress has been made in several areas of transplantation, such as selection of recipients and immunosuppression, ischemic injury remains an important limiting factor in some aspects of the procedure. During the various phases of transplantation, the liver graft is exposed to different types of ischemic stress, including normothermic ischemia prior or during harvesting, cold ischemia (at 1°–4°C) during the period of preservation, and rewarming during graft implantation (Fig. 2d).

It is currently believed that nonparenchymal cells (i.e., sinusoidal endothelial cells) are the major targets of cold ischemic injury (cold preservation). In contrast to sinusoidal endothelial cells (SECs), parenchymal liver cells (hepatocytes) remain morphologically well preserved even after extended periods of cold preservation. It is thought that later injuries to the hepatocytes are a consequence of nonparenchymal injury, the final demonstration of this relationship is not yet available.

Injuries due to low-temperature preservation are different from warm ischemic injuries in many aspects. The main feature of warm ischemia injury is massive and early (3–6 hours) apoptosis of hepatocytes after reperfusion. In contrast, cold storage is associated with diffuse SEC damage within 30–60 minutes of reperfusion, but hepatocytes are mostly preserved. The degree of SEC injury correlates with functional impairment of the graft following transplantation. The SEC injury is characterized by alteration of the extracellular matrix and cytoskeleton and detachment and rounding of endothelial cells. We have observed that similar processes also occur during the early

stage of angiogenesis. Angiogenesis is a fundamental process during wound healing, tissue collateralization, and neoplasia and occurs as a result of low oxygen tension. The manifestations of angiogenesis consist of cytoskeletal and extracellular matrix changes by matrix metalloproteinases, resulting in endothelial cell detachment, rounding, and, finally, proliferation. We have recently demonstrated that inhibition of antiangiogenic protease inhibitors (such as minocycline, interferon, and fumagillin) is able to prevent SEC detachment during cold ischemia, reducing preservation injury. Several recent reports point to matrix metalloproteinases (MMPs) as having critical roles in the mechanisms of SEC detachment during cold hepatic preservation. MMPs are important contributors to angiogenesis.

On reperfusion of the cold preserved graft, SECs undergo apoptosis, with subsequent DNA fragmentation and cell death. The strong correlation between the degree of apoptosis and graft viability suggests that SEC apoptosis is a critical process of cold ischemic injury. Although a number of intracellular mediators of apoptosis are known, the data regarding the initiating extracellular processes are scarce. Experimental studies have shown that platelets are sequestered in the ischemic liver after reperfusion and contribute to organ injury. In addition, strong evidence suggests a pivotal role of leukocytes in the mechanisms of cell death during reperfusion after cold ischemia. Sindram *et al*. have recently demonstrated that the action of platelets and leukocytes is not independent but rather is synergistic in nature. Jaeschke has shown that oxygen free radicals and proteases are involved in blood cell-mediated killing of hepatocytes. Neutralization of oxygen species may represent a promising approach to minimize preservation injury. Further studies are needed to detail how the signal from platelets and leukocytes is transferred into the cell, causing endothelial cell apoptosis.

Protective Strategies

Development of cold preservation solution by the University of Wisconsin (UW) in the mid-1980s was a breakthrough in liver transplantation, allowing safe organ preservation for up to 24–30 hours. The solution was empirically developed and contains a cocktail of substances such as lactobionate, raffinose, phosphate buffer, glutathione, colloids (hydroxyethyl starch), and various ions. Despite no definitive understanding of how the ingredients confer protection, the UW solution still remains the most widely used preservation solution. Other solutions, such as Eurocollins (with high glucose content and very high osmolarity) or Celsior (with mannitol, glutamate, and histidine buffer), have never gained wide acceptance among liver transplant surgeons.

As in the warm ischemic liver, ischemic preconditioning is potentially an effective approach to prevent preservation injury in the transplant setting. Although the impact on the outcome of clinical transplantation is not known, protection has been demonstrated in animal models and some pathways of protection have been identified. Several mechanisms appear to be similar to those of the normothermic setting, including involvement of adenosine or protein kinases. We have recently demonstrated that a moderate burst of oxygen radicals during the brief period of ischemic injury might be critical in triggering defense mechanisms. However, prospective randomized studies are needed to determine the clinical benefit and applicability of ischemic preconditioning in liver transplantation.

CONCLUSION

Ischemia and reperfusion injuries remain important in liver surgery, transplantation, and various forms of shock. Although some injurious mechanisms (e.g., apoptotic cell death) are common to the two types of ischemia injury—normothermic (warm) and cold (organ preservation)—critical differences exist, such as the resistance of hepatocytes to cold ischemic injury. Better understanding of the underlying mechanism of injury and some protective strategies, such as ischemic preconditioning, may lead to the discovery of innovative and effective protocols to prevent injury.

Acknowledgments

This manuscript was supported by grants from the National Institutes of Health (DK54048 to P. A. C.) and the Swiss National Science Foundation (SNF3200–061411). H. A. R. is supported by grants from the Roche Research Foundation (Switzerland) and the Olga Mayenfisch Foundation (Zurich).

See Also the Following Articles

Hepatic Circulation • Liver Transplantation • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Bilzer, M., and Gerbes, A. (2000). Preservation injury of the liver: Mechanisms and novel therapeutic strategies. *J. Hepatol.* **32**, 508–515.

Clavien, P. A. (1998). Sinusoidal endothelial cell injury during hepatic preservation and reperfusion. *Hepatology* **28**(2), 281–285.

Clavien, P. A., Harvey, P. R., and Strasberg, S. M. (1992). Preservation and reperfusion injuries in liver allografts. An

overview and synthesis of current studies. *Transplantation* **53**(5), 957–978.
Cursio, R., Guggenheim, J., Ricci, J., Crenesse, D., Rostagno, P., Maulon, L., Saint-Paul, M., *et al.* (1999). A caspase inhibitor fully protects rats against lethal normothermic liver ischemia by inhibition of liver apoptosis. *FASEB J.* **13**, 253–261.
Gujral, J., Bucci, T., Farhood, A., and Jaeschke, H. (2001). Mechanism of cell death during warm hepatic ischemia–reperfusion in rats: Apoptosis or necrosis? *Hepatology* **33**, 397–405.
Jaeschke, H. (1991). Reactive oxygen and ischemia–Reperfusion injury of the liver. *Chem. Biol. Interact.* **79**, 115–136.
Lemasters, J. (1999). Necrapoptosis and the mitochondrial permeability transition: Shared pathways to necrosis and apoptosis. *Am. J. Physiol.* **276**(1), G1–G6.
Lentsch, A., Kato, A., Yoshidome, H., McMasters, K., and Edwards, M. (2000). Inflammatory mechanisms and therapeutic strategies for warm hepatic ischemia/reperfusion injury. *Hepatology* **32**(2), 169–173.
Peralta, C., Closa, D., Hotter, G., Gelpi, E., Prats, N., and Rosello-Catafau, J. (1996). Liver ischemic preconditioning is mediated by the inhibitory action of nitric oxide on endothelin. *Biochem. Biophys. Res. Commun.* **229**(1), 264–270.
Selzner, M., Rudiger, H., Sindram, D., Madden, J., and Clavien, P. (2000). Mechanisms of ischemic injury are different in the steatotic and normal rat liver. *Hepatology* **32**, 1280–1288.
Sindram, D., Prote, R., Hoffmann, M., Bentley, R., and Clavien, P. (2001). Synergism between platelets and leukocytes in inducing endothelial cell apoptosis in the cold ischemic rat liver: A Kupffer-mediated injury. *FASEB J.* **15**, 1230–1232.

Liver Transplantation

PATRICIA A. SHEINER
New York Medical College and Westchester Medical Center

ascites The excessive accumulation of fluid in the peritoneal cavity.
cross-matching Test for compatibility between a donor's blood and a recipient's blood.
E Ag positive Result in which a patient's blood tests positive for the presence of hepatitis B "E" antigen.
encephalopathy A set of neurological symptoms that may occur due to damage to the brain (e.g., as a result of liver disease).
hepatorenal syndrome The development of renal failure in patients with acute or chronic liver failure in the absence of any other known cause of renal disease.
HLA matching Test that matches two people with the same human leukocyte antigens.
immunosuppression A condition in which the immune response is reduced or absent.
INR (International Normalized Ratio) A test that measures clotting impairment.

The first successful liver transplant was performed by Thomas Starzl at the University of Colorado in 1967. The patient lived approximately 1 year before dying of recurrence of the hepatocellular carcinoma for which she was transplanted. In the ensuing years, the mortality rate for patients undergoing liver transplant remained high, due to the complexity of the procedure and the lack of drugs that could selectively work against rejection without creating a state of overwhelming immunosuppression and associated life-threatening infections. In 1979, with the introduction of cyclosporine, a more selective calcineurin inhibitor, results following liver transplantation began to improve. At a 1983 National Institutes of Health consensus conference, it was declared that liver transplantation was the procedure of choice for patients with end-stage liver disease and should no longer be considered experimental.

INDICATIONS AND CONTRAINDICATIONS

The common causes of liver disease in children and adults are shown in Table I. Liver disease can be chronic or acute. Indications for transplantation in patients with chronic liver disease are the complications of end-stage liver disease, such as poor synthetic function, intractable ascites, encephalopathy, and hepatorenal syndrome. Hepatocellular carcinoma is a known complication of

Copyright 2004, Elsevier (USA). All rights reserved.

overview and synthesis of current studies. *Transplantation* **53**(5), 957–978.
Cursio, R., Guggenheim, J., Ricci, J., Crenesse, D., Rostagno, P., Maulon, L., Saint-Paul, M., *et al.* (1999). A caspase inhibitor fully protects rats against lethal normothermic liver ischemia by inhibition of liver apoptosis. *FASEB J.* **13**, 253–261.
Gujral, J., Bucci, T., Farhood, A., and Jaeschke, H. (2001). Mechanism of cell death during warm hepatic ischemia–reperfusion in rats: Apoptosis or necrosis? *Hepatology* **33**, 397–405.
Jaeschke, H. (1991). Reactive oxygen and ischemia–Reperfusion injury of the liver. *Chem. Biol. Interact.* **79**, 115–136.
Lemasters, J. (1999). Necrapoptosis and the mitochondrial permeability transition: Shared pathways to necrosis and apoptosis. *Am. J. Physiol.* **276**(1), G1–G6.
Lentsch, A., Kato, A., Yoshidome, H., McMasters, K., and Edwards, M. (2000). Inflammatory mechanisms and therapeutic strategies for warm hepatic ischemia/reperfusion injury. *Hepatology* **32**(2), 169–173.
Peralta, C., Closa, D., Hotter, G., Gelpi, E., Prats, N., and Rosello-Catafau, J. (1996). Liver ischemic preconditioning is mediated by the inhibitory action of nitric oxide on endothelin. *Biochem. Biophys. Res. Commun.* **229**(1), 264–270.
Selzner, M., Rudiger, H., Sindram, D., Madden, J., and Clavien, P. (2000). Mechanisms of ischemic injury are different in the steatotic and normal rat liver. *Hepatology* **32**, 1280–1288.
Sindram, D., Prote, R., Hoffmann, M., Bentley, R., and Clavien, P. (2001). Synergism between platelets and leukocytes in inducing endothelial cell apoptosis in the cold ischemic rat liver: A Kupffer-mediated injury. *FASEB J.* **15**, 1230–1232.

Liver Transplantation

PATRICIA A. SHEINER
New York Medical College and Westchester Medical Center

ascites The excessive accumulation of fluid in the peritoneal cavity.
cross-matching Test for compatibility between a donor's blood and a recipient's blood.
E Ag positive Result in which a patient's blood tests positive for the presence of hepatitis B "E" antigen.
encephalopathy A set of neurological symptoms that may occur due to damage to the brain (e.g., as a result of liver disease).
hepatorenal syndrome The development of renal failure in patients with acute or chronic liver failure in the absence of any other known cause of renal disease.
HLA matching Test that matches two people with the same human leukocyte antigens.
immunosuppression A condition in which the immune response is reduced or absent.
INR (International Normalized Ratio) A test that measures clotting impairment.

The first successful liver transplant was performed by Thomas Starzl at the University of Colorado in 1967. The patient lived approximately 1 year before dying of recurrence of the hepatocellular carcinoma for which she was transplanted. In the ensuing years, the mortality rate for patients undergoing liver transplant remained high, due to the complexity of the procedure and the lack of drugs that could selectively work against rejection without creating a state of overwhelming immunosuppression and associated life-threatening infections. In 1979, with the introduction of cyclosporine, a more selective calcineurin inhibitor, results following liver transplantation began to improve. At a 1983 National Institutes of Health consensus conference, it was declared that liver transplantation was the procedure of choice for patients with end-stage liver disease and should no longer be considered experimental.

INDICATIONS AND CONTRAINDICATIONS

The common causes of liver disease in children and adults are shown in Table I. Liver disease can be chronic or acute. Indications for transplantation in patients with chronic liver disease are the complications of end-stage liver disease, such as poor synthetic function, intractable ascites, encephalopathy, and hepatorenal syndrome. Hepatocellular carcinoma is a known complication of

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Common Causes of Liver Disease in Children and Adults

Chronic liver disease	**Acute fulminant hepatic failure**
Hepatitis B	Viral hepatitis
Hepatitis C	Drug-induced (e.g., halothane, sulfonamides)
Alcoholic cirrhosis	Metabolic liver disease (e.g., Wilson's disease, Reyes' syndrome)
Autoimmune disease	**Inborn errors of metabolism**
Biliary atresia	Glycogen storage disease
Primary biliary cirrhosis	α1-Antitrypsin disease
Primary sclerosing cholangitis	Wilson's disease
Cryptogenic cirrhosis	**Primary hepatic malignancies**
Budd–Chiari syndrome	Hepatoma with or without cirrhosis
Veno-occlusive disease	

end-stage liver disease but may also be seen in patients with chronic inflammation (e.g., hepatitis B or hematomochrosis) in the absence of cirrhosis. Transplant may be the treatment of choice in patients with unresectable hepatocellular carcinoma.

Acute liver failure is the new onset of liver disease with poor synthetic function and encephalopathy in a patient who 12 weeks previously had normal liver function. In patients with acute liver failure, transplant is necessary before the patient develops cerebral edema.

Prioritizing Patients on the Transplant Waiting List

The disparity between the number of organ donors and the number of patients waiting for transplants makes necessary the very difficult task of determining who should be given priority when a liver becomes available. The United Network for Organ Sharing has recently adopted a new scoring system for determining waiting list priority, based on the Model for End-Stage Liver Disease (MELD). MELD is an objective scoring system based on creatinine, bilirubin, and tests of coagulation (INR) that predicts which patients are more likely to die within 3 months without a transplant. The maximum MELD score is 40; the higher the score, the worse the prognosis is. Under the previous system, priority for transplant was based in part on the amount of time patients had spent on the waiting list. The impact of waiting time could therefore be detrimental to very sick patients who were referred to the transplant center late in the course of their disease. In an attempt to reduce the number of patients dying on the waiting list and to minimize the impact of "waiting time," patients are now prioritized according to their MELD score. Time spent on the list is used only among patients with the same MELD score.

Patients with hepatocellular carcinoma are particularly disadvantaged by the long waits for organs, because while they wait, their tumors may grow to a point that renders them ineligible for transplant. Tumor patients who are at a high risk for disease progression but who are still likely to have good outcomes with transplantation are given increasingly higher MELD scores, so they can be transplanted in a timely fashion.

One of the arguments against MELD is that it does not predict (and was not meant to predict) peri-transplant survival. The question that will remain unanswered until results of the new system can be studied is whether transplants will now be performed mainly in very ill candidates with poorer posttransplant survival.

Transplantation in Patients with Tumors

Before the development of advanced radiological techniques that could reveal small cancers during the pretransplant workup, tumors were often discovered after transplant in the explanted liver. Long-term outcomes in patients with incidental tumors were no different than outcomes in patients without incidental tumors and recurrent tumors were rare. It is now believed that transplantation in patients with tumors ≤ 3 cm is associated with an extremely low recurrence rate. Furthermore, transplantation in patients with either a single tumor < 5 cm or up to three tumors < 3 cm, without evidence of major vascular invasion or extrahepatic disease, is associated with excellent results.

Transplantation for Metabolic Problems

Infrequently, liver transplant is recommended for patients with essentially normal livers who have enzyme deficiencies that may be cured by transplant. Examples of such disorders include familial amyloidosis and oxaluria.

Transplantation for Alcoholic Liver Disease

Patients with end-stage liver disease secondary to alcoholic abuse undergo a rigorous evaluation. Patients are considered for transplant if they have a history of abstinence (most programs use 6 months as a minimum), have developed insight into their disease, maintain good family support systems, and have entered some sort of Alcoholics Anonymous or rehabilitation program. The recidivism rate after transplant is reported

to be approximately 20%; long-term survival, however, is good.

TYPES OF TRANSPLANTS

Transplants from Cadaveric Donors

The most common liver transplant is the orthotopic cadaveric transplant, in which a whole cadaveric liver with the suprahepatic and infrahepatic cava is transplanted into a recipient. What makes a good cadaveric liver? To be worthy of initial consideration, a cadaveric liver donor must only be ABO-compatible and somewhat size-matched to the recipient (i.e., a liver from a 200 lb. male is unlikely to fit into a 98 lb. female with a shrunken liver). HLA matching and crossmatching are not necessary. Other important factors in the evaluation of a potential cadaveric donor are the donor's past medical history (including history of alcoholism, diabetes, and cholesterol), length of hospital stay, and need for pressors after the declaration of brain death. Weight is important, because fatty livers do not function well. Liver function tests must be performed to evaluate damage to the liver. Viral studies are performed to rule out hepatitis C, hepatitis B, and human immunodeficiency virus (HIV).

The only absolute contraindication to use of a cadaveric donor liver is positive serology for HIV or hepatitis B surface antigen. A history of active or recent cancer in the donor (other than primary brain malignancies, which rarely metastasize to the liver) is also generally a contraindication. Livers from donors with poor liver function tests or with significant amounts of macrovesicular fat in their livers do not tolerate ischemia and do not function well. Although age is not an absolute contraindication to donation, and some centers transplant livers from donors as old as 80 years, livers from older donors do not tolerate long cold or warm ischemic times.

In children, whole cadaveric livers are often too large and the left lateral segment or left lobe alone may be transplanted. The reduction of the liver is performed *ex vivo*. The liver is kept in a basin of ice and saline to prevent rewarming while it is cut down to the right size. It is now recognized that if the left lateral segment or left lobe of a cadaveric liver is used for a child, careful splitting of the liver, with close attention to the blood supply of both sides, permits the right lobe to be used for transplantation in an adult patient. More recently, these "splits" are being performed *in vivo*, during the donor operation, so that cold and warm ischemia times are minimized.

Transplants from Living Donors

Living donor transplants in pediatric patients were first reported by Raia in Brazil in 1989 and by Broelsch in Chicago in 1991. Before the introduction of living donor liver transplants, mortality among children waiting for liver transplants was very high because small cadaver organs were extremely scarce. With the use of left lateral segments from living donors, mortality among children on the waiting list fell dramatically.

More recently, with persistent shortages of cadaver organs and continuing increases in waiting list mortality among adult transplant candidates, left lobes and right lobes have been taken from living donors for transplantation into adult recipients. The use of the left lobe was initially thought to be more desirable, as it was the smaller of the two sides and thus safer for the donor. Results with left lobe transplants were not as good, however, in part because left lobe transplants are technically more difficult but mainly because left lobes were inadequate for recipients with severe liver disease and portal hypertension. Right lobe transplants from living donors have therefore become the more viable option.

Not all recipients are good candidates for living donor donation, however. Patients with severe portal hypertension or renal failure or patients who are very ill need a full-size graft. On the other hand, it is particularly reasonable to consider living donor transplants for patients who might not receive an organ in a timely fashion under the present allocation system, such as those with larger tumors or those with poor quality of life but without corresponding poor liver function tests.

Any individual over the age of 18 can be considered for living donation. Potential living donors undergo careful screening and intensive evaluation of their general health as well as cardiac, pulmonary, and renal function. They are also evaluated for the presence of any unknown liver disease. Imaging studies (i.e., computed tomography scans, magnetic resonance imaging, and, when necessary, angiography) are performed to evaluate liver size (both right and left lobes) and arterial and venous anatomy. Donors with any significant co-morbidity or with inadequate liver volume to sustain either the recipient or themselves after resection are ruled out as candidates. They are also evaluated for their psychological ability to understand the risks and benefits of the procedure and for their desire to donate without coercion. The morbidity to the donor includes bile leaks from the cut surface, wound infection, small bowel obstruction, postoperative pain, and 8–12 weeks of disability. There is a small but very real risk of mortality to the donor (approximately 1/350).

COMPLICATIONS AND SIDE EFFECTS

Early Posttransplant Complications

Primary nonfunction is the term used to describe a situation in which a transplanted liver does not function and the patient needs to be urgently retransplanted. Primary nonfunction is diagnosed by increasingly elevated liver function tests and bilirubin, failure of the coagulapathy to normalize, and poor bile production. Renal failure and hemodynamic instability may also occur. Retransplantation is the only treatment.

Poor early graft function is diagnosed when the liver does not initially function well, with elevated liver function tests and coagulopathy for the first few days posttransplant. Poor early graft function resolves without long-term sequelae in the liver.

Technical complications are also seen early after transplant. Hepatic artery thrombosis occurs in approximately 4% of cases. Hepatic artery thrombosis (HAT) can cause abnormal liver function tests but is usually detected due to biliary complications. As the blood supply of the bile ducts is dependent on the artery, early HAT usually results in bile leaks secondary to necrosis of the anastomosis. Other complications may include liver abscesses and, rarely, liver failure. Most centers perform duplex ultrasonography on postoperative day 1 to screen for HAT. Early diagnosis before secondary complications occur can allow for successful surgical reconstruction of the artery and salvage of the liver. Late hepatic artery thrombosis (occurring more than 30 days posttransplant) usually results in bile duct strictures and retransplant is often required.

Portal vein thrombosis is a much less common but more devastating early posttransplant complication. The portal vein supplies 75% of the blood flow to the liver. Acute thrombosis results in liver necrosis, massive ascites, and hemodynamic instability. Rapid diagnosis may occasionally allow for surgical repair. More often, retransplantation is urgently needed.

Possible posttransplant complications also include bile leaks from the biliary anastomosis (which may or may not be associated with HAT).

Side Effects of Immunosuppression

The goal of immunosuppression is to prevent rejection of the transplanted organ. At the same time, immunosuppressive medications must be carefully and individually titrated to minimize side effects. Fortunately, as the interval from transplantation lengthens, most patients require less and less immunosuppression to maintain their freedom from rejection.

Immunosuppression has many major side effects. Immunosuppressive drugs in general make patients more vulnerable to infectious complications and certain malignancies. Steroids are well known to predispose patients to diabetes, hypertension, osteoporosis, and weight gain. Cyclosporine and tacrolimus, two of the mainstays of transplant immunosuppression, can cause hypertension, diabetes, neurotoxicity, nephrotoxicity, and decreased bone density.

In a study at the Mount Sinai Medical Center, Sheiner *et al.* examined 88 liver transplant recipients who had survived at least 5 years after their transplant and compared results from these patients with U. S. population statistics. After controlling for confounding variables, they found that hypertension and diabetes were significantly more prevalent in liver transplant recipients. Furthermore, more than 60% of the study patients had hypertension, making it the most common medical condition in the long-term transplant survivors. Half the patients with hypertension required more than one medication to control it.

Numerous authors have observed that transplant recipients have increased prevalences of impaired glucose tolerance and overt diabetes. In the Mount Sinai study, nearly 18% of the long-term survivors became diabetic. Almost 50% of the long-term survivors were moderately to severely obese, which may also have contributed to the increased prevalence of diabetes.

Weight gain and hypercholesterolemia have been well described in association with immunosuppressive therapy. Both conditions, common in the general public, were also common in the transplant recipients. Forty-eight percent were obese by standard criteria, compared to 33% of the general population. Although steroids may be a major risk factor for weight gain and hypercholesterolemia early after transplant, Sheiner *et al.* found that even long after transplant, when steroid doses were low, weight gain and high cholesterol levels persisted.

Despite the apparently high prevalence of obesity and hypercholesterolemia, however, the prevalence of heart disease in the liver transplant patients was not statistically different from that in the general population. This finding is in contrast to the accelerated atheromatous disease, associated with premature death, that has been reported in long-term survivors of heart and renal transplant.

Both steroids and cyclosporine are known to cause bone loss. Glucocorticoids directly suppress osteoblast function, inhibit intestinal calcium absorption, and stimulate renal calcium excretion. In heart transplant patients, cyclosporine was found to induce high bone

turnover. The Mount Sinai study patients were not screened for bone loss; osteoporosis was diagnosed only after symptoms became evident (i.e., back pain with radiological evidence of bone loss or compression fracture). Even without formal screening, there was a higher prevalence of fractures in female liver recipients than in the general population.

The association of peptic ulcers with corticosteroids was first reported more than 50 years ago. Researchers who have reported an increase in ulcer disease in transplant recipients have suggested that anti-ulcer prophylaxis be routine. There is evidence that the risk of ulcers may be greatly increased when certain medications, such as nonsteroidal anti-inflammatory drugs (NSAIDs), are taken with steroids. Liver transplant recipients are generally advised not to take NSAIDS, but mainly because of the associated nephrotoxicity.

Chronic immunosuppression can also cause renal toxicity and renal failure. In liver recipients followed for more than 2 years posttransplant, researchers in Sydney found elevated serum creatinine levels, but no patient required dialysis or renal transplantation. Among 68 pediatric heart recipients from Pittsburgh who survived >5 years, 2 required renal transplant. Results were similar in Mount Sinai's long-term survivors. Although creatinine clearance rates were generally lower after transplant, only 4 of 88 patients developed chronic renal failure requiring hemodialysis. It must be remembered, however, that the patients had already survived 5 years and were probably a particularly good risk group.

Finally, Mount Sinai analyzed the incidence of cancer in all their liver recipients (not just the 5-year survivors) to see whether malignancies were a significant source of early death. In fact, most patients who developed a malignancy went on to survive >5 years. Reported cancer incidences in transplant patients range from 4 to 18%, with the incidence apparently related to the duration of immunosuppression exposure. The most common malignancies in the Mount Sinai population were non-Hodgkin's lymphoma and nonmelanotic skin cancers. The incidences of these were similar to those reported in the general transplant literature.

The increased incidence of certain *de novo* malignancies (i.e., non-Hodgkin's lymphoma and nonmelanotic skin cancers) in transplant patients and their association with immunosuppression is well known. Direct mutagenic effects of the immunosuppressive drugs, length of exposure to these drugs, and/or decreased immune surveillance (allowing activation or proliferation of oncogenic viruses) may be responsible.

RECURRENT DISEASE AFTER TRANSPLANT

Hepatitis C

Hepatitis C virus is now the major cause of end-stage liver disease leading to transplantation. Unfortunately, hepatitis C recurs in the new liver. Although most studies have yet to show a difference in survival between patients transplanted for hepatitis C and those transplanted for other diseases, recurrent disease is a major source of morbidity and mortality after transplant.

Given the impact of recurrent disease not only on survival but on quality of life as well, it is important to look at ways to prevent or diminish the severity of recurrent disease. Treatment of recurrent hepatitis C after transplant with interferon has not been shown to be effective. Newer approaches, such as combination therapy with interferon and ribavirin or pegylated interferon, may yield better response rates but in the author's experience has not been well tolerated and the long-term effect is still unknown.

Perhaps, until better antiviral medications become available, the effects of immunosuppressive agents should be examined more closely and immunosuppression regimens should be tailored to reduce the risk of recurrent hepatitis C. It is has become increasingly clear that immunosuppression, especially steroids, plays a role in the recurrence of hepatitis C. Furthermore, augmented immunosuppression for treatment of rejection has been shown to be a particular risk factor for recurrence. Given that steroids seem to be the main culprit in recurrence of hepatitis C, the goal of transplant immunosuppression in hepatitis C patients may be the rapid withdrawal of steroids or steroid-free immunosuppression. Some centers have begun to study the use of induction immunosuppression with the rapid withdrawal of steroids in their hepatitis C population.

Hepatitis B

Results after transplantation for hepatitis B were initially poor due to recurrent hepatitis in the allograft. In fact, it was thought that patients who were DNA positive or E antigen (Ag) positive were at particularly high risk for poor transplant outcomes and these patients were generally excluded from candidacy. Eventually, however, it became apparent that passive immunization with monthly infusions of hepatitis B immunoglobulin (HBIg) improves survival, with recurrence rates ranging from 5 to 30% (recurrence is reportedly increased in the high-risk patients mentioned above). Lamivudine may be used in conjunction with

HBIg or as a rescue for patients who recur despite prophylaxis. Adefovir, a newer antiviral agent, may provide further improvement in long-term results but is still in clinical trials.

Autoimmune Hepatitis

Autoimmune hepatitis is another disease that can recur in the liver allograft despite immunosuppression. Although steroid withdrawal after transplant is generally considered beneficial, it has been suggested that steroids may need to be maintained in patients transplanted for autoimmune hepatitis, to prevent recurrence. Recurrence is treated with steroids and/or azathioprine or 6-mercaptopurine. Retransplant for recurrent autoimmune hepatitis has been necessary in a small proportion of patients.

Primary Biliary Cirrhosis and Primary Sclerosing Cholangitis

Primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC) are primary cholestatic disorders of the liver. Both diseases can recur after transplant. Biliary strictures from ischemia might mimic PSC, however, and chronic rejection might mimic PBC, making the diagnosis of recurrent disease more difficult.

THE FUTURE OF LIVER TRANSPLANTATION

Further advances in immunosuppression and in tolerization will help improve the results of transplantation while decreasing the risk of side effects from long-term immunosuppression.

Improvement in the treatment of hepatitis C, prevention of recurrent disease, and development of an effective immunoprophylaxis are necessary. Further refinements in the techniques of living donor surgery, as well as educational programs to increase public awareness of the importance and value of organ donation, will also be important.

Acknowledgment

The author thanks Nancy Ehrlich Lapid for her editorial assistance.

See Also the Following Articles

Cholangitis, Sclerosing • Cholestatic Diseases, Chronic • Cirrhosis • Fulminant Hepatic Failure • Hepatic Encephalopathy • Hepatitis C • Hepatocellular Carcinoma (HCC) • Hepatorenal Syndrome • Transplantation Immunology

Further Reading

Ginns, L. C., Cosimi, A. B., and Morris, P. J. (eds.) (1999). "Transplantation." Blackwell Science, Malden, MA.

Mor, E., Kaspa, R. T., Sheiner, P., and Schwartz, M. (1998). Treatment of hepatocellular carcinoma associated with cirrhosis in the era of liver transplantation. *Ann. Intern. Med.* **129**, 643–653.

Schiano, T. D., Kim-Schluger, L., Gondolesi, G., and Miller, C. M. (2001). Adult living donor liver transplantation: The hepatologist's perspective. *Hepatology* **33**, 3–9.

Sheiner, P. A., Magliocca, J. F., Bodian, C. A., Kim-Schluger, L., Altaca, G., Guarrera, J. V., Emre, S., Fishbein, T. M., Guy, S. R., Schwartz, M. E., and Miller, C. M. (2000). Long-term medical complications in patients surviving ≥ 5 years after liver transplant. *Transplantation* **69**, 781–789.

Liver, Anatomy

CHEN LIU AND JAMES M. CRAWFORD
University of Florida College of Medicine

bile canaliculus Minute canal between the abutting hepatocytes; passes bile from hepatocyte to the biliary system.
bile duct Structure that passes bile from the liver to the gallbladder or to the duodenum.
canals of Hering Structures lined by both cholangiocytes and hepatocytes; the most peripheral portion of the biliary system.
cholangiocyte Epithelial cell that lines the bile ducts.
Glisson's capsule External capsule of the liver.
hepatic acinus Triangular hepatic parenchymal mass with a terminal hepatic vein at the apex and portal venous channels at the base.
hepatocyte Polygonal epithelial parenchymal cells of the liver.
Kupffer cell Hepatic macrophage that presents in the lumen of hepatic sinusoids.
portal tract (portal triads) Structures composed of bile duct, hepatic artery, and portal vein, embedded within connective tissue.
sinusoid Channel between the hepatic cords.
space of Disse Region between the sinusoidal side of the hepatocytes and the sinusoidal wall.
stellate cell Located in the space of Disse; has long cytoplasmic processes.

The liver is the single largest organ in the human body. It serves as the primary regulatory site for metabolism, processing nutrients entering via the splanchnic circulation for controlled distribution to extrahepatic tissues via the systemic circulation. It synthesizes essential circulating proteins, including albumin and selected clotting factors, and is a major processor of circulating proteins such as lipoproteins. It is responsible for the detoxification of many endogenous and exogenous substances, including bilirubin and xenobiotics. In addition to urinary secretion, hepatic secretion of bile constitutes the other major route for elimination of waste substances, including cholesterol, bilirubin pigments, copper, and amphiphilic xenobiotics. The liver may also be critical in modulating the immune system through its resident macrophages and lymphocytes. The anatomic structure of the liver is well suited for these functions: it receives almost the entirety of blood derived from the intestines as well as a component of systemic arterial blood, has an extensive perfusion bed for contact of blood with hepatic cells, and then delivers its effluent to the inferior vena cava.

LIVER DEVELOPMENT

The liver is derived from the endodermal lining of the most distal portion of the foregut during the third to fourth weeks of gestation, arising as a diverticular bud off the ventral portion of the foregut. The proliferating epithelial cells in this diverticulum, termed hepatoblasts, invade the immediately adjacent septum transversum, a mesenchymal plate that separates the thoracic cavity from the abdominal cavity. The hepatoblasts form cohesive cords and plates within the mesenchyme of the septum transversum, coincident with the ingrowth of a network of vascular tributaries derived from the vitelline vein and a second vascular system derived from the hepatic artery. These vascular networks run in parallel and are invested with a mesenchyme, forming the so-called portal tracts of the liver. The terminal vascular channels run between the hepatoblast cords, forming sinusoids; scattered mesenchymal cells grow between the endothelial cells of the sinusoids and the hepatic cords to form the connective tissue cellular components of the hepatic sinusoid.

The foregut diverticulum also develops a patent luminal channel that will become the extrahepatic biliary tree. Within the liver, two layers of hepatoblasts immediately adjacent to the mesenchyme of the portal tracts mature into a cuboidal biliary-type epithelial layer. This paired epithelial layer is circumferential around the portal tract mesenchyme, and is termed the "limiting plate." This circumferential architecture reorganizes into a series of ductular channels that tether the parenchyma to mature bile ducts running along the axis of the portal tract. The mature portal tract is then achieved, containing portal vein, hepatic artery, and bile duct (Fig. 1). This pattern of maturation moves from the hilum outward and continues after birth; the biliary channels are at all times in continuity with the lumen of the developing extrahepatic biliary tree.

 Copyright 2004, Elsevier (USA). All rights reserved.

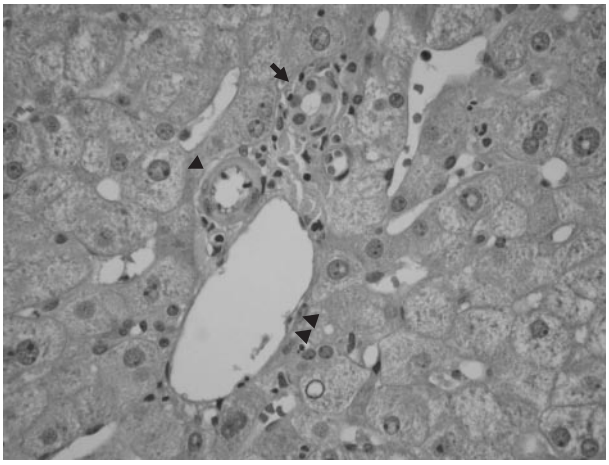


FIGURE 1 Micrograph of normal liver. The portal tract contains portal vein (double arrowheads), hepatic artery (single arrowhead), and bile duct (arrow).

The primordial liver separates from the septum transversum, leaving behind a mesenchymal remnant that will become the diaphragm. As the redundant major vessels of the embryonal circulation reorganize into the mature abdominal vascular system, the fetal liver receives some vascular supply from the umbilical vein through a vitelline vein remnant, while also receiving blood from the portal vein and hepatic artery. Most of the umbilical vein blood enters directly into the inferior vena cava through the short ductus venosus in the liver hilar area. At the time of birth, the mature hepatic circulation is achieved through closure of the ductus venosus and atrophy of the vitelline vein to produce the nonpatent falciform ligament connecting the hilum of the liver to the umbilicus.

The developing liver is an important metabolic center in the developing fetus; as the largest organ, it represents 10% of the fetal weight, falling to around 5% of body weight at the time of birth. The liver is also the main hematopoietic organ in fetal life, with hematopoietic elements constituting half of the liver mass at midterm, falling to almost zero at term. Although the hepatic organization and development begin in the third week of gestation, the functional and architectural maturation is not complete at birth and continues well into the postnatal period. The liver doubles in size in the first year of life alone; full maturation of the liver is not achieved until about 15 years of age.

GROSS ANATOMY

The liver is situated in the right hypochondrium and epigastric regions under the rib cage and extends from the right fifth intercostal space at the midclavicular line to just below the costal margin. The mean weight of the liver in the adult man and woman is 1800 g and 1400 g, respectively. It accounts for approximately 1.8–3.1% of body weight in adults. The liver is wedge shaped and is covered by Glisson's capsule (hepatic capsule), which is composed of the visceral peritoneum (a mesothelial epithelial layer), and an underlying shallow network of connective tissue. From Glisson's capsule, fibrous septa project inward into the hepatic parenchyma to a depth of about 0.5 cm. The superior convexity of the liver is not completely invested with peritoneum; a reflection of peritoneum onto the diaphragm produces a triangular "bare spot" on the most superior portion of the liver. The inferior surface of the liver has a deep concave recess with a central hilus, referred to as the liver hilum, or "porta hepatis," where major blood vessels, bile ducts, lymphatic vessels, and nerve trunks enter or leave the liver. To the right of the hilum is the gallbladder fossa, a concavity in which lies the gallbladder.

The liver is connected to the diaphragm and abdominal cavity by five ligaments. The anterior falciform ligament is a crescentic fold of peritoneum extending from the diaphragm to the umbilicus. The round ligament (ligamentum teres) is the fibrous remnant of the vitelline vein, forming the inferior border of the falciform ligament. The superior coronary ligament is composed of the peritoneal reflections onto the diaphragm. The right and left triangular ligaments are the respective lateral borders of the coronary ligament. Hence, the liver is held in place only by ligaments to the diaphragm and anterior abdominal wall, and by vasculature to the inferior vena cava posteriorly, and through the hilum to the posterior abdominal wall inferiorly. Mobilization of the liver during surgery is therefore readily accomplished.

The midline falciform ligament divides the liver into the historical right and left lobes. The protuberant caudate lobe projects posteriorly between the sulcus for the inferior vena cava and the fissure for the ligamentum venosus. The quadrate lobe lies inferiorly between the gallbladder fossa and the fissure for the round ligament.

VASCULAR ANATOMY

Liver anatomy is better defined by its vascular beds. In 1897, Cantlie noted that the right and left portal veins fed the right and left lobes, divided along a plane several centimeters right of midline, extending anterior to posterior from the gallbladder fossa to the vena cava. This plane is referred to as Cantlie's line, and constitutes the watershed between the right and left hepatic

circulations. Couinaud further described the functional anatomy of the liver and demonstrated that the vascular distribution network further divided the liver into four sectors and eight segments. By this nomenclature, the liver is divided by vertical and oblique planes, or scissurae, defined by branches of the three main hepatic veins, and a vertical plane separating the right and left portal vein branches. These four sectors are further divided by the transverse scissura into eight segments, named clockwise in the frontal plane, with the segment I corresponding to the caudate lobe; the quadrate lobe is viewed as part of the right lobe. Each segment is an independent functional unit supplied by a single vascular trunk containing both portal vein and hepatic artery branches. These anatomic subdivisions define the dissection planes for partial hepatic resections.

The liver receives approximately 25% of the cardiac output. Venous blood containing nutrients from the gastrointestinal tract is brought to the liver by the hepatic portal vein, which accounts for 75% of the liver blood supply. The hepatic artery branches off the celiac artery and provides 25% of hepatic blood. Branches of the portal vein and hepatic artery ramify within progressively smaller portal tracts, ending in terminal portal veins and terminal hepatic arteries. The volume of liver parenchyma supplied by terminal vasculature is on the order of 2 to 3 mm^3, giving rise to so-called microarchitectural units. Detailed structural studies have shown that there are between 18 and 20 orders of branches of portal tract circulation, producing an estimated 400,000 to 500,000 microarchitectural units in the adult liver.

The portal veins give rise to small venous channels that branch off at many points along the ramification of the portal vein system; these channels penetrate the liver parenchyma as the primary blood supply to sinusoids. The hepatic artery supplies the mesenchymal stroma of portal tracts, the peribiliary vascular plexus, the vasa vasorum of major blood vessels (including the vasa vasorum of major hepatic veins), and Glisson's capsule. Postcapillary blood from these distributions may enter the portal vein, the parenchyma directly, or the hepatic veins. Arterial blood may also supply sinusoids directly, although this does not appear to be a major component of arterial blood flow in humans.

Regardless of how portal vein and hepatic arterial blood reaches the parenchyma, blood percolates through the sinusoidal network to collect within terminal hepatic veins. The terminal hepatic veins collect into a conducting hepatic venous system, which drains ultimately into three major hepatic venous branches: right, middle, and left hepatic veins. The left and middle hepatic veins often join together, forming a common trunk before entering the inferior vena cava. There are also variable short venous segments that drain the posterior surface of the liver directly into the inferior vena cava.

BILE DUCTS

Bile is secreted by hepatocytes into bile canaliculi, small channels between abutting hepatocytes delineated along their length by tight junctions between hepatocytes. Bile canaliculi ramify throughout the hepatic parenchyma, depositing bile within the most peripheral portion of the biliary tree, the canals of Hering. Each canal of Hering is a "trough"-like channel bounded on one side by bile duct epithelial cells (cholangiocytes) and on the other by hepatocytes. Canals of Hering may lie circumferentially along the interface of the hepatocellular parenchyma and portal tract mesenchyme, or may penetrate the parenchyma to up to one-third the distance to the terminal hepatic vein. From a fluid dynamic standpoint, the canal of Hering provides a broad entry point for flow of bile from bile canaliculi into the biliary tree.

Canals of Hering connect to bile ductules, circular channels fully surrounded by cholangiocytes and a basement membrane. Ductules are the bridges between canals of Hering and the terminal bile ducts within portal tracts. In keeping with the variable canal of Hering anatomy, bile ductules may have both intraportal and short intraparenchymal segments. Terminal bile ducts collect ductular bile and convey it downstream to conducting bile ducts, and finally to the hepatic hilum for drainage into the gut through the extrahepatic biliary tree.

Terminal and conducting bile ducts run in parallel with hepatic arteries in a 1 : 1 relationship. The diameter of the bile duct (measured from basement membrane to basement membrane) is essentially equal to the diameter of the companion hepatic artery along most of the intrahepatic portal tree; the largest bile ducts are accompanied by multiple smaller hepatic artery branches. Hence, identification of hepatic arteries without companion bile ducts is indication of bile duct loss; a very helpful sign in assessing microscopic liver damage.

LYMPHATIC SYSTEM

Lymphatic vessels are abundant in portal tracts and are thought to begin as blind-ended channels in the terminal portal tracts. Extrasinusoidal fluid from the parenchyma draining retrograde into the portal tract mesenchyme collects in these lymphatics. Lymph travels downstream through a rich plexus of lymphatic

channels in the portal tree, converging around the major hepatic arteries, bile ducts, and portal vein to form 12 to 15 lymph vessels draining at the portal hilum. From there, lymph flows to the foraminal node at the epiploic foramen, the superior aortic nodes, and the ipsilateral aortic nodes. The lymph ultimately drains into the cisterna chyli superior to the celiac artery root, and then to the thoracic duct. Lymph collecting in lymphatics of Glisson's capsule drains either via hilar lymphatics or posteriorly and superiorly into lymphatics exiting along the hepatic veins and the inferior vena cava.

INNERVATION

The liver is innervated by both sympathetic and parasympathetic nerve fibers. They both contribute to the perivascular nerve plexus within the portal tree. Innervation of hepatocytes and sinusoids is relatively sparse and not uniform between lobules. Both sympathetic and parasympathetic nerve fibers influence hepatic metabolic activity, and play a role in regulating fluid secretion by cholangiocytes lining the biliary tree. This regulation may not be essential for normal liver function, because the transplanted liver apparently functions in the absence of regenerated nerve fibers. However, neural regulation may play a critical role in modulating cellular proliferation following hepatic injury, raising questions about the biology of liver damage following hepatic transplantation.

MICROSCOPIC ANATOMY

Microscopic Organization

Unlike the prototype nephron in the kidney, the liver does not have a defined smallest, structurally distinct, "self-sufficient" unit. Because terminal hepatic veins lie roughly 1 mm from portal tracts on several sides, nineteenth century microscopists viewed the liver "lobule" as having portal tracts at the periphery of a hexagon and terminal hepatic veins at the center. This classic lobule enables reference to parenchyma that is "centrilobular," "midzonal," and "periportal," a terminology that is well established in current practice.

In 1954, Rappaport demonstrated that blood flow from the penetrating venous channels flowed along parallel sinusoids to converge on the terminal hepatic veins, thereby defining a triangular hepatic "acinus" as having penetrating portal venous channels at the base and a terminal hepatic vein at the apex. The parenchyma of the hepatic acinus is divided into three zones, zone 1 being the closest to the vascular supply and hence best oxygenated, zone 3 abutting the terminal hepatic venule, and zone 2 being intermediate. This zonation is of considerable metabolic consequence, because a lobular gradient exists for oxygenation and solute content of sinusoidal blood and for the activity of many hepatocellular enzymes. This is called "zonation," or the "lobular gradient." Many forms of hepatic injury exhibit a zonal distribution, owing to variable exposure and susceptibility to injury of hepatocytes along the sinusoid. These microanatomic concepts have continued to undergo refinement, including efforts to define hepatic microanatomy on the basis of biliary drainage. The key concept is that hepatocytes along the sinusoids are not uniform, but behave variably on the basis of their position along the lobular gradient.

Microscopic Structure

The hepatic parenchyma is organized into cribriform, anastomosing sheets or plates of hepatocytes, seen in two-dimensional histologic sections as cords of cells (Fig. 1). The hepatocyte cords immediately abutting the portal tract are referred to as the limiting plate, forming a discontinuous rim around the mesenchyme of the portal tract. Between the hepatocyte cords is the sinusoid, conducting blood from the penetrating venous channels of the portal vein and hepatic arterial system to terminal hepatic venules. Hepatocytes are thus bathed on two sides by well-mixed portal venous and hepatic arterial blood.

Sinusoids are lined by a unique fenestrated endothelium that lacks a basement membrane. This is unlike any other vascular bed in the body. The absence of a perfusion barrier permits free exchange of plasma proteins with the perisinusoidal space, and hence with hepatocytes. This is the so-called space of Disse, between the sinusoidal endothelium and hepatocytes. Instead of a basement membrane, delicate connective tissue fibers traverse the space of Disse.

Hepatocytes

Hepatocytes constitute 80% of the cell population of the liver. There are approximately 100 billion of hepatocytes in an adult liver. An individual hepatocyte is a polygonal epithelial cell approximately 25 μm in diameter. The nucleus is centrally located and contains conspicuous nucleoli (Fig. 2). With increasing age, up to 20% of hepatocytes are binucleated and tetraploid. The hepatocyte is a polarized cell consisting of two functionally specialized membrane domains—the basolateral domain, which faces the perisinusoidal space of Disse and the lateral region between adjacent hepatocytes, and the apical domain, which forms the bile canaliculus midway between adjacent hepatocytes. Tight junctions

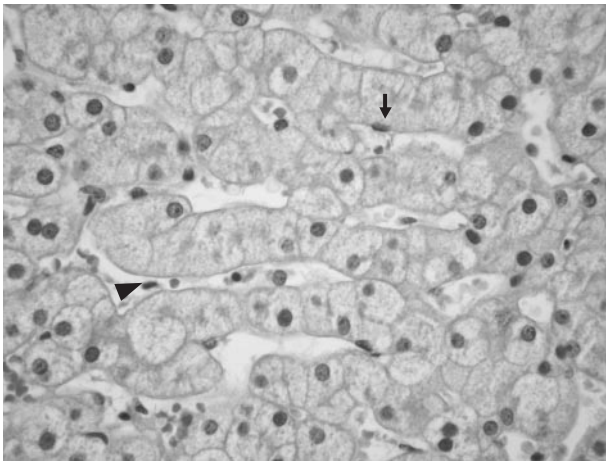


FIGURE 2 Micrograph of normal hepatocytes. The hepatocytes form hepatic cords and the sinusoids are lined with endothelial cells (arrow) and Kupffer cells (arrowhead).

separate the bile canaliculus from the lateral space between hepatocytes, thereby constraining bile to flow within the canalicular space. The basolateral domain constitutes about 88% of the hepatocyte plasma membrane surface, and the apical domain about 12%. The basolateral domain is occupied by transport and endocytic/exocytic events pertaining to exchange with sinusoidal plasma. The apical domain is responsible for the secretion of bile.

The hepatocellular cytoplasm contains numerous organelles, including rough endoplasmic reticulum (RER), smooth endoplasmic reticulum (SER), Golgi complexes, mitochondria, lysosomes, and peroxisomes. The RER is mostly involved in synthesis of membrane proteins or soluble proteins for extracellular secretion. The SER is rich in enzymes for drug metabolism, biosynthesis of cholesterol and conversion of cholesterol to bile acids, and the cytochrome P450 enzyme system. The Golgi complex of hepatocytes, located in the vicinity of the bile canaliculus, is necessary for processing of membrane proteins and secretory proteins prior to final targeting, including terminal glycosylation, and also plays a role in synthesis of very-low-density lipoproteins. There are approximately 1000 mitochondria in an individual hepatocyte, constituting between 13 and 20% of the cytoplasmic volume. Mitochondria are a major metabolic compartment, including the tricarboxylic acid cycle, fatty acid oxidation, and half of the urea cycle. They provide the essential oxidative energy for hepatic metabolism. Mitochondria are much smaller and more numerous in hepatocytes of acinar zone 3 compared to those in zone 1. Lysosomes in the hepatocyte are responsible for degradation of endocytosed proteins and are capable of discharging their contents into bile as a normal mechanism for eliminating some substances such as copper. Lysosomes can accumulate lipofuscin as a long-term by-product of autophagy, ferritin in iron-overload states, copper in Wilson disease, and lipids in a number of lysosomal storage disorders. There are approximately 300 to 600 peroxisomes in an individual hepatocyte. Peroxisomes contain oxidases and catalases and are responsible for 20% of hepatocellular oxygen consumption. Omega oxidation of long-chain fatty acids also occurs in peroxisomes.

The cytosol of the hepatocyte is occupied 80% on a volume basis by the aforementioned organelles. The remainder is a dense fluid containing cytosolic enzymes and a dynamic three-dimensional network of cytoskeletal proteins forming microfilaments, intermediate filaments, and microtubules. Each cytoskeletal element exists in a depolymerized and a polymerized state. Intermediate filaments are composed predominantly of cytokeratins 8 and 18, and are relatively stable; almost all microfilament protein is polymerized. In contrast, microfilaments and microtubules are highly dynamic structures, with rapid exchange of protein subunits between solubilized and polymerized states. The proportion of polymerized microfilaments and microtubules, and their location and polarity, are dictated on a moment-to-moment basis by the needs of the hepatocyte.

Hepatic Stellate Cells

Hepatic stellate cells lie within the space of Disse, and have extensions that wrap around the sinusoidal channel or extend into the lateral space between hepatocytes. Under normal conditions, they are scattered within the space of Disse and contain lipid droplets rich in vitamin A. They contain contractile filaments and are capable of regulating vascular tone within the sinusoid. They produce the delicate extracellular matrix of the space of Disse and secrete growth factors such as hepatic growth factor (HGF) or insulin-like growth factor-2 to regulate hepatocyte regeneration. When hepatic injury occurs, stellate cells are activated and rapidly proliferate, losing their lipid droplets in the process. In their activated state, they lay down abundant extracellular matrix and render the space of Disse fibrotic and relatively impermeant to plasma proteins. Activation of stellate cells is a major contributor to the development of liver cirrhosis.

Kupffer Cells

Scattered Kupffer cells of the monocyte–phagocyte system are attached to the luminal side of the endothelial cells (Fig. 2). These cells are the major macrophages of

the hepatic corpus and clear particulate matter and bacterial endotoxin from the splanchnic circulation. Hence, they are a major "filter" for intestinal blood before it enters the systemic circulation. Kupffer cells are a major source of cytokines in inflammatory states such as hepatitis, influencing in particular inflammatory injury of the subjacent sinusoidal endothelium and hepatocytes. Kupffer cells have an irregular stellate shape with bean-shaped nuclei. The cytoplasm of Kupffer cells often contains vacuoles and granular aggregates of ceroid pigments. The cells can actively proliferate in response to many inflammatory stimuli, including viral infections.

Lymphocytes

Liver-associated lymphocytes are abundantly present in the liver. It is estimated that there are approximately 1×10^{10} lymphoid cells in a normal human liver. So-called pit cells are present in the space of Disse; portal tracts also normally contain lymphocytes. Hepatic lymphocytes are heterogeneous in terms of the molecular markers on the cell surface. Natural killer (NK) cells and T lymphocytes are present in the liver cell population. Lymphocyte apoptosis is high in the liver, prompting the hypothesis that T lymphocytes are normally cleared by the liver and undergo apoptosis. The liver thus may be an important immunoregulatory organ through its selective removal of T lymphocytes from the circulation. The details of the physiological function of liver-associated lymphocytes are still unclear.

Bile Duct Epithelial Cells

The epithelium of the biliary tree, including the canals of Hering, is a continuous layer of tightly coupled cells, the so-called bile duct epithelial cells, or cholangiocytes. The smaller bile ducts are lined by cuboidal cells. The larger bile ducts are lined by columnar cells and may include goblet cells. Bile duct epithelial cells have prominent Golgi complexes, numerous cytoplasmic vesicles, and short luminal microvilli. Bile duct epithelial cells rest on a basement membrane and the biliary channel is highly impermeable to fluid leakage under normal conditions. The nuclei of the bile duct epithelial cells are located near the basement membrane. Bile duct epithelial cells are responsible for regulated secretion of approximately 30 to 40% of the volume of bile as a bicarbonate-rich fluid, and hence are significant modifiers of hepatic bile en route to the alimentary tract. Bile acids may be absorbed by the biliary epithelium, and so there is opportunity for intrahepatic recirculation of bile acids. This may be of value when the biliary tree is obstructed, allowing removal of bile acids from bile for redirection to urinary secretion.

CONCLUSION

Although the macroscopic anatomy of the liver is somewhat mundane, the microscopic anatomy is uniquely suited for its specialized physiologic functions: serving as a major metabolic organ for the entire body, synthesizing and secreting a critical set of plasma proteins, processing circulating plasma proteins, clearing noxious matter from the splanchnic circulation, and elimination of waste products via biliary secretion. It is small wonder that only the brain and kidneys receive more blood flow than the liver, and that the liver is a major producer of body heat.

See Also the Following Articles

Biliary Tract, Anatomy • Circulation, Overview • Gastrointestinal Tract Anatomy, Overview • Hepatic Circulation • Hepatocytes • Lymph, Lymphatics, and Lymph Flow • Lymphocytes

Further Reading

Arias, I. M., Che, M., Gatmaitan, Z., Leveille, C., Nishida, T., and St. Pierre, M. (1993). The biology of the bile canaliculus. *Hepatology* **17**, 318–329.

Burt, A. P., and Day, C. P. (2001). Pathophysiology of the liver. *In* "Pathology of the Liver" (R. N. M. MacSween, A. D. Burt, and B. C. Portmann, eds.), 4th Ed., pp. 67–106. Churchill Livingstone, London.

Comparini, L. (1969). Lymph vessels of the liver in man. Microscopic morphology and histotopography. *Angiologica* **6**, 262–274.

Crispe, I. N. (2003). Hepatic T cells and liver tolerance. *Nat. Rev. Immunol.* **3**, 51–62.

Ekataksin, W., and Wake, K. (1997). New concepts in biliary and vascular anatomy of the liver. *Prog. Liver Dis.* **15**, 1–30.

Ludwig, J., Ritman, E. L., LaRusso, N. F., Sheedy, P. F., and Zumpe, G. (1998). Anatomy of the human biliary system studied by quantitative computer-aided three-dimensional imaging techniques. *Hepatology* **27**, 893–899.

MacSween, R., and Scothorne, R. (2001). Developmental anatomy and normal structure. *In* "Pathology of the Liver" (R. N. M,. MacSween, A. D. Burt, and B. C. Portmann, eds.), 4th Ed., pp. 1–66. Churchill Livingstone, London.

Rappaport, A. M. (1958). The structural and functional unit in the human liver (liver acinus). *Anat. Rec.* **130**, 673–689.

Saxena, R., Zucker, S., and Crawford, J. (2003). Anatomy and physiology of the liver. *In* "Hepatology" (D. Zakim and T. D. Boyer, eds.), 4th Ed., pp. 1–30. W. B. Saunders, Philadelphia.

Theise, N. D., Saxena, R., Portmann, B. C., Thung, S. N., Yee, H., Chiriboga, L., Kumar, A, and Crawford, J. M. (1999). The canals of Hering and hepatic stem cells in humans. *Hepatology* **30**, 1425–1433.

Liver, Development

NITIKA ARORA-GUPTA AND CURT H. HAGEDORN
Emory University School of Medicine

fibroblast growth factors Proteins expressed by the cardiac mesoderm; extracellular signals for inducing liver progenitor cells.

homeobox Segment of nucleic acid base pairs that is conserved in genes involved in controlling the development of an organ or structure that produces factors that control gene expression.

phrygian cap Congenital gallbladder anatomical abnormality, resulting in a conical cap shape (the traditional ancient Phrygian headgear) at the gallbladder tip.

Obtaining genetic knowledge about the development and morphogenesis of the liver has been a challenging topic. Because the functions of the liver are essential to life, and many developmental mutations are lethal to the embryo. Despite the challenges, progress has been made on the molecular and cellular mechanisms of liver development. Transcription factors and signal transducers play an important role in the development of hepatoblasts from endodermal progenitor cells and in their continued growth and differentiation into hepatocytes and cholangiocytes. Abnormalities of these events may lead to developmental disorders of the liver, its vasculature, or the biliary system.

DEVELOPMENT OF THE LIVER PARENCHYMA

Functions of the liver include processing nutrients delivered by the vitelline umbilical vein or portal vein, synthesizing critical blood metabolites and serum proteins, detoxification, and supporting hematopoiesis during gestation. The liver develops from an outgrowth from the ventral foregut endoderm during the third week of intrauterine life and the biliary system originates from further outpouching of the liver bud. Early development can be divided into three stages that involve differentiation of the progenitor cells into hepatoblasts, growth of hepatoblasts, and ultimately morphogenesis of the liver.

Embryology and Gene Regulation

Early liver development starts with formation and patterning of the endoderm. The ventral endoderm differentiates into thyroid, lung, liver, and pancreas and the dorsal endoderm differentiates into pancreas and intestines (Fig. 1). The events leading to differentiation of the ventral endoderm to hepatoblasts, then the liver bud, and eventually a formed liver are due to liver specific transcriptional programs and external signaling events originating from the cardiac mesoderm. The *forkhead box A* (*FoxA*) genes are highly regulated and control gut development throughout the animal kingdom. Fox A proteins regulate almost all liver-specific genes as well as a number of genes specific to the lung and pancreas. These proteins have a 110-amino-acid DNA-binding domain and regulate gene transcription. In addition to the FoxA proteins, a subclass of the GATA zinc finger transcription factors is important for liver-specific gene expression (e.g., albumin).

Model systems have demonstrated that the cardiogenic mesoderm, which is transiently adjacent to the prospective hepatic endoderm, provides critical extracellular signals for inducing liver progenitor cells (Fig. 1). The cardiac mesoderm expresses a minimum of

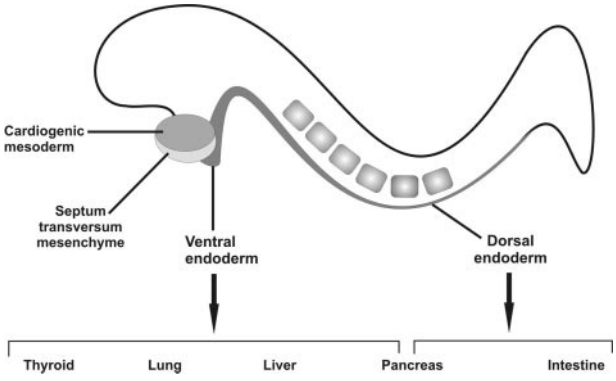


FIGURE 1 Formation and patterning of the mouse endoderm into the visceral organs at the sixth somite stage (approx. 8.25 days of gestation). The cardiogenic mesoderm sends extracellular signals to the ventral endoderm and promotes formation of the liver bud. Without the cardiogenic mesoderm, the endoderm differentiates into pancreatic tissue by default. The septum transversum initiates formation of separate body chambers and eventually forms the diaphragm.

 Copyright 2004, Elsevier (USA). All rights reserved.

three different fibroblast growth factor (FGF) proteins and the ventral endoderm expresses at least two of the four FGF receptors. Inhibitory FGF signaling blocks hepatic induction by the cardiogenic mesoderm whereas FGF treatment of endoderm explants induces early hepatic gene expression. These events occur at the seven- to eight-somite stage of the mouse embryo (approximately 8.3 days of gestation). The septum transversum mesenchyme also plays a signaling role by expressing bone morphogenetic proteins 2 and 4 (transforming growth factor-β family) before and during the induction of hepatoblasts within the endoderm. Without these signaling events, the ventral foregut endoderm differentiates by default into pancreatic tissue.

Another major step in liver development is the organization of hepatoblasts within the endoderm into a liver bud. Two genes, the hematopoietically expressed and Prospero-related homeoboxes, *Hex* and *Prox 1*, which encode transcriptional regulatory proteins, are required for early steps in liver bud development. Extracellular matrix proteins that interact with β1 integrin receptors also play a role in forming the liver bud, which occurs at day 9 to 10 of embryonic gestation in the mouse. By day 14.5, the liver becomes a large vascularized hematopoietic organ; this development does not occur if a septum transversum homeobox gene, *Hlx*, is knocked out. This indicates that soluble factors originating from the septum transversum are critical for developing the hematopoietic component of the embryonic liver.

Developmental Anomalies

Abnormalities of these early events can lead to anomalous development of the liver bud. Lobulation can be affected, leading to accessory lobes such as a supradiaphragmatic lobe or ectopic liver tissue (mesentery, or surface of the gall bladder, spleen, or adrenals). The right lobe can be abnormally enlarged (Reidel's lobe) and foregut diverticula can give rise to simple hepatic foregut cysts. However, most hepatic cysts are associated with abnormalities of the bile ducts.

DEVELOPMENT OF THE BILIARY SYSTEM

Embryology and Gene Regulation

A diverticulum (pars cystica) arises from the caudal surface of the liver bud and develops into the gallbladder, cystic, and extrahepatic bile duct. This is initially a solid cord, which develops a lumen. The adjacent parenchyma forms the developing biliary ductal epithelium and this structure migrates into the liver, following the branches of the portal vein during the sixth week of human fetal development. The hepatic artery also migrates into the liver along the portal vein. The forkhead transcription factor (Foxf1) expressed in the septum transversum mesenchyme is thought to play an important role in biliary morphogenesis. In addition, hepatocyte nuclear factors (Hnf6 and Hnf1 β) play a role in regulating hepatoblast differentiation into biliary cells. Perturbations in these transcription and signaling events can lead to defects in the biliary tree.

Developmental Anomalies

Anomalies of the biliary tract occur in 10−20% of the population. Anomalies include atresia, hypoplasia, or dilatation. Extrahepatic biliary atresia is the most frequent major anomaly and occurs in 1/10,000 live births. Although it usually presents during the first month of life, the process originates during the fifth week of fetal development. The cause is unknown but may involve perinatal viral infections. Patients with biliary atresia usually have persistent inflammation of the intrahepatic biliary tree, suggesting that the atresia may reflect a dynamic process of the entire hepatobiliary system. A hepatoportoenterostomy procedure of Kasai is the treatment of choice and has the best results when performed within 8 weeks postnatally. The short-term benefit of this surgery is decompression and biliary drainage to forestall the rapid onset of cirrhosis and sustain growth until a liver transplant can be done.

Cystic lesions are the result of developmental dilatations of the bile duct and can be either intrahepatic or extrahepatic, segmental or diffuse, and can involve the large or small bile ducts. Dilatation of the common bile duct leads to choledochal cysts, which are usually extrahepatic, but may sometimes extend into the liver. Their pathogenesis remains uncertain. Diverticulae of the common bile duct or dilatation of the intraduodenal portion of the common duct (choledochocele) can also occur.

Hepatocytes along branches of the portal vein form a biliary ductal plate that gives rise to the bile canaliculi. Error in remodeling of the embryonic ductal plate leads to fibropolycystic disorders of the biliary system. For example, Caroli's disease is an autosomal recessive condition with saccular dilatation of several segments of the large intrahepatic bile ducts. Saccular dilatation of small bile ducts can also occur with fibrosis (congenital

hepatic fibrosis) and with cystic changes in the collecting ducts of the kidney (autosomal recessive polycystic kidney disease). The gene for polycystic kidney and hepatic disease 1 (*PKHD1*), located on chromosome 6p21.1–p12, has been linked to classical forms of this disorder. Caroli's syndrome involves dilatations of the large intrahepatic bile ducts, with fibrosis and/or renal involvement. Autosomal dominant polycystic kidney disease can be also associated with liver cystic lesions, which may or may not communicate with the biliary system.

Intrahepatic biliary hypoplasia, another disorder of infancy or early childhood, involves an absence or decrease in the number of interlobular bile ducts or ductules, but the hepatic artery and portal vein maintain a normal size. The pathogenesis remains obscure. Accessory bile ducts are rare and usually are an extension of the right hepatic bile duct. Cholecystohepatic ducts are seen due to persistence of fetal connections between the liver and the gall bladder. They are usually associated with atresia of the right and left hepatic duct. Although the genetic defects explaining these biliary abnormalities remain unknown, Hnf6-null homozygous mice show an absence of the gall bladder. Other abnormalities of the gall bladder include duplication, oversized gallbladder, diverticula, or a phrygian cap. The anatomic location of the gallbladder may also vary (Fig. 2).

DEVELOPMENT OF HEPATIC VASCULATURE

Embryology and the Role of Endothelial Cells

Increasing evidence supports a critical role for endothelial cells during early development of the liver. This includes a role for angioblasts, which invade the septum transversum coincident with hepatocyte proliferation. Vascular endothelial growth factor (VEGF) signaling plays an important role in endothelial cell–hepatocyte interactions. During embryonic development, the portal vein originates from the paired vitelline veins and the hepatic artery originates from the celiac trunk.

Developmental Anomalies

Complete atresia or malformations in the shape, size, and number of the portal veins have been reported and can produce portal hypertension. Other congenital defects include clusters of dilated vessels within the periportal region and dilated portal veins, large or multiple hepatic arteries, or arterioportal anastomoses.

CONCLUSION

Much remains to be discovered about the regulation of hepatogenesis and development of the liver and biliary

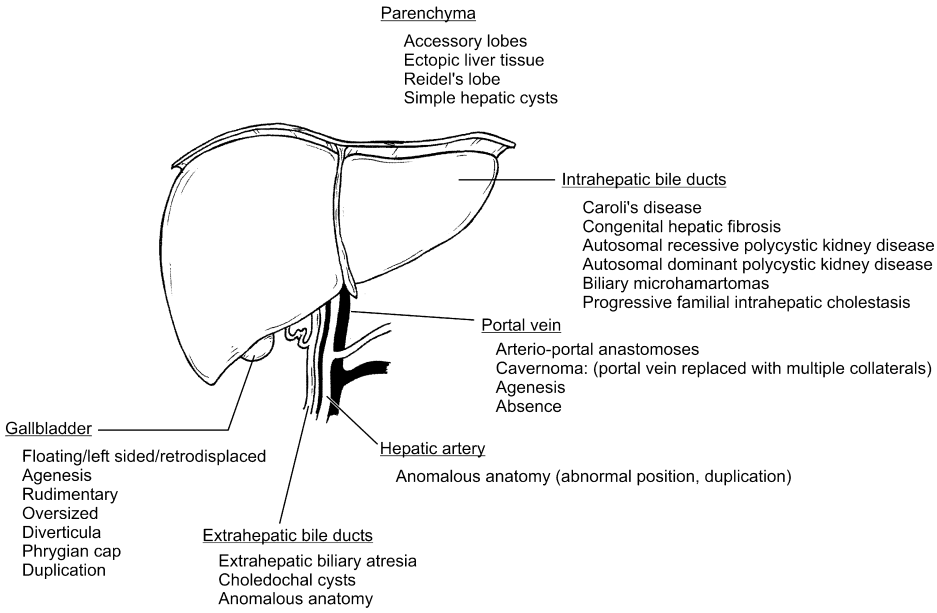


FIGURE 2 Developmental disorders of the liver arising from defects in various stages of liver development. The genetic or developmental explanations for most of these disorders remain unknown.

tree. However, studies in model systems are now making the long list of developmental defects seem more rational (Fig. 2).

See Also the Following Articles

Biliary Tract, Development • Biliary Tract, Developmental Anomalies of the • Development, Overview • Gallbladder, Pediatric • Neonatal Cholestasis and Biliary Atresia

Further Reading

Crawford, J. M. (2002) Development of the intrahepatic biliary tree. *Semin. Liver Dis.* **22**, 213–226.

Matsumoto, K., Yoshitomi, H., Rossant, J., and Zaret, K. S. (2001). Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563.

Onuchic, L. F., Furu, L., Nagasawa, Y., Hou, X., Eggermann, T., Ren, Z., Bergmann, C., Sanderek, J., Esquivel, E., Zeltner, R., Rudnik-Schoneborn, S., Mrug, M., Sweeney, W., Avner, E. D., Zerres, K., Guay-Woodford, L. M., Somlo, S., and Germino, G. G. (2002). PKHD1, the polycystic kidney and hepatic disease gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix repeats. *Am. J. Hum. Genet.* **70**, 1305–1317.

Sherlock, S., and Dooley, J. (eds.) (2002). "Diseases of the Liver and the Biliary System," 11th Ed., pp. 579–591. Blackwell Science, Oxford.

Zakim, D., and Boyer, T. (eds.) (2003). "Hepatology: A Textbook of Liver Disease," 4th Ed., pp. 1649–1698. W. B. Saunders, Philadelphia.

Zaret, K. S. (2001). Hepatocyte differentiation: From the endoderm and beyond. *Curr. Opin. Genet. Dev.* **11**, 568–574.

Zaret, K. S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nat. Rev. Genet.* **3**, 499–512.

Liver Disease, Pregnancy and

J. EILEEN HAY
Mayo Clinic and Foundation

acute fatty liver of pregnancy Acute liver failure due to microvesicular fatty infiltration of hepatocytes in the third trimester of pregnancy.

HELLP syndrome Triad of hemolysis, elevated liver tests, and low platelet count in the third trimester of pregnancy.

hyperemesis gravidarum Intractable vomiting in the first trimester of pregnancy.

intrahepatic cholestasis of pregnancy Pruritus and liver dysfunction in the second half of pregnancy.

preeclampsia triad of hypertension, edema, and proteinuria in the third trimester of pregnancy.

Pregnancy is an altered physiologic state associated with many hemodynamic and metabolic adaptations to support the placenta and fetus. Although most pregnant women remain healthy, pathophysiologic changes can occur, including five unique liver diseases seen only in the pregnant or postpartum patient. These pregnancy-associated liver diseases have characteristic clinical features and timing of onset in relation to pregnancy. Hyperemesis gravidarum is intractable nausea and vomiting in the first trimester; intrahepatic cholestasis of pregnancy, a disease of pruritus and abnormal liver tests, occurs in the second half of pregnancy; and preeclampsia and its complications of HELLP (hemolysis, elevated liver tests, and low platelets) syndrome and acute fatty liver of pregnancy are third-trimester diseases.

INTRODUCTION

Any liver disease, acute or chronic, occurring in young adult women may also occur coincidentally in pregnancy. Viral hepatitis is a common cause of jaundice in pregnancy and hepatitis A, hepatitis B, and hepatitis C occur with the same frequency and with the same clinical features as in nonpregnant individuals, with pregnancy and the viral hepatitis having little effect on each other. Gallstones are common, particularly in multiparous patients, but fortunately they are rarely symptomatic with biliary colic, pancreatitis,

Copyright 2004, Elsevier (USA). All rights reserved.

Lower Gastrointestinal Bleeding and Severe Hematochezia

GUSTAVO A. MACHICADO AND DENNIS M. JENSEN
University of California, Los Angeles and Veteran's Administration Greater Los Angeles Healthcare System

angioma Arteriovenous malformation or proliferation, with or without dilation, of the blood vessels.
coagulopathy Disorder affecting the coagulability of the blood.
hematemesis Vomiting of blood, either red, clotted, or coffee-ground appearance.
hematochezia Passage of bloody stools, in contradistinction to melena (tarry stools).
hematocrit Percentage of the blood sample volume occupied by cells.
ischemia Low blood flow to the gut with tissue injury, either due to mechanical obstruction of the blood supply or reduction in local perfusion.
lower gastrointestinal Pertaining to the colon.
shock (hypovolemic) Caused by a reduction in the intravascular volume of blood, as from hemorrhage or dehydration.
stigmata (of recent hemorrhage) Endoscopically visible signs of recent hemorrhage such as active bleeding, adherent clot, or visible vessel.
thrombocytopenia Reduction in the number of platelets in the circulating blood.
upper gastrointestinal Pertaining to the esophagus, stomach, or duodenum.
varices (varix) Dilated veins (vein) in the esophagus and/or stomach related to portal hypertension.

Hematochezia is defined as passage of red blood or clots per rectum. It may be associated with diarrhea or pain, but often is painless. Hematochezia is a common reason for adults both young and old to present for medical attention. The approach to the diagnosis and treatment depends on the acuteness and severity of the hemorrhage. Patients with recurrent or severe hematochezia, with hypovolemia, or with severe anemia causing symptoms usually require hospitalization. Some with very severe blood loss or comorbid medical or surgical problems should be considered for intensive care unit management. Those patients with severe anemia and slower bleeding may also need to be hospitalized for blood transfusions and evaluation. The type and timing of diagnostic and therapeutic interventions will also be dictated by the severity of the hemorrhage.

INITIAL EVALUATION

Patients with presumed lower gastrointestinal (LGI) bleeding may present for medical attention with severe, persistent hematochezia and hypotension or shock. The majority of patients with this type of presentation have colonic sources of bleeding but some have upper gastrointestinal (UGI) or small bowel sources (Fig. 1). More commonly, the presentation of colonic bleeding is with recurrent bouts of self-limited hematochezia or with stools positive for occult blood with or without anemia.

The initial management of patients who present with severe hematochezia is similar to that of any patient presenting with severe bleeding from any source. Initial resuscitation efforts for patients presenting in shock (hypotension, tachycardia, and obvious intravascular volume depletion) include intensive care unit monitoring, adequate intravenous access, and vigorous volume replacement. If anemia is significant and/or the patient has severe or symptomatic heart, pulmonary, or vascular disease, transfusion of packed red blood cells (RBCs) is recommended. For patients with coagulopathies such as hypoprothrombinemia, administration of fresh frozen plasma

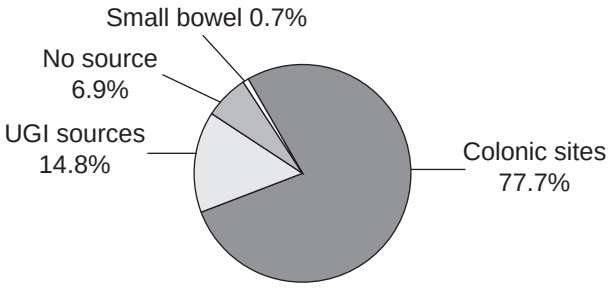


FIGURE 1 The distribution of bleeding sites for 291 consecutive patients (from the CURE Hemostasis Research Group study) with severe hematochezia, hospitalized and evaluated with urgent colonoscopy, anoscopy, panendoscopy, and/or push enteroscopy. UGI, Upper gastrointestinal.

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Resuscitation and Management of Patients with Severe Hematochezia

Establish one or preferably two large-bore intravenous lines
Assess intravenous volume and replace vigorously
Evaluate degree of blood loss and replace with packed red blood cells
Evaluate coagulation and correct with fresh frozen plasma, platelets, and/or desmopressin acetate
Place a nasogastric tube to check for a possible upper gastrointestinal source—blood or bile
Treat comorbid conditions

is indicated in cases of severe or ongoing hematochezia. Those with ongoing hematochezia and thrombocytopenia or severe renal insufficiency may require transfusions of platelets or desmopressin acetate (DDAVP). Table I summarizes resuscitation and initial management of patients with severe hematochezia.

A careful medical history and physical examination may yield important clues about the potential source of GI hemorrhage. A history of cirrhosis might suggest esophageal or gastric varices or angiomas of the gut. Severe chronic renal disease on hemodialysis could indicate angiomas. Severe heart disease can be associated with colonic angiomas or intestinal ischemia. A history of peptic ulcer disease, diverticulosis, inflammatory bowel disease, or internal hemorrhoids might direct the physician to these diagnoses as the potential bleeding site. Similarly, a history of crampy abdominal pain, anorexia, diarrhea, or fever with hematochezia may indicate a diagnosis of inflammatory bowel disease, ischemic bowel disease, or infectious colitis. A malignancy should be considered with weight loss, anorexia, constipation, or a change in bowel habits with anemia and hematochezia. A careful medical history should include questions about all medications the patient ingested recently, because many over-the-counter (OTC) and prescription drugs may cause or be associated with acute or chronic GI bleeding. Specifically, patients should be queried about recent use of aspirin or nonsteroidal antiinflammatory agents (NSAIDs), which commonly cause GI ulcerations and hemorrhage or aggravate bleeding from preexisting GI lesions. Anticoagulants or gingko may also aggravate GI bleeding from gut lesions.

DIAGNOSTIC APPROACH TO SEVERE HEMATOCHEZIA

Initial Management

While patients with severe hematochezia and anemia or hypotension are being resuscitated, the initial evaluation should include the passage of a nasogastric (NG) tube and gastric lavage to exclude the possibility of upper gastrointestinal (UGI) bleeding in patients without hematemesis. If bile is obtained in the presence of ongoing hematochezia, there is continuity with the duodenum and an upper gastrointestinal lesion is unlikely as the source of the hematochezia unless the bleeding is intermittent. When only clear fluid is obtained and no bile is present, continuity with the duodenum is not ascertained and the study should be regarded nondiagnostic. If there is evidence of blood (red or dark blood, clots, or coffee-ground appearing) in the NG aspirate, then an esophagogastroduodenoscopy (EGD) is indicated to exclude the possibility that a lesion in the upper gastrointestinal tract is responsible for the hematochezia (Fig. 1). Selected patients with suspected rectal lesions (such as internal hemorrhoids, fissures, or proctocolitis) should be given enemas and have urgent flexible sigmoidoscopy and anoscopy with a slotted anoscope before purge and urgent colonoscopy. Table I provides other general recommendations about the acute management of patients with severe hematochezia.

Urgent Colonoscopy

The CURE Hemostasis Group advocates an urgent approach to the diagnosis and treatment of patients with severe hematochezia similar to that currently used for patients with severe UGI hemorrhage. While resuscitating the patient, the approach is to purge the colon of clots, blood, and stool with a balanced electrolyte solution. Then, an urgent colonoscopy is performed for diagnosis and, if necessary, colonoscopic hemostasis during the same examination. This approach has dramatically changed the diagnosis, treatment, and cost of management of patients with severe hematochezia hospitalized in the CURE group institutions.

A polyethylene glycol (PEG)-based purge (Golytely or Colyte) is the commercially available preparation now utilized as preparation for urgent colonoscopy. If there is no evidence of UGI bleeding and the patient already has a nasogastric tube in place, the solution is administered via the tube. Alternatively, some patients can drink enough purge solution at a rapid enough rate (e.g., 1 liter every 30–45 minutes) to effectively prepare their colons. Metoclopramide (Reglan, 10 mg) is administered intravenously 10–15 minutes prior to starting the purge for its prokinetic and antiemetic effects. A liter of purge solution (Golytely or Colyte) is administered or drunk every 30–45 minutes until the rectal effluent is clear of blood, clots, and particulate matter. Usually 6–10 liters over 3–6 hours is needed to accomplish

TABLE II Colon Preparation Prior to Urgent Colonoscopy in Patients with Severe Hematochezia

Give metoclopramide (Reglan) if no contraindications, 10 mg intravenously or intramuscularly 5–30 minutes prior to starting purge
Give a polyethylene glycol-based purge for colonoscopy (Nulytely or Colyte) orally or via nasogastric tube at 1 liter every 30–45 minutes until effluent is clear of clots, stool, and blood—usually 6–10 liters required over 3–6 hours
In patients with tense ascites, perform therapeutic paracentesis to prevent respiratory compromise during colonoscopy
If patient is in congestive heart failure, treat with intravenous diurectics; if in renal failure, use concurrent hemodialysis

this goal. Following this, an urgent colonoscopy is performed at the bedside or in an in-patient endoscopy unit, while monitoring the patient.

Patients who have congestive heart failure often require diuresis during or after colonic purge. Similarly, patients with chronic renal failure who are on dialysis usually require hemodialysis during or just after the bowel preparation, because 10–20% fluid absorption is expected during the rapid purge in such patients. During colonic purge, volume overload is common with these and other severe comorbid conditions, such as end-stage liver disease with large-volume ascites. Table II provides general recommendations on this method of bowel preparation.

Results for Urgent Colonoscopy

The CURE group recently prospectively studied 291 consecutive patients who were hospitalized because of severe hematochezia. Patients who had ongoing, severe hematochezia and those who appeared to stop bleeding after hospitalization were included in this study. All patients had NG aspiration and lavage, purging of the colon and performance of urgent colonoscopy, and/or performance of upper endoscopy (or enteroscopy) for positive NG lavage or if the colonoscopy was negative.

A specific diagnosis was made by either stigmata of recent hemorrhage on the lesion (such as active bleeding, nonbleeding visible vessel, or adherent clot) at endoscopic examination or the identification of a GI lesion without stigmata, but in the absence of other significant GI lesions that might have caused the hemorrhage. A colonic bleeding site was found in 77.7% of patients admitted with hematochezia, an UGI source was present in 14.8%, a small bowel source was found in 0.7%, and no source could be found in 6.9% (see Fig. 1).

A variety of lesions were found in patients bleeding from a colonic site (77.7%). The three most common diagnoses were diverticulosis, internal hemorrhoids, and ischemic colitis. Table III lists these and other frequent colonic diagnoses.

Red Cell Scanning

Technetium-labeled red blood cell (RBC) scans can be useful to localize the bleeding site in patients with severe hematochezia. Early RBC scans (e.g., baseline and up to 1–4 hours only) are recommended in selected patients with severe, ongoing hematochezia. Selected patients include those with recurrent GI bleeding that has defied diagnosis by urgent colonoscopy, enteroscopy, and anoscopy; hospitalized patients who have ongoing severe hematochezia during colonic purging; and patients with severe ongoing hematochezia for whom a visceral angiogram is being considered.

The advantages of RBC scanning are that it can be performed while the patient is undergoing resuscitative efforts and purging. Although RBC scanning requires active bleeding to show extravasation into the gut, it is reported to be positive with a bleeding rate as low as 0.1 ml/min. In addition, the scanning can be repeated because the labeled RBCs remain in the circulation at least 24 hours. There are also disadvantages of RBC scanning. In most patients, the scans will be negative because patients have stopped bleeding before the scan is done. Another major disadvantage is limited ability to establish the location of small bowel lesions, when scans are read as positive, particularly delayed scans (after 4 hours). If the RBC scan is positive, a confirmatory test such as colonoscopy, push enteroscopy, angiography, or intraoperative enteroscopy is recommended to identify the precise location of the hemorrhage and clarify the nature of the bleeding lesion. Empiric surgical resection is not recommended based on the results of

TABLE III Colonic Sources of Severe Hematochezia[a]

Diagnosis	Colonic source (%)
Diverticulosis	29.6
Internal hemorrhoids	14.2
Ischemic colitis	12.4
Solitary rectal ulcers	9.2
Ulcerative or other colitis	8.8
Delayed post polypectomy ulcer	8.0
Colon angiomas	5.7
Colon polyp or cancer	6.2
Other lower gastrointestinal	7.0

[a] Expressed as the percent of all colonic sources of severe hematochezia. From the CURE Hemostasis Research Group Study (Jensen, 2001; Kovacs *et al.*, 2001). $N = 291$ total severe hematochezia patients.

a RBC scan alone, because localization is often poor, especially with delayed scans (more than 4 hours) that are positive. Whereas early scans (less than 4 hours after baseline) have been reported to be helpful for localization (positive yields more than 75% at surgery), delayed scans have not proved reliable for localization (yield less than 40% at surgery).

Emergency Visceral Angiography

Visceral angiography may be useful in the diagnosis and localization of bleeding sites in patients with severe or ongoing hematochezia. Its diagnostic yield has been reported to be from 12 to 69% in patients with presumed active bleeding. Angiography may also identify nonbleeding lesions such as angiomas or tumors. An advantage is that angiography can be performed without a bowel preparation or during bowel preparation. Skilled angiographers can also treat bleeding lesions by selective embolization. The disadvantages of angiography are that critically ill patients have to be transported to the radiology department for the procedure. High blood flow ($\sim$0.5 ml/minute) is also required to see extravasation into the gut lumen, and this high bleeding rate is unusual for colonic lesions. The contrast agent may cause serious complications, particularly in the elderly, such as volume overload or renal insufficiency. Emergency angiography is recommended in severely bleeding patients when urgent anoscopy, colonoscopy, and enteroscopy have failed to yield a diagnosis and there is continued or recurrent bleeding. Screening of these patients first with a RBC scan may increase the yield of angiography, if only those with positive early scans are selected for emergency angiography.

DIAGNOSTIC APPROACH TO LESS SEVERE BLEEDING

Most patients who present with self-limited hematochezia or occult blood positive stools and anemia can be managed as outpatients. The exceptions are those who present with severe comorbid conditions and symptoms (such as angina pectoris or dizziness) or signs of the severe anemia (such as hypotension, syncope, or shock). These patients usually require hospitalization and RBC blood transfusions. Patients with severe comorbid conditions such as heart disease or renal disease may also require concomitant diuresis or dialysis, during intravascular volume repletion or transfusion. Patients with slow GI bleeding who do not require RBC transfusions can be managed with iron supplements and outpatient evaluations.

ENDOSCOPIC DIAGNOSIS AND TREATMENT OF BLEEDING COLONIC LESIONS

Diverticular Hemorrhage

Diverticulosis is the most frequent colonic lesion responsible for severe hematochezia. Patients with diverticular hemorrhage do not present with slow GI hemorrhage, anemia, and Hemoccult positive stools. Instead, they present with overt bleeding and usually with hematochezia. Diverticulosis was diagnosed as the cause of hematochezia in 24.1% of all patients admitted with severe hematochezia in a recent study. In the study, the majority of patients had colonic diverticulosis, because they were over 65 years of age and diverticulosis is common in the elderly. However, for all patients with diverticulosis of the colon admitted with severe hematochezia, more than 50% were found to have bleeding from nondiverticular colonic, sources. These were reported as "incidental diverticulosis." If an urgent colonoscopy after good colonic preparation had not been performed, many of these nondiverticular lesions would have been missed or overlooked because stigmata of colonic hemorrhages tend to disappear in 1–3 days. "Definitive diverticular bleeding" was diagnosed when there were stigmata of recent hemorrhages (active bleeding, a nonbleeding visible vessel, or an adherent clot) in a diverticulum on urgent colonoscopy. Adequate colonic cleansing with the purge and target water-jet irrigation facilitated identification of stigmata of hemorrhages during colonoscopy. "Presumptive diverticular hemorrhage" was diagnosed when only colonic diverticulosis and no other potential bleeding lesions were found on colonoscopy, anoscopy, and enteroscopy (see Fig. 2).

Thirty-four patients with definitive diverticular hemorrhages have been studied to compare medical-surgical with medical-colonoscopic management. Seventeen patients received medical-surgical treatment in an initial observational study to define the natural history of definitive diverticular hemorrhage. In a subsequent study of definitive diverticular bleeding, 17 patients had medical treatment and hemostasis at the time the stigmata were diagnosed at urgent colonoscopy. Both groups of definitive diverticular bleeds were comparable in terms of age, comorbid conditions, recent aspirin and/or NSAID ingestion, and blood transfusion requirements prior to colonoscopy. Patients in the medical-surgical cohort of the study underwent emergency colonoscopy for diagnosis but did not receive any colonoscopic treatment. Medical-colonoscopic treated patients also underwent

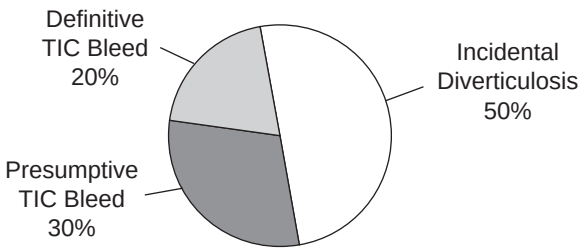


FIGURE 2 Prevalence of incidental diverticulosis and presumptive and definitive diverticular hemorrhage (56 patients with diverticulosis and severe hematochezia in a study by the CURE Hemostasis Group). The types of diverticular hemorrhage are designated in patients with severe hematochezia, evaluated by urgent colonoscopy, after colonic purge. Definitive diverticular hemorrhage is finding a stigma of recent hemorrhage (active bleeding, nonbleeding visible vessel, or adherent clot) on or in a single diverticulum, resistant to target water irrigation and suctioning. Presumptive diverticular hemorrhage is diagnosed when only colonic diverticulosis (without stigmata of hemorrhage) and no other potential bleeding lesions are found on colonoscopy, anoscopy, and push enteroscopy. Incidental diverticulosis refers to patients with colonic diverticulosis, but some other definitive source of hemorrhage is found on urgent colonoscopy, anoscopy, endoscopy, or push enteroscopy.

emergency colonoscopic treatment for stigmata of hemorrhage (active bleeding, adherent clot, or a nonbleeding visible vessel). Colonoscopic treatment consisted of epinephrine injection, bipolar cautery, or both. For both cohorts of patients, the diverticulum with the stigmata of hemorrhage was tattooed with India ink to facilitate localization at surgery or retreatment. After the urgent colonoscopy, ongoing or recurrent hemorrhage that required transfusions of two or more units of packed red blood cells was observed in 53% of the medical-surgical treated patients, in contrast to 6% of the medical-colonoscopic treated group. Severe bleeding requiring transfusion with more than three units of packed red blood cells and subsequently requiring surgery was seen in 35% of the medical-surgical group in contrast to 6% of the medical-colonoscopic treated group. No complications were seen from colonoscopic treatment of bleeding diverticulosis. The median time to hospital discharge was 5 days for the medical-surgical group and 2 days for the medical-colonoscopic group (see Table IV). None of the patients in either group had recurrent bleeding on long-term (mean 36 months) followup. In contrast, patients with "presumed diverticular hemorrhage" had a recurrent bleeding rate of 5.7% during a mean of 36 months of followup. This is in contrast to surgical series that report 38–50% recurrent bleeding rates during 3–5 months of followup.

Colonic Angiomas

Colonic angiomas are also referred to as telangiectasia, arteriovenous malformations, or angiodysplasia. These lesions may occur sporadically or can be part of a congenital syndrome such as hereditary hemorrhagic telangiectasia (Osler–Weber–Rendu syndrome). The etiology of sporadic colonic angiomas is unknown. However, they have been observed to occur with increased frequency with advanced age and in patients with severe heart disease (particularly valvular heart disease), chronic renal insufficiency, cirrhosis of the liver, and collagen vascular disorders.

In a previous prospective study of 100 patients who were hospitalized with very severe, persistent hematochezia, 30% were found to be bleeding from colonic angiomas. All of these patients were elderly and most had severe comorbid conditions, such as heart, renal, or liver disease. In contrast, in a subsequent prospective study of 291 patients hospitalized with either self-limited hematochezia or ongoing hematochezia, colon angiomas were diagnosed as the cause of hemorrhage in 4.7%. Another prospective study of patients with hemorrhage from colonic angiomas included 100 consecutive patients. Approximately 75% presented with anemia, Hemoccult positive stools, and self-limited hematochezia. In contrast, 25% had severe hematochezia requiring resuscitation and transfusions. These were hospitalized and had urgent colonoscopy. Angiomas varied in size from 1 to 30 mm, but a majority were smaller than 10 mm and could be coagulated easily via colonoscopy. The most common location was the right colon: the cecum and ascending colon accounted for 72%. A less common location for angiomas was the left colon (the descending and sigmoid colon accounted for 22%).

The CURE group randomized 100 patients with bleeding from colonic angiomas for hemostasis at

TABLE IV Outcomes of Treatments of 34 Consecutive Patients with Definitive Diverticular Hemorrhage[a]

	Med-Surg ($N = 17$)	Med-Colon ($N = 17$)
Endoscopic hemostasis	0	17 (100%)
Additional bleeding	9 (53%)	1 (6%)*
Severe bleeding	6 (35%)	1 (6%)*
Emergency surgery	6 (35%)	1 (6%)*
Median time to discharge after colonoscopy (days)	5	2
Complications	2 (12%)	0

[a] From Jensen (2000, 2001). *$P < 0.05$.

colonoscopy with either bipolar coagulation or heater probe. Both coagulation devices have been shown to be equally effective for the treatment of GI angiomas. In these patients, the clinical course significantly improved after, compared to before, colonoscopic hemostasis in terms of recurrent bouts of hemorrhage, need for blood transfusions, and serum hemoglobins. Complications occurred in 5.7%. Two patients developed postcoagulation syndrome and recovered with medical therapy alone and four others developed delayed bleeding due to ulcers at the coagulated site. Two of the latter patients with severe coagulopathies required surgery for control of recurrent colonic bleeding. No colonic perforations occurred in this group, although perforations have been reported by other groups who have treated colonic angiomas with other coagulation devices, such as monopolar probes or lasers, and also in those patients who had repeat coagulation at the same site with heater probe or bipolar coagulation.

Colonic Carcinomas, Polyps, and Postpolypectomy Bleeds

Patients with colonic cancers or polyps usually present with occult gastrointestinal bleeding and anemia and rarely with gross hematochezia. In a prospective study of 100 patients with severe, persistent hematochezia, urgent colonoscopy revealed ulcerated cancer or polyps in 9% of patients. In a second series of 291 patients admitted for hematochezia, ulcerated polyps or cancers were found in 5.2%. Bleeding from ulcerated cancers or polyps usually does not involve arterial bleeding but rather tends to be oozing bleeding. Endoscopic control of such bleeding can usually be achieved with the injection of 1 : 10,000 epinephrine with or without bipolar probe coagulation in the case of cancers. During the same colonoscopy, lesions are biopsied to obtain a histologic diagnosis. Subsequently, surgical resection is recommended. In contrast, bleeding polyps can usually be removed with a polypectomy snare and electrocautery. For large sessile polyps that are not amenable to complete endoscopic removal, hemostasis and biopsy are recommended to establish a histologic diagnosis, prior to consideration for surgical resection.

Significant bleeding can occur immediately after a polypectomy and this is usually due to ineffective or inadequate coagulation of the vessels within the polyp stalk. If bleeding occurs immediately following polypectomy, hemorrhage can be arrested with the same polypectomy snare by resnaring the stalk and applying pressure with or without electrocautery. If resnaring is not feasible, the area can be injected with epinephrine and subsequently cauterized with bipolar or heater probe, but other devices, such as the endoloop or hemoclips, have also been used.

Delayed hemorrhage occurs due to the coagulum sloughing off the polyp stalk, or the polypectomy base in the case of sessile polyps. In the series of 291 patients with severe hematochezia, postpolypectomy delayed bleeding accounted for 8.0% of colonic diagnoses. Most reported rates of postpolypectomy bleeding are less than 6%. In the original series of patients with severe hematochezia, this diagnosis was not reported. With the colonoscopy initiatives for colorectal cancer screening in the past 5 years, this is now the sixth leading colonic cause of severe hematochezia (see Table III). Delayed bleeding in the patients discussed here occurred a median of 9 days (range 2−73) postpolypectomy. The mean size of the index polyps was 20 mm. Ulcers with a mean size of 11 mm were found in the patients who presented with delayed postpolypectomy hemorrhage and underwent emergency colonoscopy following a bowel purge. Active bleeding was present in 23%, nonbleeding visible vessel was present in 23%, adherent clot was present in 38%, a flat spot was present in 8%, and clean ulcer was present in 8%. The majority (92%) of the patients achieved complete hemostasis with epinephrine injection and cautery. Only one patient rebled. Endoscopic treatment of these lesions was effective and safe.

Internal Hemorrhoids

Internal hemorrhoids are the most common lesion causing self-limited rectal bleeding in ambulatory adults. Rarely, internal hemorrhoids cause persistent bleeding needing emergency hemostasis. In the original series of 100 patients with severe, persistent hematochezia, internal hemorrhoids were the bleeding site in less than 4%. Often, a simple slotted anoscopy examination at the bedside will yield the diagnosis, obviating the need for urgent colonoscopy. Several effective methods can be used to control bleeding from internal hemorrhoids without surgery. In one randomized study, bipolar probe and heater probe coagulation were reported to be equally effective in controlling bleeding from internal hemorrhoids in a randomized study of patients who had failed medical therapy. Internal hemorrhoid banding has also been used successfully for many years. Endoscopic banding is commonly used after urgent colonoscopy, when internal hemorrhoids are diagnosed as the bleeding site, because of the effectiveness of this technique. Injection sclerotherapy has also been used in the past with a variety of sclerosants for the treatment of internal hemorrhoids. Complications such as thrombosis, infection, ulceration, or delayed

bleeding are reported for all these techniques of internal hemorrhoid hemostasis. However, complication rates are low and most physicians and patients favor this approach over surgical hemorrhoidectomy.

See Also the Following Articles

Colonoscopy • Diverticulosis • Hemorrhage • Hemorrhoids • Occult Gastrointestinal Bleeding • Upper Gastrointestinal Bleeding • Variceal Bleeding

Acknowledgments

The studies of the CURE Hemostasis Research Group included herein were funded in part by NIH 1 K24 DK02650 and Human Studies CORE of NIH Grant DK41301. The authors also thank the other members of the research group for their support: T. O. G. Kovacs, R. Jutabha, I. Gralnek, and G. Dulai. We are grateful to the research coordinators (Lana Fontana and Annie St. Amand) and the data managers (Vichuda Lousuebsakul and Mary Ellen Jensen). We also thank Julie Pham for the word processing and Ken Hirabayashi for the preparation of the figures.

Further Reading

Athanasoulis, C. A. (1980). Therapeutic applications of angiography. *N. Engl. J. Med.* **302**, 1117–1125.

Binmoeller, K. F., Thonke, F., and Soehendra, N. (1993). Endoscopic hemoclip treatment for gastrointestinal bleeding. *Endoscopy* **25**, 167–170.

Gralnek, I. M., and Jensen, D. M. (2001). An assortment of colonic lesions that present with severe hematochezia. *Tech. Gastrointest. Endosc.* **3**, 216–220.

Jensen, D. M. (2001). Diverticular bleeding: An appraisal based upon stigmata of recent hemorrhage. *Tech. Gastrointest. Endosc.* **3**, 192–198.

Jensen, D. M. (2001). Endoscopic diagnosis and treatment of severe hematochezia. *Tech. Gastrointest. Endosc.* **3**, 177–184.

Jensen, D. M., Jutabha, R., Machicado, G. A., *et al.* (1997). Prospective randomized comparative study of bipolar electrocoagulation versus heater probe for treatment of chronically bleeding internal hemorrhoids. *Gastrointest. Endosc.* **46**, 435–443.

Jensen, D. M., Machicado, G. A., Jutabha, R., and Kovacs, T. O. G. (2000). Urgent colonoscopy for the diagnosis and treatment of severe diverticular hemorrhage. *N. Engl. J. Med.* **342**, 78–82.

Jutabha, R., Miura-Jutabha, C., and Jensen, D. M. (2001). Current medical, anoscopic, endoscopic and surgical treatments for bleeding internal hemorrhoids. *Tech. Gastrointest. Endosc.* **3**, 199–205.

Kovacs, T. O. G., and Jensen, D. M. (2001). Upper or small bowel hemorrhage presenting as hematochezia. *Tech. Gastrointest. Endosc.* **3**, 206–215.

Longstreth, G. F. (1997). Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: A population-based study. *Am. J. Gastroenterol.* **92**, 419–424.

Machicado, G. A., and Jensen, D. M. (2001). Bleeding colonic angiomas and radiation telangiectasias: Endoscopic diagnosis and treatment. *Tech. Gastrointest. Endosc.* **3**, 185–191.

Peura, D. A., Lanza, F. L., Gostout, C., *et al.* (1997). The American College of Gastroenterology Bleeding Registry: Preliminary findings. *Am. J. Gastroenterol.* **92**, 924–928.

Pfenninger, J. L. (1997). Modern treatments for internal hemorrhoids. *BMJ* **314**, 1211–1212.

Rex, D. K., Blair, S. L., and Waye, J. D. (1992). Colonoscopy and endoscopic therapy for delayed post-polypectomy hemorrhage. *Gastrointest. Endosc.* **38**, 127–129.

Ryan, P., Styles, C. B., and Chmiel, R. (1992). Identification of the site of severe colon bleeding by technetium-labeled red-cell scan. *Dis. Colon Rectum* **35**, 219–222.

Uflacher, R. (1986). Transcatheter embolization for treatment of acute lower gastrointestinal bleeding. *Acta Radiol.* **28**, 425–430.

Zucherman, G. R., and Prakash, C. (1999). Acute lower intestinal bleeding. *Gastrointest. Endosc.* **49**, 228–238.

Lupus Erythematosus

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

antinuclear antibodies (ANAs) Autoantibodies that are directed against nuclear antigens and measured by indirect immunofluorescence. Specific ANAs are directed against DNA, histones, components of the spliceosome (Sm, RNP, Ro, and La), and other nuclear antigens (centromere, topoisomerase I).

drug-induced lupus erythematosus A distinct form of lupus caused by sustained exposure to certain drugs (procainamide, phenytoin, and isoniazid are the commonest) associated with the production of anti-histone antibodies and clinically with fever, arthralgia, myalgia, rash, and serositis and the absence of renal or central nervous system involvement.

major histocompatibility complex The genetic region on the short arm of chromosome 6 encoding groups of highly polymorphic surface proteins that are expressed on the surface of all living cells and that are important in antigen presentation and shaping of the immune repertoire.

Lupus erythematosus is a complex multisystem autoimmune disorder characterized by inflammation in many organ systems and by the formation of distinct autoantibodies often associated with specific end-organ involvement. There is considerable phenotypic heterogeneity accounting for a wide spectrum of disease expression ranging from mild to life-threatening forms of this disease.

CLASSIFICATION

Several interrelated and overlapping clinical disease subsets of lupus erythematosus (LE) are recognized. Historically, LE was first described as a cutaneous disease and in many patients LE is exclusively a cutaneous disease. Systemic lupus erythematosus (SLE) is the classic systemic form of the disease, typically with some combination of cutaneous, musculoskeletal, hematologic, and renal involvement. Drug-induced lupus erythematosus is a distinct form of the disease associated with the production of anti-histone antibodies and clinically with fever, arthralgia, myalgia, rash, and serositis as well as the absence of renal or central nervous system involvement.

EPIDEMIOLOGY

SLE is a relatively uncommon disease with an incidence of approximately 5 per 100,000 per year and prevalence of between 25 and 50 per 100,000. Certain racial groups are at increased risk for SLE and include African Americans and African Caribbean individuals. The peak incidence is between 15 and 45 years of age. There is a striking female predominance with a female : male ratio of 9 : 1. Twin studies indicate that SLE is predominantly a genetic disease, although several extrinsic triggers have been implicated. These include ultraviolet light, hair products, smoking, estrogens, and viral agents like Epstein-Barr virus and hepatitis C virus (HCV).

ETIOLOGY AND PATHOGENESIS

SLE appears to be a multigenic disease resulting from complex interactions between susceptibility genes and environmental factors resulting in the generation and persistence of an autoimmune state with subsequent destructive inflammatory events. It is useful to conceptualize the pathogenesis of SLE as a set of preclinical events occurring in a genetically predisposed individual characterized by the induction and expansion of autoimmunity followed by immune-mediated injury causing clinically recognizable disease.

The genes conferring susceptibility to SLE include certain allelic major histocompatibility complex molecules and several other immunoregulatory genes that have been partially identified. Acting in concert with other factors, these genes lead to the development of autoreactive T cells, followed by an expansion phase that is characterized by considerable clonal expansion of B and T cells that recognize self-antigens. Once mature, high-affinity autoantibodies form and have access to autoantigen and immune-mediated injury ensues, resulting in clinical forms of this disease. It is very likely that cell-mediated processes are also involved but their role in SLE pathogenesis is less well understood.

 Copyright 2004, Elsevier (USA). All rights reserved.

CLINICAL FEATURES

General

Fatigue is the commonest feature of SLE, sometimes reflecting disease activity, but equally prevalent even when the disease is inactive when it can represent the effects of depression, stress, fibromyalgia, or medication side effects. Fever is also a common feature of disease and is typically episodic. Constant fever is more suggestive of infection. Weight loss may precede the diagnosis of SLE, whereas weight gain may be the result of steroid therapy.

Cutaneous

The most characteristic skin lesion associated with lupus is the diffuse, macular, erythematous facial eruption in a malar distribution, also known as the "butterfly" rash. It is referred to as acute LE, since it is associated with active SLE. Other less common but equally distinct cutaneous lesions include chronic cutaneous LE (discoid LE) and subacute cutaneous LE, both of which may also occur in patients without systemic features.

Musculoskeletal

Though exceedingly common, arthritis associated with SLE is usually mild and not associated with the development of deformities. It may be migratory and most often involves the hand joints. Myalgia is common but true myositis is rare.

Renal

The majority of patients with SLE have renal involvement, often subclinical. In 50–75% of patients, this will become clinically apparent in a few years after disease onset. Renal disease appears to be the result of immune complex damage and in some cases is associated with elevated anti-DNA antibody titers and low levels of complement proteins C3 and C4. There are six different histologic patterns described in patients with SLE in whom renal biopsy is performed: normal (type I), mesangial (type II), focal proliferative (type III), diffuse proliferative (type IV), membranous (type V), and sclerosing (type VI). Although some of these histologic lesions are associated with distinct clinic syndromes, there is substantial overlap and histologic transition over time. Type IV disease is both the commonest and severest form of lupus nephritis.

Hematologic

All three cell lines are affected in patients with SLE. Leukopenia is common, but rarely is clinically significant. Autoimmune hemolytic anemia is seen in a small number of patients with SLE, with most patients who are anemic exhibiting an anemia of chronic disease. Mild thrombocytopenia is common; occasionally it is severe. Anti-phospholipid antibodies directed against a variety of clotting factors occur in up to one-third of SLE patients and may cause a prolongation in the activated partial thromboplastin time (lupus anticoagulant).

Gastrointestinal

Most gastrointestinal symptoms in SLE are related to medication-related side effects [nonsteroidal anti-inflammatory drugs (NSAIDs), corticosteroids, immunomodulators]. SLE has also been associated with direct involvement of the gastrointestinal tract. Oral ulcers are exceedingly common. Dysphagia is less common and often multifactorial (most often related to hypomotility). Vasculitis of mesenteric vessels is a rare but serious complication of SLE. It may present with an acute abdomen or may have an insidious onset. Complications include bowel necrosis and perforation. Bland mesenteric thrombosis may occur secondary to a hypercoagulable state associated with anti-phospholipid antibodies. Pancreatitis is seen in up to 8% of patients with SLE. It is occasionally secondary to mesenteric vasculitis. Liver disease in SLE is frequent. Lupoid hepatitis refers to an antinuclear antibody (ANA)-positive autoimmune hepatitis, most often associated with anti-smooth muscle antibodies and sometimes with anti-mitochondrial antibodies as well (which are rare in SLE). In contrast, the term "lupus hepatitis" is increasingly employed to refer to an autoimmune hepatitis occurring in patients with SLE. It has been associated with antibodies to ribosomal P antigen and typically occurs in patients with clinically overt SLE. Protein-losing enteropathy is an uncommon clinical feature of SLE.

Other

Neurologic disease is very common in SLE and most often presents as neuropathy, mononeuritis multiplex, meningitis, cerebritis, neurocognitive dysfunction, seizures, and stroke syndromes. Headaches are a relatively common complaint and are often not an indicator of underlying disease activity. Pulmonary manifestations include pleuritis, pleural effusions, pneumonitis, interstitial lung disease, pulmonary hypertension, and alveolar hemorrhage. Myocarditis, pericarditis, valve abnormalities, and coronary arteritis can all occur in patients with SLE.

DIAGNOSIS

Classic SLE can be readily diagnosed when a patient from a high-risk epidemiologic group (for example, young African American women) presents with the typical constellation of symptoms and signs. Diagnostic difficulties arise when the disease occurs in low-risk epidemiologic groups (such as, children or older males) or when it presents with involvement of just a few organ systems. Chronic human immunodeficiency virus and hepatitis C virus infection can phenotypically mimic SLE. Criteria have been developed for the classification of SLE by an expert consensus panel. Although the criteria were devised to create uniformity in clinical studies, the criteria are frequently used to diagnose SLE (see Table I). ANA testing in SLE is widely performed but suffers from low specificity while maintaining high sensitivity. The rational use of ANA testing is dependent on generating clinical pretest probabilities of SLE and avoiding ANA testing when the clinical probability of SLE is low. Though less sensitive, anti-DNA and anti-Sm antibodies are considered specific for SLE.

TABLE I The ACR 1982 Revised Criteria for the Classification of Systemic Lupus Erythematosus

Criterion	Definition
Malar rash	Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
Discoid rash	Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions
Photosensitivity	Skin rash as a result of unusual reaction to sunlight, reported in patient history or observed by physician
Oral ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by physician
Arthritis	Nonerosive arthritis involving 2 or more peripheral joints, characterized by tenderness, swelling, or effusion
Serositis	(1) Pleuritis—convincing history of pleuritic pain or rub heard by a physician or evidence of pleural effusion OR (2) Pericarditis—documented by ECG or rub or evidence of pericardial effusion
Renal disorder	(1) Persistent proteinuria >0.5 g per day or $>3+$ if quantitation not performed OR (2) Cellular casts—may be red cell, hemoglobin, granular, tubular, or mixed
Neurologic disorder	(1) Seizures—in the absence of offending drugs or known metabolic disorder derangements, e.g., uremia, ketoacidosis, or electrolyte imbalance OR (2) Psychosis—in the absence of offending drugs or known metabolic derangements, e.g., uremia, ketoacidosis, or electrolyte imbalance
Hematologic disorder	(1) Hemolytic anemia with reticulocytosis OR (2) Leukopenia— $<4000/mm^3$ total on 2 or more occasions OR (3) Lymphopenia— $<1500/mm^3$ on 2 or more occasions OR (4) Thrombocytopenia— $<100{,}000/mm^3$ in the absence of offending drugs
Immunologic disorder	(1) Anti-DNA: antibody to native DNA in abnormal titer OR (2) Anti-Sm: presence of antibody to Sm nuclear antigen OR (3) Positive finding of antiphospholipid antibodies based on (i) an abnormal serum level of IgG or IgM anticardiolipin antibodies, (ii) a positive test result for lupus anticoagulant using a standard method, or (iii) a false-positive serologic test for syphilis known to be positive for at least 6 months and confirmed by *Treponema pallidum* immobilization or fluorescence treponemal antibody absorption test
Antinuclear antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with "drug-induced lupus" syndrome

Note. The proposed classification is based on 11 criteria. For the purpose of identifying patients in clinical studies, a person shall be said to have systemic lupus erythematosus if any 4 or more of the 11 criteria are present, serially or simultaneously, during any interval of observation. Criterion 10 (Immunologic disorder) was updated in 1997. Sensitivity and specificity 96%. Data taken from Tan *et al*. (1982) and Hochberg (1997).

THERAPY AND PROGNOSIS

SLE is marked by substantial clinical heterogeneity with resulting variations in natural history and prognosis. Patients from lower socioeconomic backgrounds, African Americans, and those with renal disease have a worse prognosis. Therapy is thus dependent on an accurate assessment of organ involvement and disease activity. Overall, the outlook for SLE has improved remarkably over the past 50 years with an improvement in 10-year survival from 40 to $>90\%$.

Treatment involves the consideration of general issues such as photoprotection, nutrition, cardiovascular risk factor modification, osteoporosis prevention, infection control, exercise, and pregnancy counseling as well as therapies directed at specific organ systems. NSAIDs are employed for treating arthritis, antimalarial drugs (principally hydroxychloroquine) are useful in managing cutaneous and joint disease, and corticosteroids are reserved for major organ system involvement. Intravenous pulse cyclophosphamide is the treatment of choice for proliferative lupus nephritis. Azathioprine is often employed as a first-line steroid-sparing agent. Newer therapeutic approaches are focusing on interrupting immune cell–cell signaling mechanisms.

See Also the Following Articles

Dysphagia • Hepatitis C • Lymphocytes • Vasculitis

Further Reading

Boumpas, D. T., Austin, H. A., III, Fessler, B. J., Balow, J. E., Klippel, J. H., and Lockshin, M. D. (1995). Systemic lupus erythematosus: Emerging concepts. Part 1: Renal, neuropsychiatric, cardiovascular, pulmonary, and hematologic disease. *Ann. Intern. Med.* **122**(12), 940–950.

Boumpas, D. T., Fessler, B. J., Austin, H. A., III, Balow, J. E., Klippel, J. H., and Lockshin, M. D. (1995). Systemic lupus erythematosus: Emerging concepts. Part 2: Dermatologic and joint disease, the antiphospholipid antibody syndrome, pregnancy and hormonal therapy, morbidity and mortality, and pathogenesis. *Ann. Intern. Med.* **123**(1), 42–53.

Hallegua, D. S., and Wallace, D. J. (2000). Gastrointestinal manifestations of systemic lupus erythematosus. *Curr. Opin. Rheumatol.* **12**(5), 379–385.

Hochberg, M. C. (1997). Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**(9), 1725.

Ruiz-Irastorza, G., Khamashta, M. A., Castellino, G., and Hughes, G. R. (2001). Systemic lupus erythematosus. *Lancet* **357**(9261), 1027–1032.

Strand, V. (2000). New therapies for systemic lupus erythematosus. *Rheum. Dis. Clin. North Am.* **26**(2), 389–406, viii.

Tan, E. M., Cohen, A. S., Fries, J. F., Masi, A. T., McShane, D. J., Rothfield, N. F., Schaller, J. G., Talal, N., and Winchester, R. J. (1982). The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271–1277.

Wakeland, E. K., Liu, K., Graham, R. R., and Behrens, T. W. (2001). Delineating the genetic basis of systemic lupus erythematosus. *Immunity* **15**(3), 397–408.

Lymph, Lymphatics, and Lymph Flow

PATRICK TSO AND RONALD JANDACEK
University of Cincinnati Medical Center

alkaline phosphatase Enzyme (pH optimum 9.5–10.5) that hydrolyzes many orthophosphoric monoesters and serves as an important component of intracellular pH buffering.

chylomicrons Large, spherical lipoproteins made exclusively by the small intestine; their major function is to transport from the enterocytes to the lymph the large amount of dietary fat absorbed by the small intestine.

cirrhosis Degenerative disease of the liver that results in damage to hepatic parenchymal cells and decreased blood flow to the liver; characterized by the development of fibroids and nodes.

diamine oxidase Enzyme (also known as amine oxidase or histaminase) secreted by enterocytes; plays an important role in the breakdown of histamine and polyamines.

hepatic lymph hepatic lymph Derived from the perisinusoidal space, or space of Disse, it is through these channels that fluid and proteins are drained from the liver.

lipid absorption Process by which lipids are digested, mainly in the intestinal lumen, where they are absorbed by enterocytes and then packaged and secreted as chylomicrons into the lymph. Bile salt plays a critical role in the absorption. A concentrated form of energy, lipids are composed of a number of classes of compounds that are insoluble in water but soluble in organic solvents; triacylglycerols and triglycerides are the most abundant of all dietary lipids.

lymph Fluid that is collected from the tissues throughout the body and transported via the lymphatic vessels to the venous blood.

small intestine Consists of the duodenum, jejunum, and ileum and is where most digestion and absorption of nutrients takes place.

Lymphatic vessels, composed of countless microvessels of endothelial tubing, form an interconnected system that collects and transports fluids from the tissue of various organs in the body and channels them into the venous blood circulation. The gastrointestinal tract has the most extensive lymphatic system, which is essential to the absorption and digestion of nutrients and of some gastrointestinal hormones by the small intestine and their eventual transport to the general circulation via enterocyte processing. To a certain degree, lymphatics are involved in every pathological process.

ANATOMICAL CONSIDERATIONS

The historical aspects of knowledge about the gastrointestinal lymphatic system have been described in an excellent book written by Dr. James A. Barrowman in 1978. The formation of lymph begins in the lamina propria (interstitial compartment) of the small intestine. This interstitial compartment is best viewed as a complex noncellular matrix that supports the cells as well as the capillaries and lacteals. The main components of the interstitial matrix are bundles of collagen fibers and glycosaminoglycan chains, which are mechanically entangled and cross-linked, producing a gel-like structure. Hyaluronic acid, the main interstitial glycosaminoglycan, is a polymer of *N*-acetylglucosamine and glycuronic acid; it has a molecular weight that varies from a few thousand to several million. The anionic nature of hyaluronic acid causes pockets of negative charge within the gel matrix. The rate of diffusion of water and water-soluble molecules across the interstitial matrix (also called the hydraulic conductivity) depends mostly on the degree of hydration of the matrix. For example, doubling of the interstitial volume (therefore hydration) is believed to increase the hydraulic conductance more than a thousandfold. The interstitial matrix probably undergoes continuous modification, especially during periods when the matrix is markedly expanded (hydrated), e.g., during fat absorption. For instance, hyaluronic acid, which can be found in thoracic duct lymph, increases markedly during the absorption of fat.

Initial lymphatic vessels (lacteals), fine vessels found in the villus, range between 100 and 500 nm in length. They have only a single layer of endothelial cells with a fragmented basement membrane. The endothelial cells of the initial lymphatic vessel are joined together quite variably. For instance, they can be joined together by tight junctions, or there may be gaps between endothelial cells. As shown in Fig. 1, the gap between endothelial cells of the lacteals allows the chylomicrons to enter the lacteal. Three-dimensional reconstruction from electron micrographs show a special type of open junction formed by processes of

 Copyright 2004, Elsevier (USA). All rights reserved.

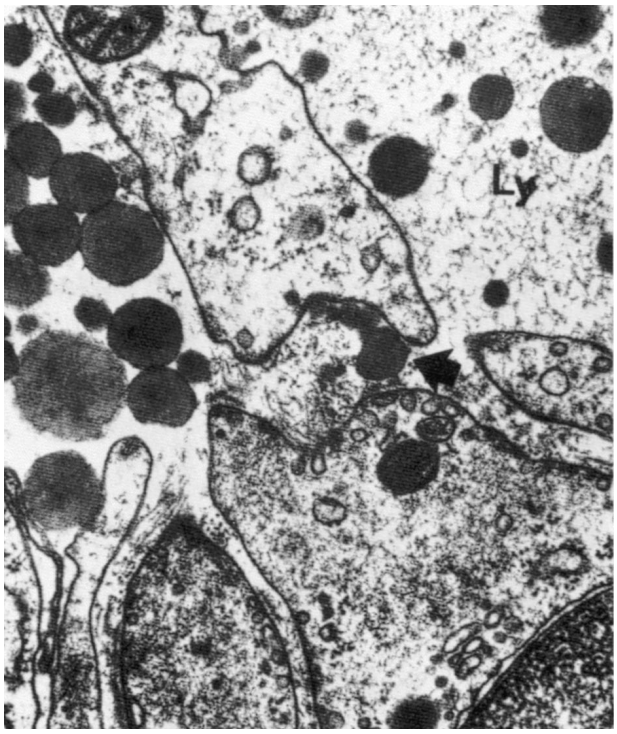


FIGURE 1 Chylomicrons produced by enterocytes entering lymphatic vessels (Ly) by moving through gaps (arrow) that develop between cytoplasmic extensions of the endothelial lining of the vessel.

lymphatic endothelium that enclose channels from the interstitial matrix (interstitium) to the lymphatic lumen along the length of the initial lymphatics.

The fine lymphatic vessels merge progressively, forming collecting vessels that transport the lymph to the main intestinal lymph duct, the cisterna chyli, and the thoracic duct. The thoracic duct then carries the lymph, which enters the general circulation by emptying into the subclavian vein. Valves are present in the main intestinal lymph duct and the thoracic duct. Although not well studied, a number of papers have reported innervation of the lymph vessels by sensory fibers.

FORMATION OF GASTROINTESTINAL LYMPH

In the small intestine, unlike most other organs, the source of fluids, electrolytes, peptides, and lipoproteins is the capillary filtrate and the intestinal absorption and secretion by enterocytes that line the intestinal epithelium. Various factors affect lymph flow as a result of changes in filtration of fluids and solutes from the capillaries. Plasma fluids, electrolytes, and proteins cross the capillary to diffuse through the interstitium and return to the circulation through the lymph. Because the plasma filtrate contributes to lymph flow, factors that affect the filtration of fluid from the capillaries also affect lymph flow. For example, in portal hypertension (elevation of venous pressure), enhanced intestinal lymph flow occurs as a result of the increase in transcapillary hydrostatic pressure gradient. In contrast, reduced hydrostatic pressure occurs during arterial hypotension (i.e., hemorrhage), which leads to reduction in lymph flow. It can also be predicted that a reduction in plasma protein concentration (hypoproteinemia) increases lymph flow by a reduction in the transcapillary oncotic pressure gradient (as explained by the Starling force equation). The lymph-to-plasma ratio, which is inversely proportional to the size of the protein, is another important factor regulating the amount of protein filtered from the capillaries. The bigger the protein, the less the amount that is filtered from the capillaries to the interstitium. For total protein, however, the lymph-to-plasma ratio is between 0.5 and 0.65.

The small intestine absorbs a large amount of fluid and solutes (especially sodium) from the intestinal lumen daily. This absorbed fluid expands the interstitial space, thereby increasing the interstitial pressure. This increase in interstitial pressure causes fluid to move from the interstitium to the initial lymphatics. Thus, the lymphatic system plays an important role in fluid absorption from the small intestine to the general circulation.

The liver also contributes to the lymph flow in the thoracic duct. Although no direct anatomical evidence is available to trace the origin of hepatic lymph, it is generally believed that hepatic lymph is formed at the space of Disse. The permeability of the endothelial lining of the liver sinusoids, unlike that of the small intestine, is very porous, allowing most proteins to travel through with little difficulty. Consequently, the protein concentration in liver lymph is almost as high as the protein concentration in the circulating blood. The protein in liver lymph is composed of protein filtered from blood and protein produced by the liver. Hepatic lymph flow increases dramatically with liver cirrhosis.

The Transport of Lipids in Lymph

Dietary lipid, chiefly triglyceride, undergoes hydrolysis in the small intestinal lumen to form 2-monoglycerides and fatty acids. The reaction is catalyzed by pancreatic lipase. After triglyceride digestion products are taken up by enterocytes, they are resynthesized to form triglycerides. These triglyceride lipid

droplets are then coated with phospholipids and apolipoproteins to form triglyceride-rich lipoproteins. The small intestine secretes both very-low-density lipoproteins (during fasting) and chylomicrons (during active fat absorption). Chylomicrons are richer in triglyceride and much larger compared to very-low-density lipoproteins (100–200 nm versus 50–80 nm). Using electron microscopy, enterocytes secrete chylomicrons by exocytosis. The interstitium is resistant to chylomicron movement. Therefore, the increased hydration of the interstitial matrix and the resulting increase in lymph flow are believed to facilitate chylomicron movement across the interstitial matrix and into the initial lymphatics. Chylomicrons are not only important in the transport of dietary fats by the small intestine, but they are also vehicles for the transport of fat-soluble vitamins and drugs. For instance, it has been demonstrated that *p*-aminosalicylic acid, tetracycline, and the highly lipophilic antimalarial, halofantrine, are transported in lymph. There is tremendous interest from the pharmaceutical industry in studying drug delivery via the lymphatic route.

Gastrointestinal Hormones in Lymph

In addition to transporting macromolecules, lymph is also a potential avenue by which gastrointestinal hormones enter the general circulation. The presence of gastrointestinal hormones in lymph has been documented previously by bioassays and radioimmunoassays of the gastrointestinal hormones. However, the relative importance of the lymphatic route versus the portal route in the transport of gastrointestinal hormones remains to be explored.

Lymphatic Transport of Proteins

As with lymph from other regions of the body, plasma provides the greatest source of proteins to intestinal lymph. There are, however, proteins in intestinal lymph derived from the small intestine. Apolipoproteins A-I, A-IV, and B48, the major apolipoproteins associated with lymph lipoproteins, are derived from the small intestine. The intestine is a major source of circulating apolipoproteins A-I and A-IV. The intestinal lymph also contains a number of enzymes secreted by enterocytes. For instance, enterocytes secrete alkaline phosphatase, which is associated with surfactant-like particles that appear in lymph during fat absorption. There is a linear relationship between the amount of fat and the amount of alkaline phosphatase in lymph. Another enzyme secreted by enterocytes is diamine oxidase (histaminase). It plays an important role in the breakdown of histamine and polyamines.

A major component of the lymphoid system, the gastrointestinal-associated lymphoid tissue (GALT), is located in the gastrointestinal tract. The GALT consists of three components: (1) Peyer's patches, (2) intraepithelial lymphocytes, and (3) plasma cells and lymphocytes in the lamina propria. The plasma cells secrete immunoglobulin A (IgA). Transported in lymph, the concentration of IgA is significantly higher in lymph than in blood.

The Transport of Xenobiotics in Lymph

In general, the degree of lipophilicity of foreign substances appears to determine the route of transport of these compounds from the small intestine, and the more lipid-soluble compounds tend to favor the lymphatic route. Clearly, the importance of lymphatic transport of xenobiotics rests on the fact that the liver is bypassed, thus allowing extensive systemic distribution of the compound. However, the small intestinal epithelial cells have the capacity to carry out both phase I and phase II reactions. Therefore, they can convert the absorbed lipophilic compounds to more hydrophilic compounds and direct more polar metabolites to the portal route. Lymphatic transport of numerous xenobiotics, including benzo[*a*]pyrene, 7,12-dimethylbenzanthracene, polychlorinated biphenyls, and dichlorodiphenyltrichloroethane (DDT), has been demonstrated. For this reason, the intestinal lymphatic transport of lipophilic toxic xenobiotics is of considerable interest, particularly as related to their potential role in cancer development.

PATHOPHYSIOLOGY OF GASTROINTESTINAL LYMPHATICS

To some extent, lymphatics are involved in every pathological process. One important example is their role in the dissemination of epithelial malignant tumors that arise in the gut and its associated structures. The role of disordered lymphatic function in the pathogenesis of selected gastrointestinal diseases is considered in the following discussions.

Primary Intestinal Lymphangiectasia

This disorder is thought to result from congenital structural abnormalities in the lymphatics of the intestinal wall or its mesenteries, causing impaired drainage of the mucosal initial lymphatics. Chylous effusions in the pleural or peritoneal cavity are often associated with intestinal lymphangiectasia. The central lymphatics of the intestinal villi are dilated and filled with protein and

chylomicron-rich lymph; the villi are clubbed as a result of mucosal edema. Gastrointestinal symptoms are generally mild, though abdominal pain, nausea, vomiting, and diarrhea with steatorrhea can occur. Severe protein depletion can lead to hypoproteinemic edema, causing impaired growth in children.

Primary intestinal lymphangiectasia is diagnosed by protein-losing enteropathy and mucosal biopsy. It can be treated in several ways: (1) by resection, if only a short segment of small intestine is affected, (2) by lymphovenous anastomosis, to correct the localized anomaly, and (3) by reduction in dietary long-chain triglycerides. This last treatment reduces the lipid required for transport by the lymphatics and reduces intestinal lymph formation, because fat feeding is a powerful intestinal lymphagogue (an agent that promotes lymph formation).

Secondary Intestinal Lymphangiectasia

Secondary intestinal lymphangiectasia is caused by any pathological process that obstructs major lymph trunks from draining the intestine. Crohn's disease, which involves the intestine or mesentery, is an example of this pathological condition. Chronic pancreatitis is another example and is caused by inflammation of the retroperitoneal tissues. Whipple's disease displays a moderate degree of secondary intestinal lymphangiectasia and results from the infiltration of lymph nodes with periodic acid Schiff (PAS)-positive macrophages. Cirrhotic patients with enteric loss of plasma proteins exhibit intestinal lymphangiectasia in small intestinal biopsies.

Inflammatory Bowel Disease

Edema is a prominent histopathological feature of all forms of inflammatory bowel disease. The pathogenesis of edema is complex. There is evidence that capillary permeability is enhanced as a result of local liberation of chemical mediators of inflammation such as histamine, serotonin, bradykinin, and other such substances. Bacterial endotoxins may also produce enhanced vascular permeability to protein, as may oxygen free radicals derived from leukocyte activity.

The extent to which abnormalities of lymphatic drainage play a part in edema of the intestine in inflammatory bowel disease is unclear. In Crohn's disease, there is lymphatic dilatation in all layers of the bowel wall, notably in the submucosa. It is suggested that an element of lymphatic obstruction from inflammatory changes in adjacent tissues and lymph nodes may be involved. Additionally, a defect in the function of intestinal lymphatic vessels may occur in Crohn's disease.

Acute Pancreatitis

Gross edema of the gland occurs in both acute interstitial (edematous) pancreatitis and necrotizing (hemorrhagic) pancreatitis. Distended lymphatics can usually be observed by morphological examination. It is unlikely that lymphatic obstruction plays a primary role in the pathogenesis of acute pancreatitis, but some evidence suggests that lymphatic obstruction may aggravate the autodigestive process by allowing toxic products to accumulate in the tissue. For instance, obstruction of lymph flow in experimental acute pancreatitis in animals increases the inflammatory process.

Chyluria

Chyluria results from situations in which a major lymph trunk communicates with the urinary tract. Patients with chyluria have afforded investigators the opportunity to study the human intestinal lymph lipoprotein and apoprotein spectrum.

See Also the Following Articles

Apoproteins • Barrier Function in Lipid Absorption • Carbohydrate Digestion and Absorption • Crohn's Disease • Lipoproteins • Lymphocytes • Pancreatitis, Acute • Protein Digestion and Absorption of Amino Acids and Peptides

Further Reading

Alpers, D. H. (1994). Digestion and absorption of carbohydrates and proteins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1723–1749. Raven Press, New York.

Barrowman, J. A. (1978). "Physiology of the Gastro-intestinal Lymphatic System." Cambridge University Press, New York.

Barrowman, J. A., and Tso, P. (1989). Gastrointestinal lymphatics. *In* "Handbook of Physiology" (J. D. Wood, ed.), Sect. 6, Vol. 1, pp. 1733–1777. American Physiological Society, Rockville, Maryland.

Nordskog, B., and Tso, P. (2001). An examination of the factors affecting intestinal lymphatic transport of dietary lipids. *Adv. Drug Deliv. Rev.* **50**, 21–44.

Phan, C. T., and Tso, P. (2001). Intestinal lipid absorption and transport. *Front. Biosci.* **6**, D299–D319.

Tso, P. (1994). Intestinal lipid absorption. *In* "Physiology of the Gastrointestinal Tract" (J. R. Johnson, ed.), 3rd Ed., pp. 1867–1907. Raven Press, New York.

Lymphocytes

RODNEY NEWBERRY
Washington University School of Medicine, St. Louis, Missouri

antigen Any molecule that can specifically bind to an antibody.
antigen-presenting cells Have the ability to internalize protein antigens, process them into peptides, and present these peptides on their surface in association with MHC II molecules. The primary antigen-presenting cells are dendritic cells, macrophages, and B lymphocytes.
B cell receptors Molecules present on B lymphocytes; contain cell surface immunoglobulin specific for antigen.
cytokines Small soluble proteins made by cells; can affect the behavior of other cells.
cytotoxins Proteins produced by cytotoxic T lymphocytes; participate in the destruction of target cells.
immunoglobulin Protein (antibody) produced by plasma cells; has the ability to bind specific antigen.
intraepithelial lymphocyte compartment Collection of T lymphocytes and NK cells interspersed between the intestinal epithelial cells.
lamina propria lymphocyte compartment Collection of bone marrow-derived cells (including lymphocytes and antigen-presenting cells) dispersed throughout the stroma, supporting the intestinal villi.
major histocompatibility complex molecules Polymorphic cell surface glycoproteins with the ability to bind peptide antigens and to stimulate T lymphocytes bearing receptors specific for these antigen/glycoprotein complexes.
Peyer's patches Organized lymphoid structures located on the antimesenteric border of the small intestine; contain lymphocytes and antigen-presenting cells and are important for the induction of immune responses against antigens from the intestinal lumen.
poly-Ig receptor Polymeric immunoglobulin receptor located on the basolateral surface of epithelial cells; binds polymeric immunoglobulins (IgA and IgM) and plays a critical role in transporting IgA produced by plasma cells in the lamina propria to the intestinal lumen.
T cell receptors Complexes on the surface of T lymphocytes; contain a disulfide-linked heterodimer of either α and β chains or γ and δ chains and are responsible for the recognition by T lymphocytes of specific antigen/major histocompatibility glycoprotein complexes.

The immune response can grossly be divided into innate and adaptive responses, which interact effectively in a complex series of coordinated events to protect higher organisms from invading pathogens. The innate immune response is important early in immune-mediated events, is present in individuals at all times, and does not increase in magnitude on repeated exposure to pathogens, whereas the adaptive or acquired immune response is directed at specific antigens, increases in magnitude with repeated exposure to those antigens, and results in immunological memory. The roles lymphocytes play in the immune response encompass a spectrum covering both innate and adaptive immune responses. Cell lineages include B lymphocytes, T lymphocytes, and natural killer cells. The B and T lymphocytes bear antigen-specific receptors and are the primary effector cells responsible for adaptive immune responses; conversely, natural killer cells lack antigen-specific receptors and play a role in innate immune responses.

INTRODUCTION

Lymphocytes are bone marrow-derived cell lineages arising from a common lymphoid progenitor (Table I). In general, lymphocytes can be classified by cell surface receptors and by the specific immune functions attributed to each cell type. In this scheme, understanding how lymphocytes are classified is fundamental to assigning a function for lymphocytes in the immune response. Lymphocytes circulate freely throughout the body and perform their functions in concert with other immunologically relevant cells and factors. An in-depth understanding of lymphocyte function requires understanding how lymphocytes circulate, interact with other cells, and perform their effector functions; although this is beyond the scope of this article, lymphocytes and their classification and function will be discussed within the structural framework of their functions in the immune response.

NATURAL KILLER CELLS

Natural killer (NK) cells, in contrast to B cells and T cells, lack antigen-specific receptors and play a role in the innate immune response. NK cells are functionally

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Lymphocyte Cell Lineages

Natural killer cells	B lymphocytes	T lymphocytes
Arise from a common lymphoid progenitor in the bone marrow	Arise from a common lymphoid progenitor in the bone marrow	Arise from a common lymphoid progenitor in the bone marrow
Non-antigen-specific inhibitory and activating receptors	Antigen-specific receptors recognizing native antigen	Antigen-specific receptors recognizing antigen in association with MHC molecules
Important in innate immune responses	Important in adaptive immune responses	Important in adaptive immune responses

defined by their ability to kill tumor cell lines without prior immunization. The mechanism by which NK cells kill their targets is similar to that of cytotoxic T lymphocytes. Cytotoxic granules are released by NK cells onto the surface of the target cell, penetrating the target cell membrane and resulting in programmed cell death of the target. NK cells bear two classes of receptors, activating receptors and inhibitory receptors. Inhibitory receptors recognize "normal" self major histocompatibility complex (MHC) class I molecules, expressed on all nucleated cells, and thus inhibit the killing of self when normal self MHC I is expressed. Activating receptors recognize a variety of ligands present on infected or malignant cells. NK cell function requires the expression of activating ligands by target cells as well as decreased or abnormal expression of self MHC I, thus allowing NK cells to receive an activating signal in the absence of an inhibitory signal. NK cell function can be enhanced by exposure to some cytokines, in particular interferon γ (IFNγ) and interleukin-12 (IL-12) produced by macrophages early in the immune response. Therefore, NK cells are a key component of the early (innate) immune response and are crucial for removal of abnormal and/or infected cells. The variety, repertoire, and restriction of the inhibitory and activating receptors on NK cells are not well understood and provide an area of active investigation.

COMMON PROPERTIES OF B AND T LYMPHOCYTES

In comparison to other cell types in the body, and to other cell types in the immune system specifically, B and T lymphocytes share many unique properties. The B and T lymphocytes are the cornerstones of the adaptive immune response, which is remarkable for its diversity, specificity, and memory. These characteristics are primarily due to the unique B and T lymphocyte surface receptors, the B cell receptor (BCR) and the T cell receptor (TCR). These receptors are generated by gene rearrangement of somatic DNA during the development of each B or T lymphocyte. This process results in the generation of billions of B and T lymphocytes, each expressing a unique receptor with its individual antigen specificity. Importantly, B and T lymphocytes express multiple copies of a single rearranged receptor on their surface, allowing them to recognize exclusively a single antigen. The development and function of lymphocytes are determined by signals received through these receptors. Through these unique antigen-specific receptors, B and T lymphocytes have the ability to respond to a diverse repertoire of antigens in a very specific way. The B and T lymphocytes also share the property of clonal expansion. When mature B and T lymphocytes encounter antigens specific for their receptors, in the appropriate environment, they respond and divide multiple times to produce clonal populations of lymphocytes expressing the same receptor. This clonal population has the ability to produce an augmented immune response on reexposure to the same antigen, and thus forms the basis for immunological memory, the principle behind vaccination.

B LYMPHOCYTES

The B lymphocytes recognize native antigen via their BCR and can participate in the immune response at multiple levels; however, one function unique to B lymphocytes is their differentiation into plasma cells, which produce secreted immunoglobulin, or antibody. After activation of mature B lymphocytes by ligation of the BCR by specific antigen, B lymphocytes may undergo additional rearrangement of their somatic DNA and differentiate into plasma cells to produce different classes, or isotypes, of immunoglobulins (IgA, IgM, IgG, IgE, and IgD). The various isotypes of immunoglobulins produced retain the antigen specificity of the progenitor B lymphocyte, although different isotypes of immunoglobulin perform distinct functions in the immune response. A prime example of these distinct functions is the role luminal IgA plays in protecting the host from invading pathogens at mucosal surfaces. In addition

TABLE II Properties of B-1 and B-2 B lymphocytes

B-1 B lymphocytes	B-2 B lymphocytes
Self-renewing from the peritoneal and pleural cavity	Renewed from the bone marrow
Spontaneous immunoglobulin production	Low or no spontaneous immunoglobulin production
Do not require cognate T lymphocyte interactions for differentiation into plasma cells	Require cognate T lymphocyte interaction for differentiation into plasma cells
Little or no somatic hypermutation	Somatic hypermutation to produce high-affinity antibody
Little or no memory response	Strong memory response
Contribute to fecal immunoglobulin production	Contribute to fecal immunoglobulin reproduction

to this unique effector function, B lymphocytes may also act as antigen-presenting cells (APCs), which, in association with class II MHC on their cell surface, present antigen to responding T cells.

The B lymphocytes can be divided into two categories, B-1 and B-2 B lymphocytes (Table II), which can be distinguished by the expression of cell surface molecules. Although both B-1 and B-2 B lymphocytes can differentiate into antibody-producing plasma cells, the pathways they take to achieve this state differ. B-2 B lymphocytes require cognate T lymphocyte interactions in order to differentiate optimally into antibody-producing plasma cells. The optimal differentiation of these B cells into high-affinity antibody-producing plasma cells is also dependent on the presence of organized lymphoid structures. These structures facilitate somatic hypermutation, a process in which B lymphocytes modify the genetic coding of the hypervariable region of the immunoglobulin gene responsible for binding antigen. This allows the development and selection of B lymphocytes expressing immunoglobulin genes with high affinity for a specific antigen. Mature B-2 B lymphocytes are continually produced by the bone marrow and circulate throughout the body.

In contrast to B-2 B lymphocytes, B-1 lymphocytes arise from a self renewing pool located primarily in the peritoneal and, to a lesser extent, pleural cavities. In comparison to B-2 B lymphocytes, B-1 B lymphocytes express a restricted repertoire of BCR, and do not require cognate T lymphocyte interactions for their differentiation into plasma cells, although they can benefit from cytokines produced by T lymphocytes. B-1 B lymphocytes do not undergo the process of somatic hypermutation and therefore are not optimal producers of high-affinity antibodies. In addition, B-1 B lymphocytes do not display a strong memory response. Collectively, these observations suggest that B-1 B lymphocytes may play a role in the nonadaptive (innate) phase of the immune response.

T LYMPHOCYTES

αβ and γδ T Lymphocytes

The T lymphocytes can be categorized by the chains comprising their TCR. The TCR present on the surface of T lymphocytes is a heterodimer composed of either α and β chains or γ and δ chains, thus dividing T lymphocytes into two distinct categories, αβ T lymphocytes and γδ T lymphocytes. The γδ T lymphocytes are preferentially enriched at some mucosal surfaces and bind antigen in association with nonclassical MHC molecules. The TCR repertoire of γδ T lymphocytes is more restricted when compared with the TCR repertoire of αβ T lymphocytes. The function of γδ T lymphocytes is poorly understood, and they are felt to represent a more primitive form of the adaptive immune response.

T Lymphocytes Recognize Antigen in the Context of MHC

In contrast to B lymphocytes, which recognize native antigen via the BCR, T lymphocytes recognize antigen peptides bound to specialized cell surface glycoproteins, MHC molecules. There are two classes of MHC molecules with different patterns of expression. Class I MHC, or MHC I, is expressed on all nucleated cells in higher organisms, whereas MHC II expression is primarily restricted to APCs. In contrast to B lymphocytes, T lymphocytes undergo maturation and selection in the thymus. During this time, T lymphocytes with gene rearrangements that produce a functional TCR with the ability to bind self MHC complexes on stromal cells in the thymus are either positively or negatively selected. Only T lymphocytes bearing TCRs binding self MHC with the appropriate affinity are positively selected. The T lymphocytes expressing TCRs binding self MHC with high affinity represent self-reactive T lymphocytes and are eliminated, as are T lymphocytes

TABLE III Properties of CD4 and CD8 T Lymphocytes

CD4+ (helper) T lymphocytes	CD8+ (cytotoxic) T lymphocytes
Undergo selection and maturation in the thymus	Undergo selection and maturation in the thymus
Recognize antigen in association with MHC II expressed on antigen-presenting cells	Recognize antigen in association with MHC I expressed on nucleated cells
Major effector functions mediated by the production of cytokines	Major effector functions mediated by the release of cytotoxins
Mediate a variety of effects, depending on the profile of cytokines produced	Kill cellular targets

bearing TCRs lacking the ability to bind self MHC, because these thymocytes will be unable to respond to antigen presented by self MHC-bearing antigen-presenting cells. This process generates a diverse population of T lymphocytes with the capacity to recognize antigens in association with self MHC.

CD4 (Helper) and CD8 (Cytotoxic) T Lymphocytes

The T lymphocytes may be further classified by their expression of the CD4 and CD8 coreceptors (Table III). The CD4 coreceptor facilitates interactions between the TCR and the class II MHC expressed on APCs, and the CD8 coreceptor facilitates interactions between the TCR and the class I MHC expressed on nucleated cells. Thus, T lymphocytes expressing the CD4 coreceptor can recognize antigens only in the context of self MHC II, and are appropriately named "helper" T lymphocytes, based on their function. T Lymphocytes expressing the CD8 coreceptor can recognize antigens only in the context of self MHC I and are thus named "cytotoxic" T lymphocytes, based on their function.

The T lymphocytes produce two classes of effector molecules, cytotoxins and cytokines. Cytotoxins are primarily produced by CD8 T lymphocytes responding to antigens presented in association with self MHC I. On the recognition of antigen/MHC complexes by the ligation of the TCR, CD8+ T lymphocytes release lytic granules, resulting in the death of their targets. In this way CD8+ T lymphocytes selectively eliminate infected or damaged target cells.

Although both CD4 and CD8 T lymphocytes produce cytokines, cytokines are the primary mechanism by which CD4 T lymphocytes mediate their effects. Previously activated CD4 T lymphocytes produce specific patterns of cytokines on reactivation, resulting in phenotypic patterns of immune responses. For example, one pattern of cytokine expression by CD4 T lymphocytes may facilitate immunoglobulin class switch by B lymphocytes and their subsequent differentiation into plasma cells, whereas another pattern of cytokine expression may promote proinflammatory cell-mediated immune responses. The production of specific patterns of cytokines is reproducible once a CD4 T lymphocyte has fully differentiated, and thus the pattern of cytokines produced by differentiated CD4 T lymphocytes is a means to further subdivide these T lymphocytes into functional groups, or subsets. The factors directing the differentiation of CD4 T lymphocytes into the different subsets are at least partially dependent on the cytokine milieu present when an undifferentiated, or naive, T lymphocyte first encounters antigen and MHC. Thus, although CD4 T lymphocytes are primary effectors of the acquired immune response, they can promote a spectrum of phenotypic responses, including tolerance, humoral (antibody-mediated) immune responses, and proinflammatory cellular immune responses.

LYMPHOCYTE FUNCTIONS RELEVANT TO THE GASTROINTESTINAL TRACT

Lymphocyte Location in the Gastrointestinal Tract

The gastrointestinal tract contains the largest number of lymphocytes in the body. These lymphocytes are distributed throughout the gastrointestinal tract in three phenotypically and physically distinct compartments: the Peyer's patches (PPs), the lamina propria lymphocyte (LPL) compartment, and the intraepithelial lymphocyte (IEL) compartment (Fig. 1).

Peyer's patches are organized lymphoid structures located on the antimesenteric border of the small intestine. These structures resemble lymph nodes in architecture and composition, however they lack afferent lymphatics and sample luminal antigens via specialized antigen-transporting cells known as microfold cells (M cells), which are located in the overlying follicle-associated epithelium. Peyer's patches contain predominantly B-2 B lymphocytes and some αβ T lymphocytes. The organization of the Peyer's patches allows optimal interactions of B lymphocytes, T lymphocytes, and

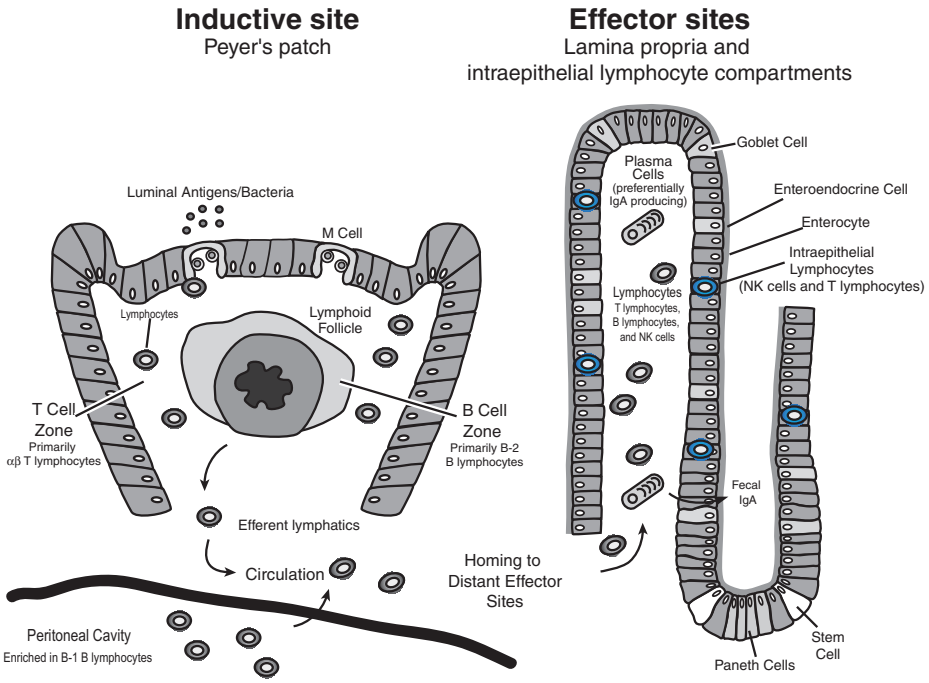


FIGURE 1 Distribution of lymphocytes in the gastrointestinal tract.

APCs for the induction of immune responses. Thus, Peyer's patches are thought to be the site of induction of intestinal immune responses. In the colon, lymphoid nodules, which are structures with an architecture similar to that of the Peyer's patch, are felt to perform analogous functions.

The lamina propria is a loosely organized lymphoid compartment within the supporting stroma of the intestine. The composition of the bone marrow-derived cells within the lamina propria resembles that of the Peyer's patch; in addition, the lamina propria contains a population of NK cells. However, in contrast to the Peyer's patch, a majority of lymphocytes within the lamina propria have a phenotype consistent with previous activation. The lamina propria is the major effector compartment for the intestinal immune response, containing a majority of plasma cells responsible for fecal immunoglobulin production, and is the site of mononuclear cell infiltration and inappropriate inflammatory responses in intestinal inflammatory disorders.

The IEL compartment contains T lymphocytes and NK cells, interspersed among the intestinal epithelial cells. In contrast to more classic lymphoid structures, an increased proportion of T lymphocytes in the IEL compartment bear the γδ TCR and the CD8αα isoform of the CD8 coreceptor, which have been associated with extrathymic T lymphocyte development. The function of intraepithelial lymphocytes is poorly understood.

Homing

The mucosal immune system in general, and the intestine specifically, have a unique ability to transmit antigen-specific immune responses selectively between distant mucosal surfaces. The mechanism allowing these coordinated immune responses is lymphocyte homing, a series of events in which lymphocytes activated by specific antigen at a mucosal surface selectively migrate to and reside in a distant mucosal surface to perform their effector function. Lymphocyte homing to the intestine is a multistep process involving antigen-specific activation of lymphocytes, resulting in the selective migration of these lymphocytes to their effector sites, the LPL and IEL compartments (Fig. 1). Homing is at least partially dependent on induced expression of cell surface molecules (integrins) by lymphocytes in the Peyer's patch. Integrins bind specific receptors on specialized venules in the lamina propria, thus allowing lymphocytes activated in the Peyer's patch to cross these venules and reside in the lamina propria. This multistep process allows for the coordinated transmission of antigen-specific responses between distant mucosal surfaces.

Fecal Immunoglobulin Production

A primary function of the intestinal immune system is the production of fecal immunoglobulin, which is

important for protection against invading organisms at mucosal surfaces. The events resulting in the production of fecal immunoglobulin are complex and involve the activation, migration, and differentiation of both B-1 and B-2 B lymphocytes (Fig. 1). B-1 B-lymphocytes contributing to the pool of plasma cells that produce fecal immunoglobulin arise primarily from the peritoneal cavity and migrate to the intestinal lamina propria to perform their effector function. These B-1 B-lymphocytes undergo differentiation into plasma cells during their migration or shortly after their arrival in the intestinal lamina propria. In contrast, B-2 B-lymphocytes contributing to the pool of plasma cells that produce fecal immunoglobulin arise from organized lymphoid structures and differentiate into plasma cells after their arrival in the intestinal lamina propria. Therefore, understanding the regulation of fecal immunoglobulin production requires accounting for factors inducing the migration and differentiation of two phenotypically and physically distinct subsets of B lymphocytes.

Fecal immunoglobulin is predominantly of the IgA isotype and to a lesser extent of the IgM isotype. Mechanistically, this arises for two reasons: (1) the need for antibody to cross the intestinal epithelium to enter the intestinal lumen and (2) the preference for lamina propria plasma cells to produce immunoglobulin of the IgA isotype. Antibody of the IgA isotype can be formed as monomers or dimers held together by a polypeptide, the J chain. Monomeric IgA is predominant in the serum whereas dimeric IgA predominates in intestinal secretions. Dimeric IgA produced by plasma cells in the intestinal lamina propria has the ability to bind a polymeric immunoglobulin (poly-Ig) receptor on the basolateral surface of intestinal epithelial cells. Binding the poly-Ig receptor allows for the uptake and transport of dimeric IgA across the epithelium into the intestinal lumen. IgM is a pentameric immunoglobulin with the ability to bind the polymeric Ig receptor as well, and therefore IgM may also be transported across the intestinal epithelium into the intestinal lumen. The preponderance of IgA-producing plasma cells in the intestinal lamina propria and the preferential production of dimeric IgA by these plasma cells contribute to making IgA the most abundant immunoglobulin isotype in the intestinal lumen.

CONCLUSION

Lymphocytes are cells comprising B lymphocytes, T lymphocytes, and natural killer cells, all lineages arising from a common lymphoid progenitor in the bone marrow. Lymphocytes circulate throughout the body and perform a wide variety of effector functions in both the early (innate) and late (adaptive) phases of the immune response. In general, lymphocytes can be subclassified into cell types by their expression of cell surface molecules, and by specific immune functions assigned to these cell types. Lymphocytes play a prominent role in host defense and are specifically responsible for carrying out coordinated immune responses between mucosal surfaces and for the production of fecal immunoglobulin. An in-depth understanding of lymphocyte function requires understanding how lymphocytes mature, differentiate into effector cells, and interact with other cells, both bone marrow derived and non-bone marrow derived.

See Also the Following Articles

Immunoglobulins • Lupus Erythematosus • Lymph, Lymphatics, and Lymph Flow • Mucosa-Associated Lymphoid Tissue (MALT)

Further Reading

Janeway, C., Travers, P., Walport, M., and Shlomchik, M. (2001). "Immunobiology," 5th Ed. Garland Publ., New York.

Ogra, P., Mestecky, J., Lamm, M., Strober, W., Bienenstock, J., and McGhee, J. (1999). "Mucosal Immunology," 2nd Ed. Academic Press, San Diego.

Stenson, W., Newberry, R. and Lorenz, R. (2001). Immunologic diseases of the gastrointestinal tract. *In* "Samter's Immunologic Diseases" (K. Austen, M. Frank, J. Atkinson, and H. Cantor, eds.), 6th Ed., Vol. II, pp. 693–711. Lippincott Williams & Wilkins, Philadelphia.

Lymphomas

HYUN SIL KIM
University of California, San Diego

enteropathy-associated T-cell lymphoma (EATL) A clinically aggressive tumor typically seen in patients with a long history of celiac disease. Unlike other gastrointestinal lymphomas, EALT is a T-cell lymphoma derived from the intestinal intraepithelial T-cell population.

mucosa-associated lymphoid tissue (MALT) lymphoma A low-grade primary B-cell gastric lymphoma associated with *Helicobacter pylori* infection; a MALT-like organization of the stomach is induced by chronic gastric inflammation and development of lymphoma.

The gastrointestinal tract is the most common site of extranodal primary lymphoma involvement. The non-Hodgkin's lymphoma is the predominant variety, usually involving the stomach and small intestine, as primary colonic lymphoma constitutes <10% of all gastrointestinal lymphomas. Primary intestinal Hodgkin's lymphoma is relatively rare. Patients who have immunologic disorders, such as collagen vascular disease, congenital immune deficiency syndromes, and acquired immunodeficiency syndrome, have an increased risk of developing gastrointestinal lymphomas. In addition, organ-transplanted patients who have received immunosuppressive agents are at risk of developing lymphomas of the intestine.

NON-HODGKIN'S LYMPHOMA

The majority of intestinal lymphomas are of the non-Hodgkin's type and, except for enteropathy-associated T-cell lymphoma, the B-cell variety predominates. These B-cell lymphomas are categorized according to their histological features: diffuse large cell or small noncleaved cell lymphomas. The stomach is often the most common site of primary lymphoma in Western countries and is seen in 50–80% of cases, followed by the small intestine (10–30%) and the colon (<10%).

Clinical Features

The initial presenting symptoms depend on the site and the extent of lymphomatous involvement, but usually patients will have nonspecific abdominal pain/cramps, weight loss, malaise, abdominal distension, malabsorption, or anemia. Intestinal perforation can also occur, particularly in small intestinal lymphomas. Other complications, such as intussuception and obstruction, have also been reported. Physical examination may be normal, but often abdominal distension or mass may be detected along with occult blood in the stool. Along with history and physical examination, the diagnostic workup usually includes endoscopies for a diagnosis and to obtain tissue specimens, abdominal and pelvic computed tomography, and barium studies of the upper and lower gastrointestinal (GI) tract for localization. Endoscopically, intestinal lymphomas have varied appearances and may present as multifocal ulcerations with irregular margins, nodules, large masses, or diffuse mucosal thickening with loss of normal intestinal folds.

Disease Course

The important considerations for the staging of lymphomas include tumor histology, extraintestinal lymph node involvement, extent of metastasis, and size of the tumor. Poor prognostic features include advanced stage, aggressive tumor growth, tumor mass >7 cm, invasion through the intestinal wall, unresectability, and bcl-2 oncogene rearrangement. The 2-year survival rate is ~80% for the early stage lymphomas, but in the advanced stage, the survival rate decreases to <30%.

Treatment

The therapy for intestinal lymphoma depends on the stage of the disease. Surgical debulking is recommended for patients with high tumor burden followed by chemotherapy and/or radiotherapy to improve survival, even for patients with stage IV disease. Surgical interventions have shown survival benefits, particularly in patients with obstructive tumors and in those who are at high risk of tumor perforation or massive hemorrhage. Since curative resection has been difficult to achieve, especially for the advanced aggressive tumor, a combined treatment plan with surgery, chemotherapy,

 Copyright 2004, Elsevier (USA). All rights reserved.

and radiotherapy has most often been utilized. The most commonly used chemotherapeutic agents for intestinal lymphoma are cyclophosphamide, doxorubicin, prednisone, methotrexate, vincristine, and bleomycin, which are used as a combination treatment. For those patients who demonstrate a complete response to chemotherapy, the predicted 4-year survival rate is 80–85%.

Lymphomas in Human Immunodeficiency Virus

In patients with human immunodeficiency virus infection, the non-Hodgkin's lymphoma of the intestine is the second most common extranodal site affected and has been known to occur in approximately 10% of patients. The majority of these B-cell lymphomas are of a highly aggressive, immunoblastic type. In addition to vague abdominal symptoms, some patients have presented with entero-colonic fistulas and bowel perforation. Epstein–Barr virus, human herpesvirus 8, and Kaposi's sarcoma-associated herpesvirus have been associated with intestinal lymphomas in acquired immunodeficiency syndrome. Treatment options are similar to those for other non-Hodgkin's lymphomas, and surgical debulking procedures, chemotherapy (cyclophosphamide, doxorubicin, vincristine, and prednisone to name a few), and radiation therapy have all been used with variable success. Successful treatment and response depend on the stage of the lymphoma and other prognostic factors, such as tumor histology and cytogenetic abnormalities.

MUCOSA-ASSOCIATED LYMPHOID TISSUE-TYPE LYMPHOMA

Of the primary gastric lymphomas, mucosa-associated lymphoid tissue (MALT) lymphoma has been of a particular interest due to its indolent clinical course and the causative agent that is associated with this type of lymphoma. *Helicobacter pylori* is a multiflagellate microaerophilic bacterium that colonizes the stomach and has been correlated with the induction of gastritis and peptic ulcer disease. In addition, *H. pylori* is associated with low-grade B-cell gastric MALT lymphoma in 80–90% of cases. Low gastric pH is an important aspect of the host defense system against bacterial infection in the stomach; however, *H. pylori* produces urease, which is active at low pH and produces ammonium, which creates a neutral environment where the bacteria can live and multiply. Although the exact pathogenesis of MALT lymphoma from *H. pylori* infection has not been defined, the bacterial infection clearly exerts a strong antigenic stimulus on the host immune response and a prolonged, chronic *H. pylori* infection most likely induces a malignant transformation of activated B cells.

In healthy individuals, MALT is found mostly in the terminal ileum—Peyer's patches being a prime example of this type of organized lymphoid follicle—but MALT lymphoid follicles do not exist in the stomach. However, in the setting of *H. pylori* infection, gastric MALT formation has been demonstrated consistently in all gastric tissues examined and the organism itself has been shown to be an important driving force in the stimulation of tumor-infiltrating lymphocytes. Due to its association with gastric MALT lymphoma and gastric cancer, *H. pylori* is now considered an infectious carcinogen.

Patients with MALT lymphoma will have clinical symptoms similar to those with peptic ulcer disease or gastritis, which include epigastric pain, nausea, dyspepsia, and bleeding. Endoscopy of the stomach may show gastritis, ulceration, or erosion. Since the antrum and distal body are common sites of *H. pylori* infection, gastric MALT lymphomas occur mostly at these sites, although any area of the stomach can be affected. Histologically, the structure of the MALT lymphoma resembles Peyer's patches and displays diffuse reactive polymorphic lymphoid follicles, which are notable for small- to medium-sized B cells infiltrating the epithelium in low-grade MALT lymphoma. Characteristic lymphoepithelial lesions with invasion of the epithelium by neoplastic B cells, or centrocytes, are prominent histologic features and immunohistologic staining of these lesions with pan-B-cell antibodies will be positive. Unlike other nodal B-cell lymphomas, MALT lymphoma has different genotypes and is characteristically negative for *bcl*-2 oncogene rearrangement. It has also been reported that trisomy 3 occurs in >60% cases of low-grade MALT lymphoma. One genetic defect, t(11;18) (q21;q21), has been observed to occur specifically in MALT lymphomas and this particular translocation defect was seen in approximately 65% of patients who do not respond to *H. pylori* eradication.

For the treatment of low-grade MALT lymphoma, eradication of *H. pylori* with various combinations of antibiotics has been shown to induce clinical remission in 70–90% of cases. Zithromycin, ampicillin, flagyl, and tetracycline along with acid suppressive agents have been commonly used for *H. pylori* treatment. Although MALT lymphoma is an indolent tumor and responds well to antibiotics for tumor regression, high-grade MALT lymphoma generally does not respond to the standard *H. pylori* therapy. Combinations of chemotherapy with cyclophosphamide, doxorubicin, vincristine, and prednisone, radiotherapy, and (or) gastric resection

have shown various survival rates ranging from 50 to 100% in those patients who did not respond to anti-*H. pylori* treatment.

ENTEROPATHY-ASSOCIATED T-CELL LYMPHOMA

In a subset of patients with a long history of celiac disease, enteropathy-associated T-cell lymphoma (EATL) has been observed. In contrast to the majority of primary intestinal lymphomas, which are of B-cell origin, EATL is strictly a T-cell lymphoma. As a complication of the disease, EATL is typically seen in the fifth decade of life and can occur even in the setting of a strict compliance with a gluten-free diet. Patients can present with abdominal pain, diarrhea, weight loss, and anemia and symptoms may resemble those with recurrent or refractory celiac disease.

The most common site of EATL involvement is the jejunum and EATL can be multifocal. Endoscopically, mucosal ulcerations of the small intestine may be observed. Immunohistologic studies of the bowel specimens from EATL demonstrate clonal rearrangements of the T-cell receptor in $>90\%$ of the tumors examined and, in some patients, showed identical T-cell receptor gene arrangements in the tumor and in the adjacent diseased intestine. Thus, EATL is believed to arise from a clonal transformation of inflammatory intestinal intraepithelial T cells into malignant lymphoma. The exact mechanism of development of enteropathy-associated T-cell lymphoma from chronic enteropathy due to gluten intolerance is unknown, although some studies have linked Epstein-Barr virus infection with the development of EATL in patients with celiac disease.

Generally, EATL follows a highly aggressive clinical course and carries a poor prognosis. One study has reported a 5-year survival of $<10\%$ with $\sim 50\%$ mortality rate within 6 months of diagnosis. Therefore, although current chemotherapeutic regimens and surgical intervention have shown some success, early clinical suspicion and detection in patients with long-standing celiac disease who develop diarrhea and other symptoms are strongly recommended.

See Also the Following Articles

AIDS, Gastrointestinal Manifestations of • *Helicobacter pylori* • Hepatic Granulomas • Mucosa-Associated Lymphoid Tissue (MALT)

Further Reading

Clayton, F., and Clayton, C. H. (1997). Gastrointestinal pathology in HIV-infected patients. Gastroenterol. *Clin. North Am.* **26**, 191–240.

Daum, S., Weiss, D., Hummel, M., Ullrich, R., Heise, W., Stein, H., Riecken, E.-O., Foss, H. D., and the Intestinal Lymphoma Study Group (2001). Frequency of clonal intraepithelial T lymphocyte proliferations in enteropathy-type intestinal T cell lymphoma, coeliac disease, and refractory sprue. *Gut* **49**, 804–812.

Gill, S. S., Heuman, D. M., and Mihas, A. A. (2001). Small intestinal neoplasms. *J. Clin. Gastroenterol.* **33**, 267–282.

Isaacson, P. G. (1994). Gastrointestinal lymphoma. *Hum. Pathol.* **25**, 1020.

Issacson, P. G., and Wright, D. H. (1983). Malignant lymphoma of mucosa-associated lymphoid tissue. *Cancer* **52**, 1410.

Koch, P., del Valle, F., Berdel, W. E., Willich, N. A., Reers, B., Hiddemann, W., Grothaus-Pinke, B., Reinartz, G., Brockmann, J., Temmesfeld, A., Schmitz, R., Rube, C., Probst, A., Jaenke, G., Bodenstein, H., Junker, A., Pott, C., Schultze, J., Heinecke, A., Parwaresch, R., Tiemann, M., and the German Multicenter Study Group (2001). Primary gastrointestinal non-Hodgkin's lymphoma. I. Anatomic and histologic distribution, clinical features, and survival data of 371 patients registered in the German Multicenter Study GIT NHL 01/92. *J. Clin. Oncol.* **19**, 3861–3873.

Mulder, C. J., Wahab, P. J., Moshaver, B., and Meijer, J. W. (2000). Refractory coeliac disease: A window between coeliac disease and enteropathy associated T cell lymphoma. *Scand. J. Gastroenterol. Suppl.* **232**, 32–37.

Wotherspoon, A. C. (1998). Gastric lymphoma of mucosa-associated lymphoid tissue and *Helicobacter pylori*. *Annu. Rev. Med.* **49**, 289–299.

Lynch Syndrome/Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

C. RICHARD BOLAND
Baylor University Medical Center, Dallas

DNA mismatch repair system Group of proteins that recognize and correct errors made by DNA polymerase during DNA replication.
germ-line mutation Parentally inherited mutation that is present in every cell of the body.
human Mut S homologue 2 Mismatch repair gene.
microsatellite Short, repetitive DNA sequence; repeats consist most commonly of mononucleotide repeats (A_n, G_n, etc.), dinucleotide repeats ($[Ca_n]$), trinucleotide repeats, or occasionally tetranucleotide or pentanucleotide repeats. Poly(A) or dinucleotide repeats are, by far, the most common of these. These repetitive sequences are highly prone to insertion or deletion mutations during DNA replication, and require DNA mismatch repair activity for faithful replication.
microsatellite instability Mutational signature in which ubiquitous errors are seen in repetitive DNA sequences as a consequence of losing mismatch repair activity.
somatic mutation Acquired mutation in a non-germ-line tissue.

Lynch syndrome is one of the familial, genetic forms of colorectal cancer. It is inherited as an autosomal dominant disease caused by a germ-line mutation in one of the genes involved in DNA mismatch. Historically, the disease has also been called hereditary non-polyposis colorectal cancer to distinguish it from the far less common, but phenotypically more distinctive disease, familial adenomatous polyposis. Familial adenomatous polyposis is characterized by hundreds to thousands of premalignant adenomatous polyps, which antedate the onset of cancer. In Lynch syndrome, the number of antecedent adenomas is relatively small, but these are prone to rapid conversion into cancers. Thus, Lynch syndrome is characterized by some precursor polyps, but not florid polyposis, and a very high risk for cancers of the colorectum and other sites.

INTRODUCTION

Colorectal cancer is a very common and highly lethal disease that affects approximately 150,000 Americans each year. Approximately one-third of these patients will succumb to this disease; however, the outcome is highly dependent on the stage of disease at the time of diagnosis. Diagnosis of the disease in stage 1 or 2 can lead to a cure in 80–95% of the cases.

Familial clusters of cancer occurring in young people raise the possibility of Lynch syndrome. The diagnosis is first suspected on clinical grounds when the "Amsterdam criteria" are met. These consist of a family with at least three members with a characteristic cancer of this disease (see later), one of whom is a first-degree relative of the other two, in which one person has developed cancer at age 50 or less, and familial adenomatous polyposis (FAP) has been ruled out. About 2–3% of all colorectal cancers are thought to be Lynch syndrome on the basis of a germ-line mutation in a DNA mismatch repair (MMR) gene. The actual number of cancers that are caused by germ-line mutations in a DNA MMR gene is probably higher than currently appreciated, because the current technologies can identify mutations only half the time, even in well-characterized families. When a definitive diagnosis can be made by finding a germ-line mutation in a DNA MMR gene, the disease can be called Lynch syndrome. Other names historically associated with this disease include cancer family syndrome and site-specific hereditary colon cancer. These terms are misleading because they imply that the colon is the exclusive organ at risk for cancer, which is not the case.

CLINICAL MANIFESTATIONS

The average age to develop a sporadic colon cancer in the United States is approximately 69 years. In the setting of Lynch syndrome, the average age is approximately 40–45 years. Patients with Lynch syndrome are at risk for cancer of several different organs; however, not every organ is at increased risk for cancer. The lifetime risk for any cancer by age 70 in this disease is estimated to be 91% for men and 69% for women. The lifetime risk of colon cancer is approximately 70% for men and 30% for women. Women with Lynch syndrome have a 40% lifetime risk of developing endometrial cancer. There are substantially increased risks for cancers of the stomach, small intestine, ovaries,

 Copyright 2004, Elsevier (USA). All rights reserved.

renal pelvis, ureter, pancreas, central nervous system, and hepatobiliary system. However, there is no increased risk for cancers of the lung or breast, for example. The premorbid phenotype is, with some exceptions, normal for carriers of a mutant DNA MMR gene. The premier feature of the disease is the sharp increase in cancer risk and the earlier age of onset of disease compared with the sporadic forms. In general, these cancers develop about 25 years earlier than would be expected for sporadic disease. Approximately 75% of sporadic colon cancers occur distal to the splenic flexure, most occurring in the sigmoid colon and rectum. In Lynch syndrome, 60–70% of colon cancers occur proximal to the splenic flexure. The biological basis of the proximal predilection is unknown.

Lynch syndrome colon cancers have a tendency to be associated with certain unique pathological features. These colon cancers have a tendency to appear poorly differentiated or medullary and to be mucinous, and a very high proportion is diploid or near-diploid. Some of these tumors show an intense Crohn-like reaction with a large number of tumor-infiltrating lymphocytes. Although none of these features is present in the majority of colon cancers, one or more of the features will be present often enough that an experienced pathologist can be suspicious of Lynch syndrome in a young patient with a proximal colon cancer that has these pathological features. The colons of patients with Lynch syndrome are not characterized by diffuse polyposis; however, there is an excess number of adenomatous polyps in young people with Lynch syndrome. The available data suggest that these benign polyps progress to malignancy at an accelerated rate compared to sporadic polyps. Thus, accelerated progression of the disease is thought to be a key clinical characteristic, which dictates the management as well. A study of patients with Lynch syndrome at a median age of 42 years found an incidence of adenomatous polyps of 30%; 18% of the patients had proximal colonic adenomas and 20% had multiple colonic adenomas. By contrast, control age-matched patients had an overall incidence of adenomatous polyps of 11%, in which 1% were proximal, and 4% of patients had multiple polyps. Additionally, there is a striking increase in polyps with advancing age as seen in Lynch syndrome patients compared with sporadic patients.

THE GENETIC BASIS OF LYNCH SYNDROME

The DNA MMR system consists of a group of proteins that cooperate to recognize and repair errors made in the DNA sequence during replication; some of these genes are linked to Lynch syndrome (see Table I). Two of the genes, *human Mut L homologue 2* and *human Mut L homologue 6* (*hMSH2* and *hMLH1*), are the "major" DNA MMR genes. The proteins coded by these genes will heterodimerize with a homologous protein, hMSH6 and hPMS2, respectively, to form working protein complexes that will both recognize mismatched DNA during its replication and lead to excision of that region followed by resynthesis, which repairs the error. hMSH2 + hMSH6 is the principal heterodimer that recognizes single base-pair mismatches and short "loop-outs" in the DNA that occasionally occur during an attempt to replicate a repetitive sequence. When the DNA polymerase encounters a mononucleotide repeat or a dinucleotide repeat, a small proportion of these, perhaps 1 in 100 to 1 in 1000, suffer a deletion of one or more repetitive elements in the newly synthesized strand, and the nascent double-stranded DNA will have a loop-out on the template strand. If not corrected, the daughter DNA strand will be shorter by one repeat element after the next round of replication. The DNA MMR system recognizes these errors, and corrects them.

The hMSH2 protein can alternatively heterodimerize with the hMSH3 protein, and this protein pair has specificity that recognizes larger loop-outs than those bound by hMSH2 + hMSH6. Thus, by having diverse proteins within the cell, a broad range of

TABLE I DNA Mismatch Repair Gene Mutations and Lynch Syndrome

Gene	Chromosome	Link with cancer
Common cause		
hMSH2	2p16	40–50% of families meeting Amsterdam criteria
hMLH1	3p21	40–50% of families meeting Amsterdam criteria
hMSH6	2p16	Causes attenuated Lynch syndrome in 8% of familial clusters of colon cancer that do not meet Amsterdam criteria
Less common		
hPMS2	7p22	Uncommon cause; mutations may be homozygous in some families (AR)
hMLH3	14q24.3	Uncommon cause of Lynch syndrome
No apparent link		
hPMS1	2q31	Not found in Lynch syndrome families
hMSH3	5q11–q13	Not found in Lynch syndrome families

replication errors can be recognized. The hMLH1 protein dimerizes with hPMS2 to complete the DNA repair process. The biochemical and biophysical details of this are not completely understood. Two additional proteins, hMLH3 and hPMS1, can also heterodimerize with hMLH1, but again, the functions of these complexes are not entirely understood; presumably, they add to the ability of the system to recognize and repair different types of replication errors.

Lynch syndrome is suspected on the basis of a strong family history in which the Amsterdam criteria are met. In about half of such families, a germ-line mutation can be found in either *hMSH2* or *hMLH1*. The proportion of mutations found in these two genes is approximately equal and depends on "founder effects" in the population. In addition, a small proportion of families with Lynch syndrome will have germ-line mutations in the *hMSH6* gene. However, because *hMSH6* is not absolutely essential for at least partial DNA MMR activity, patients with this mutation have an attenuated form of the disease. In the case of *hMSH6* germ-line mutations, the median age to develop cancer is approximately 60 years and the penetrance for colon cancer is substantially lower, but women are still at a similar risk for endometrial cancer. Thus, germ-line mutations in *hMSH6* have been called "attenuated Lynch syndrome." In one study of familial clusters of colon cancer in which the Amsterdam criteria were not met, over 8% had germ-line mutations in *hMSH6*.

Rare families have germ-line mutations in *hPMS2*; some of these families have homozygous mutations in this gene. It is not clear whether the very small number of families linked to this gene reflect the requirement for homozygous mutations, but if this is the case, it could explain recessive inheritance patterns. Germ-line mutations in *hMLH3* have been found in Lynch syndrome, but these are much less common compared to mutations in *hMSH2* or *hMLH1*. Individuals who have a germ-line mutation in a critical DNA MMR gene have one mutated and one wild-type copy of the gene. The DNA MMR activity in these cells is normal. A second mutation or other genetic event is required to inactivate the wild-type allele inherited from the unaffected parent. When this occurs, the cell becomes deficient in DNA MMR activity. MMR deficiency is not, per se, sufficient to create a neoplastic phenotype. Loss of DNA MMR activity accelerates the accumulation of mutations occurring throughout the genome, particularly at microsatellite sequences.

Most microsatellite sequences occur in introns or between genes. Mutations in these sequences are thought to be innocuous. However, point mutations can occur in critical genes, and deletion mutations are very likely to occur at microsatellite sequences in the coding regions of genes. Many genes with coding microsatellites have been identified that encode a protein critical to the regulation of the growth of epithelial cells (see Table II). Mutations in these "target genes" mediate the development of neoplasms in the colon, endometrium, and elsewhere. These genes are normal in most tissues but are mutated in the tumors.

A tumor from a patient with Lynch syndrome contains thousands of microsatellite mutations, referred to as microsatellite instability (MSI). (There are hundreds of thousands of these sequences in the genome.) An indirect way to help confirm a suspicion of Lynch syndrome is to determine whether the tumor has MSI. Finding this molecular signature in a tumor indicates that the tumor developed in the absence of DNA MMR activity. About 12–15% of all colorectal cancers have MSI, but only about 25% of these are attributable to Lynch syndrome. The other 75% are due to an acquired silencing of the *hMLH1* gene, by hypermethylation of the gene promoter. When a tumor has MSI, immunohistochemistry can detect hMSH2 and hMLH1 proteins. The absence of expression of one of these two proteins strongly suggests its involvement in causing the disease. The absence of hMSH2 expression almost always indicates Lynch syndrome. However, hMLH1 staining involves additional consideration. It may not be expressed due to the acquired silencing of its gene (which usually occurs in older individuals, making it easier to distinguish from Lynch syndrome). Also, some missense mutations in hMLH1 inactivate the protein, but do not abrogate its immunoreactivity.

Some complexity in genotype–phenotype relationships has been found. The simultaneous presence of a common and otherwise innocuous polymorphism in the cyclin D1 gene is associated with a median onset of cancer 10 years younger in some kindreds with *hMSH2* mutations, creating a more virulent situation. Germ-line mutations in the *hMSH6* gene lead to a more attenuated form of the disease, with lower penetrance, later onset, and a slight modification of target organs that develop cancer. Some patients have additional skin lesions, including multiple sebaceous neoplasms and keratoacanthomas; this is referred to as the Muir–Torre syndrome and is usually associated with germ-line mutations in the *hMSH2* gene.

MANAGEMENT OF LYNCH SYNDROME

Managing families with Lynch syndrome is a multidisciplinary undertaking. First, such families are identified by obtaining a family history; the diagnosis is confirmed with a blood test (commercially available) for a

TABLE II Target Genes of Microsatellite Instability

Repeat sequence: A10
- *TGFβ2R* (regulates colonic cell growth)
- *CASPASE-5* (programmed cell death gene)
- *MBD4* (DNA repair gene)
- *RIZ* (programmed cell death gene)
- *AIM2*
- *OGT*
- *SEC63* (also has A9)

Repeat sequence: A9
- *BLM* (DNA repair gene)
- *CHK1* (DNA repair gene)
- *MLH3* (DNA mismatch repair gene)
- *RAD50* (DNA repair gene)
- *RIZ* (programmed cell death gene; also A8)
- *TCF-4* (regulates colonic cell growth)
- *WISP-3* (regulates colonic cell growth)
- *GRB-14*
- *RHAMM*

Repeat sequence: A8
- *ACTRII* (growth factor receptor)
- *APAF* (programmed cell death gene)
- *BCL-10* (programmed cell death gene)
- *IGFIIR* (regulates colonic cell growth)
- *MSH3* (DNA mismatch repair gene)
- *hG4-1*

Repeat sequence: A6
- *PTEN* (has two A6 sequences)

Repeat sequence: C9
- *SLC23A1*

Repeat sequence: C8
- *MSH6* (DNA mismatch repair gene)

Repeat sequence: G8
- *BAX* (programmed cell death gene)
- *IGF2R* (programmed cell death gene)

Repeat sequence: G7
- *AXIN-2* (also has C6 and A6x2)
- *CDX2* (transcription factor)

Repeat sequence: T10
- *OGT*

Repeat sequence: T9
- *KIAA0977*
- *NADH-UOB*

Repeat sequence: T7
- *FAS* (programmed cell death gene)

germ-line mutation. Before patients have blood drawn for genetic testing, they should be evaluated by a genetic counselor, to ensure that they understand all of the implications of testing before the blood is drawn. Genetic counseling can be performed successfully by interviewing several family members at once; however, it may be wise to counsel members individually, because, in some cases, patients may elect not to be tested. It is not universally guaranteed that patients carrying a germ-line mutation in one of these genes will not suffer discrimination when attempting to obtain health or life insurance.

Genetic testing for germ-line mutations in the DNA MMR genes is successful in only 50% of well-characterized families. The family member most likely on a clinical basis to be a true carrier of the mutation should be tested. A positive test in this person can be used to make or exclude the diagnosis in other members of that family. A negative test is useful only once the index case is characterized. Otherwise, negative genetic tests cannot be interpreted, because it is not possible to distinguish between a mutation not being present and one that is simply not detectable.

Patients with Lynch syndrome who develop invasive colon cancers should undergo surgical resection. If surgery is to be undertaken, a subtotal colectomy is preferred over more localized surgeries to remove as much at-risk colon as possible. It is not necessary to remove the rectum in such patients, which makes the treatment more acceptable to most individuals. The rectal segment must be examined on a regular basis thereafter. Some patients may inquire about a prophylactic colectomy should they be a gene carrier. This is an acceptable approach; however, careful intensive surveillance colonoscopy may obviate the necessity for colonic resection.

Patients with advanced (stage 3 or 4) colorectal cancer should be offered chemotherapy. Interestingly, loss of the DNA MMR activity *in vitro* confers resistance to many chemotherapeutic agents. However, empirically, patients with Lynch syndrome have a better survival rate than do patients with sporadic disease, and those given adjuvant chemotherapy have a better response than is seen in the sporadic situation. Moreover, because patients are typically younger when they present with the disease, they characteristically tolerate the treatment well.

When a germ-line mutation is found in an individual who has not yet developed a colorectal cancer, colonoscopy should be performed every 1−2 years. One clinical study indicated that performance of a colonoscopy every 3 years significantly reduced the mortality due to colorectal cancer. Performing colonoscopy with careful inspection for tiny adenomas

and removing all such lesions may prevent the development of cancer, and surveillance is also likely to detect earlier stage cancers. It is not known whether the benefits of annual colonoscopy outweigh the potential hazards of such repetitive surveillance, so the optimal surveillance interval is still debatable.

Patients with Lynch syndrome are at increased risk for cancer of the endometrium and ovaries. Surveillance of these organs is not as effective in preventing invasive cancer compared to colonoscopic surveillance in reducing colorectal cancer. Thus, women with Lynch syndrome should be encouraged to have their uterus and ovaries removed once they have completed their childbearing.

Lynch syndrome patients have fourfold increase in gastric cancer, and the median onset of disease is approximately 56 years. It is not known whether surveillance endoscopic examinations will be of benefit in preventing death due to gastric cancer. This disease does not develop through an adenoma–carcinoma sequence, so there is no apparent benefit by periodic examination to remove premalignant lesions. Likewise, no preventive strategies have been reported to be useful in preventing death due to cancers of the small intestine, urinary tract, pancreas, or brain. Periodic history and physical examination and prompt attention to new symptoms are all that can be offered. Finally, it is not known whether aspirin or other nonsteroidal antiinflammatory drugs will prevent this disease or alter the natural history of patients with Lynch syndrome.

See Also the Following Articles

Cancer, Overview • Colorectal Adenocarcinoma • Colorectal Adenomas • Colorectal Cancer Screening • Familial Risk of Gastrointestinal Cancers • Genetic Counseling and Testing

Further Reading

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., *et al.* (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816.

Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Bigas-Rodriguez, M., Fodde, R., Ranzani, G. N., and Srivastava, S. (1998). A National Cancer Institute Workshop on microsatellite instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257.

Duval, A., and Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: Toward a new concept of target genes for instability. *Cancer Res.* **62**, 2447–2454.

Jarvinen, H. J., Aarnio, M., Mustonen, H., Aktan-Collan, K., Aaltonen, L. A., Peltomaki, P., De La Chapelle, A., and Mecklin, J. P. (2000). Controlled 15-year trial on screening for colorectal cancer in families with hereditary non-polyposis colorectal cancer. *Gastroenterology* **118**, 829–834.

Magnetic Resonance Imaging (MRI)

JEFFREY C. WEINREB*, DANIEL KIM†, AND JONATHAN BARKER†
*Yale University School of Medicine and †New York University School of Medicine

gadolinium chelate Magnetic resonance imaging intravenous contrast agent consisting of a complex of an organic ligand and a lanthanide metal.
in-phase gradient echo Magnetic resonance pulse sequence in which signals from lipid and water are additive.
opposed-phase gradient echo Magnetic resonance pulse sequence in which signals from lipid and water are subtractive. Used to demonstrate the presence of lipid.
pulse sequence Technical parameters consisting of a set of radiofrequency pulses, magnetic gradient pulses, and the time intervals between pulses used to create the image and image contrast on magnetic resonance images.
short tau inversion recovery (STIR) Magnetic resonance pulse sequence in which T1- and T2-dependent contrasts are additive and signal from fat is suppressed.
susceptibility Ratio of the intensity of magnetization induced in a substance to the intensity of the magnetic field to which the substance is exposed.
T1 weighted Magnetic resonance pulse sequence designed to distinguish tissues with differing T1 relaxation times.
T2 weighted Magnetic resonance pulse sequence designed to distinguish tissues with differing T2 relaxation times.

Objectives of liver, biliary, and pancreatic imaging with magnetic resonance include screening for the presence of disease, characterization of lesions, anatomic localization, evaluation of interval change during the course of treatment, and assessment of vascular anatomy and patency. Magnetic resonance imaging offers excellent soft tissue contrast, versatility, multiplanar capability, and the possibility of a comprehensive exam that can obviate the need for other studies. In clinical practice, magnetic resonance imaging may be utilized as a primary diagnostic imaging exam or as a problem solver when other imaging studies are inconclusive.

INTRODUCTION

Because of the versatility and complexity of magnetic resonance imaging (MRI), exams must be tailored for specific clinical indications. Routine MRI examinations are designed to utilize different techniques to maximize tissue contrast and characterize tissue composition. The basic pulse sequences include a T1-weighted sequence, a T2-weighted sequence, and dynamic contrast-enhanced T1-weighted sequences. Tissues are described according to their signal intensity on the different sequences. For instance, fluid will demonstrate hyperintensity on T2-weighted sequences and hypointensity on T1-weighted sequences. Adipose tissue will demonstrate hyperintensity on T1-weighted sequences. A signal from lipid may be eliminated with a variety of fat-suppression techniques.

LIVER

Technique

In general, a hepatic MRI exam consists of the following pulse sequences:

1. T2 weighted or short tau inversion recovery (STIR).
2. In-phase and opposed-phase breath-hold T1-weighted gradient echo (GRE).
3. Dynamic gadolinium-chelate-enhanced breath-hold fat-suppressed T1-weighted gradient echo obtained during the arterial dominant, portal venous, and equilibrium phases. Three-dimensional volume acquisitions permit multiplanar reconstruction, volumetric renderings, and MR angiograms (MRAs), which may be useful for diagnosis and surgical planning.
4. Additional pulse sequences may used for specific indications.

Specialized "liver specific" intravenous contrast agents, such as various forms of superparamagnetic iron oxide (SPIO) particles and mangafodipir trisodium, may improve the detection and characterization of some liver tumors.

Diffuse Liver Disease

Fatty Infiltration

Diffuse fatty infiltration, or hepatic steatosis, is a common condition characterized by accumulation of lipid within hepatocytes. The most common causes include alcohol, obesity, diabetes mellitus, and malnutrition. Although fatty infiltration is a benign condition, a subset of patients may a progress to a hepatitis-like

 Copyright 2004, Elsevier (USA). All rights reserved.

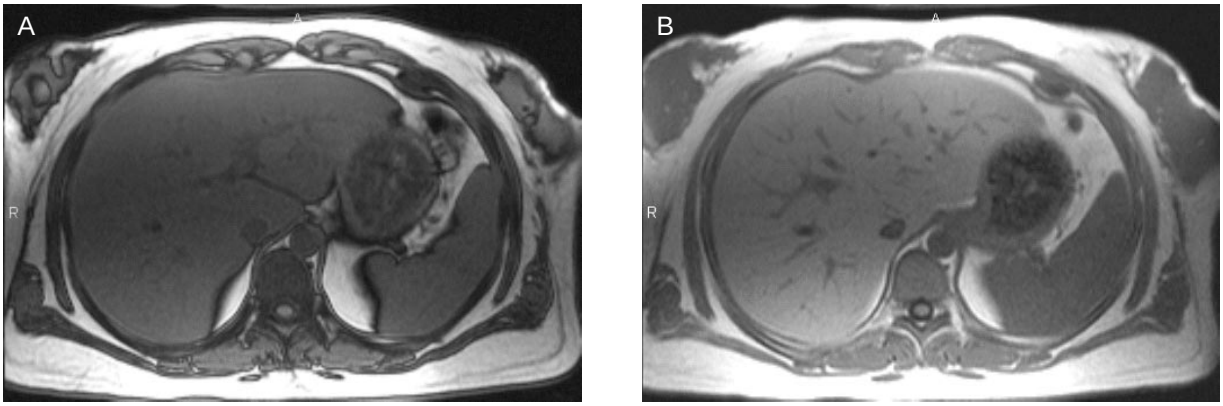


FIGURE 1 Opposed-phase (A) gradient echo image, showing a decrease in liver signal compared with in-phase image (B), which is diagnostic of diffuse fatty infiltration.

picture and ultimately cirrhosis. Consequently, MRI may play a role in establishing the diagnosis as well as in monitoring patients for complications.

Diffuse fatty infiltration is diagnosed by the loss of liver signal on opposed-phase images, compared to that on in-phase ones (Fig. 1). Focal forms of fatty infiltration, which can simulate a tumor on computed tomography (CT) or ultrasound, can be readily distinguished from a mass lesion on MRI by the absence of associated mass effect and the demonstration of lipid using chemical shift techniques.

Iron Deposition

MRI characteristics of iron deposition rely on the magnetic susceptibility effects of iron, resulting in loss of signal intensity on MR images with certain pulse sequences (Fig. 2). Primary or hereditary hemochromatosis is an autosomal recessive disorder exhibiting incomplete penetrance. Commonly involved organs include the liver, pancreas, myocardium, pituitary gland, and synovium. In the liver, the iron is deposited in hepatocytes. Untreated, patients may develop cirrhosis and liver failure. MRI clarifies the distribution of iron deposition within the body and monitors for the progression of complications. MRI can demonstrate the findings of cirrhosis and development of hepatocellular carcinoma. It is not uncommon for hemochromatosis to be diagnosed initially by MRI performed for unrelated reasons.

Hemosiderosis occurs secondary to dyserythropoiesis or transfusional iron overload states. Excess iron accumulates in the reticuloendothelial system, including the spleen, bone marrow, and Kupffer cells in the liver.

Cirrhosis

Cirrhosis is a chronic diffuse liver disease characterized by hepatic regeneration and fibrosis. The most common causes include alcohol abuse, chronic viral hepatitis, hemochromatosis, autoimmune chronic hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, and drug toxicity. MRI can be used to establish the diagnosis and monitor for the development of complications, including ascites, varices, portal vein thrombosis, and hepatocellular carcinoma.

Although the morphologic changes in cirrhosis are quite variable, depending on the etiology, early cirrhosis may include "central atrophy" characterized by enlargement of the hilar periportal space and expansion of the major interlobar fissure and gallbladder fossa. As cirrhosis progresses, morphologic changes include nodularity of the surface contour and parenchyma secondary to regenerative nodules and proliferating hepatocytes surrounded by fibrous septa (Fig. 3).

Regenerative nodules have characteristic features on MRI that usually allow confident distinction from hepatocellular carcinoma. Most commonly, they are isointense with other background nodules on both T1-weighted and T2-weighted images. Occasionally, they may be hyperintense on T1-weighted images and

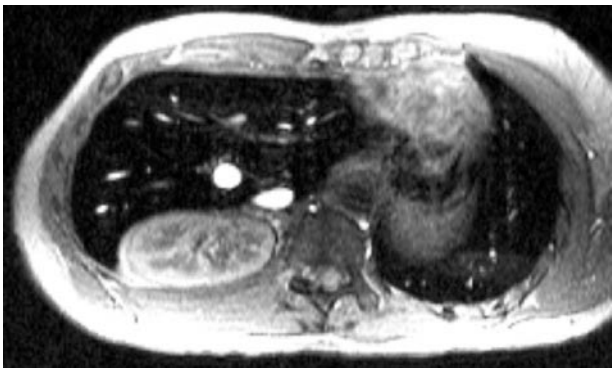


FIGURE 2 T2-weighted image, showing extremely dark liver due to iron overload.

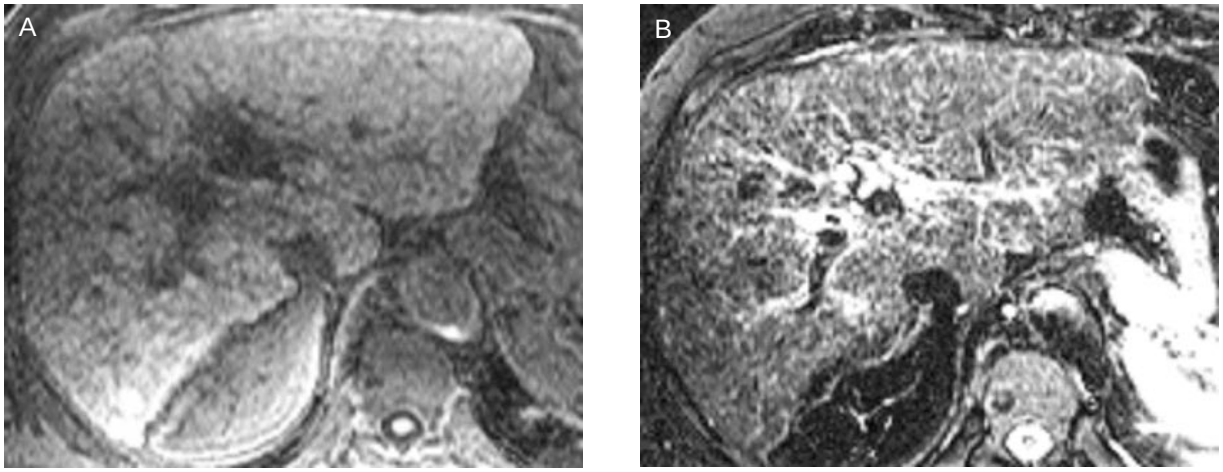


FIGURE 3 T1-weighted (A) and T2-weighted (B) images, showing nodular contour of a cirrhotic liver containing innumerable regenerative nodules.

hypointense on T2-weighted images. However, unlike hepatocellular carcinoma, they are rarely hyperintense on T2-weighted images with the exception of those that occur in the setting of chronic Budd–Chiari syndrome. Some regenerative nodules contain iron and demonstrate low signal intensity on all nonenhanced sequences. However, these benign siderotic regenerative nodules cannot be reliably differentiated from iron-containing dysplastic nodules, which may be premalignant lesions. Because regenerative nodules invariably have a portal venous blood supply with minimal contribution from the hepatic artery, they rarely enhance more than background liver during the hepatic arterial phase of gadolinium-enhanced study.

Another manifestation of cirrhosis, which may mimic a mass lesion, is confluent hepatic fibrosis. On MRI, confluent hepatic fibrosis may be isointense or hypointense on T1-weighted images and hyperintense on T2-weighted images. The contrast enhancement pattern is variable. The presence of capsular retraction from volume loss and scarring helps differentiate this process from a hepatoma (Fig. 4).

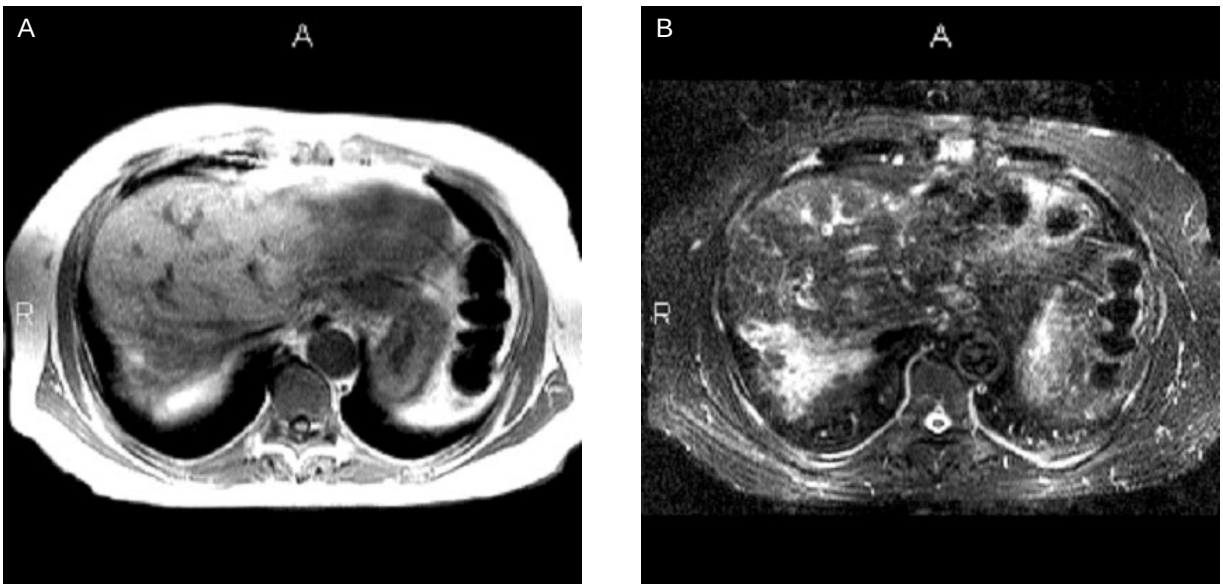


FIGURE 4 T1-weighted (A) and fat-suppressed T2-weighted (B) images of a liver with cirrhosis secondary to sclerosing cholangitis, showing liver capsule retraction adjacent to an area of confluent hepatic fibrosis in the posterior segment of the right lobe.

Focal Liver Disease

Cavernous Hemangioma

Cavernous hemangiomas are the most common benign hepatic neoplasm. They are often discovered incidentally with no clinical consequences. On MRI, cavernous hemangiomas are well-defined round or lobulated lesions. They are usually hypointense on T1-weighted sequences and very hyperintense on T2-weighted sequences, approaching the signal characteristics of fluid. The contrast-enhanced features of cavernous hemangiomas differentiate them from hypervascular or cystic malignant neoplasms. Most commonly, cavernous hemangiomas demonstrate peripheral nodular enhancement with progressive centripetal filling in. In some cavernous hemangiomas smaller than 1.5 cm, immediate homogeneous enhancement may be observed. In larger cavernous hemangiomas, the central portion may not enhance owing to fibrosis, thrombosis, or calcification.

Hepatic Cyst

Hepatic cysts are common benign lesions that occur in 2–10% of the population. They may be solitary or multiple and range in size from a few millimeters to several centimeters. Hepatic cysts are usually discovered incidentally and have no malignant potential. On MRI, they are well-defined round or oval lesions with thin, almost imperceptible walls. Simple cysts are very hypointense on T1-weighted sequences and very hyperintense on T2-weighted sequences, corresponding to the signal characteristics of fluid. Additionally, simple cysts demonstrate no internal architecture and no gadolinium contrast enhancement. Biliary hamartoma may be indistinguishable from simple hepatic cysts but are also benign entities.

Hepatic Abscess

Hepatic abscesses may result from bacterial, fungal, or parasitic infection. In the United States, bacterial abscesses are most common. Worldwide, however, parasitic abscesses, in particular, are most common. On MRI, hepatic abscesses have nonspecific imaging characteristics. In the appropriate clinical setting, multiple lesions demonstrating hyperintensity on T2-weighted images with perilesional edema and rim enhancement are suggestive of abscesses.

Focal Nodular Hyperplasia

Focal nodular hyperplasia (FNH) is the second most common benign hepatic neoplasm. It is believed to result from a hyperplastic response to a preexisting vascular malformation. FNHs contain disorganized nodules of hepatocytes and Kupffer cells that are separated by small bile ducts lacking communication with the normal biliary tree. FNHs are usually solitary and less than 5 cm in diameter. They are most often found in women.

MRI characteristics of FNH include isointensity or slight hypointensity compared with normal hepatic parenchyma on the T1-weighted sequences and slight hyperintensity or isointensity on the T2-weighted sequences (Fig. 5). FNH is usually homogeneous in signal intensity with the exception of the central scar. The central scar is hypointense to the lesion and surrounding liver on the T1-weighted sequences (T1-Ws) and hyperintense on the T2-weighted sequences.

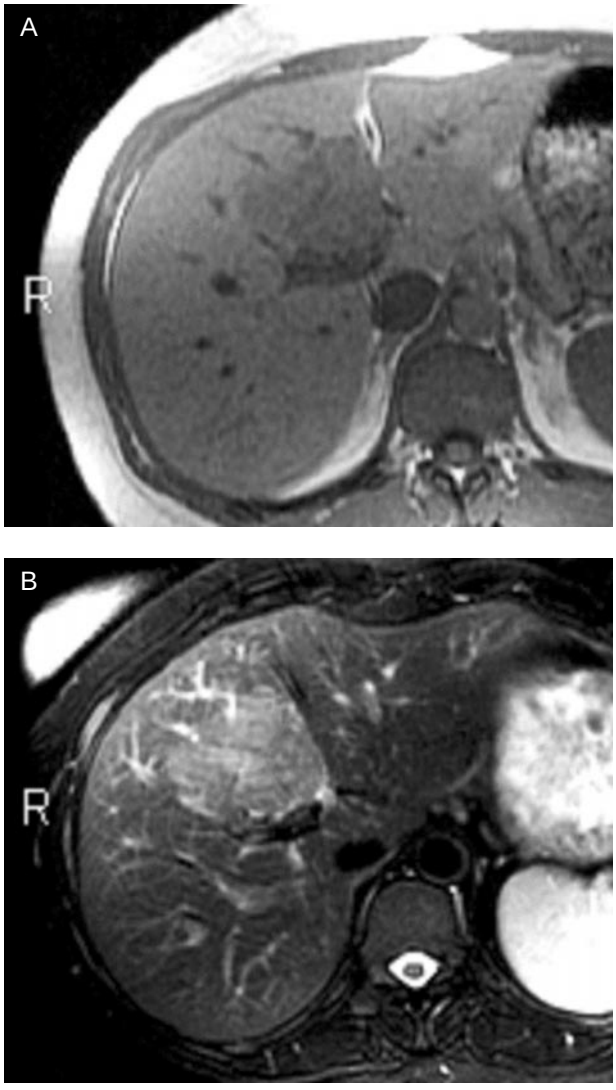


FIGURE 5 Characteristic MRI appearance of focal nodular hyperplasia that is slightly hypointense on the T1-weighted image (A) and slightly hyperintense on the T2-weighted image (B).

Following gadolinium chelate administration, the lesion demonstrates marked homogeneous intense enhancement in the arterial phase, homogeneous washout in the portal venous phase, and isointensity or mild hyperintensity in the portal venous phase. The central scar demonstrates delayed enhancement. Only approximately 40% of FNHs exhibit all of these MRI characteristics and can be definitively characterized with MRI.

Hepatocellular Adenoma

Hepatocellular adenoma is a benign hepatic neoplasm composed of hepatocytes with an absence of normal portal structures. This rare neoplasm is seen in association with oral contraceptive use, anabolic–androgenic steroid use, and glycogen storage disease types I and III. Complications include acute hemorrhage and rare malignant transformation.

On MRI, hepatic adenomas demonstrate heterogeneous signal intensity on T1-weighted and T2-weighted sequences (Fig. 6). On T1-weighted images, foci of hyperintensity may be present from prior hemorrhage. Foci of hypointensity may represent necrosis or calcification. Adenomas commonly contain lipid. With dynamic gadolinium-enhanced imaging, intense heterogeneous arterial phase enhancement is often seen.

Dysplastic Nodules

Dysplastic nodules (DNs) are found in 14–38% of cirrhotic livers and are considered by many to be premalignant lesions. On gross pathological examination, DNs are distinguished from regenerative nodules (RNs) on the basis of size (usually >8 mm), color, texture, or the degree to which they bulge from the cut surface of the liver.

DNs are categorized as low- and high-grade lesions. Histologically, low-grade DNs contain hepatocytes that are minimally abnormal without architectural or cytological atypia. They usually contain portal tracts and hepatic venules and may contain iron or copper. High-grade DNs have cellular and architectural atypia. Some high-grade dysplastic nodules may contain subnodules of overt hepatocellular carcinomas (HCCs), the majority of which are well differentiated.

There is significant overlap in the MRI appearance of DNs and small hepatocellular carcinomas. Most commonly, DNs are hyperintense on T1-Ws and hypointense on T2-Ws (Fig. 7), a pattern that is also present with some small HCCs. Close MRI monitoring of cirrhotic patients who have known DNs is recommended in order to detect changes that may herald transformation into hepatocellular carcinoma.

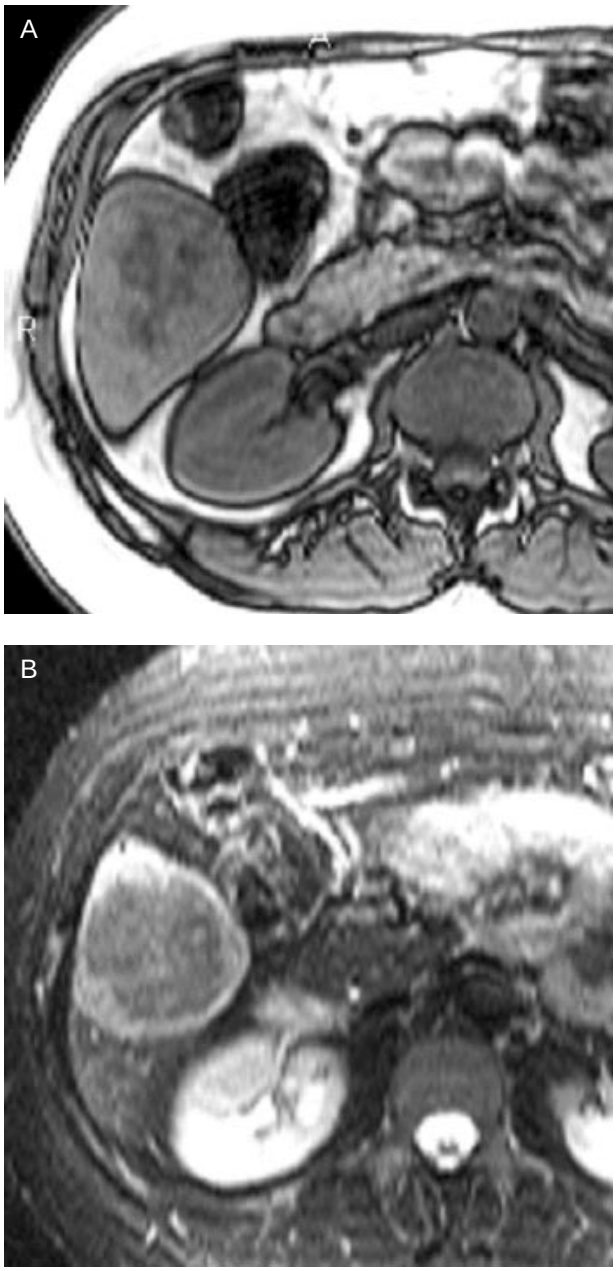


FIGURE 6 Characteristic MRI appearance of hepatocellular adenoma, which demonstrates heterogeneous signal intensity on T1-weighted (A) and T2-weighted sequences (B).

Hepatocellular Carcinoma

Hepatocellular carcinoma is the most common primary hepatic malignancy. It usually occurs in the setting of chronic liver disease and cirrhosis secondary to hepatitis B and C, alcohol abuse, hemochromatosis, and aflatoxin exposure. On MRI, hepatocellular carcinoma may present as a solitary mass, multifocal masses, or a diffusely infiltrating lesion. Vascular invasion is common.

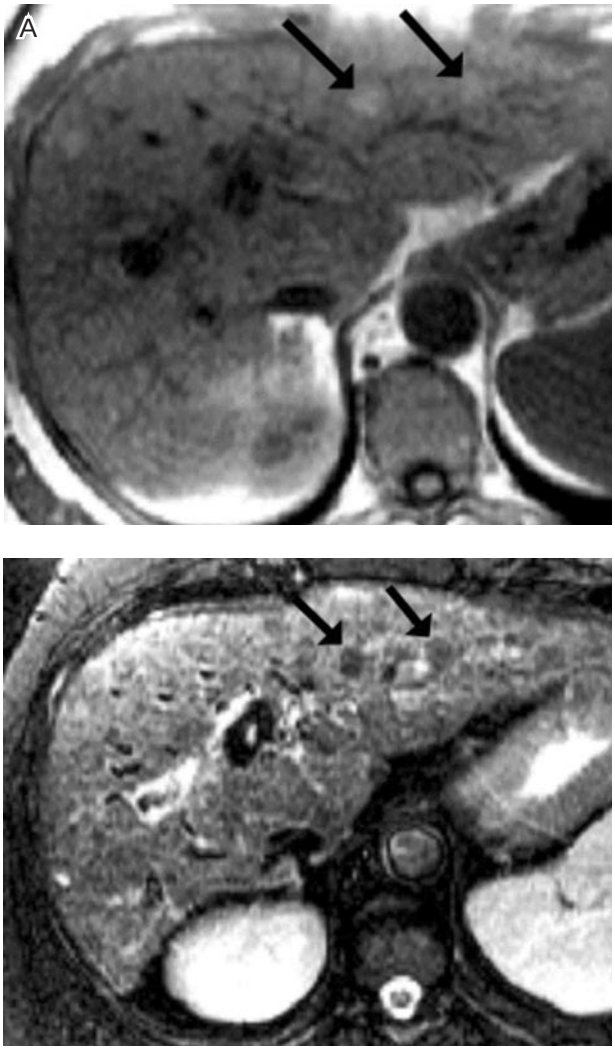


FIGURE 7 Characteristic MRI appearance in a cirrhotic liver of dysplastic nodules (arrows) that are hyperintense on T1-weighted images (A) and hypointense on T2-weighted images (B).

The signal intensity characteristics on T1-weighted and T2-weighted images are highly variable, but they are most commonly moderately hyperintense on T2-weighted images (Fig. 8).

The contrast enhancement pattern is dependent on the degree of tumor differentiation and vascularity. The most common enhancement pattern for small HCCs (<2.0 cm) is diffuse, homogeneous enhancement during the arterial phase with rapid washout during the portal venous phase (Fig. 8). Enhancement of a nodule or a focus within a nodule during the arterial phase favors a diagnosis of hepatocellular carcinoma, because DNs usually have a predominantly portal venous blood supply. Well-differentiated hepatomas demonstrate minimal arterial enhancement that washes out on the portal venous phase. Moderate and poorly differentiated tumors demonstrate marked arterial enhancement. In a cirrhotic liver, any focus of abnormal arterial phase enhancement should be considered a hepatocellular carcinoma until proved otherwise. However, as many as 4% of low-grade DNs and 32% of high-grade DNs may have an arterial blood supply. Thus, enhancement of a hepatic nodule during the arterial phase, once thought to be diagnostic of malignancy, can also occur with DNs.

In one study, triphasic helical CT detected 17% of HCCs 2–10 mm in size, 29% 11–20 mm in size, and 63% that were >20 mm; 8% had a false positive CT. In another study, dynamic contrast-enhanced MRI had an overall sensitivity of 37% for HCCs due to the inability to detect the large number of tumors less than 10 mm in size. However, the sensitivity for detecting HCCs >10 mm was 89%.

Metastasis

On MRI, hepatic metastases often demonstrate hypointensity on T1-weighted images and hyperintensity on T2-weighted images (Fig. 9). Occasionally they are difficult to detect on non-contrast-enhanced images. Hypervascular metastases may be identified only on arterial phase images. On the portal venous phase, metastases may demonstrate inhomogeneous or ring enhancement with early peripheral washout (Fig. 9).

BILIARY SYSTEM

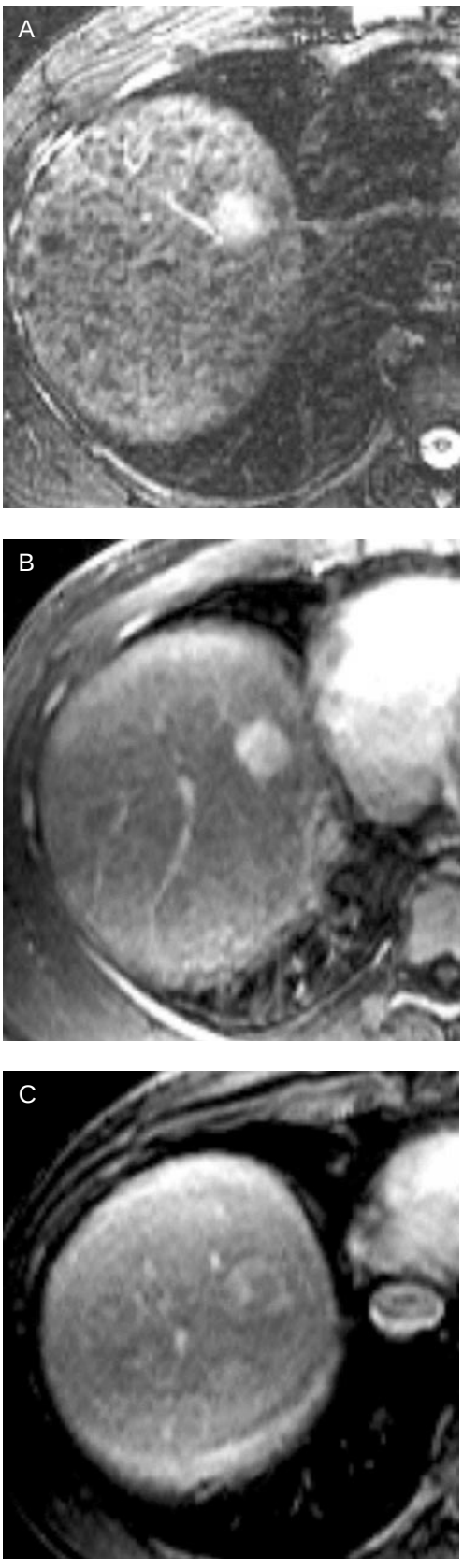
Technique

With MRI, the biliary system is usually evaluated with MR cholangiopancreatography (MRCP). The goal with MRCP is to produce a "hydrogram" in which the entire biliary system is bright due to the presence of bile fluid (Fig. 10). There are numerous techniques for producing T2-weighted images in which fluids are bright. The trade-offs between spatial resolution, image acquisition time, and image quality will determine the preferred technique on a particular scanner.

Cholelithiasis and Choledocholithiasis

Cholelithiasis and choledocholithiasis are common conditions. Although many patients remain asymptomatic, others may develop cholecystitis, pancreatitis, or biliary tract obstruction. Additionally, patients with biliary tract calculi may have a higher risk of developing subsequent cholangiocarcinoma.

MRCP demonstrates the biliary and pancreatic ductal anatomy by using T2-weighted sequences that depict fluid-containing structures as hyperintense.



Consequently, calculi will be seen as intraluminal filling defects. Other causes of filling defects that must be distinguished from calculi include pneumobilia, blood clots, and tumor. Most patients with suspected biliary calculi are adequately evaluated with ultrasound and nuclear medicine studies. However, ultrasound fails to detect many common bile ducts stones in nondilated ducts. MRCP can depict the presence or absence of biliary tract obstruction in 91–100% of cases. For choledocholithiasis, the reported sensitivity is >90% and specificity >94%. MRCP also provides a road map of the biliary tree and a more comprehensive assessment of the extrabiliary structures.

Primary Sclerosing Cholangitis

Primary sclerosing cholangitis (PSC) is a disease of unknown etiology characterized by inflammation, fibrosis, and destruction of the intrahepatic and extrahepatic biliary tree. Approximately 70% of patients with PSC have inflammatory bowel disease, most commonly ulcerative colitis.

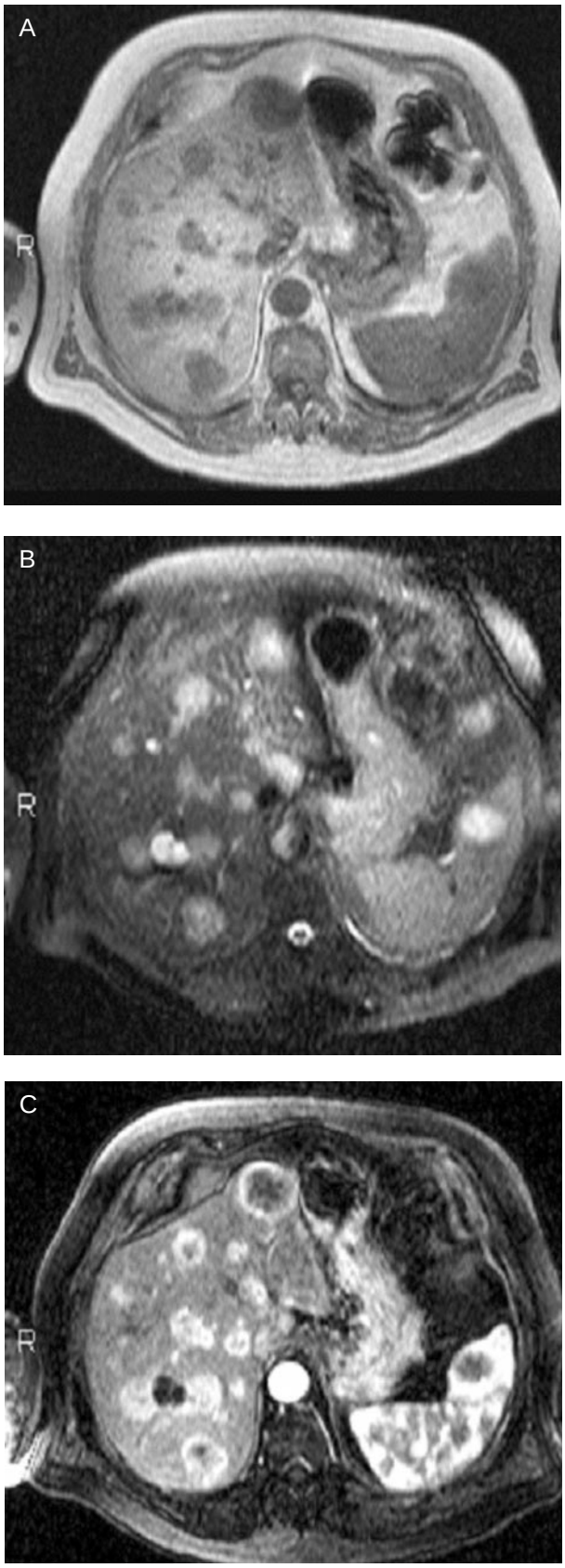
MRCP characteristics of PSC include beading characterized by focal dilatation of the intrahepatic and extrahepatic bile ducts alternating with segmental stenoses (Fig. 11). As the fibrosis progresses, the peripheral bile ducts become obliterated and the biliary tree assumes a characteristic pruned appearance. Contrast-enhanced MRI may demonstrate thickening and enhancement of the bile ducts. The hepatic parenchyma may demonstrate concomitant peripheral patchy areas of atrophy with abnormal signal intensity and enhancement.

As an adjunct imaging modality to endoscopic retrograde cholangiopancreatography (ERCP), MRCP offers several potential advantages. It is noninvasive, can depict the entire biliary tree (even in the presence of strictures and obstruction), and may detect early cholangiocarcinoma.

Primary Biliary Cirrhosis

Primary biliary cirrhosis (PBC) is a chronic autoimmune cholecystic liver disease characterized by the destruction of small intrahepatic bile ducts. It affects predominantly middle-aged women. The diagnosis of PBC is based on a combination of findings, including cholestatic liver enzymes, a positive antimitochondrial

FIGURE 8 MRI appearance of a hepatoma. The short tau inversion recovery image (A) shows a hyperintense mass in the hepatic dome. On the arterial-phase contrast-enhanced image (B), the mass enhances briskly. On the equilibrium-phase image (C), the mass is isointense with the liver, but there is an enhanced capsule.



antibody (AMA) assay, and characteristic liver biopsy findings.

In the early stages, the liver appears normal on MRI. The hallmark of the disease is lymphadenopathy in unusual locations and out of proportion to the degree of liver disease. The presence of the "periportal halo" sign is considered specific for the disease (Fig. 12). This appears as low-signal rounded lesions centered on a portal venous branch, 5–10 mm in size, involving all segments of liver. It corresponds pathologically to the inflammatory process involving portal triads with development of dense infiltrate and mild fibrosis.

Cholangiocarcinoma

Most cholangiocarcinomas are ductal adenocarcinomas that arise from the intra- and extrahepatic bile duct epithelium. Many predisposing conditions have been described, most of which involve preexisting biliary tract disease. Cholangiocarcinomas exhibit several different growth patterns, including polypoid, infiltrative, and exophytic. There is almost always some degree of biliary dilatation, which may be localized in the case of focal intrahepatic disease or generalized if the common hepatic or common bile ducts are involved. The tumors may be hypointense on T1-weighted images and mildly hyperintense on T2-weighted ones, but they are often difficult to identify on non-contrast-enhanced images. Dynamic gadolinium-enhanced imaging may show progressive enhancement of larger tumors. This enhancement may appear and be most conspicuous only on scans delayed for 10–15 minutes.

PANCREAS

Technique

MRI assessment of the pancreas usually includes an evaluation of the pancreatic parenchyma as well as the ductal system using an MRI protocol similar to that for MRI of the liver and the biliary system. Fat-suppressed T1-weighted images followed by dynamic contrast-enhanced fat-suppressed images with thin slices (on the order of 2–3 mm) are particularly useful for diagnosis of pancreatic carcinoma.

Normal Pancreas

Normal pancreatic tissue demonstrates uniform high signal on fat-suppressed T1-weighted images, possibly due to the presence of aqueous protein within the

FIGURE 9 MRI appearance of lung metastases that are hypointense on T1-weighted images (A), hyperintense on T2-weighted images (B), and show ring enhancement (C).

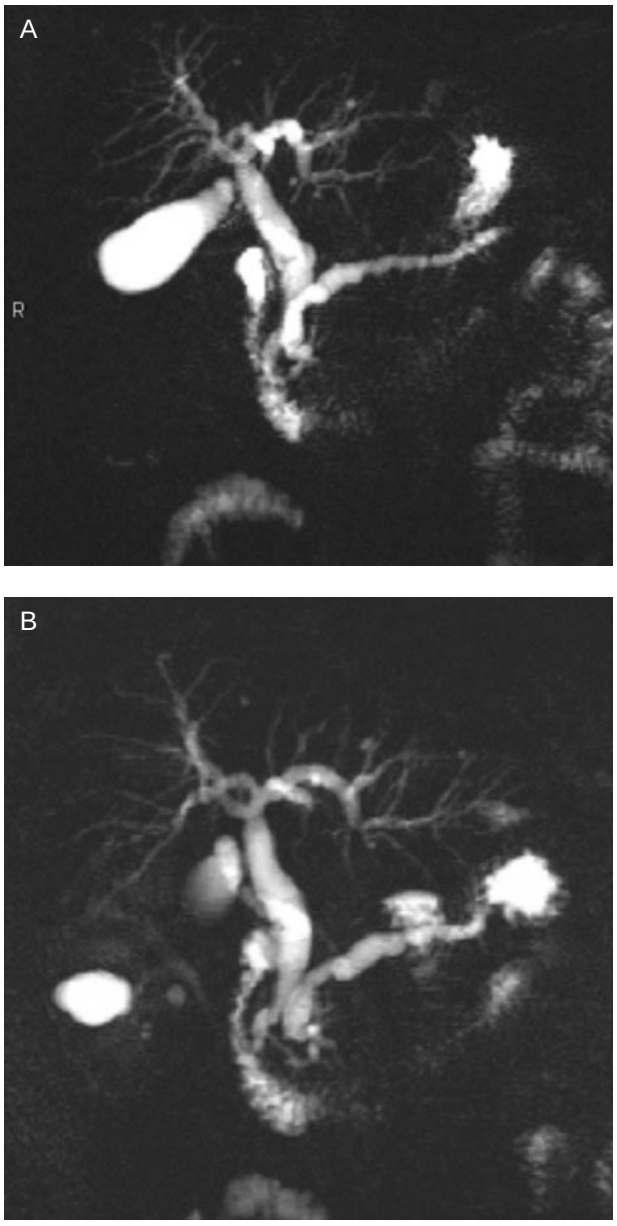


FIGURE 10 MR cholangiopancreatogram in two views (A and B) shows a dilated common bile duct and pancreatic duct secondary to metastatic ovarian carcinoma in the pancreatic head.

acini. Immediately after the administration of gadolinium chelate, there is a uniform parenchymal blush.

Pancreas Divisum

Pancreas divisum is the most common anatomic variant of the pancreas. It results from failure of fusion of the dorsal and ventral pancreatic ducts. The major regions of the pancreas, including the superior–anterior parts of the head, the body, and the tail, are drained by the long dorsal pancreatic duct through the accessory papilla. The inferior–posterior parts of the head and uncinate process are drained by the short ventral pancreatic duct that joins the common bile duct in the ampulla. Although it is usually an incidental finding, pancreas divisum is thought to be a causative factor in some cases of recurrent pancreatitis as a result of a functional obstruction of the minor papilla. Pancreas divisum may be missed with ERCP because the standard procedure opacifies only the ventral duct after cannulation.

On MRCP, pancreas divisum can be diagnosed when a dominant dorsal duct is visualized from body

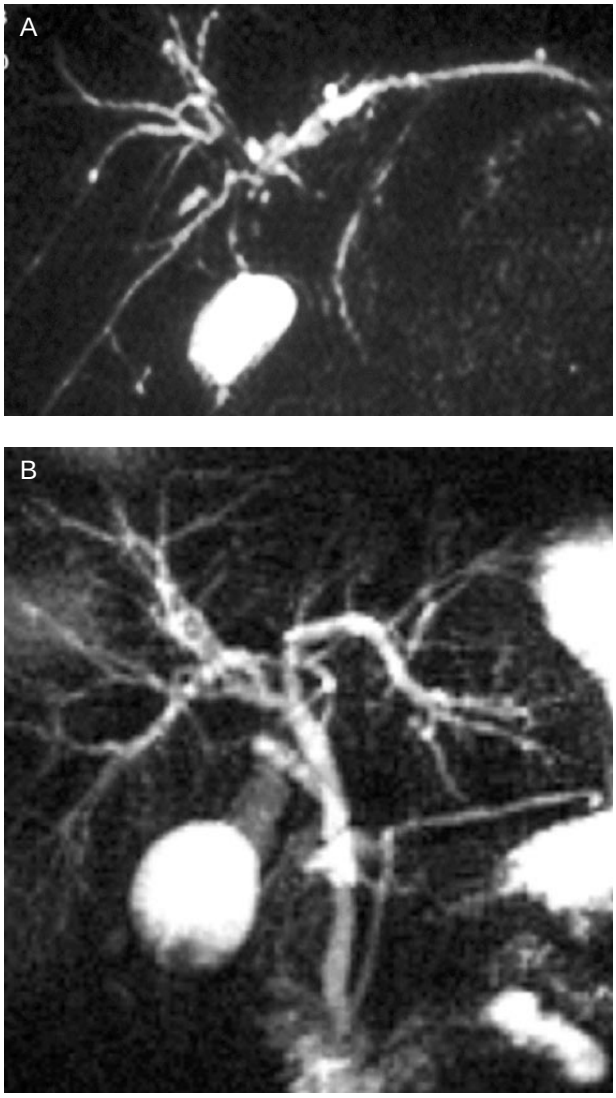


FIGURE 11 MR cholangiopancreatogram in two views (A and B) shows the beaded appearance of intrahepatic bile ducts secondary to alternating dilatation and strictures in primary sclerosing cholangitis.

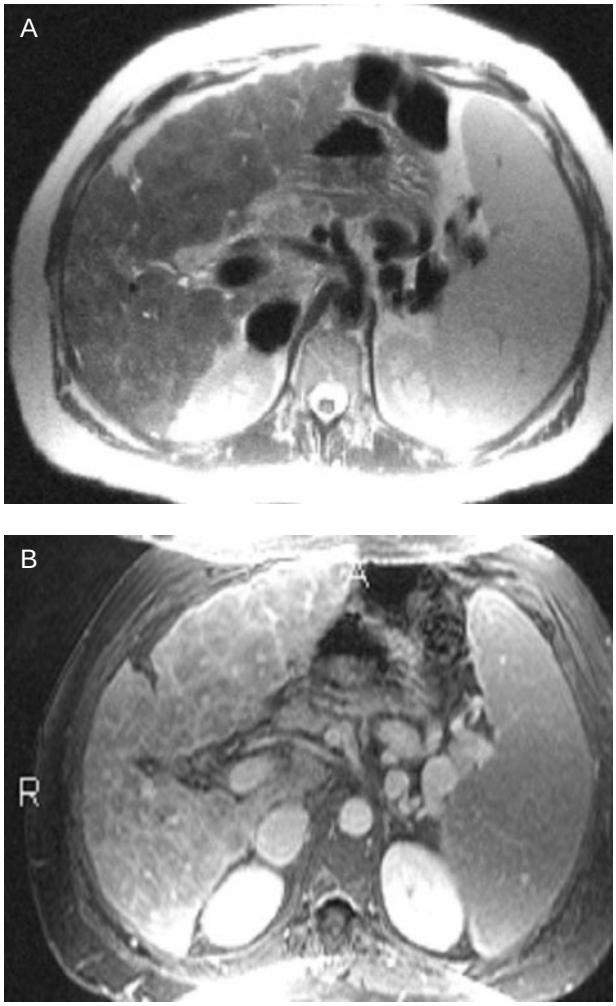


FIGURE 12 T2-weighted (A) and postcontrast fat-suppressed T1-weighted (B) images show a nodular contour and periportal halo sign (low-signal rounded lesions centered on a portal venous branch, 5–10 mm in size, involving all segments of liver) secondary to primary biliary cirrhosis.

to head crossing anterior to the common bile duct (CBD) and draining superiorly and separately from the CBD. Although it is helpful to identify the ventral duct draining separately into the major papilla with the CBD, this finding is not required for a diagnosis.

Pancreatitis

Pancreatitis is an inflammatory disease of the pancreas with acute and chronic clinical presentations. The most common etiologies include alcoholism and cholelithiasis. When diagnostic imaging studies are required, CT scanning remains the study of choice. In those patients unable to receive iodinated contrast and in selected cases in which a biliary etiology is suspected, MRI provides a valuable alternative imaging modality.

In patients with pancreatitis, the pancreas may have a diffuse low signal and decreased contrast enhancement on T1-weighted images. As with CT, contrast-enhanced scans can be used to assess necrosis. A focal form of pancreatitis may mimic a mass. Pancreatic duct dilatation, pseudocysts, and peripancreatic fluid are readily identified on T2-weighted images, and inflammatory pseudocysts are the most common small pancreatic cystic lesions.

Carcinoma

Pancreatic adenocarcinoma is the most common malignant neoplasm of the pancreas. It is also the most common cause of malignant biliary obstruction. Patients typically present with advanced local disease or metastases at the time of diagnosis. Imaging studies assist in making the diagnosis and determining which patients are candidates for surgical intervention.

On MRI, pancreatic adenocarcinomas are typically hypointense compared to normal pancreatic parenchyma on fat-suppressed T1-weighted images, and this may be more apparent immediately after contrast injection (Fig. 13).

The role of imaging is to determine the presence of disease and of local invasion and metastases. The major potential benefits of MRI compared to CT are that it is easier to detect very small non-contour-deforming tumors with MRI, easier to determine if a contour abnormality is due to a tumor or an anatomic variant, and it is easier to detect and characterize concomitant liver lesions. MRI is also useful in patients who have diminished renal function or iodine contrast allergy. CT and MRI are considered equivalent for the assessment of vascular invasion and adenopathy, and both have limitations in the detection of peritoneal and liver metastases. In most patients, there is no demonstrable advantage of MRI compared to high-quality CT for the detection or staging of pancreatic neoplasms.

Pancreatic Cystic Tumors

Pancreatic cystic tumors are classified into benign microcystic adenomas and mucinous cystic neoplasms that may be potentially malignant or malignant. Microcystic adenomas (formerly called serous cystadenoma) are well-defined, lobulated cystic neoplasms containing numerous small cysts, often having a honeycombed appearance. Each cyst measures less than 2 cm in diameter. The presence of a central stellate scar is a characteristic feature. On dynamic contrast-enhanced sequences, this scar demonstrates delayed contrast

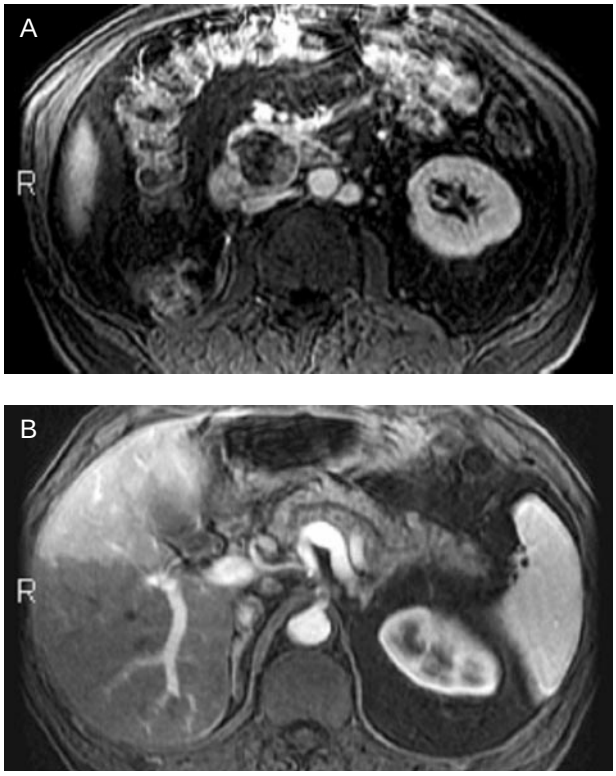


FIGURE 13 Fat-suppressed arterial-phase contrast-enhanced images, showing a pancreatic head carcinoma (A) with proximal pancreatic duct dilatation (B).

enhancement. Central calcifications may be seen on CT but are usually not appreciated on MRI.

Mucinous cystic neoplasms (formerly called mucinous cystadenoma and cystadenocarcinoma) are thick-walled, unilocular or multilocular neoplasms containing a small number of cysts measuring greater than 2 cm in diameter. Features of malignant transformation include the presence of solid nodular enhancing components within the cysts as well as adjacent vascular encasement or occlusion. They usually occur in the pancreatic tail, most commonly in middle aged women. In many cases, it is impossible to reliably differentiate between microcystic and mucinous cystic neoplasms on the basis of imaging features alone.

Intraductal Papillary Mucinous Neoplasms

Intraductal papillary mucinous neoplasms (IPMNs) have been designated in the literature under various names, including mucinous duct ectasia, mucin-producing tumor of the pancreas, intraductal mucin-hypersecreting neoplasm of the pancreas, mucus-hypersecreting tumor of the pancreas, mucus cell hypertrophy, nonpapillary hyperplasia, and papillary hyperplasia. An IPMN is characterized by epithelial proliferation in the main pancreatic duct or its major tributaries or both, with focal or diffuse mucinous transformation of the ductal epithelium. The changes can range from cellular atypia without malignancy to carcinoma-*in-situ*. When the mucin production is extensive, there is cystic dilatation of the duct, which can result in the appearance of a cystic mass on MRI. An IPMN is considered to be a premalignant lesion, and malignant transformation ranges from a dilated pancreatic duct without a stricture or mass to papillary projections into the duct lumen.

When symptoms are present, patients may present with recurrent attacks of pancreatitis, presumably secondary to intermittent obstruction of the pancreatic duct either by polypoid tumor or inspissated mucus. Currently, most IPMNs are found incidentally on abdominal MRI or CT exams. On endoscopic examination, a gaping ampulla extruding mucus is a pathognomonic finding. IPMNs differ from mucinous cystic neoplasms in that the latter are thought to result form the mucinous neoplastic transformation of the epithelium lining parenchymal ductules, whereas IPMNs arise in the main or intermediate-size ducts, most commonly in the head or uncinate. Mucinous cystic neoplasms rarely communicate with the main duct, and when they do it is in the form of a fistula.

ERCP establishes the diagnosis of IPMN by demonstrating a bulging papilla with mucinous extrusion and communication of the lesion with the pancreatic ductal system. Although ERCP remains the diagnostic imaging modality of choice, MRCP may yield sufficient diagnostic information in some patients unsuitable for ERCP or when technically difficult cases arise from the inability to opacify the pancreatic ductal system secondary to excessive mucin production or a patulous papilla.

On MRI, IPMNs demonstrate low signal intensity on T1-weighted images and high signal intensity on T2-weighted images secondary to mucinous contents (Fig. 14). Communication with the pancreatic ductal system is highly suggestive of the diagnosis.

Pancreatic Islet Cell Tumor

Pancreatic islet cell tumors develop from the amine precursor uptake and decarboxylation cell line and may be functioning or nonfunctioning. The most common tumors, in order of frequency, are insulinoma, gastrinoma, and nonfunctioning islet cell tumor.

On MRI, insulinomas are usually less than 2 cm in diameter. These tumors are commonly hypointense on fat-suppressed T1-weighted images and hyperintense on T2-weighted sequences. Because these tumors are often hypervascular, they may be best seen on the

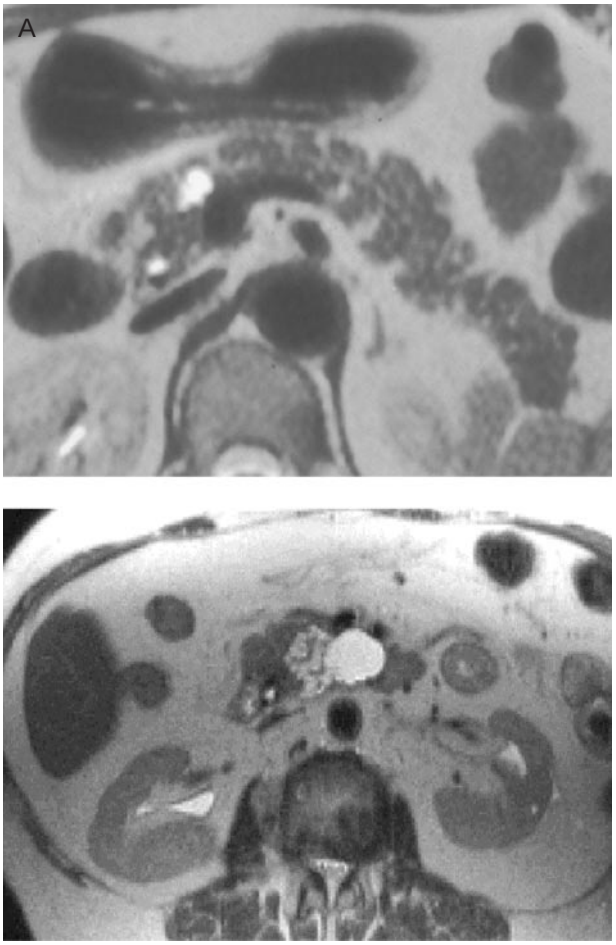


FIGURE 14 T2-weighted images, showing side branch intraductal papillary mucinous neoplasms with no main duct dilatation in two patients.

arterial phase of the dynamic contrast-enhanced sequences. Gastrinomas and nonfunctioning islet cell tumors share imaging features with insulinomas but typically present as larger tumors with a higher incidence of metastases.

Functional Studies

Occasionally, the pancreatic duct, particularly its tail and side branches, is difficult to visualize and characterize on MRCP. This can result in false positive examinations. Intravenous secretin can ameliorate this problem. In healthy subjects, secretin induces pancreatic fluid secretion and is followed by duodenal filling. The fluid secretion results in a transient increase in the main pancreatic duct caliber, improving its visualization. Side branches are not usually depicted in healthy subjects.

Dynamic secretin-stimulated MR pancreatography can also be useful for diagnosing pancreatic papillary stenosis or dysfunction and for detecting reduced pancreatic exocrine reserve. In patients with chronic pancreatitis, there may be reduced duodenal filling due to a decrease in pancreatic exocrine reserve. Acinar filling during dynamic secretin MR pancreatography may be a specific but insensitive finding of early pancreatitis.

SUMMARY

MRI is a versatile diagnostic tool that can be used to diagnose and evaluate a wide spectrum of diseases of the liver, biliary system, and pancreas.

See Also the Following Articles

Alimentary Tract, MRI of the • Computed Tomography (CT) • Picture Archiving and Communication Systems (PACS) • Radiology, Interventional

Further Reading

Horton, K. M., Bluemke, D. A., Hruban, R. H., Soyer, P., and Fishman, E. K. (1999). CT and MR imaging of benign hepatic and biliary tumors. *RadioGraphics* **19**, 431–451.

Hussain, S. M., Zondervan, P. E., Ijzerman, J. N. M., Schalm, S. W., de Man, R. A., and Krestin, G. P. (2002). Benign versus malignant hepatic nodules: MR imaging findings with pathologic correlation. *RadioGraphics* **22**, 1023–1039.

Mitchell, D. G. (1992). Focal manifestations of diffuse liver disease at MR imaging. *Radiology* **185**, 1–11.

Procacci, C., Megibow, A. J., Carbognin, G., Guarise, A., Spoto, E., Biasiutti, C., and Pistolesi, G. F. (1999). Intraductal papillary mucinous tumor of the pancreas; a pictorial essay. *RadioGraphics* **19**, 1447–1463.

Szklaruk, J., Silverman, P. M., and Charnsangavej, C. (2003). Imaging in the diagnosis, staging, treatment, and surveillance of hepatocellular carcinoma. *Am. J. Roentgenol.* **180**, 441–454.

Vitellas, K. M., Koegan, M. T., Spritzer, C. E., and Nelson, R. C. (2000). MR cholangiopancreatography of bile and pancreatic duct abnormalities with emphasis on the single-shot fast spin-echo technique. *RadioGraphics* **20**, 939–957.

Malabsorption

CHARLES M. MANSBACH
University of Tennessee and Veterans Administration Medical Center, Memphis

enterocytes Mature absorptive cells that line the villi of the small intestine.
malabsorption Failure to digest and absorb dietary nutrients.
sprue A word anglicized by Patrick Manson in 1880 from the Dutch *sprouw*, meaning "thrush"; used in seventeenth-century Europe to describe a diarrheal disorder with oral aphthous ulcerations.
steatorrhea Increased fat in stool due to malabsorption.

When considering malabsorption, most people think of the malabsorption of dietary fat, or steatorrhea. It should be realized that humans also commonly malabsorb carbohydrates, especially lactose-containing foods, and less commonly, specific metals (calcium or iron) and vitamins (the fat-soluble vitamins or vitamin B_{12}). All of these dietary components play a role in nutrition and disease. Although ideas of nutrition with respect to heart disease or dietary vitamin requirements change over time, what does not change is the body's built-in ability to absorb almost all of what we eat under normal conditions.

BRIEF OVERVIEW OF FAT ABSORPTION

Humans normally absorb 95% of their dietary fat intake. Truly remarkable levels of intake can be achieved without overloading the system. For example, because of their caloric density, fat-laden foods are carried by Arctic explorers, who may consume up to 500 g of fat per day without malabsorption. The normal American diet contains 100–150 g of fat per day. Although fat is absorbed efficiently, cholesterol is not. Only about half of the cholesterol that is eaten is absorbed.

The absorption of dietary fat is quite complex, with the primary reason being the water insolubility of the lipid. Fat does not interact with water and if it is not processed to do so, it will be excreted, with only small amounts being absorbed. The chemical structure of fat, as triacylglycerol (TAG), can be graphically visualized as having the shape of the letter E, with the glycerol backbone forming the vertical part of the E and the three fatty acids (FAs), which are attached to glycerol in ester linkages, appearing as the horizontal lines. Digestion of TAG starts in the stomach with the gastric enzyme lipase and progresses further in the duodenum, where the lipid interacts with pancreatic lipase. The end product of this is monoacylglycerol (MAG), which can be visualized as the middle FA attached to the glycerol backbone, and two FAs. These products, which can interact with water, are absorbed. The rapidity of absorption is enhanced greatly by the presence of bile salts, which are excreted from the gallbladder, where they are stored after their synthesis in the liver. Both gallbladder contraction and the elaboration of pancreatic lipase are under the control of the hormone cholecystokinin. Specialized cells localized to the duodenum sample duodenal luminal contents and secrete this hormone. The absorbed FAs and MAGs are resynthesized to TAG in the intestinal absorptive cell and are packaged in a chylomicron (the major lipid transport lipoprotein of the intestines), which is ultimately secreted into the lymph and then into the circulation.

THE CAUSE OF DIARRHEA IN STEATORRHEA

Normally, humans produce 100 to 200 g of stool per day. Anything over 250 g per day is considered diarrhea. There are there reasons for the diarrhea in patients who malabsorb lipid: (1) The malabsorption of fat is usually, but not always, associated with the malabsorption of carbohydrates and proteins as well. These osmotically active particles enter the colon and require the colon to secrete water to maintain equivalent osmolality. (2) FAs entering the colon cause water and salts to be excreted rather than absorbed because they induce Cl^--driven water excretion. (3) Ricinoleic acid, the active laxative in castor oil (glyceryl triricinoleate), is 12-hydroxy stearate. Oleic acid, a common FA constituent of fat, is easily converted to 10-hydroxy stearate by bacterial hydroxylation of the double bond at the 9–10 position of oleate.

 Copyright 2004, Elsevier (USA). All rights reserved.

DISEASES CAUSING MALABSORPTION

Celiac Sprue

Celiac sprue, a relatively common genetically and immunologically based disease, has a prevalence rate of $\approx 1:200$. Most patients have either the DQ2 or DQ8 human leukocyte antigen (HLA) haplotype, which will interact with the inciting dietary protein, gliadin, found in wheat, rye, and barley, to cause the disease. Gliadin is an unusual protein in that 30% or more of the protein is made up of glutamine residues and there are large amounts of proline as well. The peptide component of gliadin that appears to be the proximal causative agent has been discovered by hypothesizing that the inciting gliadin peptide would resist the normal proteolysis occurring in the intestine. This turns out to be a peptide composed of 33 amino acids. The peptide resists hydrolysis because of its prolyl residues, which are known to resist proteolysis by gastro/pancreatic peptidases. This peptide interacts with tissue transglutaminase, the autoantigen of celiac patients, and its incubation with intestinal T cells from celiac patients causes their induction.

Classically, celiac sprue causes steatorrhea, weight loss, and abdominal cramps. Currently, more subtle signs of the disease are recognized as common presenting symptoms. These may be unexplained osteoporosis due to vitamin D deficiency and/or calcium malabsorption, iron deficiency without intestinal blood loss, or easy bruising associated with vitamin K depletion. The disease is also associated with other autoimmune disorders, such as type 1 diabetes or Hashimoto's thyroiditis.

Causes of Fat Malabsorption in Celiac Sprue

There are multiple causes of fat malabsorption in celiac disease. Because the condition induces enlargement of the intestinal crypts, both vertically and horizontally, there is an increase in salt and water secretion from the crypts, the normal site of such secretion. Normally, flow out of the crypts is sensed by the tips of the villi that are further up the villus than the crypts. These mature villus cells are the sites of normal salt and water absorption. In celiac disease, the villus tip cells are relatively immature, resulting in impaired absorption. The result is that the intestine is in a secretory rather than in the normal absorptive state. The resulting increase in salt and water in the lumen of the intestine reduces the concentration of bile salts in the lumen. In order to function properly in the solubilization of the products of TAG hydrolysis, bile salts must self-associate to form micelles. This is concentration dependent and does not begin in humans until bile salt concentrations exceed 1.8 m*M*. A concentration of 10 m*M* bile salts is seen postprandially in normal persons. Therefore, a reduction of bile salt concentration by dilution would have a greater effect on bile salt micelle formation than what would have been predicted by the reduction in bile salt concentration alone.

Because of villus atrophy, there is a reduction in the surface area of the intestine. This is magnified by the fact that the microvilli, which jut out into the intestinal lumen from the surface of the absorptive cells, also become atrophic. The microvillus surface area forms the majority of the total surface area of the intestines. This results in a reduction of the total surface available for nutrient absorption.

Another reason for fat malabsorption in celiac sprue is that the endocrine cells in the duodenum, which release cholecystokinin (CCK), become dysfunctional or atrophic. Normal stimuli to CCK release, such as essential amino acids, result in a reduced amount of hormone production in celiac sprue. However, if intravenous CCK is given to these patients, their gallbladder contracts normally, showing that the defect is not in CCK receptors in the gallbladder. A reduction in CCK in the duodenal mucosa can be demonstrated directly. This results in a further reduction of bile salt concentration in the intestinal lumen and a reduction of lipase and proteolytic enzymes released by the pancreas in response to a meal as well.

Finally, the cells at the intestinal surface are more immature than normal in celiac sprue. This results in the impaired processing of absorbed lipids and likely a reduction in the absorption of amino acids, peptides, and carbohydrates that depend on specific active transporters found in mature enterocytes for their normal uptake.

Other Malabsorbed Nutrients in Celiac Sprue

Because celiac sprue affects the most proximal intestine the most severely, vitamins and nutrients that are absorbed in this area of the gut can be malabsorbed even when the overall absorption of fat is not. Fat absorption can occur more distally in the less affected part of the bowel because adequate absorption does not require specialized transporters. For this reason, calcium, iron, and/or folate may be malabsorbed in the absence of steatorrhea. By contrast, vitamin B_{12}, which is absorbed in the distal ileum, is usually absorbed normally in celiac sprue patients. The fat-soluble vitamins, A, D, E, and K, may be malabsorbed due either to an insufficient concentration of bile salts in the gut lumen or to a reduction in the products

of TAG hydrolysis that aid in the absorption of the vitamins.

Diagnosis of Celiac Sprue

The classic way to diagnose celiac sprue is by intestinal biopsy. This is not always easy because of the possibility of poorly oriented tissue or other problems, leading to either a false positive or a missed diagnosis in the hands of an inexperienced pathologist. At best, because there are no specific pathologic characteristics of celiac sprue on biopsy, the pathologist can only state that the tissue is consistent with a diagnosis of celiac disease. Two new blood tests, antitissue transglutaminase and antiendomysial antibodies, are now available and both appear to have few false positives and false negatives. When suspicion is high that the patient has celiac sprue, such as with a primary member of a family known to have the disease, a positive blood test is all that should be required. As more experience is acquired with the blood tests, it is likely that a positive biopsy result will become a lesser part of the diagnosis and that it will be reserved for those patients who do not respond to the gluten exclusion diet or who have other specialized issues. Some patients have a dermatological condition, dermatitis herpetiformis, a pruritic herpetic-like lesion appearing most commonly over extensor surfaces.

The diagnosis is not secured until the patient responds to a gluten exclusion diet. It is no longer current practice to challenge patients with gluten to see if they reacquire symptoms. Usually, however, the patient will assume after a period of time that they are cured and will have a dietary indiscretion. Eating gluten will rapidly cause the recrudescence of symptoms. In young children, eating gluten after a period of abstention can cause an illness similar to serum sickness, which can be quite severe.

Treatment of Celiac Sprue

Because the dietary cause of the celiac sprue is known, treatment is directed toward removal of gluten from the diet. All patients with celiac sprue respond to a diet devoid of wheat, rye, and barley. Oats are usually excluded as well until the patient has a response to the diet. Symptomatic improvement is usually seen within 2 weeks, but histologic improvement of the intestinal mucosa takes much longer and may never completely revert to normal. If symptomatic improvement does not occur, the most likely problem is dietary indiscretion. The labels of all food bought at grocery stores must be read carefully to exclude the introduction of gluten-containing material such as stabilizers or other nonspecific components used in foodstuffs such as ice cream. If the diet is correct, then the nonresponsive patient may have pancreatic insufficiency or stasis syndrome with bacterial overgrowth. If these are ruled out, the patient may have ulcerative ileojejunitis, collagenous sprue, or α chain disease, all of which are not responsive to gluten exclusion.

Stasis Syndrome

Diseases and Conditions Associated with Stasis Syndrome

Stasis syndrome, or blind-loop syndrome, is caused by the lack of forward propulsion of intestinal contents. In this environment, bacteria proliferate, causing damage to enterocytes. Diseases that reduce intestinal propulsion, such as scleroderma and diabetes, or idiopathic problems can cause stasis. Other potential causes are small bowel diverticulosis, in which bacteria proliferate within the diverticulae, which then seed the intestinal luminal contents. Narcotics, which slow intestinal propulsion, do not cause symptomatic stasis syndrome. The proton pump inhibitors, which reduce gastric acid production, also increase small bowel bacterial content, but likewise do not cause symptoms.

Pathophysiology of Stasis Syndrome

The proliferating bacteria in stasis syndrome release glycosidases, which digest some of the protective glycosides from the surface of the enterocytes. Proteases and pancreatic phospholipase, which are plentiful in the intestinal lumen, are then postulated to gain access to the cell interior. This results in damage to the enterocyte that can be seen at the electron microscopic level as dilated endoplasmic reticulum (ER), a sign of cellular distress. Impaired lipid absorption is suggested by the finding that the endoplasmic reticulum is engorged with prechylomicrons, with few chylomicrons in the Golgi. This suggests a blockade in the normal movement of chylomicrons from the ER to the Golgi. A modest steatorrhea may result. Albumin is also thought to leak into the bowel in the stasis syndrome, resulting in protein-losing enteropathy and a reduction in serum albumin concentrations. Although there is bacterially caused deconjugation of bile salts, with a resultant increase in bile salt pK_a that leads to bile salt protonation passive absorption, the potential reduction in intestinal luminal bile salt concentration is minimal, and normal amounts of hydrolyzed lipids are found in the aqueous phase of intestinal contents. Because of bacterial folate synthesis, folate levels are usually increased in the serum of these patients. Vitamin B_{12} levels may be low due to the binding of B_{12} to the surface of the bacteria, its metabolism by

the bacteria to inactive metabolites, and the utilization of B_{12} by the bacteria for their own metabolism.

Diagnosis

The diagnosis of stasis syndrome can be suspected in patients who have multiple diverticulae as seen on small-bowel follow-through X rays or who are shown to have a Billroth II anastomosis with an enlarged afferent loop. Establishing the diagnosis is more difficult. The classical way is to collect intestinal luminal fluid for quantitative bacterial counts, especially anaerobic bacteria. Counts greater than 10^5/ml anaerobic bacteria suggest the diagnosis. Bile acid deconjugation can be documented in fluid collected from the upper intestine by separating the bile salts using thin layer chromatography. Although a modest steatorrhea may be seen, this is a nonspecific finding.

Treatment

Treatment is directed toward ridding the bowel of anaerobic bacteria. This can be done using antibiotics such as tetracycline, which are given for 3 weeks at 1 g/day and then treatment is stopped. The patient is observed and retreated if required. Some patients will need to be on chronic antibiotic therapy for the least number of days per month that will reduce symptoms. Surgical revision of a blocked afferent loop or removal of small bowel diverticulae, if feasible, can be helpful as well.

Chronic Pancreatitis and Pancreatic Insufficiency

Diseases Associated with Pancreatic Insufficiency

Chronic pancreatitis is the condition that most frequently leads to pancreatic insufficiency. A diagnosis of pancreatic insufficiency requires showing that the pancreas is dysfunctional at the exocrine or endocrine level. Because of the great functional reserve of the pancreas, it takes 90–95% glandular destruction before the resulting lipase deficiency leads to steatorrhea.

By far the most common cause of pancreatic insufficiency in the United States is alcoholic pancreatitis. Minor amounts of pancreatic insufficiency are caused by cystic fibrosis or hereditary pancreatitis. Painless pancreatitis leading to pancreatic insufficiency is well described. Not all diseases that cause acute pancreatitis are associated with the development of chronic pancreatitis, however. For example, biliary pancreatitis never leads to the chronic form of the disease.

Symptoms of Chronic Pancreatitis

Symptoms are related to the development of either exocrine or endocrine insufficiency or both. In the case of exocrine insufficiency, steatorrhea may be marked, with up to 70% of ingested fat excreted. That there is not more is due to the action of gastric lipase. The amount of fat may be so great that oil droplets ooze from the rectum. The weight of the stool is not as great as one would expect for the amount of lipid in the stool because there is poor hydrolysis of proteins and (to a lesser extent) carbohydrates, leading to fewer osmotically active particles in the stool. Endocrine insufficiency manifests by the development of diabetes. This type of diabetes usually requires insulin, but not as much as in type 1 diabetes, because there is no immunological component. The side effects associated with diabetes, such as renal failure or neurological problems, are rarely seen in pancreatogenous diabetes for unclear reasons. A possibility is the reduced life span of these patients following diagnosis, compared to patients with type 1 diabetes.

Diagnosis

Chronic pancreatitis can be diagnosed by the observation of calcifications in the pancreas either by standard X ray, computed tomography (CT) scan, or ultrasound. A dilated pancreatic duct at CT scan is also helpful. In general, however, the diagnosis is not made until late in the course. The most sensitive test is the pancreatic secretin test, in which the hormone secretin is injected intravenously and pancreatic juice is collected by a tube placed in the duodenum. The volume of fluid and its bicarbonate concentration are measured. More than 70% of the gland must be destroyed before the test becomes abnormal. In Europe, similar studies are performed after a Lundth test meal. An abnormally low serum trypsinogen is also an indicator of chronic pancreatitis.

Treatment

The simplest treatment is for the diabetic complication of chronic pancreatitis, for which insulin is usually required rather than oral medication. Treating the steatorrhea is more difficult. The major problem is that the enzymes that are given orally are in part acid inactivated in the stomach. It is estimated that only 10% of these are still active in the duodenum. Some preparations, such as Viokase, are given as active enzymes in a capsule form. Others, such as Creon, have the active enzymes embedded in a resin so that the active enzymes are not released until the pH of the surrounding luminal fluid is more alkaline. Combining enzyme therapy with a proton pump inhibitor is another way of preserving enzyme

activity. Despite these steps, it is rare to completely reverse the steatorrhea of chronic pancreatitis. The associated pain experienced by patients with chronic pancreatitis is a therapeutic dilemma if medications for pain relief, i.e., narcotics, are supplied to persons with an addictive personality. For this reason, narcotics should not be used, especially over the long term. Milder analgesics such as acetaminophen are more appropriate. An occasional patient can be helped by giving Viokase, with the theory being that the active enzymes reduce pancreatic secretion by hydrolyzing cholecystokinin-releasing peptide, which controls pancreatic secretion in a feedback loop fashion.

Cholestasis

Cause and Physiological Effects

Because bile salts play an important role in the efficient absorption of lipids and are required for the absorption of fat-soluble vitamins, any process that limits the delivery of bile salts to the intestinal lumen in response to a meal reduces the effectiveness of bile salts in the absorption process. This would include conditions that are localized to the liver, such as intrahepatic cholestasis, or obstruction of the extrahepatic bile ducts, as in pancreatic cancer or gallstones. Because extrahepatic conditions that obstruct bile flow are usually dealt with by relieving the obstruction (failure to do so usually leads to the death of the patient within a short period of time), the physiological effects of a reduction in bile salt delivery to the intestine in terms of lipid or vitamin absorption are usually only seen in patients with intrahepatic cholestasis. For example, in primary biliary cirrhosis, in which there is failure to excrete bile salts before the development of cirrhosis, osteoporosis due to impaired vitamin D absorption may be seen. The steatorrhea found under these conditions is not severe, usually in the range of 10 to 20 g of fat per day.

Treatment

Treatment is directed at the disease causing the bile flow restriction. If the problem is extrahepatic, the treatment is usually surgical; if intrahepatic, treatment depends on the cause. For example, some drugs cause intrahepatic cholestasis that is relieved on removal of the drug. Primary biliary cirrhosis is helped by the administration of ursodesoxycholic acid.

Whipple's Disease

Description and Pathophysiology

Whipple's disease, a relatively rare condition affecting mostly middle-aged men, is invariably fatal if not detected and treated. It is caused by the bacterium *Tropherema whippelii*. Symptoms usually start with multiple joint pains and swelling, which may occur years before the more localizing symptoms of fever, steatorrhea, lymph node swelling, and weight loss. Other organs can be involved, including the heart or the brain. The neurological symptoms, which can be ocular or a progressive dementia, are particularly insidious because they are usually irreversible. Neurological symptoms may occur after antibiotic treatment for Whipple's disease if the antibiotic used does not cross the blood–brain barrier.

Tropherema whippelii is able to exist in macrophages for many years, both in the intestine and in draining lymph nodes. The lymph nodes may become blocked, resulting in impaired lymph flow from the intestinal villi. This leads to swollen, clublike villus structures and dilated lymphatics that can be seen at both the light and electron microscopic levels. Rarely, the mucosa can become flattened and the biopsy may be confused with celiac sprue. The lymphatic blockade leads to a reduction in the ability of the intestine to transport chylomicrons to the thoracic duct. Because there are other mechanisms of transporting lipid into the body besides the normally used lymphatic system, the amount of steatorrhea seen in this condition is modest, at 20 to 25 g of fat per day.

Diagnosis

A small bowel biopsy is the usual way that the condition is diagnosed. Once the condition is suspected and a biopsy is obtained, the biopsy results are usually straightforward. The striking finding is that the lamina propria is packed with macrophages, which have enlarged regions of cytoplasm with a foamy appearance. With periodic acid–Schiff (PAS) stain, the macrophages are stained a bright red due to the numerous bacteria contained within them. Some of these bacteria can be seen with electron microscopy to be dividing, a sign that the bacteria are alive and proliferating. In patients who are HIV positive and have the AIDS complex, PAS-positive macrophages can also be found due to the presence of *Mycobacterium avium-intracellulare*. These organisms can be distinguished from *T. whippelii* on electron microscopy because of differences in morphology.

Treatment

The current treatment of choice is sulfamethoxazole–trimethoprim (SM–TM). This antibiotic eradicates *T. whippelii* in the intestines and elsewhere in the body. Importantly, SM–TM crosses the blood– brain barrier and attacks the bacteria in the central nervous system (CNS). Rarely, patients treated with SM–TM

will develop CNS symptoms. Treatment should be given for 1 year. If an antibiotic that does not cross the blood–brain barrier is chosen, CNS symptoms may occur 1 to 3 years later, even though other systemic or intestinal symptoms have cleared. In patients who have been followed for long periods (up to 25 years) and who have become asymptomatic, PAS-positive macrophages can still be seen in the intestine, although at a greatly reduced number over that which was originally present.

Intestinal Ischemia

Presentation and Diagnosis

There are three levels of intestinal ischemia. The first is intestinal angina, in which the patient experiences upper abdominal, usually severe, crampy pain associated with meals. The pain is so severe that the patient begins to eat progressively less, resulting in weight loss. The weight loss may be extensive. Early in the course, when the patient is eating well, steatorrhea may occur as a result of poor intestinal perfusion and oxygenation. Ultrasonic Doppler evaluation of the three arteries supplying the small intestine, i.e., the celiac, the superior mesenteric, and the inferior mesenteric arteries, should show reduced flow in at least two of the three arteries. This is usually followed by an aortogram, in which radio-opaque dye is injected into the aorta to outline the lumens of the three arteries. To substantiate the diagnosis, there must be significant obstruction of at least two of the three arteries. The second level of ischemia is mucosal infarction and the third level is infarction of the entire thickness of the gut. In either case, steatorrhea does not result because the patient is too sick to eat.

Treatment

Treatment is surgical, either bypassing or endarterectomizing the obstructed arteries.

Zollinger–Ellison Syndrome

Presentation and Diagnosis

Zollinger–Ellison syndrome is caused by a tumor that elaborates the hormone gastrin. Gastrin drives the stomach to produce large amounts of acid, with the result that patients usually present with a severe peptic ulcer, a perforated ulcer, or an ulcer that bleeds. The tumor is discovered on investigation because of the unusual location of the ulcer (for example in the distal duodenum) or because of the persistent recurrence of the ulcer after therapy. Occasionally, diarrhea with malabsorption is seen. This is due to the acidic environment of the duodenum, where a pH of 1.5 can be found. This pH irreversibly denatures lipase, with the result that the patient is rendered essentially lipase insufficient.

The diagnosis is made by measuring gastrin in the serum; if the results are inconclusive, the hormone secretin is given intravenously and the gastrin response in the serum is followed. If gastrin rises to a level of 200 pg/ml or more above fasting levels, the patient is likely to have the Zollinger–Ellison syndrome.

Treatment

With the advent of proton pump inhibitors, symptomatic relief is available. The objective is to reduce gastric acid output to 10 mEq H^+/hour or less; below 10 mEq/hour, ulceration seems not to occur. Approximately 60% of the tumors are malignant at the outset and they are slow growing. If the tumor can be shown to be localized, surgical resection is an option.

Systemic Mastocytosis

Symptoms, Diagnosis, and Treatment

Systemic mastocytosis is manifested by tan to brown macules (urticaria pigmentosa) distributed over the trunk and extremities. The hallmark of the disease is the raising of a pruritic wheal on trauma to the skin, caused by mast cell release of histamine. This is a condition in which mast cells may proliferate in the skin, bone marrow, liver, spleen, lymph nodes, and intestine. The cause of the malabsorption in this condition is not known, but may be related to histamine release and overproduction of gastric acid. The diagnosis is suggested by the classic skin lesions but occasionally submucosal nodules can be seen on barium X ray of the small intestine. A small bowel biopsy shows dilated villi packed with eosinophils. Urinary levels of a metabolite of prostaglandin D2 are increased. Treatment is nonspecific. Disodium cromoglycate may be helpful in relieving the intestinal symptoms.

Giardiasis

Symptoms, Diagnosis, and Treatment

Giardia lamblia is a common parasitic contaminant of drinking water in various parts of the world, including the United States. A famous outbreak occurred in Aspen, Colorado due to cross-contamination of the freshwater line with the sewage line. *Giardia* also commonly contaminates freshwater streams because of infected beavers. Symptoms are usually abdominal distension, crampy abdominal pain, excessive flatus, diarrhea, and, less often, steatorrhea. There is no known cause for the steatorrhea. The diagnosis is made by finding cysts in the stool or, more accurately, trophs in the

duodenal contents or next to the surface of the duodenum on a biopsy. When cysts or trophs are found, treatment is a 10-day course of metronidazole. Antibiotic resistance is rare.

Short Bowel Syndrome

Diagnosis and Treatment

Short bowel syndrome is usually the result of surgical resection of the bowel. The extent of the resection can be estimated from a small bowel follow-through examination (a barium X-ray study). Alternatively, the amount of bowel removed at surgery may have been recorded and the records may be available. Malabsorption occurs because of the reduction in bowel surface area. Malabsorption is usually worse with distal resections because of the lack of ileal absorption of bile acids. Further, the ileum adapts to the absence of the jejunum, but the jejunum does not adapt to the absence of the ileum. By adaptation, it is meant that the intestinal villi become enlarged and elongated.

Treatment is difficult, but predigested meals using casein hydrolysates may be helpful. The protein hydrolysates are preferred over the crystalline amino acid preparations because di- and tripeptides are better absorbed than are single amino acids. A low-fat diet is usually best. Patients can also be given a small amount of oil prior to a meal; this is thought to slow intestinal transit of food, thus prolonging contact time between the food and the intestinal absorptive cells.

Lactose Intolerance

Symptoms, Diagnosis, and Treatment

Lactose intolerance is one of the commonest forms of intestinal "malfunction." In most mammals, loose stools occur if milk is ingested after weaning. In humans, this is also true for 90% of Asians, for 70% of African Americans, and for many tribes in Africa, with symptoms usually occurring at age 7 to 9 years. People of northern European descent and a few tribes in Africa retain the ability to drink milk throughout their lives without ill affect.

Symptoms of lactase insufficiency usually are abdominal cramps, flatus, and acidic loose stools shortly after milk ingestion. The milk sugar, lactose, must be split into its component monosaccharides, glucose and galactose, by the enzyme, lactase, which is present at the surface of intestinal absorptive cells. The decrease in lactase activity at weaning is a programmed enzymatic depletion, because feeding milk continuously does not change the developmental stage at which lactose becomes malabsorbed. The malabsorbed lactose enters the colon, where bacteria can digest the sugar and produce short-chain fatty acids. Treatment is either avoiding milk or foods containing dairy products, or treating milk prior to drinking with lactose-containing bacteria (Lactaid) or live bacteria containing lactase, such as Lactinex granules.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Celiac Disease • Cholestatic Diseases, Chronic • Exocrine Pancreatic Insufficiency • Giardiasis • Intestinal Ischemia • Mastocytosis, Gastrointestinal Manifestations of • Pancreatitis, Chronic • Short Bowel Syndrome • Whipple's Disease

Further Reading

Calam, J., Ellis, A., and Dockray, G. J. (1982). Identification and measurement of molecular variants of cholecystokinin in duodenal mucosa and plasma. Diminished concentrations in patients with celiac disease. *J. Clin. Invest.* **69**, 218–225.

Dietrich, W., Laag, E., Schopper, H., Volta, U., Ferguson, A., Gillett, H., Riecken, O., and Schuppan, D. (1998). Autoantibodies to tissue transglutaminase as predictors of celiac disease. *Gastroenterology* **115**, 1317–1321.

DiMagno, E. P., Go, V. L. W., and Summerskill, W. H. J. (1973). Relations between pancreatic enzyme outputs and malabsorption in severe pancreatic insufficiency. *N. Engl. J. Med.* **288**, 813–815.

Dreiling, D. A., and Janowitz, H. D. (1961). The measurement of pancreatic secretory function. *In* "The Exocrine Pancreas" (A. V. S. de Reuck, and M. P. Cameron, eds.), pp. 225. Ciba Foundation Symposium. Little, Brown and Co., New York.

Jacobson, D. G., Curlington, C., Connery, K., and Toskes, P. (1984). Trypsin-like immunoreactivity as a test for pancreatic insufficiency. *N. Engl. J. Med.* **310**, 1307–1399.

Lin, H. C., Zhao, X.-T., and Wang, L. (1996). Fat absorption is not complete by midgut but is dependent on load of fat. *Am. J. Physiol.* **271**, G62–G67.

Morrow, J. D., Guzzo, C., Lazurus, G., Oates, J. A., and Roberts, L. J. (1995). Improved diagnosis of mastocytosis by measurements of the major urinary metabolite of prostaglandin D2. *J. Invest. Dermatol.* **104**, 937–940.

Redman, D. A., Schmidt, T. M., McDermott, R. P., and Falkow, S. (1992). Identification of the uncultured bacillus of Whipple's disease. *N. Engl. J. Med.* **327**, 393–398.

Riepe, S., Goldstein, J., and Alpers, D. H. (1980). Effect of secreted bacteriode proteases on human intestinal brush border hydrolases. *J. Clin. Invest.* **66**, 314–322.

Shan, L., Molberg, O., Parrott, I., Felix, H., Filiz, F., Gray, G. M., Sollid, L. M., and Khosla, C. (2002). Structural basis for gluten intolerance in celiac sprue. *Science* **297**, 2275–2279.

Volta, V., Molinaro, N., deFranceschi, L., Fratangelo, D., and Biandi, F. B. (1995). IgA antiendomysial antibodies on human umbilical cord tissue for celiac disease screening. *Dig. Dis. Sci.* **40**, 1902–1905.

Mallory–Weiss Tear

BRYAN T. GREEN
Duke University Medical Center

angiography The radiographic visualization of blood vessels following introduction of contrast material.
gastrointestinal endoscopy Visual inspection of the lumen of the gastrointestinal tract by means of an endoscope passed through the mouth, anus, or ostomy.
hematemesis The vomiting of blood.
hematochezia The passage of red blood and/or clots from the rectum.
hemostasis The arrest of bleeding.
Mallory–Weiss tear Linear, nonperforating mucosal tear at or near the gastroesophageal junction.
melena The passage of black, tarry stools due to altered blood.
retching A strong involuntary effort to vomit.
upper gastrointestinal bleeding Blood loss from any portion of the digestive tract proximal to the ligament of Trietz (fourth portion of the duodenum).

Mallory–Weiss tears are linear, nonperforating mucosal tears near the gastroesophageal junction, classically induced by retching. They usually involve the gastric mucosa of the cardia but may involve the esophageal mucosa. They account for 5–15% of all cases of upper gastrointestinal bleeding. Although bleeding will stop spontaneously in the majority of cases, some cases will require therapeutic intervention.

CLINICAL FEATURES

Associated Conditions

In 1929, Mallory and Weiss described a syndrome of hematemesis and gastric hemorrhage that followed a long and intense alcoholic debauch. Early descriptions of the Mallory–Weiss syndrome emphasized the association with chronic and/or binge alcohol use and persistent vomiting, retching, and increased abdominal pressure. Although these features are present in many cases, numerous patients lack them. Conditions predisposing to Mallory–Weiss tears may be broadly grouped in two categories: conditions that produce increased intra-abdominal pressure and conditions that weaken the gastroesophageal (GE) junction. Mallory–Weiss tears have been produced by numerous conditions that increase intra-abdominal pressure, including the following: retching, nausea, vomiting, prolonged or paroxysmal coughing, hiccups, straining efforts, status asthmaticus, seizures, blunt abdominal trauma, and cardiac resuscitation. Conditions that weaken the GE junction include hiatus hernia and esophagitis. Nonsteroidal anti-inflammatory agent use is a risk factor, likely due to the adverse effect on platelet function. Portal hypertension is also associated with Mallory–Weiss tears and its presence portends a poor prognosis.

Iatrogenic Causes

Mallory–Weiss tears are uncommon during upper gastrointestinal (GI) endoscopy but are reported to occur in 0.07 to 0.4% of procedures. They have occurred in all types of upper GI endoscopy including standard esophogastroduodenoscopy, endoscopic retrograde cholangiopancreatography, percutaneous endoscopic gastrostomy tube placement, pneumatic balloon dilation, and standard dilation. They are induced by either direct trauma or more commonly retching and/or struggling during the procedure. The use of large-caliber or side-viewing endoscopes may be a risk factor. Mallory–Weiss tears induced during endoscopy typically have a benign course with spontaneous resolution of bleeding and only rarely require blood transfusion. If bleeding persists, the standard treatment modalities may be utilized.

Nasogastric and nasoduodenal tubes have been reported to produce Mallory–Weiss tears, usually due to retching while the tube is being placed. Polyethylene glycol electrolyte lavage solution ingestion to prepare the bowel for colonoscopy has been reported to produce Mallory–Weiss tears due to the vomiting it may induce, but such incidents are rare.

Clinical Presentation

Mallory–Weiss tears typically present with hematemesis (vomiting blood), classically after the patient has had several nonbloody episodes of emesis. The hematemesis may be frank red blood, suggesting an acute tear, or resemble coffee grounds, suggesting blood that has been altered while in the stomach for

 Copyright 2004, Elsevier (USA). All rights reserved.

several hours. Up to one-half of patients may have hematemesis on first emesis. Anticedent nausea, vomiting, retching, regurgitation, or abdominal pain is present in over one-half of patients. Less commonly, patients may have no hematemesis but pass melena (dark black tarry stools). Rarely, Mallory–Weiss tears may cause hematochezia (red, bloody bowel movements). This indicates a very rapidly bleeding lesion and predicts a poor outcome. One-third of patients will have hemodynamic instability at presentation, particularly those who present with hematochezia.

MANAGEMENT

Diagnosis

Mallory–Weiss tears were initially diagnosed either at surgery or autopsy but the advent of flexible endoscopy has made gastrointestinal endoscopy the gold standard for diagnosis. Although barium radiographs may occasionally be diagnostic, their use for this purpose is of historical interest only. Through the endoscope a Mallory–Weiss tear will appear as a linear laceration at or near the GE junction. Often they are seen only when the endoscope is retroflexed and the gastric side of GE junction is carefully inspected. The length of the tear may range from several millimeters to several centimeters, with the majority of lesions measuring near the shorter extreme. Roughly one-third of patients will have either active bleeding or stigmata of recent bleeding (visible vessel, adherent clot, or pigmented protuberance) at the time of endoscopy. These characteristics predict a complicated course (rebleeding, high transfusion requirements, or death) and endoscopic treatment with a variety of means (described below) is recommended. Patients without active bleeding or stigmata of recent bleeding, coagulopathy, portal hypertension, or hemodynamic instability can usually be managed by a brief period of observation.

Treatment

Resuscitation

As with all forms of GI bleeding, initial resuscitation with intravenous fluids and blood transfusion is essential. Anti-emetics should be administered and any coagulopathy corrected. Intravenous histamine 2 receptor antagonists are generally given but their benefit is unproven.

Endoscopy

Roughly one-third of patients with Mallory–Weiss tears will require endoscopic therapy. This is generally performed at the time of diagnosis with the same endoscope. A variety of means have successfully been employed, including electrocoagulation, injection, band ligation, and hemoclipping. Bipolar cautery or heater probe coaptive coagulation was initially used to treat bleeding from ulcers but can be effective at achieving hemostasis and reducing rebleeding rates for Mallory–Weiss tears. A variety of substances have been injected into and around Mallory–Weiss tears to achieve hemostasis. Epinephrine at a concentration of 1:10,000 is most commonly used and has been proven successful. It may be used alone or prior to electrocautery. The combination may be more effective than either agent alone. Injection with 98% alcohol, a sclerosing agent, has also been effective. Endoscopic band ligation and hemoclipping have been successful but are not yet widely used for Mallory–Weiss tears.

Radiology

Angiography with embolization or selective intra-arterial vasopressin infusion is effective when endoscopy fails to achieve hemostasis.

Surgery

Although once the only available treatment, surgery is now rarely required and utilized only when other measures have failed.

OUTCOME

Two-thirds of patients with a Mallory–Weiss tear will not require endoscopic therapy and will heal spontaneously. Blood transfusion is required in roughly one-half of patients. Less than 10% of patients will rebleed during the index hospitalization and usually within 24 h. Hematochezia, hemodynamic instability, portal hypertension, coagulopathy, a low admission hematocrit, and active bleeding or stigmata of recent bleeding are predictors of a complicated course. The mortality of Mallory–Weiss tears approaches 5%.

See Also the Following Articles

Emesis • Hiccups (Singultus) • Nausea • Upper Gastrointestinal Bleeding • Upper Gastrointestinal Endoscopy

Further Reading

Bharucha, A. E., Gostout, C. J., and Balm, R. K. (1997). Clinical and endoscopic risk factors in the Mallory–Weiss syndrome. *Am. J. Gastroenterol.* **92**, 805–808.

Carsen, G. M., Casarella, W. J., and Spiegel, R. M. (1978). Transcatheter embolization for treatment of Mallory–Weiss tears of the esophagogastric junction. *Radiology* **128**, 309–313.
Chung, I. K., Kim, E. J., Hwang, K. Y., *et al.* (2002). Evaluation of endoscopic hemostasis in upper gastrointestinal bleeding related to Mallory–Weiss syndrome. *Endoscopy* **34**, 474–479.
Graham, D. Y., and Schwartz, J. T. (1978). The spectrum of the Mallory–Weiss tear. *Medicine* **57**, 307–318.
Kortas, D. Y., Haas, L. S., Simpson, W. G., *et al.* (2001). Mallory–Weiss tear: Predisposing factors of a complicated course. *Am. J. Gastroenterol.* **96**, 2863–2865.
Llach, J., Elizalde, J. I., and Guevara, M. C. (2001). Endoscopic injection therapy in bleeding Mallory–Weiss syndrome: A randomized controlled trial. *Gastrointestinal Endosc.* **54**, 679–682.
Montalvo, R. D., and Lee, M. (1996). Retrospective analysis of iatrogenic Mallory–Weiss tears occurring during upper gastrointestinal endoscopy. *Hepato-Gastroenterology* **43**, 174–177.
Yamaguchi, Y., Yamato, T. Y., Katsumi, N., *et al.* (2001). Endoscopic hemoclipping for GI bleeding due to Mallory–Weiss syndrome. *Gastrointestinal Endosc.* **53**, 427–430.

Malnutrition

SASSAN PAZIRANDEH, JUSTIN A. MAYKEL, AND BRUCE R. BISTRIAN
Harvard University Medical School and Beth Israel Deaconess Medical Center

bacterial translocation Migration of various bacteria from the gastrointestinal mucosal surfaces to local lymphatic vessels and blood vessels.

bioimpedance analysis Using imperceptible electric currents of various frequencies, the resistance and reactivity of body tissues, reflecting total body water, can be used to assess body cell mass and lean body mass indirectly.

body cell mass The actively metabolic component of lean body mass, representing approximately 40% of body mass (weight) and consisting principally of skeletal muscle and visceral organs.

body mass index (BMI) Body weight in kilograms divided by height in meters squared; this index provides a good measure of nutritional states including malnutrition, normal status, overweight, and obesity. The BMI is not gender-specific and is relatively independent of height.

counterregulatory hormones The various hormones of intermediary metabolism, including growth hormone, catecholamines, cortisol, and glucagon, that counter the effects of insulin to lower glucose levels.

creatinine–height index The 24 h urinary excretion of creatinine is an indirect measure of skeletal muscle, which, when related to height and gender, can be used to estimate the degree of protein calorie malnutrition.

kwashiorkor Also known as protein malnutrition, this condition is due to a moderate to severe systemic inflammatory response with some degree of accompanying semistarvation.

lean body mass The fat-free compartment of the body, which contains nitrogen, is termed the lean body mass and makes up approximately 75% of body mass or weight. It includes the body cell mass, the extracellular fluid compartment, and the supporting tissues.

marasmus Also known as protein calorie malnutrition, this condition is primarily due to prolonged periods of inadequate intake of nutrients. It is often seen in the chronically ill with little or no underlying systemic inflammation.

mid-arm muscle circumference (MAMC) This index is derived by measuring the triceps skinfold (TSF) and mid-upper arm circumference (AC) midway between the shoulders and the elbow according to the following formula: $\text{MAMC} = \text{AC} - 3.14 \times (\text{TSF in centimeters})$.

nitrogen balance The net difference between intake of nitrogenous sources of nutrients (i.e., proteins) and nitrogen losses, principally in the urine. Nitrogen intake is estimated by protein intake in grams divided by 6.25. The 24 h urinary urea nitrogen excretion in grams plus 4 is used to estimate nitrogen output. Each gram of nitrogen lost or gained represents 28 to 30 g of lean tissue loss or gain.

subjective global assessment A comprehensive assessment of nutritional status using both objective and subjective clinical findings.

upper arm anthropometry An indirect means of measuring one's fat and muscle mass using triceps and subscapular skin folds (indexes of body fat) and the mid-upper arm circumference (an index of skeletal muscle).

Copyright 2004, Elsevier (USA). All rights reserved.

Carsen, G. M., Casarella, W. J., and Spiegel, R. M. (1978). Transcatheter embolization for treatment of Mallory–Weiss tears of the esophagogastric junction. *Radiology* **128**, 309–313.
Chung, I. K., Kim, E. J., Hwang, K. Y., *et al.* (2002). Evaluation of endoscopic hemostasis in upper gastrointestinal bleeding related to Mallory–Weiss syndrome. *Endoscopy* **34**, 474–479.
Graham, D. Y., and Schwartz, J. T. (1978). The spectrum of the Mallory–Weiss tear. *Medicine* **57**, 307–318.
Kortas, D. Y., Haas, L. S., Simpson, W. G., *et al.* (2001). Mallory–Weiss tear: Predisposing factors of a complicated course. *Am. J. Gastroenterol.* **96**, 2863–2865.
Llach, J., Elizalde, J. I., and Guevara, M. C. (2001). Endoscopic injection therapy in bleeding Mallory–Weiss syndrome: A randomized controlled trial. *Gastrointestinal Endosc.* **54**, 679–682.
Montalvo, R. D., and Lee, M. (1996). Retrospective analysis of iatrogenic Mallory–Weiss tears occurring during upper gastrointestinal endoscopy. *Hepato-Gastroenterology* **43**, 174–177.
Yamaguchi, Y., Yamato, T. Y., Katsumi, N., *et al.* (2001). Endoscopic hemoclipping for GI bleeding due to Mallory–Weiss syndrome. *Gastrointestinal Endosc.* **53**, 427–430.

Malnutrition

SASSAN PAZIRANDEH, JUSTIN A. MAYKEL, AND BRUCE R. BISTRIAN
Harvard University Medical School and Beth Israel Deaconess Medical Center

bacterial translocation Migration of various bacteria from the gastrointestinal mucosal surfaces to local lymphatic vessels and blood vessels.

bioimpedance analysis Using imperceptible electric currents of various frequencies, the resistance and reactivity of body tissues, reflecting total body water, can be used to assess body cell mass and lean body mass indirectly.

body cell mass The actively metabolic component of lean body mass, representing approximately 40% of body mass (weight) and consisting principally of skeletal muscle and visceral organs.

body mass index (BMI) Body weight in kilograms divided by height in meters squared; this index provides a good measure of nutritional states including malnutrition, normal status, overweight, and obesity. The BMI is not gender-specific and is relatively independent of height.

counterregulatory hormones The various hormones of intermediary metabolism, including growth hormone, catecholamines, cortisol, and glucagon, that counter the effects of insulin to lower glucose levels.

creatinine–height index The 24 h urinary excretion of creatinine is an indirect measure of skeletal muscle, which, when related to height and gender, can be used to estimate the degree of protein calorie malnutrition.

kwashiorkor Also known as protein malnutrition, this condition is due to a moderate to severe systemic inflammatory response with some degree of accompanying semistarvation.

lean body mass The fat-free compartment of the body, which contains nitrogen, is termed the lean body mass and makes up approximately 75% of body mass or weight. It includes the body cell mass, the extracellular fluid compartment, and the supporting tissues.

marasmus Also known as protein calorie malnutrition, this condition is primarily due to prolonged periods of inadequate intake of nutrients. It is often seen in the chronically ill with little or no underlying systemic inflammation.

mid-arm muscle circumference (MAMC) This index is derived by measuring the triceps skinfold (TSF) and mid-upper arm circumference (AC) midway between the shoulders and the elbow according to the following formula: $\text{MAMC} = \text{AC} - 3.14 \times (\text{TSF in centimeters})$.

nitrogen balance The net difference between intake of nitrogenous sources of nutrients (i.e., proteins) and nitrogen losses, principally in the urine. Nitrogen intake is estimated by protein intake in grams divided by 6.25. The 24 h urinary urea nitrogen excretion in grams plus 4 is used to estimate nitrogen output. Each gram of nitrogen lost or gained represents 28 to 30 g of lean tissue loss or gain.

subjective global assessment A comprehensive assessment of nutritional status using both objective and subjective clinical findings.

upper arm anthropometry An indirect means of measuring one's fat and muscle mass using triceps and subscapular skin folds (indexes of body fat) and the mid-upper arm circumference (an index of skeletal muscle).

Copyright 2004, Elsevier (USA). All rights reserved.

whole-body neutron activation *In vivo* neutron activation analysis that uses a neutron beam to activate body nitrogen, which can then be measured in a whole-body radiation counter. This is the gold standard for body composition analysis, but it is largely a research tool for the investigation of important clinical problems.

Malnutrition remains one of the oldest and most prevalent global challenges for medicine, the scientific community, and public health organizations. The World Health Organization reported in 1999 that approximately one billion humans were either undernourished or malnourished. In many developing countries, where food insecurity and socioeconomic limitations are severe, malnutrition is endemic. It has been estimated that worldwide 40,000 deaths per day are attributable to malnutrition. Until the early 1970s, malnutrition was only clearly recognized in the developing world and during cataclysmic events, such as war, in the developed world. However, a number of surveys in the mid-1970s identified a high prevalence of malnutrition among certain populations in the developed world and even among the more "affluent" adults. This is protein calorie malnutrition (PCM), which is related to underlying illness. Grossly defined as unintentional weight loss of more than 10%, PCM is particularly common among hospitalized patients. Even in the era of modern medicine, malnutrition remains prevalent and is often unrecognized. Surveys comparing hospitalized patients in the 1970s to those in the late 1980s show that despite improved recognition of malnutrition, approximately 50% of surgical and medical patients are still identified as malnourished. No matter what the underlying cause, malnutrition is associated with poorer medical and surgical outcomes (greater risks of postoperative complications and poorer wound healing), longer hospital stays, and increased risk of death. A number of gastrointestinal diseases are associated with malnutrition. Some chronic illnesses, such as inflammatory bowel disease, malabsorptive conditions, and cirrhosis, have easily identified signs of malnutrition, whereas more acute conditions, such as severe pancreatitis, biliary sepsis, and toxic colitis, often have occult malnutrition, where significant losses of lean tissue and associated nutritional derangements commonly occur but are not realized. In order to optimize clinical outcome, it is therefore imperative for all clinicians and especially gastroenterologists to assess all patients and recognize those at risk for malnutrition.

DEFINITION

Malnutrition is a rather broad description of various disorders of poor dietary intake and/or enhanced catabolic losses. It has been used to describe the underfeeding conditions of marasmus and kwashiorkor seen in underdeveloped countries. It has also been applied to the overfeeding condition of obesity in the developed world and increasingly in the developing world. It has also been used to refer to deficiencies of various vitamins and minerals that accompany states of inadequate nutrient intake surrounding illness. This article focuses on states of suboptimal caloric and protein intakes leading to significant weight loss along with physiologic and functional impairments.

Protein Calorie Malnutrition

Protein calorie malnutrition (PCM) is the state of inadequate intake of food (as a source of protein, calories, and other essential nutrients) occurring in the absence of significant inflammation, injury, or another condition that elicits a systemic inflammatory response. This is also known as protein energy malnutrition (see Table I). It ranges from complete deprivation of nonfluid dietary intake (starvation) to a more common state of semistarvation, in which the intake of protein and energy is suboptimal. This has classically been referred to as marasmus. Common clinical conditions in which mainly starvation or semistarvation leads to PCM include anorexia nervosa, severe malabsorptive syndromes, states of gastrointestinal dysmotility, and upper gastrointestinal obstructions (i.e., esophageal strictures).

Protein Malnutrition

A condition that is more commonly encountered in the hospitalized setting is that of hypoalbuminemic malnutrition (also known as protein malnutrition or kwashiorkor). This generally occurs as a consequence of a moderate to severe systemic inflammatory response following acute illness or injury. Although dietary deprivation of protein alone could theoretically lead to the classic condition of kwashiorkor, the hallmark of protein malnutrition seen in the hospitalized patient is that of a systemic inflammatory response (SIR). This SIR reduces dietary intake due to anorexia, enhances

TABLE I Differentiating Features of the Two Major Types of Malnutrition

Protein malnutrition (kwashiorkor)	Protein calorie malnutrition (marasmus)
Develops over weeks	Develops over months
May not have weight loss	Significant weight loss
Due to systemic inflammation and semistarvation	Due to low intake of nutrients
Hypoalbuminemia	Serum albumin is normal
Excess body water	Edema is not a typical feature

catabolic losses of energy and protein stores, and invariably leads to hypoalbuminemia. A moderate systemic inflammatory response is seen with major bleeding, the inflammation seen in conditions such as severe pancreatitis or inflammatory bowel disease, and various systemic infections such as peritonitis or pneumonias. Common clinical conditions that induce a severe systemic inflammatory response include major burns, closed head injury, multiple trauma, and severe sepsis often associated with multiorgan dysfunction.

Although for descriptive purposes, marasmus and kwashiorkor are considered separate entities, in general most states of malnutrition have some element of a systemic inflammatory response. For instance, many patients suffering from chronic diseases such as cirrhosis, renal failure, chronic pancreatitis, and chronic cardiac or pulmonary failure exhibit a persistent, low-grade inflammatory response and are usually malnourished prior to their hospital admission. Meanwhile, others with normal nutritional status on hospital admission, such as those undergoing elective surgeries, can quickly develop protein malnutrition, making them susceptible to postoperative infectious complications.

Cachexia Syndrome

In a number of chronic medical conditions, a third form of malnutrition, known as the cachexia syndrome, develops. This is characterized by severe weight loss (more than 20%). This is usually seen at the end stages of medical diseases such as cirrhosis, renal failure, congestive heart failure, COPD, autoimmune deficiency syndrome, and various forms of malignancies, particularly those involving the upper gastrointestinal tract (esophagus, stomach, and pancreas). In these conditions, a mild to moderate systemic inflammatory response, manifested by the presence of enhanced cytokine production, elevated acute-phase protein levels, and hypoalbuminemia, generally accompanies the severe loss of lean tissue and fat.

METABOLIC RESPONSE TO MALNUTRITION

Body Compartments

To appreciate the physiologic changes occurring in starvation, protein calorie malnutrition, and protein malnutrition, one must have an understanding of normal body composition. Energy provision in the form of ATP, which is the body's means of converting chemical energy to mechanical work, is the net result of the metabolism of three basic fuels: glucose, protein, and fat. The three body compartments where these energy substrates are stored are the body fat, extracellular mass (ECM), and the body cell mass (BCM) (see Table II). The mobilizable pool of protein primarily derives from the BCM (approximately 40% of body mass), consisting of skeletal muscle (75%) and visceral tissues (25%). The visceral tissues include the highly metabolic organs like the liver, which has high energy requirements (i.e., 200 kcal/kg of tissue). The fat compartment is quantitatively the largest storage of energy, but unlike the BCM, it has limited metabolic activity and energy requirements. The ECM is principally a supporting compartment with limited metabolic and energy requirements.

TABLE II Body Compartments

Fat compartment (25%)
Body cell mass (40%)
Skeletal muscle (30%)
Viscera (10%)
Extracellular mass (35%)
Skeleton (10%)
Plasma proteins (5%)
Extracellular fluids (20%)

Physiology of Starvation and Semistarvation

Under normal physiologic conditions, immediately after feeding, insulin secretion is increased in response to absorbed carbohydrates and amino acids. Insulin stimulates the uptake of amino acids and glucose by liver and skeletal muscle, where glucose is stored as glycogen and the amino acids are used for the synthesis of proteins. Once feeding ends, the systemic pool of glucose is gradually reduced. This is paralleled by a fall in circulating insulin levels, a change that leads to: (1) a reduction in glucose uptake by peripheral tissues for energy production and repletion of glycogen stores, (2) an increase in gluconeogenesis and glycogenolysis by the liver to release glucose to the circulation as a fuel, and (3) increases in the oxidation of fatty acids released from adipose tissue. Finally, after an overnight fast, in a period referred to as the postabsorptive state, there is even greater mobilization of the liver glycogen and increased gluconeogenesis (from skeletal muscle release of amino acids) and enhanced fat oxidation, in order to maintain circulating glucose levels. In fact, during the postabsorptive state, fat oxidation is the major source of fuel for energy production.

In pure starvation, in the absence of a systemic inflammatory response, there is a shift over time toward preserving lean tissues by diminishing protein losses. One way such adaptation is achieved is by lowering the body's basal energy expenditure by 25 to 30% through

TABLE III Metabolic Responses to Starvation and Injury

Starvation	Injury and stress
Decreased basal metabolic rate	Increased basal metabolic rate
Increased gluconeogenesis (+)	Increased gluconeogenesis (+++)
Increased glycogenolysis (+)	Increased glycogenolysis (+++)
Increased lipolysis (++)	Increased lipolysis (++)
Increased amino acid breakdown	Increased amino acid breakdown (+++)
Increased free fatty acid use	Increased free fatty acid use
Ketone bodies (++)	Ketone bodies (+)
Low insulin levels	High insulin levels
Counterregulatory hormones (low to normal)	Very high levels of counterregulatory hormones
Typical clinical situations: COPD, chronic liver and renal disease, CHF	Typical clinical situations: trauma, severe sepsis, ARDS, SIRS

Note. COPD, Chronic obstructive pulmonary disease; CHF, Congestive heart failure; ARDS, Acute respiratory distress syndrome; SIRS, Severe inflammatory response syndrome.

changes in thyroid and catecholamine metabolism (see Table III). Meanwhile, as fasting continues beyond 2 to 3 days, the glycogen stores in the liver become completely depleted and almost half of the muscle glycogen is exhausted. The body then relies on the mobilization of amino acids (from skeletal muscle) and glycerol (from fat) to the liver in order to make glucose via gluconeogenesis. This is essential for tissues such as the red blood cells, renal medulla, and peripheral nerves that can use only glucose as fuel. The lowering of insulin, together with a rise of glucagon, further stimulates fatty acid oxidation to produce the ketone bodies acetoacetate and β-hydroxybutyrate. In this condition, called starvation ketosis, ketone bodies become an important source of fuel for skeletal muscle, visceral tissue, and brain. This adaptation diminishes the requirement for peripheral amino acid mobilization by providing an alternate fuel (ketone bodies) for those organs that are not glucose-dependent. In states of semistarvation, intakes of at least 100 g of carbohydrates will prevent starvation ketosis.

Physiologic Response to Injury and Critical Illness

What distinguishes the body's response to injury compared to that of starvation is the activation of the hypothalamic–pituitary–adrenal axis as well as the induced presence of a number of inflammatory mediators (cytokines). Irrespective of the cause of injury (i.e., severe trauma, sepsis, burn, inflammation, infection, or major surgery), there is a sudden increase in systemic cortisol concentrations as stimulated by adrenocorticotropin hormone release from the pituitary gland. Cortisol secretion, partially mediated and enhanced by cytokine production [interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor (TNF)], has principally catabolic effects. Meanwhile, cytokine-mediated resistance to anabolic hormones including insulin, growth hormone, insulin-like growth factor-I, and androgens produces net protein catabolism. The resulting net mobilization of amino acids from skeletal muscles and connective tissue provides circulating substrates for hepatic gluconeogenesis, wound healing, and support of the immune system. Meanwhile, the adrenal medulla and sympathetic nervous system are also stimulated with the systemic inflammatory response, via such triggers as pain, volume depletion, and fever. Survival becomes the priority for the body at the temporary expense of skeletal muscle mass.

In summary, supraphysiologic levels of these counterregulatory hormones lead to the net catabolism of protein, glycogen, and fat, in order to meet the heightened demands of the severe systemic inflammatory response. For the short term (7 to 10 days) in the previously well-nourished host with a mild to moderate degree of systemic inflammation, this trade-off provides a net benefit in terms of morbidity and mortality.

Cytokines and Their Role in Injury

The systemic inflammatory response, whether due to gram-negative sepsis or in response to severe injury or infection, leads to the production of a number of polypeptide mediators that are synthesized by the liver, fixed tissue, and circulating cytokine-producing cells (macrophages, endothelial cells, monocytes, and lymphocytes). Those most studied include the principal proximate cytokines (TNF, IL-1, and IL-6) that result from the activation of innate immunity. Their role in the metabolism of the critically ill is important due to their effects on insulin and counterregulatory hormone

release, as well as their direct effects on amino acid, fat, and glucose metabolism. Insulin is the principal hormone involved in glucose homeostasis, acting to regulate glucose production in the liver and to foster peripheral glucose uptake mainly by skeletal muscle. In normal circumstances, insulin also inhibits lipolysis, proteolysis, and glycogenolysis. However, in critical illness, despite relatively higher levels of insulin, the counterregulatory hormones and other mediators such as IL-1, IL-6, and TNF oppose these inhibitory effects of insulin, resulting in insulin resistance. This affects insulin's ability to reduce hepatic glucose production and to stimulate muscle glucose uptake.

CONSEQUENCES OF MALNUTRITION

In critical illness, the increase in resting energy expenditure is often overcompensated for by a dramatic reduction in activity-related energy expenditure, so that total energy expenditure may be less than in normal healthy individuals of similar body size. However, anorexia generally leads to a negative energy balance, and when combined with reduced activity, increased catabolism, and reduced protein synthetic efficiency, it produces an accelerated loss of body cell mass. After 7 to 10 days of combined semistarvation and moderate inflammatory response in an individual with an initially normal nutritional status, the body's ability to heal wounds and to support immune function begins to be impaired. Such effects are also seen with shorter periods of semistarvation in an individual with either preexisting malnutrition or severe inflammation alone. In such circumstances, likely complications include dysfunction of most aspects of immune function. Such impairments consist of diminished innate immunity, complement and immunoglubulin production, and cellular immunity as well as impairment of various aspects of leukocyte action including chemotaxis, phagocytosis, oxidative burst, and killing function. Other consequences are poor wound healing, as well as reduced performance of cardiac, respiratory, gastrointestinal, and other vital organ function. Indeed, numerous studies have clearly shown increased infectious risk, longer hospital stay, poorer outcome, and increased mortality in patients with significant unintentional weight loss (more than 10%) or significant hypoalbuminemia (<3.0 g/dl) prior to their acute illness.

NUTRITIONAL ASSESSMENT

To identify patients at high risk for protein malnutrition and those with prior protein calorie malnutrition, a systematic and comprehensive nutritional assessment is crucial. Such an assessment begins with a thorough history and physical examination. Through this evaluation, the clinician should be able to estimate the body composition and analyze the pertinent laboratory values, with an effort to compile all of the subjective and objective data required to define the patient's nutritional state.

Clinical History and Physical Examination

A nutritional history should include the patient's recent state of health and current complaints and symptoms related to recent dietary intake, signs of systemic inflammation, and voluntary motor activity. Special attention should be given to any gastrointestinal symptoms, such as nausea, vomiting, diarrhea, and dysphagia, as well as to the patient's dietary and weight history. An unintentional weight loss of approximately 10% in the recent past indicates a significant degree of semistarvation and a mild degree of PCM. Moderate PCM is usually characterized by a 10 to 20% weight loss and severe PCM should be considered when there is weight loss of more than 20% from baseline. Generally, in unintentional weight loss that develops with disease, approximately 50% of the loss is from body fat stores. Prior to hospitalization, if a premorbid weight is not available, the patient's weight can be compared to the established standards of weight in relation to height. A weight of 85% or less of the desirable range for a particular height indicates the likelihood of PCM. Body mass index (BMI), which is weight in kilograms divided by the height in meters squared, is another helpful parameter for diagnosing malnutrition, because it is independent of the patient's gender and height. The normal range is 20 to 25 kg/m^2 and a value less than 18.5 kg/m^2 represents significant PCM. Death from PCM is likely when the BMI falls in the range of 12 to 14 kg/m^2. A BMI of 25 to 30 kg/m^2 is considered to represent overweight and values greater than 30 kg/m^2 indicate the presence of obesity. In two common settings, body weight becomes a less specific marker of malnutrition. First, in many chronically ill patients who have a relative increase of their total body water due to decompensated hepatic, renal, or cardiac function, the increase in weight represents overhydrated lean tissue. Up to one-half of such patients will have severe PCM when assessed by specific measures such as upper arm anthropometry or total body nitrogen. Second, after hospitalization, particularly in critically ill patients, an increase in the total body water is expected due to elevated antinatriuretic and antidiuretic activity characteristic of the systemic inflammatory response related to the increased secretion of aldosterone, vasopressin, and insulin. This

state of overhydration is often compounded by massive volume resuscitation in the most critically ill patients or more commonly by the failure to restrict total fluid intake in the critically ill particularly when it is considered that 1.5 to 2 liters/day is necessary for adequate nutritional support in the absence of increased fluid losses.

Assessments of the Body Compartments

Anthropometry

During starvation alone or with accompanying stress, the contribution of the lean tissue to body weight loss ranges from 25 to 50%, respectively. Although individuals who suffer from more critical illnesses (i.e., closed head injury, multiple trauma, severe sepsis, or major burns) sustain lean tissue to fat loss ratios of 75 : 25, when not fed, their lean tissue and weight loss is usually masked by fluid retention. Simple assessment of body compartments can still be a valuable and an important task in determining an individual's nutritional status. One simple and inexpensive estimation of the body reserves is via upper arm anthropometry, particularly in patients with marasmus and in patients with volume overload states, such as those with chronic hepatic, renal, and cardiac failure. The triceps skinfold (TSF) and subscapular skinfold thickness are generally used as indexes of body fat. Such measurements have fallen out of favor due to their low level of sensitivity, lack of correction for hydration status, and heterogeneity in normal individuals. By themselves, they are not very sensitive, since many healthy individuals have a body fat composition of less than 5%. However, due to their relative simplicity, such measurements should be considered with the intent of identifying the most severely malnourished patients, a group who can certainly benefit from nutritional support. Upper arm anthropometry consists of the measurement of TSF thickness with skinfold calipers (in centimeters), measurement of the mid-arm circumference with a tape, and the calculation of the mid-arm muscle circumference (MAMC) using the following calculation: MAMC (centimeters) = arm circumference $-$ (3.14) $\times$ (TSF in centimeters). MAMC is used to estimate the total skeletal muscle mass. A value of less than the fifth percentile of the normal standard indicates the presence of severe PCM and losses of at least 30% of normal lean tissue.

Creatinine–Height Index

Another means of estimating lean tissue stores, specifically the muscle mass, is by measuring urinary creatinine excretion over a 24 h period. Once normalized for height, this measure is a reasonable index of the skeletal muscle mass, which is the body's largest compartment of actively metabolizing lean tissue, and a good estimate of body cell mass and lean body mass. What makes this a difficult parameter to rely on is the relative inconvenience and frequent inaccuracy of the collection in ambulatory patients. It is also not as reliable in patients with diminished creatinine clearance, with serum creatinine in the range of 4 to 6 mg/dl. Creatinine–height indexes (CHIs) of less than 75% and less than 60% of normal indicate moderate and severe protein depletion, respectively. A CHI of 60% of normal also correlates well with a MAMC below the fifth percentile, again indicating severe PCM.

Other Methods

Methods of total body water measurement, directly using isotope dilution or indirectly by bioimpedance analysis and underwater weighing, are much more accurate means of estimating the lean body mass and body cell mass. These techniques are based on the concept that fat-free body mass (lean body mass) and body water maintain a relatively close and stable relationship, which is true in health but not in disease. One can calculate the fat compartment by subtracting the fat-free body mass from the body weight. The value of these methods of body composition analysis is most evident in patients with marasmus, but they are of little value in patients with chronic cardiac, renal, or hepatic disease with fluid retention or in patients with critical illness leading to distortions in tissue composition. Dual-energy X-ray absorptiometry (DEXA) and total body nitrogen by neutron activation are more recent and more accurate techniques of estimating body compartments. The latter technique is valid even in the critically ill. Despite the accuracy and the sensitivity of DEXA and the neutron activation method, due to their expense and limited availability, as well as lack of a general knowledge of their role among most clinicians, they remain essentially research tools.

Serum Markers

Serum albumin is the most studied and cited value associated with the assessment of nutritional status. It is the most abundant serum protein and quantitatively the largest synthetic product of the liver. A low serum albumin can help differentiate between marasmus and hypoalbuminemic malnutrition (protein malnutrition). In marasmus, the albumin level is always above 3 g/dl and often normal. However, as the sole indicator of nutritional status, albumin is not a very sensitive marker. In fact, with ongoing stress, even with adequate nutritional support, serum albumin levels cannot be restored to normal. A low albumin level is seen in all

conditions of chronic inflammation and levels are severely depressed when the systemic inflammatory response is severe (i.e., major burns, severe sepsis, closed head injury, and multiple trauma). An understanding of the physiology of albumin clarifies these changes. The decline in serum albumin in the setting of injury, stress, and severe infection is due to several forces that are cytokine-mediated: (1) extravasations of the albumin molecule into the extravascular space as a result of an increased capillary permeability, (2) cytokine-induced reduction of hepatic albumin synthesis, (3) an increase in the breakdown of albumin, and (4) a reduction in hepatic production due to low protein intake. In addition, untreated edematous states due to volume overload are almost always manifested with serum albumin levels below normal. And since albumin is a negative acute-phase protein, when abnormally low, it indicates the recent presence of a systemic inflammatory response that also causes a reduction in dietary intake, increased rates of lean tissue catabolism, and reduced rates of protein synthesis. Thus, hypoalbuminemia results as a consequence of the systemic inflammatory response, which then commonly leads to protein calorie malnutrition. Studies have demonstrated a strong association between hypoalbuminemia and increased medical and surgical complications in hospitalized patients and poorer prognosis in outpatients with disease. A low serum albumin is an excellent predictor of morbidity, mortality, and longer hospital stay. Although not a specific index of malnutrition, it is a valuable tool in identifying the group of individuals with the greatest need for nutritional support—those with severe metabolic stress who are likely to have significant loss of lean tissue.

A number of other serum proteins (transferrin, retinol-binding protein, and prealbumin) have been used in the evaluation of malnutrition. Due to their shorter half-lives, they have been studied as potential alternatives to albumin. Like albumin, they are also negative acute-phase proteins; they have not been shown to be superior to albumin as indicators of the systemic inflammatory response. As with albumin, their response to nutritional support is limited until the systemic inflammatory response remits.

FUNCTIONAL TESTS FOR MALNUTRITION

To assess the physiologic and functional consequences of malnutrition and to evaluate the effects of nutritional support, a number of tests have been designed to address this issue on more subjective grounds. Delayed hypersensitivity skin testing can be used to assess cell-mediated immunity, which is often impaired in malnutrition of many varieties, including protein, calorie, iron, zinc, essential fatty acid, vitamin B12, and folate deficiencies. Even though the skin anergy of malnutrition is correctable with nutritional repletion, the multifactorial nature of this type of immune suppression makes skin testing a very insensitive tool. Other functional tests, for instance, hand-grip strength using an ergometer, electrophysiologic studies of the adductor pollicis muscle relaxation, and assessment of respiratory muscle strength, have all been shown to be good indices of nutritional changes due to illness, as well as reasonable predictors of postoperative morbidity. However, the use of such techniques is limited to those individuals who can cooperate and have good cardiopulmonary status and they have not found wide applicability as measures of response to nutritional support.

SUBJECTIVE GLOBAL ASSESSMENT

In an effort to combine all of the objective data along with the clinical and historical elements, an approach termed subjective global assessment (SGA) has been proposed. This systematic approach tries to incorporate the relevant clinical factors influencing the patient's nutritional state. The SGA is a bedside clinical assessment focusing on five areas: severity, duration, and the pattern of weight loss; dietary intake; presence of gastrointestinal symptoms; the individual's functional capacity; and finally, the metabolic demands of the patient's illness. Furthermore, several aspects of the physical examination, such as estimation of the subcutaneous fat and muscle bulk, cutaneous and hair lesions, and the fluid status, are all incorporated in the assessment to finally categorize the patient as being well nourished (category A), moderately malnourished (category B), or severely malnourished (category C). Studies show that the SGA is a reproducible way of categorizing one's nutritional state and it is a more accurate predictor of postoperative complications than serum albumin, anthropometry, creatinine–height index, and many other traditional markers of nutritional status taken individually.

NUTRITIONAL SUPPORT

Due to the prevalence of protein calorie malnutrition in hospitalized patients and the well-documented association between poor nutrition and increased morbidity and mortality, invasive nutritional therapies are

commonly employed in hospitalized patients. In order to identify the patients at risk and to provide appropriate and effective nutritional support, a thorough and early evaluation by a team of physicians and dieticians is crucial. The timing of initiation of nutritional support depends on the patient's nutritional assessment. In the absence of prior malnutrition and with an on-going moderate systemic inflammatory response, initiating invasive feeding 5 to 7 days or, even more conservatively, 7 to 10 days after a period of fasting, if oral intake has not resumed, is a reasonable and well-accepted guide. Those with a poor premorbid nutritional status (i.e., more than 10% weight loss or a BMI of less than 18.5 kg/m^2) and superimposed illness are not as able to mount an adequate inflammatory response. They often have nutritionally related impairment of their protein synthetic capacity and are more prone to organ failure. It is therefore important not to delay nutritional therapy in such individuals and to intervene within the first several days. Although it is even more effective to feed moderately malnourished individuals for 7 days prior to an elective major operative procedure, present-day reality is that early postinjury feeding is generally the rule. On the other hand, those with the most severe injuries, such as multiple trauma, major burns, closed head injury, and severe sepsis, can have lean tissue losses of 600 to 900 g/day and early feeding is essential. In this group of patients, even in the absence of pre-illness weight loss, adjuvant nutritional therapy should be started soon after the acute resuscitation and metabolic issues are resolved. Although many patients with major traumatic or burn injury are not malnourished at the time of presentation, their heightened degree of catabolism and increased metabolic demands quickly lead to severe protein malnutrition.

Thus, the goals of nutritional provision should be to provide adequate amino acids and energy along with essential nutrients, so as to allow optimal protein synthesis for the support of the immune system, wound healing, and vital organ function. The initial goal is to diminish but not to completely reverse protein loss, since this is not possible initially, given the intensity of the systemic inflammatory response and the major reduction in voluntary muscle activity. It is important to keep in mind that, in order to safely provide nutritional support for optimal protein synthesis, energy intake to meet balance is not necessary. Additionally, it should not be the goal of therapy in the first weeks of most injuries, as efforts to match and particularly to exceed the high metabolic energy needs of the injured and the critically ill have not been shown to be beneficial. In fact, they can lead to overfeeding complications such as hyperglycemia, hepatosteatosis, and immune suppression.

Modes of Nutritional Delivery

With a functional gastrointestinal tract, the traditional approach has been to deliver nutritional therapies enterally. Enteral feeding is a more physiologic route; in addition, theoretical reasoning has been used in support of this method. Bacterial translocation, in the absence of lumenal gut content, has been postulated to lead to septic complications. Although bacterial translocation has been demonstrated to be significant in animal models, confirmatory studies in humans have been inconclusive. To date, no studies in human have convincingly shown that such translocation is reduced via enteral feeding or promoted by parenteral nutrition. Although fasting does lead to minor architectural changes in the intestinal mucosa, the gut mucosal barrier in human does not seem to be impaired. Furthermore, in the critically ill and in many postsurgical patients, intolerance of enteral feeding, particularly when feeding into the stomach, is common. Total parenteral nutrition (TPN) is a well-established means of providing nutritional support in such patients and there is no clear evidence that this method has more risks or is less effective than enteral nutrition. Although many clinicians recommend the use of the enteral route versus the parenteral route based on the belief that parenteral nutrition is more likely to produce infectious complications, it appears from a closer review of the literature that many of the complications were more clearly related to overfeeding. When provided appropriately and with care, there are no significant differences between the two different modes of delivery. When it is possible to use enteral feeding to meet nutritional goals, this method is preferred, but TPN exclusively or in combination with enteral feeding should be considered when such goals are not achievable by enteral feeding alone. An added benefit to combination feeding is that such individuals tend to transition to full enteral feeding more quickly than those who are fed exclusively via the parenteral route.

CONCLUSION

Malnutrition in clinical practice is very common and, although it is an independent risk factor for morbidity, when accompanied by disease it can have an even more negative impact on outcome. The recognition of malnutrition requires an understanding of its physiological and functional consequences as well as an appreciation of its effects on body composition. There is a sizable body of suggestive evidence demonstrating that adjuvant nutritional support, when administered adequately, to the right group of patients, and in a timely fashion, can significantly improve outcomes.

See Also the Following Articles

Anorexia Nervosa • Bulimia Nervosa • Enteral Nutrition • Nutritional Assessment • Nutrition in Aging • Parenteral Nutrition • Protein–Calorie Deficiency—"Kwashiorkor"

Further Reading

Bistrian, B. R. (2000). Nutritional assessment. *In* "Textbook of Medicine" (L. Goldman and J. Bennett, eds.), 21st Ed., p. 1145. W. B. Saunders, Philadelphia, PA.

Heymsfield, S. B., Baumgartner, R. N., and Pan, S. F. (1999). Nutritional assessment of malnutrition by anthropometric methods. *In* "Modern Nutrition in Health and Disease" (M. E. Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 903–922. Lippincott Williams & Wilkins, Philadelphia, PA.

Hill, G. (1992). Body composition research: Implications for the practice of clinical nutrition. *J. Parenteral Enteral Nutr.* **16**, 197.

Hoffer, L. J. (1999). Metabolic consequences of starvation. *In* "Modern Nutrition in Health and Disease" (M. E Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 645–666. Lippincott Williams & Wilkins, Philadelphia, PA.

Jeejeebhoy, K. N. (2000). Nutritional assessment. *Nutrition* **16**, 585–589.

Stack, J. A., Babineau, T. J., and Bistrian, B. R. (1996). Assessment of nutritional status in clinical practice. *The Gastroenterologist Suppl.* **1**, S8–S14.

Sternberg, J., and Bistrian, B. R. (2000). Nutritional support in the patient with gastrointestinal disease. *In* "Therapy of Digestive Disorders" (M. Wolfe, ed.), pp. 842–850. W. B. Saunders, Philadelphia, PA.

Torun, B. and Chew, F. (1999). Protein-energy malnutrition. *In* "Modern Nutrition in Health and Disease" (M. E. Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 963–988. Lippincott Williams & Wilkins, Philadelphia, PA.

Malrotation

ROBERT W. CHANG, STEVEN M. ANDREOLI, AND MORITZ M. ZIEGLER
Children's Hospital Boston

Ladd's bands Peritoneal attachments to an abnormally positioned cecum. They can cause obstruction of the duodenum by extrinsic compression.

short bowel syndrome A spectrum of intestinal malabsorption that occurs after major resection of the small intestine.

volvulus A twisting or displacement of the intestine around a fixed focal point that can cause obstruction and/or strangulation.

Malrotation is defined as an abnormal position of bowel created by a variance in its embryologic development. Normal development of the human intestine includes rotation and fixation of the embryonic midgut. Abnormalities of this process create a spectrum of anomalies, which vary in their clinical significance from asymptomatic presentations to a twist of the midgut (volvulus) and abdominal catastrophe.

INTRODUCTION

Embryologic variation in midgut formation has long been recognized. Calder, in 1752, first described congenital duodenal obstruction due to malrotation. In 1898, Mall described the embryology of abnormal intestinal rotation. Dott, in 1923, correlated embryologic observations with clinical evidence and suggested a surgical treatment for duodenal obstruction. In his landmark article in 1932, Ladd described the surgical procedure that bears his name and remains the gold standard treatment for this condition.

The true incidence of malrotation of the midgut is difficult to determine. The autopsy prevalence may be as high as 0.5 to 1% of the total population, but the incidence in patients with clinical symptoms may be much less. The clinical presentation in the infant may occur in 1 in 6000 live births. Malrotation accounts for

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Anorexia Nervosa • Bulimia Nervosa • Enteral Nutrition • Nutritional Assessment • Nutrition in Aging • Parenteral Nutrition • Protein–Calorie Deficiency—"Kwashiorkor"

Further Reading

Bistrian, B. R. (2000). Nutritional assessment. *In* "Textbook of Medicine" (L. Goldman and J. Bennett, eds.), 21st Ed., p. 1145. W. B. Saunders, Philadelphia, PA.

Heymsfield, S. B., Baumgartner, R. N., and Pan, S. F. (1999). Nutritional assessment of malnutrition by anthropometric methods. *In* "Modern Nutrition in Health and Disease" (M. E. Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 903–922. Lippincott Williams & Wilkins, Philadelphia, PA.

Hill, G. (1992). Body composition research: Implications for the practice of clinical nutrition. *J. Parenteral Enteral Nutr.* **16**, 197.

Hoffer, L. J. (1999). Metabolic consequences of starvation. *In* "Modern Nutrition in Health and Disease" (M. E Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 645–666. Lippincott Williams & Wilkins, Philadelphia, PA.

Jeejeebhoy, K. N. (2000). Nutritional assessment. *Nutrition* **16**, 585–589.

Stack, J. A., Babineau, T. J., and Bistrian, B. R. (1996). Assessment of nutritional status in clinical practice. *The Gastroenterologist Suppl.* **1**, S8–S14.

Sternberg, J., and Bistrian, B. R. (2000). Nutritional support in the patient with gastrointestinal disease. *In* "Therapy of Digestive Disorders" (M. Wolfe, ed.), pp. 842–850. W. B. Saunders, Philadelphia, PA.

Torun, B. and Chew, F. (1999). Protein-energy malnutrition. *In* "Modern Nutrition in Health and Disease" (M. E. Shils, J. A. Olson, M. Shike, and A. C. Ross, eds.), 9th Ed., pp. 963–988. Lippincott Williams & Wilkins, Philadelphia, PA.

Malrotation

ROBERT W. CHANG, STEVEN M. ANDREOLI, AND MORITZ M. ZIEGLER
Children's Hospital Boston

Ladd's bands Peritoneal attachments to an abnormally positioned cecum. They can cause obstruction of the duodenum by extrinsic compression.

short bowel syndrome A spectrum of intestinal malabsorption that occurs after major resection of the small intestine.

volvulus A twisting or displacement of the intestine around a fixed focal point that can cause obstruction and/or strangulation.

Malrotation is defined as an abnormal position of bowel created by a variance in its embryologic development. Normal development of the human intestine includes rotation and fixation of the embryonic midgut. Abnormalities of this process create a spectrum of anomalies, which vary in their clinical significance from asymptomatic presentations to a twist of the midgut (volvulus) and abdominal catastrophe.

INTRODUCTION

Embryologic variation in midgut formation has long been recognized. Calder, in 1752, first described congenital duodenal obstruction due to malrotation. In 1898, Mall described the embryology of abnormal intestinal rotation. Dott, in 1923, correlated embryologic observations with clinical evidence and suggested a surgical treatment for duodenal obstruction. In his landmark article in 1932, Ladd described the surgical procedure that bears his name and remains the gold standard treatment for this condition.

The true incidence of malrotation of the midgut is difficult to determine. The autopsy prevalence may be as high as 0.5 to 1% of the total population, but the incidence in patients with clinical symptoms may be much less. The clinical presentation in the infant may occur in 1 in 6000 live births. Malrotation accounts for

Copyright 2004, Elsevier (USA). All rights reserved.

approximately 1 in 25,000 of all hospital admissions. Fifty to 75% of these patients will present with symptoms in the first month of life, with 90% of all patients presenting within the first year of life. Although volvulus and bowel ischemia present more frequently before 1 year of age, they can occur at any age. In the neonatal age group, males predominate in a 2 : 1 ratio, but this difference disappears with age. Sixty-two percent of patients with malrotation have associated gastrointestinal anomalies including intestinal atresia, duodenal web, Meckel's diverticulum, intussusception, Hirschsprung's disease, extrahepatic biliary system anomalies, and gastric volvulus.

EMBRYOLOGY AND DEVELOPMENTAL PATHOLOGY

The primitive gut is initially a tubular structure composed of endoderm. In the human, the intestinal segment that opens ventrally into the yolk sac is defined as the midgut (see Fig. 1). This portion will come to be supplied by the superior mesenteric artery (SMA), which serves as its axis, with the omphalomesenteric duct located at its apex. The primary loop rotates 180° counterclockwise such that the proximal "prearterial" loop lies posterior to the SMA. The cranial section of this limb will become the duodenum and its proximal position is fixed to the right of the SMA. The more distal portion of the prearterial loop becomes fixed to the left of the SMA and becomes the third and fourth portions of the duodenum, attached by the ligament of Treitz to the body wall to the left of the aorta, at or just above the level of the gastric outlet. Thus, the distal portion of the duodenum ultimately completes a 270° turn counterclockwise (Fig. 1).

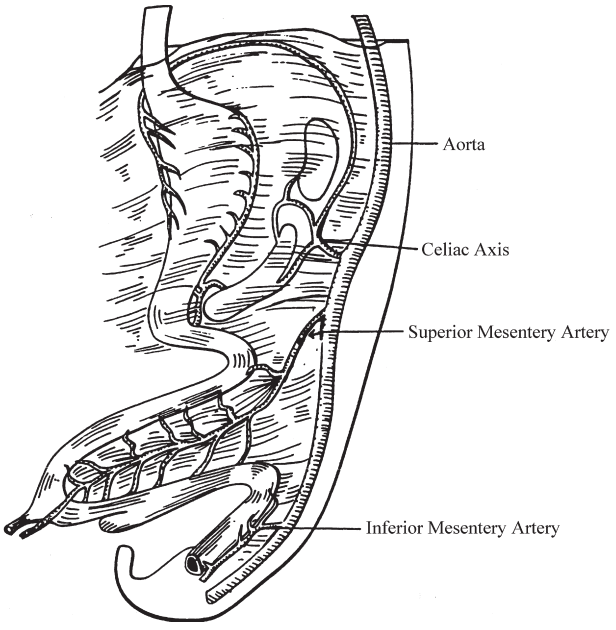


FIGURE 1 The primitive midgut grows to form an intestinal loop and its blood supply is from the superior mesenteric artery. Figure is from the personal collection of M. M. Ziegler.

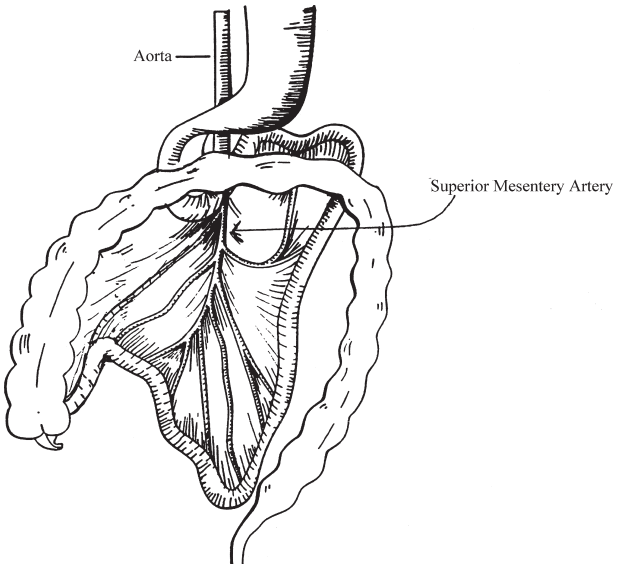


FIGURE 2 The correct anatomical location of the midgut after it rotates 270° counterclockwise around the axis of the superior mesenteric artery. Figure is from the personal collection of M. M. Ziegler.

The jejunum and proximal ileum undergo elongation to eventually form the primary loops of bowel at birth. The postarterial limb of the midgut becomes the terminal ileum, the cecum, and the proximal transverse colon. This segment also rotates 270° counterclockwise but does so anterior to the SMA.

Following the initial growth and rotation phases of the midgut outside of the fetal body stalk, the midgut retracts into the abdominal cavity. The duodenal–jejunal junction will come to lie posterior to the SMA, the small bowel will lie both to the right and to the left of midline, and the cecum and ascending colon will end up anterior to the SMA.

The final step in midgut positioning is fixation to the posterior body wall (see Fig. 2). After reduction into the left abdomen, the cecum descends and is fixed normally in the right iliac fossa. The intestinal mesentery is fixed in an oblique orientation from this point to the left upper quadrant where the ligament of Treitz secures the duodenum. This broad-based fixation is critical for the stability of the intestine and mesentery (Figs. 2 and 3). When considering rotational abnormalities, it is useful

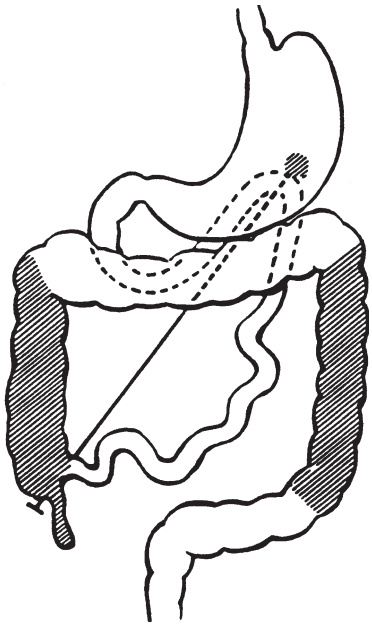


FIGURE 3 The normal attachment of the small bowel mesentery. Its broad base extends from the ligament of Treitz to the cecum. The ascending colon and descending colon are fixed retroperitoneally. Figure is from the personal collection of M. M. Ziegler.

to think of embryologic development in three main stages: herniation, rotation and return, and fixation. Herniation usually commences during the 5th to 10th weeks of gestation. Failure to go beyond this stage results in an omphalocele. Although not usually considered a true malrotation, the embryologic abnormality of intestinal orientation in this entity is a form of nonrotation.

The return of the intestinal contents to the abdomen is usually accomplished during the 10th to 12th weeks and associated abnormalities create a variety of pathologic entities. One form of rotational anomaly (nonrotation) occurs when the gut fails to rotate; that is, the gut turns less than 90° instead of the normal 270°. The distal cecocolic limb enters the abdomen first instead of last. Thus, nonrotation occurs when the duodenum descends inferiorly to the right of the SMA. The small bowel lies to the right and the colon lies on the left. The cecum is located at or near the midline and the small intestine lies to the right of midline. This condition may produce subtle symptoms and volvulus can occur. This variant has been reported in 0.2% of gastrointestinal contrast studies (Fig. 4).

Incomplete rotation is characterized by the arrest of normal rotation at 180° and may cause duodenal obstruction by either extrinsic compression or volvulus. The prearterial segment fails to complete the positioning posteriorly and to the left of the SMA. The cecum comes to lie in the upper abdomen usually to the left of the SMA and posterior wall attachment is accomplished by peritoneal bands, known as Ladd's bands, that can potentially cross and obstruct the duodenum. These bands are derived from embryonic dorsal mesogastrium and usually serve to secure the cecum and mesocolon to the posterior body wall (Fig. 5). As with nonrotation, the resulting narrow vascular pedicle provides an opportunity for volvulus to occur (Fig. 6).

Reverse or mixed rotations are less common forms of malrotation that can be caused by rotational arrest anywhere in midgut development. This creates a wide spectrum of anomalies that range from an incompletely descended cecum to more serious conditions that predispose to volvulus. For example, reverse rotation occurs when there is an abnormal 90° clockwise turn instead of the opposite 270° counterclockwise rotation. The distal limb enters the cavity first and rotates in a counterclockwise manner 180° about the SMA base. The duodenum will then lie in front of the transverse colon, producing a retroarterial tunnel, which may result in vascular or lymphatic occlusion (Fig. 7).

Paraduodenal or mesocolic hernias are a rare group of anomalies that occur when the small bowel is trapped behind the mesocolon as the colon attaches

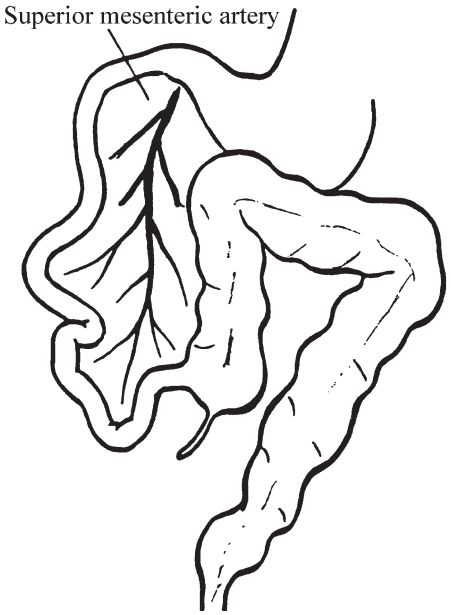


FIGURE 4 Nonrotation. Return of the postarterial segment first leaves the colon on the left and the small intestine on the right. Note that the ileum enters the cecum from the right side. Figure is from the personal collection of M. M. Ziegler.

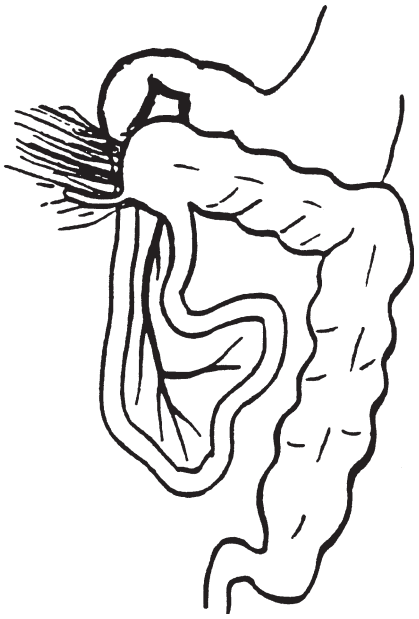


FIGURE 5 Malrotation. Duodenal obstruction can occur from Ladd's bands. Figure is from the personal collection of M. M. Ziegler.

to the retroperitoneum. Failure of proper mesocolic–posterior body wall fixation offers potential space on either side for bowel sequestration and entrapment. Right-sided hernias occur as the prearterial limb fails to rotate around the SMA and is entrapped by the mesocolon of the cecum and ascending colon. A left-sided hernia occurs when the prearterial limb rotates to the left and invaginates the mesocolon of the descending colon. The colon continues its normal positioning and the small bowel becomes encased in a mesenteric sac. As with any hernia, these situations present a potential risk for obstruction, incarceration, and strangulation of the intestine.

Fixation normally begins after the 12th week of gestation. Failure of this process results in a mobile or subhepatic cecum or a retrocecal appendix. The subsequent mesenteric mobility may increase the risk for volvulus.

PRESENTATION AND DIAGNOSIS

The symptoms of malrotation generally result from either partial duodenal obstruction or midgut volvulus. The manner of presentation usually follows an age distribution.

Presentation in the newborn period is usually acute. Bilious emesis is a cardinal manifestation of neonatal intestinal obstruction and demands that malrotation be included in the differential diagnosis. Less common symptoms include coffee ground emesis, distension, pain, and bloody stools. Duodenal obstruction is most often caused by extrinsic compression from Ladd's

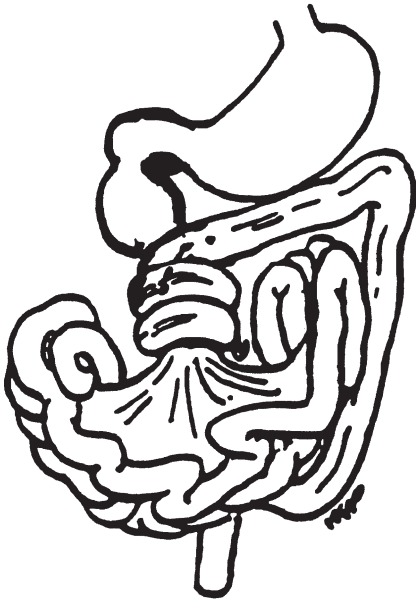


FIGURE 6 Malrotation. Duodenal obstruction can occur from volvulus. Figure is from the personal collection of M. M. Ziegler.

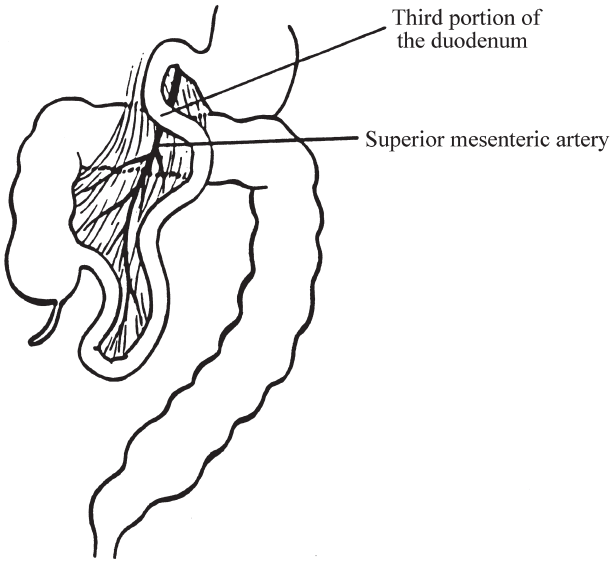


FIGURE 7 Reverse rotation. The duodenum is anterior to the superior mesenteric artery, which in turn is anterior to the transverse colon. Figure is from the personal collection of M. M. Ziegler.

bands. Ladd's bands cross anteriorly and laterally to the descending duodenum; thus, a postampullary site is the typical point of obstruction with characteristic bilious vomiting. There is a low but significant incidence of associated intrinsic duodenal obstruction, which may occur in the form of a web, stenosis, or true atresia. This is seen in 8–12% of patients who undergo operation for malrotation and it should be excluded by either careful inspection or by passage of a trans-lumenal balloon-tipped catheter. The physical examination may be normal in up to 50% of patients. Among older patients, presentations include chronic abdominal pain, weight loss or failure to thrive, chronic pancreatitis, or other relatively nonspecific gastrointestinal complaints. A number of patients present asymptomatically, in which the lesion is discovered incidentally.

The most pressing clinical concern is the presence of SMA pedicle torsion and volvulus. Midgut volvulus is the presentation in approximately half of the patients who come to operation for rotational abnormalities. Guiac-positive stool or hematochezia resulting from mucosal injury is a relatively common early finding. The patient may present *in extremis* and midgut volvulus has the tragic sequelae of bowel necrosis, sepsis, hypotension, acidosis, respiratory failure, and coagulopathy. The outcome for midgut volvulus is time-dependent, which is the reason that all suspected neonatal intestinal obstruction is aggressively diagnosed and treated.

For pediatric intestinal obstruction, the adjunct diagnostic algorithm generally begins with a plain abdominal radiograph. Classic early findings may reveal a nonspecific bowel gas pattern or may show gastric or duodenal distension. The "duodenal triangle," a triangular gas shadow in the right upper quadrant, thought to be the liver edge overlying the duodenum, has been described as a reliable radiologic sign for malrotation. Plain films, however, cannot rule out malrotation and may not distinguish between intrinsic and extrinsic duodenal compression.

The study of choice when malrotation is suspected is the upper gastrointestinal contrast series. Typically, it will show an abnormally located duodenojejunal junction; normally this junction lies to the left of the midline, at the level of the pylorus and fixated posteriorly. In malrotation, the junction is seen lower, more anterior, and in the midline or to the right. The presence of distal contrast will rule out duodenal atresia, but it does not differentiate malrotation from duodenal stenosis or an intrinsic web. Typically, in the presence of volvulus, narrowing of the duodenum will produce signs of torsion or compression, with image descriptors such as "bird beak," "corkscrew," or "coiled" used. Diagnostic uncertainty requires immediate operation.

The use of contrast enemas may define a high, abnormally located cecum and it may help in investigating more distal neonatal intestinal obstruction. It has limited utility in the diagnosis of malrotation.

Other modalities such as ultrasonography, computed tomography, and angiography have shown some efficacy in identifying axial relationships of the SMA pedicle. The typical finding is a reversed anatomic relationship of the SMA and SMV, which in malrotation cases would find the SMA in an aberrant anterior or left position. Angiographic findings of volvulus are described as a "barber pole" sign, due to the whirling of the mesenteric vessels and the dilated SMV with developed, dilated collateral vessels. As with plain films, a normal axial investigation does not rule out malrotation. These studies are helpful in diagnosis, but have no role in the acute setting, especially in the newborn, when time is of the essence. The most prudent diagnostic and therapeutic evaluation may be exploratory laparotomy.

TREATMENT

Neonates and infants who are diagnosed with malrotation require emergent operative management. Prompt evaluation, aggressive resuscitation, and timely preoperative preparation are essential for successful outcomes. The urgent nature of this disease must be stressed to the clinician. If the midgut is found to be unsalvageable at laparotomy, the survival rate falls to 50% and lifelong short bowel syndrome requiring parenteral nutrition is indicated.

The modern operative procedure is still based on the classic description by Ladd and proceeds as follows: abdominal entry is made through a wide transverse supraumbilical incision. Rapid exploration is then carried out with complete evisceration of the midgut. The usual finding at operation is either midgut volvulus or an aberrant location of the cecum and right colon. After the root of the mesentery is inspected, the volvulus, if present, is relieved by counterclockwise rotation of the affected segment (Fig. 8). The bowel may require a period of intraoperative warming and observation to assess viability. Ladd's bands stretching across the duodenum from the right upper quadrant are then divided (Fig. 9). To prevent recurrence, the base of the mesentery is broadened by dividing the peritoneal attachments of the cecum, small bowel mesentery, mesocolon, and duodenum (Fig. 10). The second, third, and fourth parts of the duodenum are then straightened and positioned along the right abdomen. The ileocecal junction is positioned in the left lower quadrant. If the bowel is deemed to be viable, intestinal contents are milked distally to

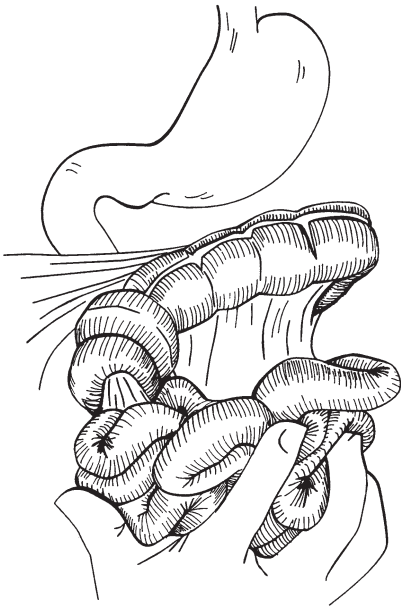


FIGURE 8 Counterclockwise detorsion of midgut volvulus. Figure is from the personal collection of M. M. Ziegler.

assess lumenal patency. An appendectomy is performed before the cecum is placed in the left lower quadrant (Fig. 11). Cecopexy has not been shown to offer any benefit.

Mesocolic hernias are addressed by dividing the lateral peritoneal attachments. Left hernias require mobilization of the inferior mesenteric vein and reduction and closure of the hernia sac. If properly performed, recurrence is rare. The overall incidence of adhesive small bowel obstruction following surgery approaches 10%.

If a limited segment of bowel is necrotic, resection and primary anastomosis is performed. In some cases, the severity of bowel edema prevents primary closure and silo or Gortex closure should be used. After appropriate resuscitation and hydration, a second-look laparotomy is performed in 36–48 h. In tragic situations where most or all of the midgut is necrotic, closure and family discussion should take place before a total enterectomy is performed.

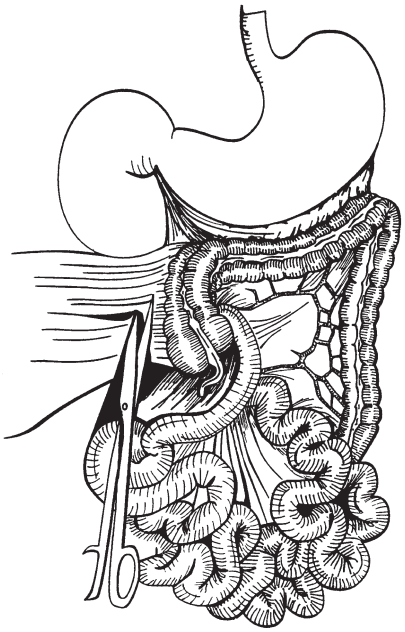


FIGURE 9 Incision of Ladd's bands is critical in order to relieve duodenal obstruction. Figure is from the personal collection of M. M. Ziegler.

OUTCOME AND FUTURE DIRECTION

In Ladd's original report, 8 of 35 patients died and there were no recurrences in the remaining patients (27 of 35). In current practice, the Ladd procedure is the therapy of choice and carries an overall mortality of 3–9%. The mortality rate increases with the presence of intestinal necrosis and prematurity. Eighteen percent of patients with short bowel syndrome had midgut volvulus and subsequent bowel resection as the etiology of their disease.

In patients found to have malrotation as an incidental finding on radiologic studies or at laparotomy for other reasons, a Ladd procedure should be performed. There is no way to accurately predict which patients are at risk for potentially catastrophic volvulus; therefore, early treatment minimizes morbidity and mortality.

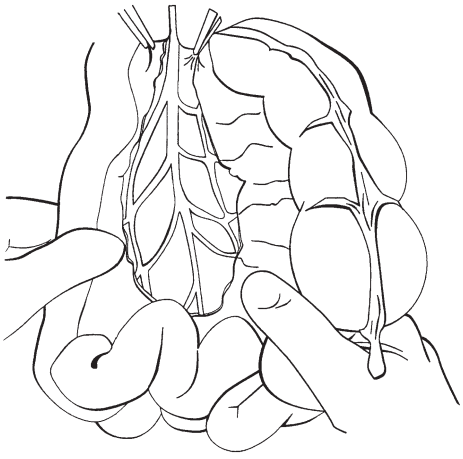


FIGURE 10 The peritoneal layers on the anterior surface of the small bowel mesentery are incised in order to broaden the base of the mesentery. Figure is from the personal collection of M. M. Ziegler.

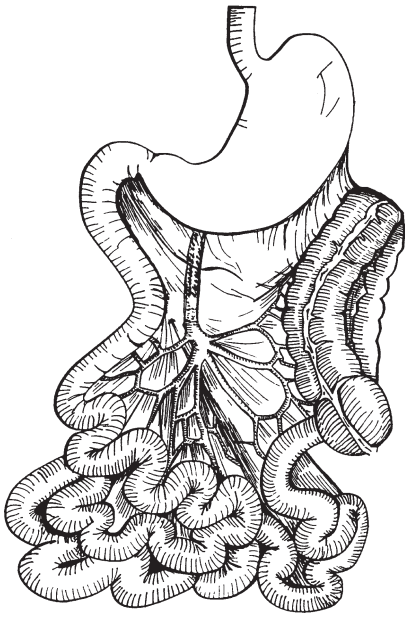


FIGURE 11 The operative findings at the end of a Ladd's procedure. Note that the duodenum is straightened out in the right flank. The cecum is now in the left lower quadrant and an appendectomy has been performed. Figure is from the personal collection of M. M. Ziegler.

In the past 10 years, several groups have shown the safety and efficacy of laparoscopic Ladd's procedures in cases of malrotation without volvulus. The advantage of this technique includes minimal morbidity from trocar sites compared with the generous supraumbilical incision. Return to full feeding can occur as soon as 24 h after surgery compared with several days following open surgery, leading to an overall decrease in hospital stay. There is general agreement that this approach should be limited to cases without volvulus due to the friability and delicate nature of the compromised bowel.

See Also the Following Articles

Duodenal Obstruction • Hernias • Neonatal Intestinal Obstruction • Short Bowel Syndrome • Volvulus

Further Reading

Bax, N. M., and van der Zee, D. C. (1998). Laparoscopic treatment of intestinal malrotation in children. *Surg. Endosc.* **12**, 1314–1316.

Clark, L. O. K. (2000). Malrotation. *In* "Pediatric Surgery" (K. W. Ashcraft, ed.), 3rd. Ed., pp. 425–434. W. B. Saunders, Philadelphia, PA.

Ford, E. G., Senac, M. O., Jr., Srikanth, M. S., and Weitzman, J. J. (1992). Malrotation of the intestine in children. *Ann. Surg.* **215**, 172–178.

Nair, R., and Hadley, G. P. (1996). Intestinal malrotation—Experience with 56 patients. *South African J. Surg.* **34**, 73–76.

Powell, D. M., Othersen, H. B., and Smith, C. D. (1989). Malrotation of the intestines in children: The effect of age on presentation and therapy. *J. Pediatr. Surg.* **24**, 777–780.

Torres, A. M., and Ziegler, M. M. (1993). Malrotation of the intestine. *World J. Surg.* **17**, 326–331.

Manometry

FERNANDO AZPIROZ
University Hospital Vall d'Hebron, Autonomous University of Barcelona

achalasia Ineffective esophageal peristalsis and defective relaxation of the lower esophageal sphincter in response to swallowing, due to depletion of inhibitory nerves in the esophageal smooth muscle.

diffuse esophageal spasm Disordered esophageal peristalsis; diagnosed by manometry; frequently associated with chest pain or dysphagia of unknown origin.

Hirschsprung's disease Congenital disorder characterized by severe constipation due to absence of neurones in the myenteric plexus of the distal gut.

hypertensive lower esophageal sphincter Manometric finding of uncertain clinical significance; frequently found in patients with noncardiac chest pain or nonorganic dysphagia without detectable underlying cause.

intestinal pseudo-obstruction Clinical symptoms, either recurrent or chronic, resembling a mechanical obstruction of the gut without demonstrable luminal compromise.

nutcracker esophagus Vigorous peristaltic contractions of the esophageal body; frequently associated with chest pain or other symptoms, such as depression, anxiety, and somatization, without a clear underlying cause and with uncertain clinical significance.

Gut manometry, the measurement of intraluminal pressures in a segment of the gut in order to study motor function. Pressures in the alimentary canal can be measured using an intraluminal tube, either with microsensors mounted over the tube or with a multilumen tube in which each lumen has a distal intraluminal opening and is proximally connected to an external pressure transducer.

ESOPHAGEAL MANOMETRY

Definition and Technique

Esophageal manometry is a well-established and standardized test that measures the motor activity of the esophagus and the sphincters during basal conditions and in response to swallowing. The esophagus is a muscular tube with sphincters on each end. It is separated from the pharynx by an upper sphincter with a basal contraction of about 40–120 mmHg, and is separated from the stomach by a lower sphincter with a basal contraction of about 10–45 mmHg above intragastric pressure. The body of the esophagus is intrathoracic and has negative intraluminal pressure. Pharyngeal contraction during swallowing is associated with a complete relaxation of both the upper and the lower esophageal sphincters, and once the bolus has entered the esophagus, a contraction of the upper sphincter is followed by a peristaltic contraction that migrates distally at a velocity of 2–4 cm/sec. The magnitude of the esophageal contraction increases distally (up to 40–130 mmHg in the lower third). Once the contraction reaches the lower sphincter, it recontracts.

Outcome Measures

Esophageal manometry has several possible outcomes:

1. Achalasia. The manometric diagnosis is based on 100% failed peristalsis (no peristaltic contractions), usually in the form of simultaneous contractions of the esophageal body in response to swallowing. Secondary criteria are incomplete or absent relaxation of the lower sphincter and positive basal pressure in the esophageal body due to retention.

2. Diffuse esophageal spasm. This disorder is defined by normal peristalsis alternating with more than 30% of the swallows with failed peristalsis, usually with simultaneous contractions. No characteristic lesion has been uniformly found in association with this entity.

3. Nutcracker esophagus. This disorder is defined by hypertensive but otherwise normal peristalsis (distal contractions >180 mmHg).

4. Hypertensive lower sphincter. This disorder is defined by increased basal pressure (>45 mmHg) with otherwise normal function.

5. Nonspecific motor disorders. These disorders are represented by a series of manometric findings that seem to occur more frequently in patients with unexplained symptoms, such as dysphagia or noncardiac chest pain, than in healthy subjects; their true clinical significance is not known. These disorders are defined by normal peristalsis plus one or more of the following symptoms: nontransmitted contractions ($>20\%$), prolonged contractions (>6 sec), triple-peak contractions, retrograde contractions, and weak contractions (<35 mmHg).

 Copyright 2004, Elsevier (USA). All rights reserved.

Indications

Suspected achalasia, suspected diffuse esophageal spasm, and symptoms of presumed esophageal origin without detectable abnormalities by conventional radiologic and endoscopic tests; however, the clinical significance of nonspecific motor disorders in this context is not clear. In cases involving systemic diseases, particularly connective tissue diseases, diabetes, and somatic myopathies, manometry may help to establish esophageal involvement by identifying weak peristalsis, weak lower sphincter, and/or nonspecific motor disorders. In patients with gastroesophageal reflux disease, manometry may be indicated before surgery to definitively rule out primary motor disorders and to assess peristaltic function, although the impact of the latter on treatment remains debatable. Because manometry locates the lower esophageal sphincter, it may be used to guide positioning of intraluminal devices, particularly electrodes for pH monitoring. Finally, abnormal esophageal motility may mimic various disorders, such as anorexia nervosa, that may be correctly identified by manometry.

INTESTINAL MANOMETRY

Definition and Technique

Intestinal manometry measures the motor activity of the upper small intestine either with or without simultaneous recording of antral activity. Evaluation of intestinal activity requires long observation periods ($\geq$5 hours) both during fasting and in response to ingestion of a meal. Intubation of the small bowel is usually guided by a steerable catheter or by endoscopy. Normally, during fasting, the upper gut exhibits a cyclical motility pattern of alternating periods of quiescence and activity. The periods of activity develop at variable intervals of about 100 min, initiate in the antrum, and migrate into the intestine at about 5–10 cm/min. After a meal, this cyclical motility pattern is interrupted and converted into a fed pattern, with regular antral activity at three contractions per minute and continuous irregular intestinal activity. The duration of the fed pattern depends on the type and volume of the meal.

Outcome Measures

Intestinal manometry has several possible outcomes:

1. Myopathic pattern. Impaired contractile capability of the intestinal wall produces low-amplitude contractions with otherwise preserved motor patterns. This type of pattern corresponds either to a myopathy (hollow visceral myopathy, myotonic dystrophy, or dermatomyositis) or to an infiltrative process (scleroderma or amyloidosis).

2. Neurophatic pattern. This pattern is characterized by contractions of normal strength but abnormal organization, due to an alteration of the neural mechanisms that control motility. The neuropathy may affect neural pathways at any level between the brain and the gut (brain tumor, autonomic neuropathy, or enteric neuropathy), without distinguishable manometric characteristics. The types of abnormalities are defined as abnormal migration or configuration of the activity periods during fasting (simultaneous appearance, retrograde propagation over 30 cm, or discrete elevation of the baseline of $>$30 mmHg for $>$3 min), bursts of nonpropagated activity (periods $>$2 min with waves $>$20 mmHg at 10–12/min registered in one or more consecutive manometric ports simultaneously), sustained uncoordinated phasic pressure activity ($>$30 min of intense and frequent contractions in one intestinal site uncoordinated with the rest of the gut), and failure to convert to a fed pattern.

3. Obstructive pattern. This is a form of intermittent repetitive activity that develops in the postprandial period, with clusters of waves or prolonged waves repeating rhythmically every 1–3 min. This pattern suggests a partial mechanical obstruction of the gut that may have been overlooked in the previous workup of the patient.

4. Antral hypomotility. Reduced amplitude or frequency of antral waves after a meal is frequent in patients complaining of symptoms without demonstrable cause. The significance of this pattern is doubtful, and hence, the indication for antral recording is debatable.

Indications and Clinical Value

Manometry facilitates management of patients with intestinal pseudo-obstruction by showing either a myopathy, a neuropathy, or an obstructive pattern. In patients with gastroparesis, i.e., gastric retention due to impaired gastric function, it is important to determine whether the intestine is also affected. The same applies to patients with colonic inertia, i.e., severe, intractable constipation due to impaired colonic motility. In patients with functional gastrointestinal disorders, i.e., symptoms such as diarrhea, bloating, or pain without demonstrable cause, intestinal manometry may be indicated if the symptoms are severe, to rule out intestinal neuropathy.

ANORECTAL MANOMETRY

Definition and General Technique

Anorectal manometry evaluates the function of the rectum and anal canal, which is maintenance of continence and controlled evacuation of fecal residues. This function is achieved by a combined action of the rectal reservoir function and regulated anal closure. Evaluation of anorectal function consists of the following battery of tests:

1. Anal sphincters. The anal canal is surrounded by two concentric muscular cylinders. The internal anal sphincter is a smooth muscle that exerts a continuous tonic contraction, thereby maintaining anal closure during basal conditions. The external anal sphincter is a peripheral cylinder of striated muscle that produces additional contraction at will during squeeze. Sphincteric function is evaluated by measuring the pressure at different levels of the anal canal: basal pressures largely reflect the activity of the internal sphincter and voluntary squeeze contractions reflect the activity of the external sphincter.

2. Reflexes. Sphincteric function is controlled by a relatively complex series of neural circuits. The internal sphincter, which can be considered a thickening of the circular muscle layer of the intestinal wall, is innervated by the myenteric plexus, which can be evaluated using the anorectal inhibitory reflex: rectal distension, e.g., by means of a balloon, normally produces a relaxation of the internal anal sphincter and hence a pressure drop in the anal canal. The external sphincter is controlled in part by spinal reflexes that can be evaluated using the cough reflex: an intraabdominal pressure increment, e.g., coughing, normally produces a reflex contraction of the external anal sphincter and hence a pressure increment in the anal canal.

3. Defecatory maneuver. Rectal evacuation, i.e., defecation, is normally produced by an abdominal compression associated with anal relaxation. This maneuver can be evaluated by asking the patient to attempt defecation while simultaneously measuring the pressure in an intrarectal balloon (intraabdominal pressure) and the pressures at different levels in the anal canal.

4. Rectal reservoir function. The rectal wall has a distensibility and a sensitivity that allow the rectum to fill and elicit conscious perception. This function is evaluated by metered rectal distension, using an intrarectal bag or balloon, and by measuring both the compliance (pressure at different distending volumes) and the intensity of conscious perception (for instance, the distending levels that induce first perception, urge to defecate, and discomfort).

Outcome Measures

Isolated dysfunctions may not have clinical relevance, and symptoms may occur only when multiple parameters are affected. The following parameters are usually evaluated:

1. Anal pressures. Decreased resting pressure indicates a weak or disrupted internal anal sphincter. Increased resting pressure may indicate internal anal hypertony, a key pathophysiological factor in anal fissure, or striated muscle spasm, frequently associated to anoperineal pain syndromes. Decreased voluntary squeeze pressure reflects impaired external sphincter contraction.

2. Reflexes. The presence of an anorectal inhibitory reflex rules out Hirschsprung's disease. Impaired cough reflex (anal pressure increment in response to cough smaller than the voluntary squeeze) indicates a defect in the sacropudendal reflex arc.

3. Defecatory maneuver. Impaired relaxation during attempted defecation (incomplete pressure drop) indicates a functional outlet obstruction, which is frequently associated with excessive straining (excessive abdominal compression). A weak abdominal compression, i.e., ineffective straining, may be due to a neuromuscular disease.

4. Rectal evaluation. Increased rectal capacity (megarectum) may be associated with functional outlet obstruction; decreased capacity may be associated with fecal incontinence. Increased rectal sensitivity may be associated with urge and fecal incontinence. Patients with severely reduced sensitivity may be at risk of incontinence due to impaired perception of rectal filling.

Indications

1. Constipation. Manometry can identify a functional outlet obstruction, and these patients may be taught to produce a normal defecatory maneuver. Suspected Hirschsprung's disease is excluded by a normal anorectal inhibitory reflex; if the reflex is repeatedly absent, a deep rectal wall biopsy should be done for histologic evaluation of the status of the enteric nervous system.

2. Anal incontinence. Any type of leakage, soiling, gas, liquid, or solid incontinence should be investigated, particularly in the absence of gross anatomical defects, because these patients may respond to functional treatment.

3. Pelvic floor disorders. Some cases may be related to functional outlet obstruction and perineal stress by excessive straining. In other cases, the disorder affects multiple systems, including the anorectum.

4. Diseases associated with neuromyopathies when anorectal affectation is suspected.

5. Pre- and postanorectal surgery, particularly if anorectal dysfunction is suspected.

6. Anoperineal pain or discomfort syndromes. Manometry may detect a neuromuscular dysfunction or a functional outlet obstruction with defecatory trauma, and these patients may respond to appropriate treatment.

See Also the Following Articles

Achalasia • Anal Sphincter • Barostat • Constipation • Defecation • Gastric Motility • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Pseudoobstruction

Further Reading

Azpiroz, F., Enck, P., and Whitehead, W. E. (2002). Anorectal functional testing. Review of collective experience. *Am. J. Gastroenterol.* **97**, 232–240.

Diamant, N. F., Kahn, M. A., Wald, A., and Whitehead, W. E. (1999). American Gastroenterological Association technical review on anorectal testing techniques. *Gastroenterology* **116**, 735–760.

Kahrilas, P. J., Clouse, R. E., and Hogan, W. J. (1994). American Gastroenterological Association technical review on the clinical use of esophageal manometry. *Gastroenterology* **107**, 1865–1884.

Katz, O. P. (1992). Disorders of increased esophageal contractivity. *In* "The Esophagus" (D. O. Castell, ed.), pp. 261–275.

Malagelada, J. R., Camilleri, M., and Stanghellini, V. (1986). "Manometric Diagnosis of Gastrointestinal Motility Disorders." Thieme Stratton, New York.

Marginal Ulcer

RICHARD H. TURNAGE, LESTER W. JOHNSON, AND QUYEN CHU

Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center

marginal ulcers Also termed stomal or anastomotic ulcers; those peptic ulcers that occur at or near a surgically created gastroduodenostomy or gastrojejunostomy.

Marginal ulcers (also known as stomal or anastomotic ulcers) are peptic ulcers that occur at or near a surgically created gastroduodenostomy or gastrojejunostomy. The original operative procedure may have been performed for the treatment of peptic ulcer disease or benign or malignant gastric, duodenal, or pancreatic diseases. This article discusses the etiology, clinical presentation, diagnostic strategy, and management of patients with marginal ulcers.

INCIDENCE AND TIME TO OCCURRENCE

The incidence of marginal ulcers in patients undergoing vagotomy, resection, and gastroenteric anastomoses for peptic ulcer disease is less than 5%. In these patients, ulcer recurrence occurs most often within the first 3 to 5 postoperative years. Marginal ulcers following gastroenterostomy for diseases other than peptic ulcer disease may occur many years following the original procedure. Marginal ulcers have also been reported in as many as 8% of morbidly obese patients undergoing gastric bypass as an adjunct to weight loss.

ETIOLOGY AND PATHOGENESIS

The most common causes of marginal ulcers are (1) incomplete gastric acid suppression (e.g., incomplete vagotomy or absence or failure of appropriate ulcer prophylaxis with histamine 2 receptor antagonists (H2-RA) or proton pump inhibitors (PPIs) and (2) ulcerogenic medications, particularly aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs). Less common causes include hypersecretory states, such as gastrinoma and retained/excluded gastric antrum. In

Copyright 2004, Elsevier (USA). All rights reserved.

4. Diseases associated with neuromyopathies when anorectal affectation is suspected.

5. Pre- and postanorectal surgery, particularly if anorectal dysfunction is suspected.

6. Anoperineal pain or discomfort syndromes. Manometry may detect a neuromuscular dysfunction or a functional outlet obstruction with defecatory trauma, and these patients may respond to appropriate treatment.

See Also the Following Articles

Achalasia • Anal Sphincter • Barostat • Constipation • Defecation • Gastric Motility • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Pseudoobstruction

Further Reading

Azpiroz, F., Enck, P., and Whitehead, W. E. (2002). Anorectal functional testing. Review of collective experience. *Am. J. Gastroenterol.* **97**, 232–240.

Diamant, N. F., Kahn, M. A., Wald, A., and Whitehead, W. E. (1999). American Gastroenterological Association technical review on anorectal testing techniques. *Gastroenterology* **116**, 735–760.

Kahrilas, P. J., Clouse, R. E., and Hogan, W. J. (1994). American Gastroenterological Association technical review on the clinical use of esophageal manometry. *Gastroenterology* **107**, 1865–1884.

Katz, O. P. (1992). Disorders of increased esophageal contractivity. *In* "The Esophagus" (D. O. Castell, ed.), pp. 261–275.

Malagelada, J. R., Camilleri, M., and Stanghellini, V. (1986). "Manometric Diagnosis of Gastrointestinal Motility Disorders." Thieme Stratton, New York.

Marginal Ulcer

RICHARD H. TURNAGE, LESTER W. JOHNSON, AND QUYEN CHU

Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center

marginal ulcers Also termed stomal or anastomotic ulcers; those peptic ulcers that occur at or near a surgically created gastroduodenostomy or gastrojejunostomy.

Marginal ulcers (also known as stomal or anastomotic ulcers) are peptic ulcers that occur at or near a surgically created gastroduodenostomy or gastrojejunostomy. The original operative procedure may have been performed for the treatment of peptic ulcer disease or benign or malignant gastric, duodenal, or pancreatic diseases. This article discusses the etiology, clinical presentation, diagnostic strategy, and management of patients with marginal ulcers.

INCIDENCE AND TIME TO OCCURRENCE

The incidence of marginal ulcers in patients undergoing vagotomy, resection, and gastroenteric anastomoses for peptic ulcer disease is less than 5%. In these patients, ulcer recurrence occurs most often within the first 3 to 5 postoperative years. Marginal ulcers following gastroenterostomy for diseases other than peptic ulcer disease may occur many years following the original procedure. Marginal ulcers have also been reported in as many as 8% of morbidly obese patients undergoing gastric bypass as an adjunct to weight loss.

ETIOLOGY AND PATHOGENESIS

The most common causes of marginal ulcers are (1) incomplete gastric acid suppression (e.g., incomplete vagotomy or absence or failure of appropriate ulcer prophylaxis with histamine 2 receptor antagonists (H2-RA) or proton pump inhibitors (PPIs) and (2) ulcerogenic medications, particularly aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs). Less common causes include hypersecretory states, such as gastrinoma and retained/excluded gastric antrum. In

Copyright 2004, Elsevier (USA). All rights reserved.

contrast to the well-known pathogenic role that *Helicobacter pylori* plays in primary ulcer disease, the relationship between *H. pylori* infection and marginal ulcer formation is unclear.

CLINICAL PRESENTATION

The most frequent signs and symptoms of marginal ulceration are abdominal pain or evidence of a complication such as hemorrhage, perforation, or obstruction. Schirmer *et al*. found that 41% of their patients with marginal ulcer presented with hemorrhage, a rate twice that of patients initially presenting with peptic ulcers. Although perforation into the peritoneal cavity is uncommon in patients with marginal ulcers due to inadequate inhibition of gastric acid secretion, recent studies have found that marginal ulcers due to aspirin, and perhaps other NSAIDs, are associated with severe complications including obstruction, bleeding, and even perforation.

DIAGNOSTIC EVALUATION

History and physical examination will identify symptoms and signs of ulcer disease or evidence of complications, (e.g., hematemesis and melena, anemia, vomiting, or weight loss). The history may also provide evidence of aspirin or NSAID use. In the absence of perforation, the diagnosis of marginal ulceration is best made by esophagogastroduodenoscopy, which will document the presence and location of the ulcer and the presence of *H. pylori*. Multiple ulcers and those in atypical locations strongly suggest a hypersecretory syndrome or the use of NSAIDs. Retained food, a gastric bezoar, or a stenotic anastomosis suggests associated impairment of gastric emptying. Endoscopic biopsy of gastric ulcers is of particular importance in detecting cancer in these patients who are at a greater risk than the general population in developing gastric adenocarcinoma. Upper gastrointestinal endoscopy is invaluable in the evaluation and treatment of patients bleeding from a marginal ulcer.

Barium contrast studies of the stomach are of limited value in diagnosing marginal ulcers because of the difficulty in detecting ulcerations in the presence of postoperative deformities of the stomach, duodenum, and small intestine. Contrast radiographs do, however, provide important information in the evaluation of patients with suspected gastric outlet obstruction and in those whose postoperative anatomy is uncertain.

A fasting serum gastrin level should be measured in all patients with marginal ulcers to identify patients with gastrin-mediated hypersecretory syndromes. A normal fasting serum gastrin concentration essentially excludes the possibility of Zollinger-Ellison syndrome, whereas a serum gastrin level greater than 1000 pg/ml, combined with a fasting gastric pH < 2.5, is virtually diagnostic. Patients with lesser degrees of hypergastrinemia require a secretin stimulation test to confirm the diagnosis of gastrinoma.

MEDICAL MANAGEMENT

In the setting in which a patient presents with upper abdominal pain and endoscopy demonstrates a marginal ulcer, appropriate medical therapy consists of a PPI or H2-RA and antibiotics directed at *H. pylori*, if present. The optimal duration of treatment with a PPI or H2-RA is unknown although most evidence supports long-term maintenance PPI or H2-RA therapy. Although no data directly support the eradication of *H. pylori* in patients with postoperative ulcer recurrence, the critical role this infection plays in primary peptic ulcer disease (PUD), the possibility that *H. pylori* promotes NSAID-induced mucosal injury, and the minimal risks associated with *H. pylori* eradication strongly support the use of antibiotic therapy when *H. pylori* is present.

Patients should be carefully counseled regarding the important relationship between NSAID use (including over-the-counter preparations) and ulcer disease, and the necessity of discontinuing these medications. NSAID use, including low-dose aspirin, should be discontinued during the period of ulcer healing, if feasible. Once ulcer healing has been documented endoscopically, non-NSAID analgesic agents should be substituted for the NSAIDs and if these are ineffective, then nonacetylated NSAIDs or a cyclooxygenase 2-specific NSAID should be considered. Consideration should be given for substituting clopidogrel for low-dose aspirin for cardiovascular prophylaxis. If patients are to be continued on low-dose aspirin, they should receive ulcer prophylaxis with misoprostol or a PPI.

SURGICAL MANAGEMENT

The adequacy of prior medical management and the elimination of risk factors for ulcer recurrence (e.g., NSAIDs) are important considerations in determining whether a patient is an appropriate candidate for operative therapy. The choice of a particular operation for a patient with a marginal ulcer should be that with the lowest recurrence risk that the patient is medically able to tolerate and tailored according to the adequacy of a prior vagotomy and drainage procedure, medical comorbidities, presence of complications of the ulcer,

and the extent of the prior resection. Non-acid-reducing procedures, such as local ulcer excision, revision of drainage procedures alone, and closure of perforations, are associated with extremely high failure rates.

The operative approach to patients with marginal ulcer formation after a vagotomy and antrectomy or a distal gastrectomy is partial gastric resection. Recurrent ulcer disease following vagotomy and antrectomy or partial gastrectomy strongly suggests NSAID use or a hypersecretory state. Several recent reports have demonstrated the importance of aspirin abuse in the etiology of such virulent ulcer disease and very high rates of ulcer recurrence following even very aggressive operative intervention. In this setting, further gastric resection may necessitate subtotal or even total gastrectomy, procedures associated with severe postgastrectomy symptoms in as many as 20–40% of patients.

Patients with marginal ulcer and an incomplete vagotomy may benefit from a thoracoscopic truncal vagotomy so long as NSAID use has been eliminated and there is no evidence of gastric stasis or gastric outlet obstruction. This approach avoids a difficult dissection in the upper abdomen and eliminates a likely cause of ulcer recurrence. In those instances of ulcer recurrence accompanied by gastric outlet obstruction, vagotomy and revision of the anastomosis are necessary to relieve the mechanical obstruction and further reduce gastric acid secretion. Also, patients who underwent resection for gastric ulcers appear to have a threefold increase in the risk of gastric adenocarcinoma when followed for more than 25 years following a Billroth II reconstruction. Clearly, an aggressive diagnostic approach including endoscopic biopsy is necessary in patients with gastric ulcers.

The frequency of postoperative ulcer recurrence varies from 8 to 22% depending upon the remedial procedure performed and the persistence of aspirin, and probably NSAID, use. The highest rates of recurrent ulcer formation occur after repeat vagotomy alone, revision of a gastroenterostomy alone, or any other non-acid-reducing operation. Lower rates are associated with re-vagotomy and antrectomy or distal gastrectomy. Patients who continue to take aspirin following the operative management of their marginal ulcer are at particularly high risk of recurrence (80–90%) regardless of the operative procedure performed. The virulence of the ulcer disease is suggested by the high rate of gastric outlet obstruction, perforation, and death in these patients. The mortality rate of operations for recurrent ulcer disease is 2–4%. The presence of complications of ulcer disease necessitating an emergent operation appears to significantly increase the risk of operative mortality. The chronic postoperative complications of both primary and remedial acid-reducing operations include diarrhea, dumping, chronic abdominal pain syndromes, impaired gastric emptying, osteoporosis, and anemia. Most authors report that 60 to 80% of patients who undergo remedial operations for recurrent ulcer disease will have either no or only occasional mild symptoms, whereas 20 to 40% will have severe or incapacitating symptoms including ulcer recurrence. Excluding ulcer recurrence, approximately 10–20% of patients undergoing remedial operative procedures will have chronic symptoms that significantly impact the patients' life.

See Also the Following Articles

Duodenal Ulcer • Gastric Acid Secretion • Gastric Outlet Obstruction • Gastric Surgery • Gastric Ulcer • Receptor Antagonists • *Helicobacter pylori* • NSAID Induced Injury • Proton Pump Inhibitors

Further Reading

Hirschowitz, B. I., and Lanas, A. (1998). Intractable upper gastrointestinal ulceration due to aspirin in patients who have undergone surgery for peptic ulcer. *Gastroenterology* **114**, 883–892.

Jordan, P. H., Jr., and Thornby, J. (1994). Twenty years after parietal cell vagotomy or selective vagotomy antrectomy for treatment of duodenal ulcer: Final report. *Ann. Surg.* **220**, 283–296.

Maclean, L. D., Rhode, B., Nohr, C., Katz, S., and McClean, P. H. (1997). Stomal ulcer after gastric bypass. *J. Am. College Surg.* **185**, 1–7.

McFadden, D. W., and Zinner, M. J. (1991). Reoperation for recurrent peptic ulcer disease. *Surg. Clin. North Am.* **71**, 77–92.

Schirmer, B. D., Meyers, W. C., Hanks, J. B., Kortz, W. J., Jones, R. S., and Postlethwait, R. W. (1982). Marginal ulcer: A difficult problem. *Ann. Surg.* **195**, 653–661.

Stabile, B. E., and Passaro, E., Jr. (1976). Recurrent peptic ulcer. *Gastroenterology* **70**, 124–135.

Thirlby, R. C. (1994). Postoperative recurrent ulcer. *Gastroenterol. Clin. North Am.* **23**, 295–311.

Toftgaard, C. (1989). Gastric cancer after peptic ulcer surgery: A historic prospective cohort investigation. *Ann. Surg.* **210**, 159–164.

Turnage, R. H., Sarosi, G., Cryer, B., Spechler, S., Peterson, W., and Feldman, M. The evaluation and management of patients with recurrent ulcer disease following acid-reducing operations: A systematic review. *J. Gastrointest. Surg.* in press.

Mast Cells

ALISON P. WATT AND MADELEINE ENNIS
Queens University of Belfast, United Kingdom

anaphylatoxin A substance (e.g., C5a, C3a) that is capable of directly inducing mast cell mediator release.
anaphylaxis Hypersensitivity reaction triggered by immunological or anaphylatoxin-mediated mast cell mediator release, resulting in vasodilation and smooth muscle contraction. Anaphylactic shock is often fatal.
Fc (fragment crystallizable) Non-antigen-binding fragment of an immunoglobulin molecule.
FcεRi High-affinity receptor for immunoglobulin E. Fc receptors are cell surface receptors that bind the Fc portion of a particular immunoglobulin.
heterogeneity Generally mast cells exhibit significant variation in biochemical, histochemical, and functional aspects of their phenotype with varying tissue location.
histamine Vasoactive amine found in mast cells and basophils. On release, it induces smooth muscle contraction and increased vascular permeability.
immunoglobulin E One of the five major classes of antibodies. Immunoglobulins are globular proteins produced by B cells. They are designed to recognize and bind to specific molecules known as antigens. IgE is associated with allergic reactions.
metachromasia Result obtained when a stain is applied to cells or tissues and gives a color different from that of the stain solution.

Mast cells are mononuclear cells distributed ubiquitously throughout the body, particularly in submucosal tissues that interface with the external environment. These include the gastrointestinal, genitourinary, and respiratory tracts. Mast cells may also be found in close association with blood vessels and peripheral nerves. Mast cells do not represent a homogenous population. Evidence of morphological, histochemical, and functional differences has been observed in different anatomical locations, suggesting that differentiation of mast cells is not complete until they reach their target tissue, where microenvironmental factors are crucial in their development and maturation. Activation of mast cells results in the release of both preformed and newly synthesized mediators, the best known being histamine. Hence, the mast cell is considered, along with the basophil, to be a main effector of acute allergic and inflammatory responses. However, the mast cell is now implicated in a diversity of other biological functions, which include angiogenesis, chronic inflammation, and immunological disorders. Mast cell activity may be modulated by a variety of endogenous and exogenous factors.

HISTORY OF THE MAST CELL

In 1863, Friedrich von Recklinghausen first observed the mast cell. Fifteen years later, a German medical student, Paul Ehrlich, described a cell with cytoplasmic granules, which took up blue aniline dyes in histological sections, but stained a pinkish purple color. Ehrlich called this process metachromasia; he believed that these large cytoplasmic granules represented phagocytosed material and hence named these cells "*Mastzellen*," meaning well-fed cells.

The first important discovery regarding the mast cell granule was in 1937 when Holmgren and Wilander established the presence of heparin in mast cell granules by correlating the heparin concentration in tissues with the number of mast cells present. The mystery of metachromasia was also solved around this period, when it was demonstrated that sulfated glycosaminoglycan heparin changed the color of blue aniline dyes to a reddish purple color.

Richet and Portier first described anaphylaxis in 1902. However, a link between the mast cell and anaphylaxis was not made until 1941, when Jacques and Walters proved that anticoagulation in dogs undergoing anaphylactic reactions was due to the release of heparin. In 1927, the identification of histamine in normal tissues was carried out by Best and co-workers. The first important experiments designed to show an actual release of histamine in association with anaphylactic shock were those of Watanbe, showing differences in the histamine content of guinea pig lung before and after shock. In fact, it was not until 1953 that Riley and West identified mast cells as the major source of tissue histamine after they observed a striking correlation between the histamine content of a tissue and the number of mast cells present.

In the early 1970s, Ishizaka and Ishizaka demonstrated that immunoglobulin E (IgE) induced histamine

 Copyright 2004, Elsevier (USA). All rights reserved.

release from mast cells. Injection of a purified ragweed IgE preparation into normal skin resulted in sensitization of the skin, which disappeared on removal of the IgE antibody. It is now well established that IgE molecules have a high affinity for specific IgE receptors expressed on the mast cell surface and that cross-linkage of these receptors results in mast cell mediator release, thus establishing a central role for mast cells in IgE-mediated hypersensitivity.

ORIGIN, DISTRIBUTION, AND DEVELOPMENT OF MAST CELLS

Origin and Distribution

Mast cells are widely distributed in vertebrates, where they are found within organs, especially in connective tissues and in areas interfacing the external environment, such as the skin, lungs, and gastrointestinal tract. Transplantation studies with beige (C57BL-Bg^J/Bg^J) mice provided the first evidence that mast cells are of hematopoietic origin. Beige mice possess two mutant alleles at the Bg locus and are characterized by the presence of mast cells with giant granules. Bone marrow cells from beige mice injected into normal irradiated (C57BL- +/+) recipients revealed typical mast cells containing giant granules developing in the recipient within 2–3 months of bone marrow transplantation.

Further studies using mast cell-deficient mice provided new insights into the origin and differentiation of mast cells. Mast cell-deficient mice have a double dose of mutant alleles at either the W (W/W^v) or the Sl (Sl/Sl^d) locus. These mice manifest macrocytic anemia, sterility, and lack of hair pigmentation. In addition, they are virtually devoid of mature morphologically identifiable mast cells in all organs and anatomical sites examined.

Transplantation of bone marrow from normal littermates or beige mice into W/W^v mice corrected the mast cell deficiency, suggesting that the defect in mast cell development was in the early progenitor cells of mast cell lineage. However, bone marrow transplantation did not correct mast cell deficiency in Sl/Sl^d mice, reflecting an abnormality in microenvironmental factors that promote mast cell differentiation and maturation in the tissue.

Subsequent studies demonstrated that mast cells originate from multipotential hematopoietic stem cells as demonstrated by the formation of other hematopoietic cells in addition to mast cells from spleen colony-forming units. In sharp contrast to other cells of hematopoietic stem cell lineage, mast cells do not circulate as mature mast cells. Morphologically identifiable mast cell precursors, originating from pluripotential $CD34^+$c-kit^+ CD13 cells in humans, circulate in the blood and lymphatics before invading connective and mucosal tissues. Once in the tissues, these precursors gain their mature morphologic and functional characteristics under the influence of local microenvironmental factors.

Stem Cell Factor and the c-kit Ligand

Stem cell factor (SCF) is a glycosylated protein with a molecular weight of 2–35 kDa; the main sources of SCF *in vivo* are fibroblasts and bone marrow stromal cells. SCF exists in membrane-bound and soluble forms, both of which are biologically active. Molecular analysis of the W locus revealed that it encoded the tyrosine kinase receptor c-kit. SCF exerts its effects by binding to the c-kit receptor. The c-kit receptor consists of an extracellular domain with five immunoglobulin-like repeats and a tyrosine kinase split into two domains by an insert sequence. This structure is well preserved in both rats and humans.

Mast cell deficiency in W/W^v mice is now known as a deficiency of the c-kit receptor, whereas mast cell deficiency in Sl/Sl^d mice is due to an inability to produce functional SCF. The work of Kitamura and colleagues with mast cell-deficient mice clearly demonstrates the critical importance of SCF/c-kit interactions. In addition to promoting the proliferation and development of mast cells, SCF has been implicated in the process of chemotaxis, adhesion, survival, and the synthesis mediators. Thus, SCF is a multifunctional mast cell growth factor.

MAST CELL HETEROGENEITY

Histochemical Characteristics

Mast cells from different tissues exhibit heterogeneity in terms of their morphological, biochemical, and functional properties. Enerbäck reported differences in histochemical and secretory properties between rodent mast cells of the intestinal mucosa and rodent skin mast cells. He discovered that conditions of fixation and histochemical staining could discriminate between mast cell subtypes. The classification of rodent mast cell subtypes is based on the phenotypical differences between connective tissue mast cells (CTMC) of the skin, peritoneal cavity, and other sites in the rats and mucosal mast cells (MMC) observed in the intestinal lamina propria. By analogy to rodent mast cells, variations in fixation properties provided the first documented evidence of the existence of different types of human mast cells. Human intestinal MMC display metachromatic

staining after fixation in Carnoy's solution but not when fixed in neutral buffered formalin, whereas CTMC found in the intestinal submucosa retain metachromatic staining properties regardless of the fixative used. The variation in dye-binding properties observed between the two types of mast cell reflects the different proteoglycans in the two subpopulations. Metachromatic staining occurs as a result of interaction of the cationic dyes with the anionic proteoglycan matrix. The different charges of the proteoglycans account for the diverse staining characteristics of MMC and CTMC. Glycosaminoglycan chondroitin sulfate found in MMC is less sulfated and has a lower charge than the highly sulfated heparin, which is the major proteoglycan in the CTMC. However, subpopulations of mast cells are not confined to particular anatomical regions; consequently, it is inaccurate to designate human mast cells as "mucosal mast cells" or "connective tissue mast cells" on the basis of the cells' sensitivity to formalin.

Biochemical Characteristics

The two principal types of human mast cells are also distinguished by the observation of distinct protease patterns in mast cells in different anatomical locations. Mast cells containing only tryptase are named MC_T; these cells predominate in the alveolar septa of the lung and in the intestinal mucosa. Other human mast cells contain measurable levels of tryptase, chymase, a cathepsin G-like protease, and carboxypeptidase; these cells are designated MC_{TC} for their tryptase and chymase and predominate in the skin and intestinal submucosa. However, all of the tissues mentioned contain representatives of both mast cell subpopulations and therefore cannot be classified as MC_T or MC_{TC} on the basis of tissue locality alone. There have also been isolated reports of mast cells expressing chymase but not tryptase (MC_C).

Heterogeneity in rats may also be distinguished by protease content. Rat CTMC contain the chymotrypsin-like at mast cell protease I (RMCP I) and carboxypeptidase A, whereas rat MMC contain another chymotrypsin-like neutral protease, RMCP II. The tissue localization of human MC_T corresponds most closely to that of rodent MMC, whereas the tissue localization of MC_{TC} corresponds most closely to that of rodent CTMC.

Functional Characteristics

Variation in functional characteristics is also observed between mast cell subpopulations. This is demonstrated by varying sensitivity to drugs and secretagogues. For example, human skin mast cells have been demonstrated to respond to stimulation with the neuropeptide substance P but not to inhibition by the anti-allergy drug disodium cromoglycate. In contrast, mast cells isolated from the lung were insensitive to substance P, yet histamine release was inhibited by disodium cromoglycate.

Similarly, in the rat, there appears to be a correlation between mast cell protease phenotype and histamine release in response to stimulation with compound 48/80 or substance P and between the phenotype and susceptibility to inhibition by disodium cromoglycate. Rat CTMC containing RMCP I are sensitive to compound 48/80 and substance P and also to inhibition by disodium cromoglycate, whereas rat MMC containing RMCP II are insensitive to these compounds.

Origins of Heterogeneity

Factors influencing the development of heterogeneous populations of mast cells have been the subject of extensive investigation. The failure to detect granulated mast cells in the circulation under normal conditions, coupled with the characterization of progenitor cells for both murine and human mast cells, supports the hypothesis that the mast cell does not establish functional and phenotypic maturity until it reaches its resident tissue or microenvironment. Mast cell phenotype is considered to be regulated by microenvironmental factors from a variety of sources. Mast cells are in contact with a number of immune cells, which may synthesize and secrete mediators such as cytokines, chemokines, growth factors, and histamine-releasing agents, all of which may affect the mast cell population. The mast cell itself, through cytokine production, may participate in the regulation of its own numbers and phenotype, particularly during immunological or disease processes.

MAST CELL ACTIVATION

Activation of mast cells may occur via two possible mechanisms: immunological and nonimmunological.

Immunological Activation

The presence of a multivalent allergen stimulates the production of specific IgE antibodies. These antibodies bind with high affinity to FcεRi receptors on the mast cell surface. Following cross-linkage of antigen-bound IgE–FcεRi complex, receptor aggregation occurs and mast cell exocytosis is triggered (Fig. 1).

The FcεRI receptor is a tetrameric member of the antigen receptor family, consisting of one α-chain (45 kDa), one β-chain (33 kDa), and two γ-chain subunits (9 kDa). The α-chain contains one

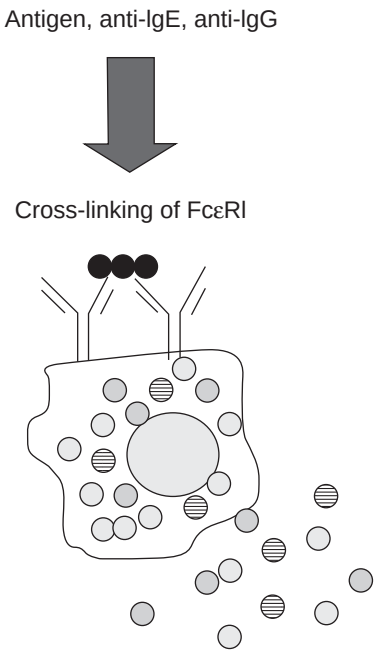


FIGURE 1 Cross-linking of the FcεRI by a specific antigen, IgE or IgG, results in mast cell degranulation.

membrane-spanning region and a single IgE-binding site and does not contribute to the signal transduction properties of the receptor. The β-chain has four membrane-spanning domains, with its amino- and carboxy-termini protruding into the cell cytoplasm, and the γ-chains have one membrane-spanning domain and a long cytoplasmic tail but no extracellular amino-terminal. The γ-chain is thought to belong to the same gene/protein family as the ζ-chain of the T-cell receptor and so the two probably share a common role in signaling.

Nonimmunological Activation

Many other substances and conditions are capable of triggering mast cell secretion. These nonimmunological stimuli can be subdivided into two classes based on whether the mode is cytotoxic or noncytotoxic. Cytotoxic compounds such as the detergent Tween 20 disrupt the plasma membrane, thereby releasing intracellular contents. Noncytotoxic compounds stimulate mediator secretion while the plasma membrane remains intact. Compound 48/80 and the calcium ionophores A23187 and ionomycin are three such known substances. Numerous physiological compounds have been observed to induce mast cell secretion *in vitro*. These include growth factors, cytokines, neuropeptides, and anaphylatoxins. Cold, pressure, stress, and free radicals are also known to initiate mast cell mediator secretion.

MAST CELL MEDIATORS

On activation, mast cells secrete a diverse array of potent biologically active mediators, some of which are stored preformed in the cells' cytoplasmic granules and others of which are synthesized on appropriate cell activation.

Preformed Mediators

Biogenic Amines

Histamine is the single amine known to be stored in human mast cells. Mast cells of other species are known to store additional amines; for example, the rat mast cell stores serotonin. Histamine is formed in the Golgi apparatus of mast cells and basophils during decarboxylation of histidine by the pyridoxal phosphate-dependent enzyme, histidine decarboxylase (Fig. 2). Histamine is then stored by ionic linkage with the carboxyl groups of proteins and proteoglycans of the secretory granules at an acidic pH. During degranulation, histamine dissociates from the proteoglycan–protein complex by cation exchange for extracellular sodium at neutral pH. Rat peritoneal mast cells contain 10–30 pg of histamine per cell, whereas rat mucosal cells contain only 1–3 pg of histamine per cell. In mast cells isolated from human lung, skin, lymphoid tissue, and small intestine, approximately 3–8 pg of histamine per cell has been reported. On release, histamine is rapidly metabolized into methylhistamine, methylimidazole acetic acid, or imidazole acetic acid. For this reason, it is likely that histamine usually influences events at or close to the site of its release. Specific histamine receptors, of which there are four classes, H2, H2, H3, and H4, mediate the wide-ranging biological activities of histamine. One or more of these receptor subtypes are expressed on many different tissues and cell types, including smooth muscle of the respiratory tract, cardiac atrium, uterus, T and B cells, monocytes, neutrophils, glial cells, and nerve cells. H1 receptor signaling differs from H2

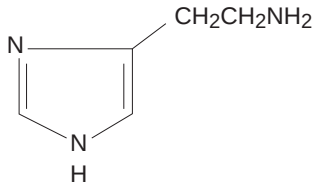


FIGURE 2 Structure of histamine.

signaling; in the H1 pathway, enhanced breakdown of phosphatidylinositol results in the mobilization of intracellular Ca^{2+}, whereas in H2 signaling, the second messenger is cyclic AMP. Although the H1 and H2 receptors were cloned a decade ago, cloning of the H3 receptor did not occur until 1999 and this elucidation quickly led to the discovery of H3 receptor subtypes. Molecular studies have identified $H3_A$, $H3_B$, and $H3_C$ mRNA isoforms in the rat and $H3_L$ and $H3_S$ in the guinea pig. However, similar variants in humans have yet to be identified. A fourth histamine receptor subtype, H4, has recently been cloned and characterized with a gene sequence and imidazole-based pharmacology similar to that of the H3 receptor.

Proteoglycans

Proteoglycans are the major granular constituent of all mast cells and are composed of a central protein core of repeating serine and glycine residues with extended, unbranched carbohydrate side chains of repeating disaccharide subunits constructed of uronic acid and hexosamine moieties. The presence of sulfate groups on each disaccharide gives proteoglycans a high charge, which enables proteoglycans to act as extracellular mediators and as storage matrices for other preformed mediators.

The predominant type of proteoglycan stored in mast cells varies among subpopulations. For example, rat peritoneal and skin mast cells, mouse CTMC, and human cutaneous mast cells all contain heparin, whereas rat intestinal MMC and rat basophilic leukemia cells all contain chondroitin sulfate di-B. The proteoglycans heparin and chondroitin sulfate E are associated with human mast cells. Both stabilize mast cell proteases and alter the biological activity of many enzymes.

Mast cells are thought to be the only source of heparin in the body. Heparin has numerous functions beyond the regulation of mediator storage in the cytoplasmic granule. In addition to being an anticoagulant, heparin can bind certain cytokines, chemokines, and growth factors as well as modulate the proliferation and migration of cells involved in atherosclerosis, wound healing, and tissue remodeling.

Neutral Proteases

Mast cell populations in different species and at varying stages of development may be defined by their granule content of serine proteases. The distinction is most obvious in rodent mast cells but has also been observed in humans and sheep. RMCP I is known to have properties similar to those of α-chymotrypsin, in that it cleaves ester and peptide bonds of carboxy-terminal aromatic amino acids, whereas RMCP II is chemically and antigenically distinct and is localized to mucosal mast cells of the gastrointestinal and respiratory tracts.

The major proteases in human mast cells are tryptase and chymase. Tryptase is the predominant enzyme, composed of a tetramer of 134 kDa, with two subunits of 31−34 kDa and associated with all human mast cells examined. Tryptase is stored bound to heparin and is released as a complex with this proteoglycan. This protease cleaves fibrinogen, activates latent collagenase, hydrolyzes some neuropeptides, and may cause mucus secretion and be mitogenic.

Chymase is stored in the granules of some but not all human mast cells. This monomer of 30 kDa is unaffected by heparin and its function is inhibited by biological inhibitors of serine proteases, such as α1- antichymotrypsin. Biological functions of chymase include conversion of angiotensin I to II, stimulation of mucus secretion, neuropeptide degradation, and conversion of interleukin-1β (IL-1β) to an active form.

Newly Synthesized Mediators

Lipid-Derived Mediators

The most important mast cell-derived lipid mediators are believed to be the cyclooxygenase and lipoxygenase metabolites of arachidonic acid. When mast cell activation occurs, arachidonic acid is mobilized from storage pools by various phospholipases and subsequently converted to prostanoids by cyclooxygenases and to leukotrienes by 5-lipoxygenase. Prostaglandin D_2 (PGD_2) is the major cyclooxygenase product of mast cells, whereas the major lipoxygenase products are the sulfidopeptide leukotrienes (LTs; LTC_4, LTD_4, LTE_4,). Human mast cells also have the capacity to produce LTB_4, but in smaller quantities than either PGD_2 or LTC_4.

PGD_2 is generated following immunological activation of human mast cells and its many functions include inhibition of platelet aggregation, mediation of neutrophil accumulation in human skin, and chemokinesis of human neutrophils. PGD_2 has been detected in supernatants of dispersed and purified human pulmonary mast cells after IgE-mediated activation.

LTB_4, LTC_4, LTD_4, and LTE_4 are all derived from 5-hydroperoxyeicosatetraenoic acid. LTC_4 is formed from LTA_4 by the enzymatic addition of glutathione. Once formed, LTC_4 is released into the extracellular fluid and converted to LTD_4 and LTE_4 by the release of glutamic acid and glycine, respectively. Leukotrienes stimulate prolonged bronchoconstriction with an effect 10−1000 times more potent than that of histamine. They also have the ability to enhance venular permeability, promote bronchial mucus secretion, and induce

the constriction of arterial, arteriolar, and intestinal smooth muscle. LTB_4 is derived from LTA_4 by the enzymatic addition of water and has been demonstrated to possess potent chemotactic activity for neutrophils and eosinophils, to enhance lysosomal enzyme release, and to augment superoxide anion production.

Platelet-activating factor (PAF) has also been detected in some mast cell populations. PAF has been detected following activation of mouse bone marrow-derived mast cells, rabbit basophils, and human mast cells. The synthesis of PAF is closely linked with that of the leukotrienes. The first step of PAF synthesis involves the cleavage of alkylarachidonyl-GPC, and subsequent liberation of lyso-PAF and arachidonic acid. When lyso-PAF is acetylated in the 2-position of the glycerol residue, PAF is formed. Arachidonic acid may be channeled down the 5-lipoxygenase pathway, resulting in leukotriene production. PAF is a potent chemotactic agent for neutrophils and eosinophils, and in addition, it causes eosinophils to adhere to endothelial cells and synthesize LTC_4 and more PAF. PAF causes dose-related aggregation and calcium influx in platelets as well as neutrophil exocytosis, superoxide generation, and chemiluminescence. PAF also induces bronchoconstriction in many species including humans, guinea pig, rhesus monkey, and baboon.

Cytokines

Cytokines are a group of protein or glycoprotein molecules that possess a wide range of bioactivities and have been observed to play roles in cell growth and repair, inflammation, and the immune response. The first evidence that mast cell populations produced cytokines was the observation that cells transformed with Abelson murine leukemia virus constitutively produced granulocyte/macrophage colony-stimulating factor (GM-CSF) and IL-4 mRNA and released GM-CSF and IL-4. It is unclear whether mast cells produce several cytokines simultaneously or whether phenotypes exist with regard to cytokine production. Immunohistochemical studies demonstrate that some mast cells synthesize combinations of cytokines and others appear to produce only one type of cytokine. Immature HMC-1 cells fail to secrete IL-3 and IL-13 and up-regulate IL-4 mRNA only transiently, whereas human cutaneous mast cells fail to secrete IL-3, IL-4, IL-5, and IL-13 after stimulation with anti-IgE, substance P, or compound 48/80.

Rodent and human mast cells secrete a wide range of multifunctional cytokines. Following IgE-mediated stimulation, mast cells produce a broad range of cytokines (Table I), mainly of a T_H2 profile, which include IL-4, IL-5, IL-13, GM-CSF, and tumor necrosis factor α (TNFα), all of which are up-regulated in allergic inflammation. Mast cells are also capable of responding to nonimmunological stimuli by synthesis of pro-inflammatory cytokines. The release of cytokines by the mast cell can influence the recruitment and/or function of additional effector cells, such as vascular endothelial cell, fibroblasts, epithelial cells, and nerves.

Differential Release of Mediators

It is well known that mast cells secrete their mediators by degranulation and compound exocytosis as typically seen in anaphylactic reactions. However, mast cells may, by a more subtle intragranular change known as piecemeal degranulation, selectively release particular mediators. Piecemeal degranulation is characterized by intragranular changes showing different electron densities and crystalline substructures in association with the appearance of small vesicles frequently containing electron-dense material.

TABLE I The Major Mast Cell-Derived Cytokines and Their Primary Biological Function

Cytokine	Function
IL-1	Activates lymphocytes; stimulates macrophages; increases leukocyte/endothelial adhesion; induces acute-phase proteins
IL-3	Mast cell colony-stimulating factor
IL-4	B-cell growth factor; isotype switching of IgE and IgG1; mast cell growth cofactor
IL-5	B-cell growth and differentiation factor; eosinophil differentiation factor
IL-6	B-cell differentiation induces acute-phase proteins
IL-8	Chemotactic for neutrophils and basophils; activates neutrophils
IL-13	B-lymphocyte isotype switching to IgE production
TNFα	Activates macrophages, granulocytes, and cytotoxic cells; increases leukocyte/endothelial cell adhesion; induces acute-phase proteins; stimulates angiogenesis; enhances MHC class I production
TGFβ	Stimulates connective tissue growth and collagen formulation; inhibits most immune functions
GM-CSF	Stimulates proliferation of granulocyte and macrophage precursors

Differential mediator release may be regulated by small differences in the cytoplasmic levels of free calcium ions.

SCF has been demonstrated to induce bone marrow-derived cultured mouse mast cells to release IL-6 accompanied by little or no detectable release of TNFα, leukotriene C_4, histamine, or serotonin. Substance P selectively induces TNFα mRNA expression and secretion from a murine cell line. High concentrations of progesterone trigger the release of serotonin from RPMC, without histamine. Piecemeal degranulation has also been observed in mast cells from patients with interstitial cystitis and irritable bowel syndrome.

MAST CELLS AND THE GASTROINTESTINAL TRACT

Mast cells have been proposed to play a role in the inflammatory response in several gastrointestinal diseases including celiac disease and inflammatory bowel disease. Mucosal mast cells are present in all layers of the gastrointestinal tract, where they are located predominantly in the lamina propria of the mucosal layer and the submucosal layer. Mast cell activation and the subsequent release of mediators induce epithelial ion secretion and enhance permeability of the intestine. This may occur through direct action of mast cell mediators on the epithelial cells or indirectly via cross talk between the mast cells and the nervous system.

Mast cells are frequently found in close proximity to the epithelial surface of the gastrointestinal mucosa, where they are strategically located to ensure early interaction with an invading pathogen. The capacity of the mast cell to release both preformed and newly synthesized mediators and to undergo multiple cycles of release permits a rapid and sustained response to bacterial invasion. Mast cells not only can phagocytose bacteria, they also have the ability to recognize infectious agents through specific receptors present on the mast cell surface. The long life span of mast cells means that it is likely that they will encounter a particular pathogen more than once in their lifetime.

Mast cells also contribute to aspects of the adaptive immune response to pathogens, including the well-known response to parasitic infection. Mast cells are known to accumulate in large numbers in the intestinal mucosa in certain parasitic infections and many of these parasite infections are associated with the development of parasite-specific IgE. *In vivo* studies of rats and mice infected with the intestinal nematodes *Nippostronglylus brasiliensis* or *Trichinella spiralis* demonstrate marked mastocytosis in the intestinal lamina propria. Increased mast cell numbers and elevation of mast cell products at the site of infection may cause direct damage to the invading parasite while also recruiting other effector cells such as eosinophils. For example, rat mast cells have been demonstrated to kill *Schistosoma mansoni* by the secretion of RMCP II.

CONCLUSION

Mast cells are regarded as the main tissue-based effector cells in allergic diseases and it is this aspect of mast cell biology that has driven many investigators in attempts to modify mast cell behavior and thus modify allergic inflammation. Despite their detrimental role in allergic disease, mast cells also have an important protective role in host defense mechanisms against parasites, a process that shares many of the characteristic features of allergic inflammation including IgE production and basophil and eosinophil recruitment. An increased understanding of the multifaceted nature of the mast cell has led to the acceptance of a much broader role for the cell in physiological and pathophysiological processes (Fig. 3). The discovery that mast cells could generate and release cytokines in addition to their preformed and newly synthesized inflammatory mediators indicates that mast cells play an important role in immunoregulation. Mast cells have been implicated in a diversity of biological responses including angiogenesis, wound healing, bone remodeling, rheumatoid arthritis, and inflammatory bowel disease.

Immunological and other disease processes may alter mast cell numbers or phenotype by inducing changes in the microenvironment, which alters the mast cell populations residing in that area. The mast cell itself may serve as an important regulator of its own microenvironment. Mast cell mediators can augment local vascular permeability, allowing them to influence the levels of regulatory factors present in the microenvironment. Furthermore, they are able to recruit other cells that may also modulate the mast cell population. Considering the capacity of the mast cell to influence its own growth, differentiation, and activation state by regulating the local microenvironment, investigation of the modulatory effects of autocrine, paracrine, and neuroendocrine agents on mast cell activity is crucial in gaining a comprehensive understanding of the role of the mast cell in health and disease. The mast cell is a unique and complex cell that participates in a number of diverse biological functions. It is well over 100 years since Friedrich von Recklinghausen first observed the mast cell, yet many aspects of mast cell biology remain to be elucidated.

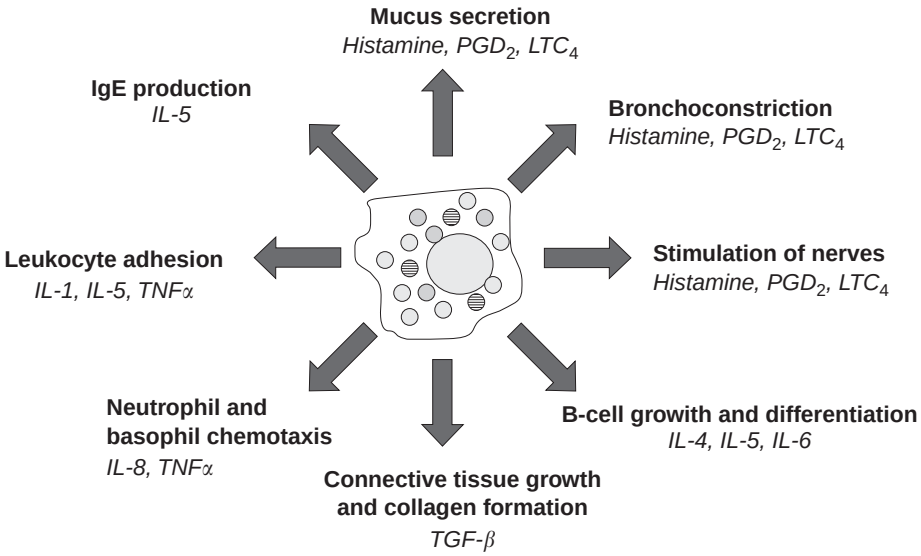


FIGURE 3 Biological functions of mast cell performed on release of mediators (italics).

See Also the Following Articles

Exocytosis • Histamine • Mastocytosis, Gastrointestinal Manifestations of • TH1, TH2 Responses

Further Reading

Abraham, S. N., and Arock, M. (1998). Mast cells and basophils in innate immunity. *Semin. Immunol.* **10**, 373–381.

Bloom, G. D. (1984). A short history of the mast cell. *Acta Otolaryngol. Suppl.* **414**, 87–92.

Bradding, P. (1996). Human mast cell cytokines. *Clin. Exp. Allergy* **26**, 13–19.

Emanuel, M. B. (1999). Histamine and the antiallergic antihistamines: A history of their discoveries. *Clin. Exp. Allergy* **29**, 1–11.

Foreman, J. C. (1993). Non-immunological stimuli of mast cells and basophils. *In* "Immunopharmacology of Mast Cells and Basophils" (J. C. Foreman, ed.), pp. 57–69. Academic Press, London.

Galli, S. J. (1990). New insights into "The riddle of the mast cell": Microenvironmental regulation of mast cell development and phenotypic heterogeneity. *Lab. Invest.* **62**, 5–33.

Hill, P. B., and Martin, R. J. (1998). A review of mast cell biology. *Vet. Dermatol.* **9**, 145–166.

Metcalfe, D. D., Baram, D., and Mekori, Y. A. (1997). Mast cells. *Physiol. Rev.* **77**, 1033–1079.

Pearce, F. L. (1986). Functional differences between mast cells from various locations. *In* "Mast Cell Differentiation and Heterogeneity" (D. A. Befus, J. Bienenstock, and J. A. Denberg, eds.), pp. 215–222. Raven Press, New York.

Pearce, F. L. (1989). Non-IgE mediated mast cell stimulation. *In* "IgE, Mast Cells and the Allergic Response" (D. Chadwick, D. Evered, and J. Whelan, eds.), Ciba Foundation Symposium, Vol. 147, pp. 74–92. Wiley, Chichester, UK.

Sagi-Eisenberg, R. (1993). Signal-transmission pathways in mast cell exocytosis. *In* "Immunopharmacology of Mast Cells and Basophils" (J. C. Foreman, ed.), pp. 71–88. Academic Press, London.

Wershil, B. K. (2000). Lessons from genetically engineered animal models. IX. Mast cell deficient mice and intestinal biology. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G343–G348.

Williams, C. M. M., and Galli, S. J. (2000). The diverse potential effector and immunoregulatory roles of mast cells in allergic disease. *J. Allergy Clin. Immunol.* **105**, 847–859.

Yong, L. C. J. (1997). The mast cell: Origin, morphology, distribution, and function. *Exp. Toxicol. Pathol.* **49**, 409–424.

Yu, L. C. H., and Perdue, M. H. (2001). Role of mast cells in intestinal mucosal function: Studies in models of hypersensitivity and stress. *Immunol. Rev.* **179**, 61–73.

Mastocytosis, Gastrointestinal Manifestations of

ROBERT T. JENSEN AND DEAN D. METCALFE
National Institutes of Health

histamine A vasoactive amine, vasodilator, and smooth muscle constrictor that is found in high concentrations in mast cells.

mast cell A cell of hematopoietic origin that matures within tissues and functions in both innate and acquired immunity through the release of mediators of inflammation following specific activation.

mastocytosis A term collectively used for a group of rare disorders characterized by an abnormal accumulation of mast cells in one or more organ systems.

Systemic mastocytosis is a disorder characterized by mast cell infiltration of organs other than the skin, especially bone marrow, liver, spleen, lymph nodes, and the gastrointestinal (GI) tract. Gastrointestinal disease in the form of peptic ulcer disease and /or malabsorption is often diagnosed in patients with mastocytosis. Because mastocytosis remains without a cure, the selection of appropriate therapy for GI disease remains a persistent challenge as disease progresses. Involvement of the liver through mast cell infiltration with subsequent fibrosis and venopathy further complicates the picture.

INTRODUCTION

Systemic mastocytosis is a disorder characterized by mast cell infiltration of organs other than the skin, especially bone marrow, liver, spleen, lymph nodes, and the gastrointestinal (GI) tract. Systemic mastocytosis is not a uniform entity and is now divided into clinical variants: cutaneous mastocytosis, indolent systemic mastocytosis (ISM), systemic mastocytosis with associated clonal hematologic non-mast-cell lineage disease (SM-AHNMD), aggressive systemic mastocytosis (ASM), and mast cell leukemia. The involvement of the GI tract varies with the different clinical patterns, with more aggressive forms of disease often having more severe GI disease. The GI involvement is clinically important because it is frequently chronic and can cause significant morbidity.

FREQUENCY OF GI INVOLVEMENT

The overall frequency of symptomatic involvement of the GI tract varies greatly in different series with a range of 14–85% and a mean of 51% (16 series). Most series since 1985 report that 60–80% of patients with mastocytosis have GI symptoms and in the only prospective study of GI manifestations in these patients, 80% had GI symptoms. In this series, GI symptoms were almost as frequent as pruritis (88%) and twice as common as flushing (43%). The frequency of various GI symptoms/signs is summarized from 13 series in Table I. Abdominal pain and diarrhea are the most common symptoms and an enlarged liver and (or) spleen are frequent physical findings, especially in patients with SM-AHNMD or ASM.

GI ABNORMALITIES IN PATIENTS WITH SYSTEMIC MASTOCYTOSIS

Esophageal Abnormalities

Esophageal abnormalities (reflux, motility disorders) are uncommon in these patients, occurring in less than 25%.

Gastric Abnormalities

Between 85 and 100% of patients with mastocytosis have increased production of histamine, which is a potent stimulant of gastric acid secretion. In two studies,

TABLE I Frequency of Different Gastrointestinal Symptoms/Signs in Patients with Systemic Mastocytosis

Clinical feature	Frequency mean (range)
Symptom	
Abdominal pain	51% (12–100)
Diarrhea	43% (14–100)
Nausea /vomiting	28% (0–57)
Gastrointestinal bleeding	11% (0–24)
Peptic ulcer disease	23% (5–44)
Sign	
Steatorrhea /malabsorption	13% (5–24)
Hepatomegaly	48% (12–83)
Splenomegaly	45% (11–86)

Note. Data are from 13 different series (Jensen, 2000).

20–40% of these patients had increased acid secretion and a mean of 23% developed peptic ulcers in various series (Table I). On endoscopic studies, up to 30% of patients have gastric mucosal abnormalities including nodules, ulcers, and urticarial lesions. In one prospective study, almost 60% of the abdominal pain was categorized as dyspeptic due to increased acid production.

Small Intestinal Abnormalities

Malabsorption occurs with a mean frequency of 13% in different series of patients with mastocytosis (Table I); however, in a prospective study 31% had impaired small intestinal (SI) function. In general, the malabsorption is mild. The malabsorption, in addition to fat and carbohydrate absorption, can involve fat-soluble vitamins, vitamin B_{12}, and elements such as calcium /iron. In patients with malabsorption, the SI villi are frequently abnormal (blunted, atrophic) and infiltrated by mast cells. With barium X-ray studies, up to 29–73% of these patients have SI mucosal abnormalities, especially small nodules, edema with thickened folds, and polypoid lesions.

Colon/Rectal Abnormalities

Except for diarrhea (Table I), colonic symptoms are not frequently reported in these patients with mastocytosis. However, on barium X-ray or endoscopic studies, 10–20% had abnormalities including nodules, edema, thickened folds, and polypoid lesions.

Liver Abnormalities

Liver involvement is frequent in these patients with mastocytosis, whether assessed by abnormal liver function studies (16–78%), by detection of hepatomegaly (mean 48%; range 12–83%; Table I), or by clinical evidence of liver malfunction (portal hypertension /ascites: 5–40%). In various studies, hepatomegaly was presentin 0–22% of patients with ISM and in 40% of patients with SM-AHNMD or ASM. Liver biopsies frequently show increased fibrosis (14–100%), cirrhosis (0–15%), mast cell infiltration (34%), inflammatory infiltrates (100%), focal hepatocyte necrosis (59%), and evidence of portal venopathy (12%). The hepatomegaly correlates with the degree of mast cell infiltration. The portal hypertension is primarily due to intrahepatic venous obstruction secondary to fibrosis and possibly to increased blood flow from the splenic vein.

Splenic Abnormalities

Splenomegaly occurs with a mean frequency of 45% in patients with mastocytosis (Table I), varying from 11 to 80% in different series. Mast cell infiltration occurs in 34–100% and is likely one of the main mechanisms causing the splenomegaly.

PATHOGENESIS OF THE PRINCIPAL GI SYMPTOMS

General

Numerous studies show that GI, as well as other symptoms commonly occurring in patients with mastocytosis, can be precipitated in some individuals by drugs (antibiotics, narcotics, aspirin, procaine, anti-inflammatory agents), changes in temperature, stress (trauma, medical procedures), exercise, or ingestion of various foods (chocolate, nuts, red wine, alcohol). Rarely, these agents can cause vascular collapse, hypotension, or syncope.

Pathogenesis of Abdominal Pain

Abdominal pain has been attributed to mast cell infiltration in the GI tract, peptic ulcer disease, a motility disorder, and release of mediators from mast cells resulting in altered bowel function. In a prospective study of the 80% of patients with abdominal pain, in 45% the pain was dyspeptic in nature (relieved by acid suppressants, usually epigastric in location) and in 35% it was nondyspeptic (usually lower GI in location, not relieved by acid suppressants). Most patients (85–100%) with systemic mastocytosis have increased histamine release, which can stimulate acid secretion, and this likely plays an important role in leading to pain in the dyspeptic patient. The nondyspeptic pain has been shown not to be due to increased acid secretion and its mechanism remains unclear, although it is frequently proposed that abnormal motility may be an important factor.

Pathogenesis of Diarrhea

Diarrhea in patients with mastocytosis is frequently an episodic increase in the number of bowel movements per day, which can be accompanied by increased stool weight (>200 g/day) with or without malabsorption. In one prospective study, 50% of the patients with diarrhea had increased stool weight and in most of these patients this was associated with gastric acid hypersecretion, which can cause diarrhea. Other proposed causes of diarrhea in these patients include overproduction of prostaglandin D2, fat malabsorption secondary to mast cell infiltration of the small intestine resulting in generation of hydroxy fatty acids, or altered GI transit due to mediators released by the mast cells.

Pathogenesis of GI Malabsorption

Fat malabsorption is not only uncommon in patients with mastocytosis (5–24%, Table I), it is usually mild (< 20 g/day) (normal: < 7 g/day). The most important mechanism is likely proximal SI dysfunction. Abnormal morphology of villi (blunted, atrophy, mast cell infiltrates) in the proximal SI has been described and in barium X-ray studies, 29–73% of patients with mastocytosis have abnormalities that may contribute to the SI absorptive dysfunction. In various patients it has been proposed that the malabsorption might also be due to gastric acid hypersecretion, inflammatory infiltration in the SI, association with celiac sprue, or disturbed SI motility.

TREATMENT OF GI MANIFESTATIONS OF SYSTEMIC MASTOCYTOSIS

The treatment of GI symptoms (Table I) is dictated by the severity of the abdominal pain, diarrhea, malabsorption, and nausea. The possible contribution of gastric acid hypersecretion can be treated by using proton pump inhibitors or histamine-2 (H2)-receptor antagonists. For nondyspeptic pain, H2-receptor antagonists or a combination of H1- and H2-receptor antagonists are frequently effective. The mast cell membrane-stabilizing drug, disodium cromoglycate, has been used in a number of studies to treat GI manifestations including diarrhea, pain, nausea, and vomiting. In patients with prostaglandin D2 contributing to the diarrhea, it is reported that low doses of aspirin (40 mg/day) or other anti-inflammatory agents may be helpful. In rare patients with severe malabsorption, systemic steroids may be effective. Ascites and/or portal hypertension may be difficult to control with the usual treatment of beta-blocking agents, diuretics, and salt reduction. Portal hypertension was successfully treated in one patient by portacaval shunting. Finally, interferon-α-2b resulted in partial responses in a small number of patients with the malignant forms of mastocytosis.

See Also the Following Articles

Diarrhea • Duodenal Ulcer • Emesis • Gastric Acid Secretion • Histamine • Malabsorption • Mast Cells • Nausea

Further Reading

Anonymous (2002). Mast cell disorders. *Hematol. Oncol. Clin. North Am.* **14**, 497–701.

Cherner, J. A., Jensen, R. T., Dubois, A., O'Dorisio, T. M., Gardner, J. D., and Metcalfe, D. D. (1988). Gastrointestinal dysfunction in systemic mastocytosis. *Gastroenterology* **95**, 657–667.

Golkar, L., and Berhard, J. D. (1997). *Mastocytosis. Lancet* **349**, 1379–1385.

Jensen, R. T. (2000). Gastrointestinal abnormalities and involvement in systemic mastocytosis. *Hematol. Oncol. Clin. North Am.* **14**, 579–623.

Marone, G., Spadaro, G., Granata, F., and Triggiani, M. (2001). Treatment of mastocytosis: Pharmacologic basis and current concepts. *Leukemia Res.* **25**, 583–594.

Mican, J. M., DiBisceglie, A. M., Fong, T. L., Travis, W. D., Kleiner, D. E., Baker, B., and Metcalfe, D. D. (1995). Hepatic involvement in mastocytosis: Clinicopathologic correlations in 41 cases. *Hepatology* **22**, 1163–1170.

Valent, P., Horny, H. P., Escribano, L., Longley, B. J., Li, C. Y., Schwartz, L. B., Marone, G., Nunez, R., Akin, C., Sotlar, K., Sperr, W. R., Wolff, K., Brunning, R. D., Parwaresch, R. M., Austen, K. F., Lennert, K., Metcalfe, D. D., Vardiman, J. W., and Bennett, J. M. (2001). Diagnostic criteria and classification of mastocytosis: A consensus proposal. *Leukemia Res.* **25**, 603–625.

Valent, P., Horny, H.-P., Li, B., Longley, B. J., Metcalfe, D. D., Parwaresch, M., and Bennet, J. M. (2001). Mastocytosis. *In* "World Health Organization Classification of Tumours: Pathology and Genetics in Tumours of Haematopoietic and Lymphoid Tissues" (E. S. Jaffe, N. L. Harris, H. Stein, and J. W. Varderman, eds.), pp. 291–302. IARC Press, Lyon, France.

Worobec, A. S. (2000). Treatment of systemic mast cell disorders. *Hematol. Oncol. Clin. North Am.* **14**, 659–687.

Meckel's Diverticulum

CYNTHIA D. DOWNARD, SARA LEWICKE, AND MORITZ M. ZIEGLER
Children's Hospital Boston

diverticulum Pouch protruding from a tubular organ.
ectopic Out of place, not in a proper position.
intussusception Infolding of one part of the intestine into another.

Meckel's diverticulum is a 2- to 6-cm out-pocketing of the ileum along the antimesenteric border, usually located 50 to 180 cm proximal to the ileocecal valve. It is the most common congenital aberration of the gastrointestinal tract, occurring in an estimated 2–3% of the population, more commonly in men than women. It is a true diverticulum in that it is composed of tissue from all sections of the ileal wall. Meckel's diverticulum is often described by the "rule of twos"—it may contain of two types of ectopic tissue (gastric and pancreatic), is usually located within 2 feet of the ileocecal junction, occurs in 2% of the population, is close to 2 inches in length, and patients are usually symptomatic before 2 years of age.

INTRODUCTION

History

Meckel's diverticulum was first described in 1598 by Fabricius Hidanus as an "unusual diverticulum" of the small intestine. Littre reported presence of a diverticulum of small bowel present in a hernia in 1742. In 1809, Johann Friedrich Meckel the Younger described the embryological aspects and pathologic nature of an ileal diverticulum, which led to the subsequent adoption of his name for this anomaly.

Embryology

Meckel's diverticulum results from the abnormal persistence of the vitelline (or omphalomesenteric) duct during development. In embryonic life, the vitelline duct is a stalk connecting the midgut of the developing embryo to the yolk sac, and is involved in nutrient transfer and hematopoiesis. Typically, at the end of the sixth week of gestation, the vitelline duct degenerates and the yolk sac detaches from the midgut loop. It is thought that a Meckel's diverticulum develops when the proximal portion of this connection fails to degrade properly. If the entire vitelline duct fails to obliterate, the midgut can remain attached to the umbilicus as a fistula, also called a patent omphalomesenteric duct. Fibrous cords from the diverticulum, with or without cysts, may also remain attached to the umbilicus.

PRESENTATION

General Presentation

Although autopsy series estimate that 2% of the population has a Meckel's diverticulum, less than 20% of people with this anomaly will ever come to clinical attention. A vast majority of Meckel's diverticula are discovered incidentally during laparotomy, during contrast studies for unrelated reasons, or at autopsy.

Symptomatic Presentation

Intestinal Hemorrhage

Intestinal hemorrhage is the most common presentation of symptomatic Meckel's diverticulum in children less than 18 years of age. Ectopic gastric tissue is estimated to be present in 60% of symptomatic Meckel's diverticula. This functional gastric mucosa can produce acid, in turn causing ulceration of the surrounding tissue. Ulcers of this type usually occur at the junction between gastric mucosa within the diverticulum and normal ileal mucosa or opposite the diverticulum on the ileal wall, and can erode into surrounding blood vessels. Interestingly, *Helicobacter pylori*, although often implicated in peptic ulcer disease, is almost never associated with ileal ulceration in Meckel's diverticulum.

Bowel Obstruction

Bowel obstruction is the most common presentation of symptomatic Meckel's diverticulum in the adult population, occurring in over 50% of cases. Intestinal obstruction frequently develops when a fibrous band of a vitelline artery remnant from the Meckel's diverticulum to the

 Copyright 2004, Elsevier (USA). All rights reserved.

umbilicus provides a fixed point for intestinal torsion and closed-loop bowel obstruction. Obstruction may also be due to intussusception with an inverted diverticulum serving as the lead point. This occurs more often in children than in adults. In addition, obstruction may also result due to incarceration of a Littre's hernia, with a Meckel's diverticulum contained within the hernia sac.

Meckel's Diverticulitis

Meckel's diverticulitis occurs in up to one-third of presenting cases, attributable to the irritating effect of peptic acids produced by ectopic gastric mucosa within the diverticulum. Often the symptoms of Meckel's diverticulitis will imitate those of acute appendicitis, highlighting the necessity of examining the ileum for the presence of an irritated Meckel's diverticulum during a negative exploration for presumed appendicitis.

DIAGNOSIS

Meckel's diverticulum is among the most difficult conditions to diagnose preoperatively. Although Meckel's diverticulum is frequently within a differential diagnosis of abdominal pain, only 5–6% of cases are correctly definitively identified preoperatively. Diagnostic tests rely on the unique features of a Meckel's diverticulum, namely, ectopic gastric tissue and ulceration with bleeding, for conclusive identification.

Radionuclide Scanning

The best tool for diagnosing Meckel's diverticulum in children is scintigraphy using [^{99m}Tc]pertechnetate, an isotope taken up by mucus-secreting cells of the gastric mucosa and ectopic gastric tissue within a Meckel's diverticulum. The high incidence of ectopic gastric tissue within symptomatic Meckel's diverticula, particularly in children, makes this scan a useful diagnostic tool. Several agents, such as pentagastrin to increase gastric mucosal absorption, histamine-2 (H2) blockers to decrease excretion of the isotope, and glucagon to decrease peristalsis, can be used to help increase the accuracy and precision of this test.

When intestinal hemorrhage is the presenting symptom of a Meckel's diverticulum, [^{99m}Tc]pertechnetate scintigraphy loses accuracy. Alternatively, erythrocytes can be labeled with a ^{99m}Tc–sulfur colloid to pinpoint a site of hemorrhage in patients losing blood at a rate of at least 0.1 ml/min. If an identified site of bleeding correlates with the terminal ileum, an ulcer associated with a Meckel's diverticulum is the most likely cause. This test is sensitive for identifying bleeding, but is not specific for Meckel's diverticulum.

Angiography

If scintigraphy is not definitive in a patient with continued intestinal bleeding, angiography may be used to identify a site of hemorrhage. When a Meckel's diverticulum is the suspected source, selective superior mesenteric arteriography can be diagnostic. Features highly suspicious for a Meckel's diverticulum are visualization of a vascular blush at the expected site of a diverticulum or persistence of the vitelline artery. False positive results may be obtained on angiography if there is irregular arterial branching from a distal ileocolic artery.

Other Studies

Other radiologic tools, such as ultrasound and computed tomography, have a limited role in the diagnosis of Meckel's diverticulum, but may be employed when other studies are inconclusive. Contrast evaluation may demonstrate a characteristic triangular folding pattern at the site of ileal attachment, a gastric rugal pattern within the diverticulum, or, when the diverticulum is inverted, a polyplike defect within the lumen of the intestine. Air-contrast enema is both diagnostic and therapeutic in cases of intussusception, but it rarely results in a definitive diagnosis of Meckel's diverticulum.

TREATMENT

Symptomatic Diverticula

Symptomatic Meckel's diverticula should be surgically resected. Simple diverticulectomy may suffice when the diverticular opening is small and no ectopic tissue is present. Ileal wedge resection and primary anastomosis are recommended in cases of hemorrhage, ulceration, neoplasia, and chronic inflammation. Long diverticula, with ectopic gastric tissue at the distal end, can be removed by simple transverse resection at the diverticular neck. Shorter diverticula, with gastric tissue on any area, are best removed by wedge resection and end-to-end anastomosis to avoid retention of ectopic gastric mucosa. Simple ligation of the diverticulum and stump invagination are not recommended because the neck of a diverticulum is often too wide to accomplish this effectively.

Asymptomatic Diverticula

Considerable controversy surrounds the treatment of an asymptomatic Meckel's diverticulum because most

Meckel's diverticula are incidentally discovered during laparotomy. It is agreed that the likelihood of a Meckel's diverticulum ever becoming symptomatic is low, but conflicting evidence exists as to whether this usually happens in the earlier decades of life. In addition, varying rates of complication have been reported following incidental diverticulectomy. Prior studies have attempted to offer intraoperative guidelines regarding age of the patient, length of the diverticulum, and presence or absence of ectopic mucosa to determine which asymptomatic diverticula should be resected. In an extensive epidemiologic, population-based project, Cullen and colleagues utilized comprehensive records over a 42-year period from Olmsted County, Minnesota to evaluate incidence and complications arising from Meckel's diverticulum. After assessing all cases of symptomatic and asymptomatic Meckel's diverticulum requiring operation, it was determined that the risk of a Meckel's diverticulum becoming symptomatic does not decrease with age. In addition, the rate of morbidity of incidental removal was sufficiently low (2%) and the rate of early postoperative complication after symptomatic diverticulectomy was adequately high (12%) to warrant a recommendation of routine removal of incidentally discovered Meckel's diverticula. Because this is the most comprehensive review of this subject to date, these suggestions are currently the best supported criteria.

OUTCOME

Full recovery should be expected following diverticulectomy for symptomatic or asymptomatic Meckel's diverticulum. Although up to 12% of patients develop complications such as adhesions, sepsis, and wound infection following symptomatic Meckel's diverticulectomy, these are usually easily treated nonoperatively with antimicrobials and other agents. Mortality, even for symptomatic Meckel's diverticulectomy, is rare.

See Also the Following Articles

Diverticulosis • Hemorrhage • Occult Gastrointestinal Bleeding

Further Reading

Brown, R. L., and Azizkhan, R. G. (1999). Gastrointestinal bleeding in infants and children: Meckel's diverticulum and intestinal duplication. *Semin. Pediatr. Surg.* **8**, 202–209.

Cullen, J. J., Kelly, K. A., Moir, C. R., Hodge, D. O., Zinsmeister, A. R., and Melton, L. J. (1994). Surgical management of Meckel's diverticulum. *Ann. Surg.* **220**, 564–569.

Leijonmarck, C. E., Bonman-Sandelin, K., Frisell, J., and Raf, L. (1986). Meckel's diverticulum in the adult. *Br. J. Surg.* **73**, 146–149.

Rossi, P., Gourtsoyiannis, N., Bezzi, M., Raptopoulos, V., Massa, R., Capanna, G., Pedicini, V., and Coe, M. (1996). Meckel's diverticulum: Imaging diagnosis. *Am. J. Roentgenol.* **166**, 567–573.

Yahchouchy, E. K., Marano, A. F., Etienne, J. C. F., and Fingerhut, A. L. (2001). Meckel's diverticulum. *J. Am. Coll. Surg.* **192**, 658–662.

Megacolon: Neuromuscular Enteric Abnormalities

SAMUEL NURKO
Children's Hospital Boston

Hirschsprung's disease Disease with varying degrees of distal intestinal aganglionosis.
megacolon A condition in which the colon has abnormally large dimensions.

Megacolon in children is a poorly defined entity. Even though functional constipation is the most common etiology, there are some rare neuromuscular enteric abnormalities that present with megacolon. These occur because there are problems with the enteric nervous system, the enteric smooth muscle, or both. This article focuses primarily on the neuromuscular enteric disorders that primarily affect the gastrointestinal tract.

HIRSCHSPRUNG'S DISEASE

Hirschsprung's disease is the most common neuroenteric disorder. This article focuses mostly on the clinical aspects of the disease.

Hirschsprung's disease is a congenital problem in which varying degrees of aganglionosis occur in distal segments of the intestinal tract. The length of the aganglionic segment varies. It is limited to the rectum and sigmoid in 75% of cases, involves the whole colon in 8% of cases, and rarely involves the small bowel. The aganglionosis in Hirschsprung's disease results in a lack of intrinsic enteric inhibitory nerves. Intestinal obstruction is produced because there is a lack of nonadrenergic, noncholinergic innervation in the aganglionic segments with a resultant failure to relax during peristalsis. The segment proximal to the obstruction then becomes dilated.

The incidence of Hirschsprung's disease varies from 1 in 5000 to 1 in 10,000 live births. There seems to be a male preponderance with a ratio of 3 : 1 to 5 : 1, particularly in those with short segments. However, the incidence in both sexes seems to be the same in those with long-segment disease.

The mean age at the time of diagnosis has been decreasing over the years. The diagnosis is established in 15% of patients within the first month of life, in 40–50% of patients in the first 3 months, in 60% of patients at the end of the first year of age, and in 85% of patients by 4 years. The diagnosis of Hirschsprung's disease is rarely delayed until adulthood.

The symptoms vary with the age of the patient and the extent of the disease. In the newborn period, bilious emesis, abdominal distension, and failure to pass meconium or abnormal stool frequency are common. Complete intestinal obstruction may occur and perforation of the cecum or the appendix occurs in 3–5% of cases. If the diagnosis is not established in the newborn, the infant may present with mild constipation that may be followed with acute obstruction, frequent episodes of fecal impaction, or the development of acute life-threatening enterocolitis. Enterocolitis develops in 15 to 50% of cases and may be the presenting feature of Hirschsprung's disease in up to 12% of patients. It remains the main cause of death and the mortality rate can reach 20 to 50%.

From infancy until adulthood, mild to severe constipation may be the only symptom, so Hirschsprung's disease must be differentiated from functional constipation. Because clinical features do not allow a complete differentiation between these problems, the diagnosis of Hirschsprung's disease must always be considered in any child, adolescent, or adult with severe intractable constipation.

Diagnosis

Once the diagnosis of Hirschsprung's disease is suspected, confirmation is necessary. The final diagnosis needs to be based on the pathologic demonstration of aganglionosis. Because obtaining biopsies involves risks, other less invasive techniques, such as barium enema and anorectal manometry, can be used to select those patients who require a biopsy.

The barium enema (BE), although not diagnostic can be strongly suggestive and supportive. Single-contrast barium enemas are used and the colon is not prepared. In infants with Hirschsprung's disease, a transition zone from the distal nondilated colon is

 Copyright 2004, Elsevier (USA). All rights reserved.

usually easily detected. The absence of a transition zone, however, does not exclude the diagnosis. BE may be less helpful in the newborn because a visible transition zone is often not present.

Anorectal manometry is a noninvasive technique that allows the detection of the recto-anal inhibitory reflex. In normal individuals, distension of the rectal ampulla causes relaxation of the internal anal sphincter. Intrinsic nerves mediate this effect, which is absent in patients with Hirschsprung's disease. After the newborn period, manometry has been shown to accurately diagnose Hirschsprung's disease in 90 to 100% of the patients, with a specificity of 97% and a sensitivity of 79%. Therefore, in this age group, anorectal manometry is the diagnostic study of choice.

Confirming the absence of ganglion cells in the diseased segment is a necessary step in the diagnosis of Hirschsprung's disease. Rectal suction biopsy, which produces a specimen of mucosa and submucosa, is a commonly employed tool. When ganglion cells are present, the diagnosis of Hirschsprung's disease is clearly excluded. The use of immunohistochemical stains for acetylcholinesterase (AChE) in combination with the hematoxylin and eosin stain can increase the accuracy. There is a marked increase in AChE activity in the lamina propria and muscularis of the patient with Hirschsprung's disease.

For those cases in whom the suction biopsy has produced equivocal results, it may be necessary to obtain a full-thickness biopsy of the rectal wall.

Treatment

The treatment of Hirschsprung's disease is surgical. Initial medical management is important, however, in stabilizing the patient before surgical therapy is undertaken. This includes the correction of fluid and electrolyte imbalances, antibiotic therapy if enterocolitis is present, and rectal decompression with the use of rectal irrigations and rectal tubes until the time of surgery.

The basic principle for the definitive surgical therapy is resection of the aganglionic segment followed by a pull-through of ganglionic bowel down to the anus. The three most common operations performed include the Swenson pull-through (rectosigmoidectomy), the Duhamel pull-through (retrorectal transanal pull-through), and the Soave pull-through (endorectal pull-through). It is difficult to compare the results obtained with the three techniques, because the incidence of complications may be closely related to the skill of the individual surgeon, to the institution, or to the year of the study. Nevertheless, the long-term outcome of these procedures appears to be similar. Independent of the type of surgery, long-term survival is excellent, although death from enterocolitis may occur years after successful surgical reconstruction. Many studies show a higher than anticipated incidence of problems postoperatively, in particular, persistent obstruction, fecal incontinence, or enterocolitis.

TABLE I Common Problems after Hirschsprung's Surgical Repair

Obstructive symptoms and constipation
Anatomic
Anal stenosis, strictures
Functional
Aganglionosis: Residual or new
Neuronal intestinal dysplasia
Dismotility
Internal anal sphincter hypertonicity
Functional fecal retention
Fecal incontinence
Overflow incontinence from constipation
Abnormal sphincteric function
Diarrhea
Bacterial overgrowth syndrome
Enterocolitis
Clostridium difficile
Other pathogens
Growth problems
Diarrhea and malabsorption
Specific nutrient deficiencies

The most common problems encountered after surgical repair are shown in Table I. The treatment will vary depending on the diagnosis and may include laxatives, antibiotics, rectal decompression, surgery, or administration of botulinum toxin to the internal anal sphincter.

OTHER NEUROMUSCULAR DISEASES THAT AFFECT THE INTESTINE

Other disorders that mimic the distal obstruction seen in Hirschsprung's disease are mentioned in Table II. They encompass clinical entities in which there is disruption of the normal control mechanisms of the motor function of the intestine and they include abnormalities of the myenteric plexus or the smooth muscle. Those abnormalities produce ineffective antegrade peristalsis and recurrent symptoms of bowel obstruction.

The clinical presentation will vary depending on the type and extent of the abnormality. It usually includes functional obstruction, delayed intestinal transit, and abdominal pain and there may be associated urinary problems or bacterial overgrowth.

Diseases of the enteric nervous system or the smooth muscle may be familial or sporadic and may be limited to

TABLE II Etiology of Megacolon in Children

- Functional constipation
- Anatomic problems
  - Anal stenosis
  - Anorectal malformations
- Neuromuscular abnormalities
  - Disorders of the enteric nervous system
    - Hirschsprung's disease
    - Primary visceral neuropathies
      - Familial
        - Autosomic dominant
        - Autosomic recessive with intranuclear inclusions
        - Autosomic X-linked
        - Associated with multiple endocrine neoplasia type 2B
      - Nonfamilial
        - Intestinal ganglioneuromatosis
        - Neuronal intestinal dysplasia
        - Hypoganglionosis
        - Hyperganglionosis
        - Other qualitative abnormalities
    - Secondary visceral neuropathies
      - Infections
      - Autoimmune
      - Inflammatory
      - Metabolic/endocrine
      - Drugs
  - Disorders of the smooth muscle
    - Primary visceral myopathies
      - Familial
        - Autosomic dominant (hollow visceral myopathy and others)
        - Autosomic recessive (mitochondrial–gastrointestinal encephalomyopathy)
      - Nonfamilial
        - Childhood visceral myopathy
        - Other mitochondrial disorders
        - Megacystis–microcolon–hypoperistalsis syndrome
    - Secondary visceral neuropathies

the colon or they may be part of a more generalized disorder that can affect the whole gastrointestinal (GI) tract. In some cases, there may also be extraintestinal involvement or the enteric dysfunction may be a result of abnormal extraintestinal regulation.

Of the primary diseases that affect the smooth muscle, the most common are either hollow visceral myopathy or megacystis microcolon hypoperistalsis syndrome (MMIHS). MMIHS is the most severe form of pseudo-obstruction and affects predominantly female infants. It is characterized by marked dilation of the bladder, a dilated nonperistaltic small bowel, and a malrotated microcolon. Most patients require multiple surgical procedures and are dependent on total parenteral nutrition. The prognosis is poor and most patients die at a very young age from renal insufficiency or sepsis.

Of the primary diseases that affect the enteric nervous system, Hirschsprung's disease is the most common. After Hirschsprung's disease, the most common disorders are those in which there are malformations of intestinal neurons. Among those conditions, neuronal intestinal dysplasia (NID) has been the most controversial. This term has been used to describe qualitative and quantitative abnormalities of the myenteric plexus. These include hyperganglionosis, with the presence of giant ganglia, heterotopic ganglion cells, and abnormal findings using acetylcholinesterase stain. Different types of NID have been recognized and it may be associated with an aganglionic segment. Part of the problem in deciding if NID is a clinically distinct entity stems from the fact that there are no clear and validated criteria for its diagnosis. Recent prospective, controlled studies have tried to validate pathologic criteria for innervation abnormalities like Hirschsprung's disease or NID using rectal biopsies. These investigations demonstrated that even though there is complete agreement in the diagnosis of Hirschsprung's disease, there is great interobserver disagreement and a lack of concordance in making a diagnosis of NID even by a group of experienced pathologists. This lack of agreement probably stems from the lack of information about the normal developmental process that occurs in the myenteric neurons after birth. Further studies of age-dependent normal variation in the morphology of the myenteric plexus are needed.

NID is a pathologic diagnosis and does not seem to represent a distinct clinical problem. When the diagnosis has been made, the clinical symptoms have included intractable constipation and megacolon. Multiple studies have shown, however, that there does not seem to be a correlation between the pathologic findings and the clinical presentation, abnormal functional testing, or outcome. Therefore, caution needs to be exercised in the clinical management of those children who are diagnosed with NID.

Neuromuscular enteric problems can also be secondary to systemic disorders. These include mitochondrial and neurologic diseases, endocrinologic and metabolic abnormalities, connective tissue diseases, neurologic diseases, or after the administration of drugs.

Evaluation and Treatment

The differential diagnosis of megacolon in children includes the problems mentioned in Table II. Most children will have functional constipation and their diagnosis is established clinically after a careful history and physical examination. It is important to exclude anatomic malformations, such as anal stenosis, imperforate

anus, or a mechanical obstruction. The presence of a neuromuscular enteric disorder needs to be excluded in patients with intractable constipation or those with severe obstructive symptoms. Initially, Hirschsprung's disease needs to be excluded as mentioned previously.

If there is no Hirschsprung's disease and there is evidence of a generalized motility disorder, systemic problems need to be excluded. Objective measurements of gastrointestinal transit may be necessary. These include gastric emptying, transit, and motility studies. Antroduodenal manometry may be indicated to evaluate upper GI motility and may help to establish whether the patient's problem is myopathic or neuropathic in nature. If the problem seems to be confined only to the large intestine, a colonic motility test may be necessary. This may reveal evidence of neuropathy or myopathy. A full-thickness biopsy of the rectum may be indicated when there is evidence of a neuroenteric motility disorder of the colon.

In most cases, there is no specific treatment. Nutritional support is important, particularly patients with a generalized problem. Enteral feedings should be tried, but total parenteral nutrition may be necessary. Bacterial overgrowth should be treated. The use of prokinetic agents may be beneficial, as long as there is remaining function of the enteric nerves. Cisapride may be obtained under special programs and octreotide may be used, particularly if there is small bowel dysfunction.

Patients with Hirschsprung's disease need to have surgical resections, and for those patients with normal ganglion cells and a nonrelaxing internal anal sphincter, botulinum toxin or myectomy is indicated. Other surgical therapy should be reserved only for patients with involvement of small and specifically affected segments. The creation of venting ostomies is necessary to relieve pain in those with more severe obstructive symptoms. An ileostomy may be necessary when there is severe colonic dysfunction and a cecostomy may be indicated for the administration of colonic lavage. For patients with severe small bowel dismotility, an intestinal transplant may be the only option for long-term survival.

See Also the Following Articles

Constipation • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Pseudoobstruction • Manometry

Further Reading

Baker, S., Liptak, G., Colletti, R., Croffie, J., Dilorenzo, C., Extor, W., and Nurko, S. S. (1999). Constipation in infants and children. Evaluation and treatment. A medical position statement of the North American Society for Pediatric Gastroenterology and Nutrition. *J. Pediatr. Gastroenterol. Nutr.* **29**, 612–626.

Holschneider, A. M., Meier-Ruge, W., and Ure, B. M. (1994). Hirschsprung's disease and allied disorders: A review. *Eur. J. Pediatr. Surg.* **4**, 260–266.

Holschneider, A. M., and Puri, P. (eds.) (2000). "Hirschsprung's Disease and Allied Disorders," 2nd Ed. OPA (Overseas Publishers Association) N.V. Harwood Academic, Reading, UK.

Koletzko, S., Jesch, I., Faus-Kebler, T., Briner, J., Meier-Ruge, W., Muntefering, H., *et al.* (1999). Rectal biopsy for diagnosis of intestinal neuronal dysplasia in children: A prospective multicentre study on interobserver variation and clinical outcome. *Gut* **44**, 853–861.

Moore, S. W., Millar, A. J., and Cywes, S. (1994). Long term clinical, manometric, and histologic evaluation of obstructive symptoms in the postoperative Hirschsprung's patient. *J. Pediatr. Surg.* **29**, 106–111.

Nurko, S. S. (2000). Complications after gastrointestinal surgery: A medical perspective. *In* "Pediatric Gastrointestinal Disease" (W. A. Walker, *et al.*, eds.), 3rd Ed., pp. 1843–1876. B. C. Decker, Philadelphia, PA.

Nurko, S. S. (2000). New advances in the evaluation and treatment of constipation. *Curr. Gastroenterol. Rep.* **2**, 234–240.

Nurko, S. S. (2000). Treatment of Hirschsprung's disease. *In* "Therapy of Digestive Disease" (M. M. Wolfe, S. Cohen, G. Davis, *et al.*, eds.), pp. 609–616. W. B. Saunders, Philadelphia, PA.

Oguzkurt, P., Senocak, M. E., Akcoren, Z., and Buyukpamukcu, N. (2000). Diagnostic difficulties in neuronal intestinal dysplasia and segmental colitis. *J. Pediatr. Surg.* **35**(3), 519–521.

Puri, P. (1997). Variant Hirschsprung's disease. *J. Pediatr. Surg.* **32**, 149–157.

Reid, J., Buonomo, C., Morera, C., Kozakevich, H., and Nurko, S. S. (2000). The barium enema in constipation: Comparison with rectal manometry and biopsy to exclude Hirschsprung's disease after the neonatal period. *Pediatr. Radiol.* **30**, 681–684.

Rudolph, C., and Benaroch, L. (1995). Hirschsprung disease. *Pediatr. Rev.* **16**, 5–11.

Ménétrier's Disease

J. Steven Burdick*, Stephen H. Settle†, and Robert J. Coffey†
*University of Texas Southwestern Medical Center, Dallas, and †Vanderbilt University

antral sparing Pattern of gastropathy in which the proximal stomach (fundus and body) is primarily involved and the distal stomach (antrum and pylorus) is relatively or totally free of disease ("spared").

epidermal growth factor Peptide that is tropic for the growth of epithelial cells in the gastrointestinal mucosa and elsewhere.

gastropathy Disorder of the stomach that may or may not be associated with mucosal inflammation (gastritis).

Ménétrier's disease Form of gastropathy characterized by hyperplasia of the lining epithelium and hypertrophy of the mucosa, with or without inflammation (gastritis).

Ménétrier's disease (hypoproteinemic hypertrophic gastropathy) is gastropathy characterized by decreased acid production (hypochlorhydria), excessive mucin production, hyperproliferation of the gastric folds, and hypoproteinemia associated with selective loss of serum proteins across the gastric mucosa into the gastric lumen. The disease linkage with Ménétrier dates to 1888 with the description of enlarged gastric folds, "polyadenomes en nappe rugae," that resembled cerebral convolutions. The disease is rare, acquired, and a premalignant disorder. Gastric cancer has been reported at diagnosis and during followup of patients with hypertrophic gastropathy. No specific therapy is predictive of response to symptoms. The disease has been linked with ligands of the epidermal growth factor receptor, thus antagonism of this receptor as a therapeutic modality is being investigated.

ETIOLOGY

The cause of Ménétrier's disease is unknown. *Helicobacter pylori* and cytomegalovirus infection have been associated with a reversible hypertrophic gastropathy. Cytomegalovirus infection is noted predominantly in pediatric presentations but also has been reported in immunocompetent adults. Eradication of *H. pylori* has correlated with resolution of hypertrophic gastropathy and clinical expression in some patients. The hypothesis that up-regulation of transforming growth factor-α is a feature of this disease is supported by studies of transgenic mice and human data. Increased expression of transforming growth factor-α in the stomach can recapitulate the phenotype of Ménétrier's disease, including reduced acid production, excessive mucin production, and hyperproliferation of mucosa. Transforming growth factor-α is one of seven mammalian ligands of the epidermal growth factor (EGF) receptor. Thus, increased expression of any of these ligands may contribute to the disease. Blockade of EGF receptor in a patient with Ménétrier's disease results in decreased clinical and biochemical expression of the disease. There is an accompanying decrease in the proliferation of the mucosa and emergence of parietal cells after EGF blockade. The collective data from transgenic mice, humans, and response to EGF blockade are supportive of a role of enhanced epidermal growth factor receptor signaling in the pathogenesis of Ménétrier's disease.

PATHOLOGY

Ménétrier's disease is a form of hypertrophic gastropathy characterized by foveolar hyperplasia and glandular atrophy. The findings of mucosal inflammation and glandular atrophy are quite variable. The lymphocytic inflammatory response and the foveolar hyperplasia have been used to separate the gastropathy into two groups, hypertrophic lymphocytic gastritis and massive foveolar hyperplasia. Hypertrophic lymphocytic gastritis is associated with severe inflammation with numerous intraepithelial lymphocytes and mild foveolar hyperplasia, whereas massive foveolar hyperplasia has greater foveolar hyperplasia, thicker mucosa, and greater mucosal edema. Hypertrophic lymphocytic gastropathy is considered a spectrum of lymphocytic gastritis by several researchers.

CLINICAL

Patients with Ménétrier's disease frequently present with complaints of epigastric pain, fatigue, weight loss, edema, and vomiting. Severe presentations of Ménétrier's disease include hypoalbuminemia and iron deficiency anemia. Giant rugae are the hallmark of the disease. Normal folds are around 5 mm in width and are parallel to the long axis of the stomach. The folds in Ménétrier's are thick, tortuous, and convoluted. The

 Copyright 2004, Elsevier (USA). All rights reserved.

enlarged folds are particularly noted in the fundus and body of the stomach (Fig. 1). However, the entire stomach can be involved in approximately 50% of cases and duodenal involvement may occur in 3%. Excessive mucus production and hyposecretion of acid occur. The selective loss of proteins across the gastric mucosa may lead to hypoalbuminemia. Thick gastric fold changes are not specific to Ménétrier's disease and it is necessary to consider other diseases, including Zollinger–Ellison syndrome, *H. pylori* infection, gastritis, cytomegalovirus infection, gastritis, lymphoma, eosinophilic gastritis, sarcoid, syphilis, histoplasmosis, and infiltrative diseases, including neoplasms and amyloid. Large-forcep biopsies and snare biopsies may aid the pathologic evaluation. Endoscopic ultrasound changes include 5-mm thickenings of the second hypoechoic layer and cystic changes.

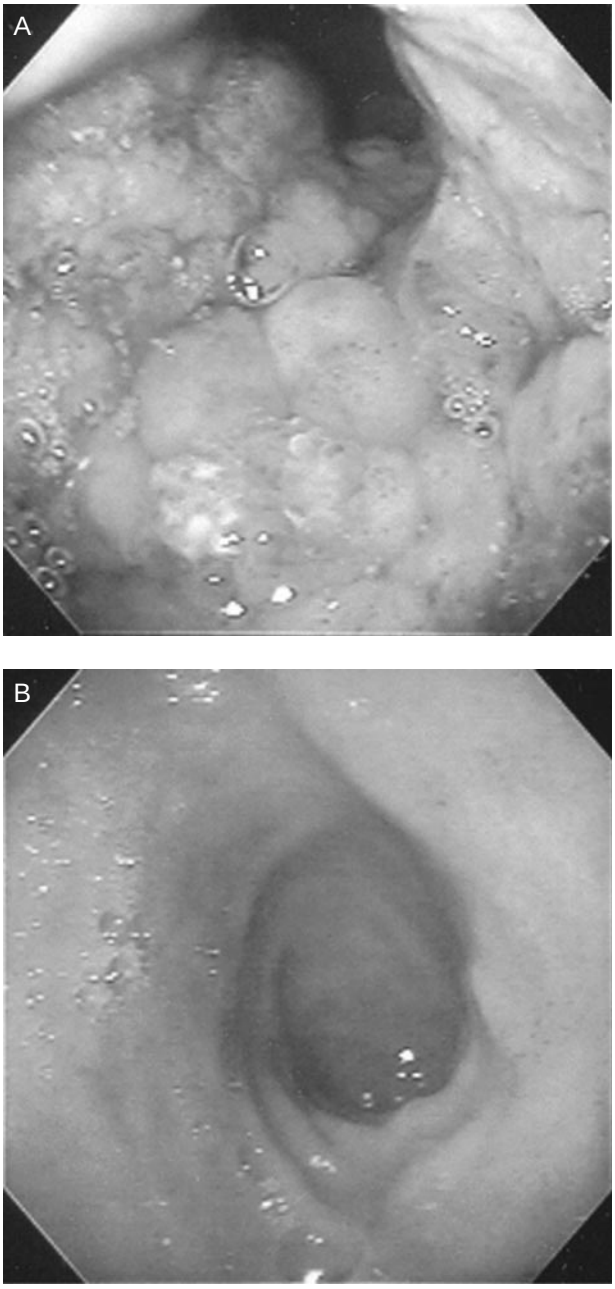


FIGURE 1 Gastroscopic appearance of Ménétrier's disease. (A) Note the enlarged folds, with excessive mucus production, in the gastric body and fundus. In the same patient, the distal stomach (B) appears fairly normal ("antral sparing").

Partial regression of abnormalities may occur but spontaneous resolution is rare. In a review of 40 patients, 55% improved with or without medical therapy at a mean followup of 7.6 years. No randomized trials are available to compare treatments. However, retrospective reviews and reports of various therapies are available. This experience indicates that no single agent uniformly leads to a clinical response. Anticholinergic therapy, the somatostatin analogue octreotide, histamine-2 (H2) receptor antagonists, and proton pump inhibitors have been used. Surgery is an option for severe disease and is required in around 10% of cases. Currently, an investigative protocol using an epidermal growth factor receptor antagonist is available for patients with severe disease. Multiple case series studies report the concomitant diagnosis of gastric cancer and Ménétrier's disease a initial presentation. Additionally, on followup, gastric cancer developed in over 7% of cases. The risk of developing gastric cancer suggests a need for continued endoscopic surveillance.

See Also the Following Articles

Cytomegalovirus • *Helicobacter pylori* • Transforming Growth Factor-β (TGF-β)

Further Reading

Badov, D., Lambert, J. R., Finaly, M., and Balazs, N. D. (1998). *Helicobacter pylori* as a pathogenic factor in Menetrier's disease. *Am. J. Gastoenterol.* **93**, 1976–1979.

Burdick, J. S., Chung, E. K., Tanner, G., Sun, M., Paciga, J. E., Cheng, J. Q., Washington, K., Goldenring, J. R., and Coffey, R. J. (2000). Treatment of Menetrier's disease with a monoclonal antibody against the epidermal growth factor receptor. *N. Engl. J. Med.* **343**, 1697–1701.

Meuwissen, S. G. M., Ridwan, B. U., Haper, H. J., and Innemee, G. (1992). Hypertrophic protein losing gastropathy. *Scan. J. Gastroenterol.* **27**, 1–7.

Sferra, T., Pawl, B. R., Qualman, S. J., and Li, B. U. K. (1996). Menetrier disease of childhood: Role of cytomegalovirus and transforming growth factor alpha. *J. Pediatr.* **128**, 213–219.

Wolfsen, H. C., Carpenter, H. A., and Talley, N. J. (1993). Menetrier's disease: A form of hypertrophic gastropathy or gastritis? *Gastroenterology* **104**, 1310–1319.

Microflora, Overview

E. T. T. L. TJWA* AND L. A. DIELEMAN†

**University Medical Centre St. Radboud, Nijmegen, The Netherlands, *University of North Carolina at Chapel Hill, and †University of Alberta, Canada*

commensalism State of colonization that results in either no damage or clinically inapparent damage to the host, though it can elicit an immune response.

germ free Gnotobiotic (Greek for "known flora") state in which all microflora are absent.

probiotics Living bacteria that colonize a host in amounts sufficient to alter the normal host microflora, resulting in a beneficial health effect for the host.

The normal intestinal microflora, which are mostly bacterial, consist of the resident populations of microbial species that inhabit the digestive tract of their host. Because "normal" is often undefinable, "indigenous" microflora, or biota, is a better term. The colonizing microbes are diverse: bacteria are in the majority, but fungi and protozoa are also present. Understanding the complex interplay between the indigenous biota, their associated metabolisms, and the host digestive tract forms the basis of understanding the microbial role in disease, but also in maintaining health. There are more than 400–500 bacterial species in the human colon of a single healthy individual, amounting to about 10^{12} viable bacteria per gram of colonic content. Considering this, the intestinal microflora can be viewed as a potential metabolic organ.

INTRODUCTION

Environmental and microbial factors contribute to the pathogenesis of many intestinal diseases. The role that indigenous microflora play in the health of their hosts has been point of controversy since Pasteur hypothesized a necessary but sometimes detrimental role, and Metchnikoff argued a more beneficial one. Most of the published work on (development of) the human gastrointestinal flora has been restricted to analysis of the fecal flora using conventional methods such as cultures. However, culture results from fecal microflora reflect only 50–80% of the total microscopic bacterial count. Studies using novel tools to identify greater numbers of microflora by molecular methods are emerging and have the potential to provide new insights on how microflora influence their host in health and disease.

IDENTIFICATION TECHNIQUES

Most of our knowledge of the microflora of the human digestive tract is derived from studies using traditional techniques of culture and microscopy. Most of the studies of intestinal microflora use fecal cultures, assuming that fecal microflora and large bowel microflora are equivalent. Microflora of the upper digestive tract have been studied through a sampling method, described by Gorbach, using a flexible plastic tube to aspirate samples. It is unlikely that all of the important organisms that are present in the intestinal crypts or gastric mucosa are present in such collected samples. The interpretation of microflora data is further complicated considering sampling, storage, and culture methods. Sampling in the highly inaccessible digestive tract is complex, and most samples are gathered in nonphysiologic situations, such as during surgery or postmortem. Ideally, the bulk of microflora should be stored in absolutely anaerobic conditions, which cannot typically be met in a standard setting. Only 10–20% of intestinal bacteria can be cultured, meaning that 80–90% of the microorganisms are uncultivatable. Novel techniques to identify microflora, without requiring culture, have therefore been established. One of these improved species identification techniques involves detection of bacterial 16S ribosomal RNA. This small ribosomal subunit contains the hypervariable V regions, which define the signature of phylogenetic groups and species of bacteria. Identification of genera, species, and even strains of the microflora of the digestive tract using detection methods for 16S ribosomal bacterial RNA sequences will substantially increase our knowledge of the composition of the intestinal microflora. In addition, molecular techniques such as denaturation or temperature gradient–pulsed gel electrophoresis (D/TGGE), fluorescence *in situ* hybridization (FISH), flow cytometry, and genetic fingerprinting with DNA microarrays have been used as novel microflora identification methods.

 Copyright 2004, Elsevier (USA). All rights reserved.

MICROFLORA OF THE DIGESTIVE TRACT

Microbial commensalism occurs within a few days of birth, with development of a succession of organisms as local microenvironments change. Commensalism begins perorally, as demonstrated in infants with congenital small bowel obstruction, in which a fecal-type flora is present proximal of the site of obstruction, whereas the distal bowel remains sterile. The first bacteria that colonize the human gut are derived from the aerobic and anaerobic flora of the birth canal; in neonates, *Escherichia coli*, *Clostridium* spp., *Streptococcus* spp., *Lactobacillus* spp., *Bacteroides* spp., and *Bifidobacterium* spp. predominate. After 4–6 days, the microflora of vaginally delivered, formula-fed, full-term newborns consist of mostly anaerobes, mainly *Bacteroides fragilis*, whereas the fecal microflora of breast-fed newborns consist of significantly more bifidobacteria. In contrast, the digestive tract of newborns delivered by cesarean section contains only 60% anaerobes (of which only 9% is *B. fragilis*). These results indicate the effects of delivery route and dietary constituents (breast milk or formula), but other early exogenous factors, such as eruption of teeth, gestational age, hospitalization, geographic variations, and sanitary conditions, can also determine the composition of the intestinal microflora. A few weeks after, birth, numerous colonizing organisms are present, maintaining an orderly and stable presence in the niches of the stomach, small intestine, and large intestine.

The stomach is often considered to be "sterile," defined as $<10^3$ colony-forming units (CFUs)/ml of gastric aspirate. The most important factor controlling the gastric flora is acidity. The bactericidal nature of gastric secretions is affected by alterations in pH. In health, when stomach pH is less than 2, gastric microflora consist of predominantly gram-positive aerobic bacteria, including streptococci, staphylococci, lactobacilli, and various fungi. Oral anaerobes such as *Peptostreptococcus, Fusobacterium*, and *Bacteroides* species are present in very low numbers. Hypochlorhydria and gastric surgery can increase the number of bacteria found in the stomach, which can include oral microflora as well as fecal-type bacteria. Since *Helicobacter pylori* was identified, it has become clear that many gastric microbes are *Helicobacter* species or closely related microaerophilic spiral organisms. *Helicobacter pylori* is considered a pathogen and is associated with peptic ulcers. In contrast to the pathogenic role of *H. pylori*, some non-*pylori* species are indigenous to the human stomach and could even have a beneficial role.

The small intestine serves as a transition zone between the relatively "sterile" upper digestive tract and the densely populated lower digestive tract. The key factor regulating commensalism in the small intestine is peristalsis. The composition of microflora of the proximal small intestine is similar to that of the stomach, although a higher concentration of the various microorganisms can be detected (10^5 CFUs/ml of jejunal aspirate). Anaerobic and coliform bacteria are present in relatively low concentrations.

TABLE I Dominant Genera of the Human intestinal Microflora (Most Common Species)

Genus	Species
Bacteroides	*thetaiotamicron, fragilis, vulgatus, ovatus*
Fusobacterium	*mortiferum, necrophorum*
Eubacterium	*aerofaciens, lentum, contortum*
Bifidobacterium	*infantis, adolescentus, longum*
Lactobacillus	*acidophilus, fermentum, plantarum*
Clostridium	*perfringens, bifermentans, ramosum*
Peptostreptococcus	*productus*
Enterococcus	*faecium, faecalis*

The microflora composition of the distal small intestine resembles that of the large intestine. In the colon, the microbial concentration is approximately 10^{12} CFUs/ml, whereas in the distal small bowel this is 10^7 to 10^8 CFUs/ml. The large intestine of individuals consuming a Western diet contains mostly obligate anaerobic bacterial species (Table I). These include gram-negative rods such as *Bacteroides* (predominantly), *Fusobacterium, E. coli, Enterobacter*, and *Klebsiella*, as well as gram-positive bacteria such as *Eubacterium*, but bifidobacteria, lactobacilli, clostridia, and gram-positive cocci such as *Peptostreptococcus, Enterococcus, Streptococcus*, and *Staphylococcus* are also found. Several bacterial species have been associated with pathogenic properties (*B. fragilis, Enterococcus faecium, Clostridium difficile, E. coli* O157:H7) as well with nonpathogenic characteristics (*Lactobacillus fermentum*).

ESTABLISHING A STABLE MICROFLORA

Each individual has a unique composition of indigenous microflora; within this population, there is both competition and cooperation. The microfloral makeup is shaped by various events exclusively connected to the specific host and its environment. Once established, many of the biota persist for years if not for the life of the host. However, there can be a marked variation in the complexity and stability of microflora between hosts. The stability of the microflora is remarkable considering the complexity of the ecologic environment of the digestive tract and the perpetual introduction of

microorganisms. This persevering microbial stability is the result of regulation of microflora and host interactions, even after multiple environmental insults.

The ability to colonize the intestinal tract correlates with the capacity to adhere to mucosal surfaces. Adherence capacity is not only bacterial species specific but is also dependent on host intestinal epithelial cell factors. Bacterial adherence is mediated by binding of microbial adhesins or lectins to complementary host epithelial cell receptors. These receptors consist of a variety of glycolipids, glycoproteins, and carbohydrate chain structures. However, not all microorganisms are capable of colonizing, because host defense mechanisms have evolved in order to block adherence of various species. Microbe-specific factors (production of specific secretory immunoglobulins, such as IgA; epithelial receptor analogues) as well as local environmental host factors (e.g., motility, the availability of essential nutrients, redox potential, and enteric secretions such as gastric acid and unconjugated bile) are means of controlling commensalism, maintaining the status quo, and preventing bacterial overgrowth. Interactions between the various microbes also form an important regulating factor. Resident flora may interact to promote or prevent the growth of other species and of themselves. Production of bacteriocin or short-chain fatty acids results in direct prevention of adherence and of growth of gram-positive bacteria such as the streptococci. Stimulation of the phagocytic system and other immunologic clearance systems has also been demonstrated to be a host resident flora defense mechanism against potential pathogens.

Environmental factors can alter the composition of the intestinal microflora. Chronic use of alcohol can induce a higher prevalence of gram-negative microorganisms in the upper digestive tract. Stress is also known to influence the gastrointestinal microflora, and *Lactobacillus* populations are among the species that are affected. Reports demonstrating a marked influence of diet on the gastrointestinal flora contradict studies comparing standardized institutional diets to random diets, which fail to demonstrate a significant effect of diet on the stable resident microflora.

Selective decontamination of (parts of) the digestive tract has been used in neutropenic patients and in many clinical situations to prevent postoperative wound infection. Although selective bowel decontamination can prevent opportunistic infections and bacterial translocation, the resident flora can shift toward a relative overgrowth of certain species, notoriously *C. difficile* and *Candida albicans*, resulting in disease. This suggests that some species of the indigenous flora can protect against bacterial and fungal pathogens. Antimicrobial agents can have a long-lasting effect on the normal gastrointestinal flora; amoxicillin treatment for only 7 days results in predominant antibiotic-resistant strains for up to 6 months.

BENEFICIAL MICROFLORA

Probiotic components of the intestinal bacterial flora are beneficial for the host. Probiotics include bacteria such as lactobacilli, bifidobacteria, and nonpathogenic *E. coli*. Lactobacilli play an important role in the maintenance of colonization resistance and in prevention of overgrowth of enteric pathogens. Lactobacilli species, albeit not predominant, are present throughout the gastrointestinal tract of healthy humans. Among the *Lactobacillus* species able to colonize the digestive tract are *Lactobacillus plantarum*, *Lactobacillus acidophilus*, *Lactobacillus fermentum*, and *Lactobacillus casei* strain GG. Probiotics can reduce fecal bacterial β-glucuronidase, a marker of intestinal bacterial metabolism, and produce antimicrobial substances that inhibit growth of anaerobic bacteria *in vitro*. Lactobacilli confer many other nutritional and protective benefits, ranging from fiber fermentation, removal of toxins, and enhancement of intestinal IgA secretion and macrophage responses. Probiotics are also capable of preventing adherence of pathogenic bacteria to intestinal epithelial cells. Several human clinical trials have shown the efficacy of several species of probiotics in viral- and bacterial-induced infectious colitis as well as in antibiotic-associated diarrhea. Rectal administration of native murine *Lactobacillus* spp. prevented colitis in interleukin (IL-10)-deficient mice and a similar effect was shown by oral prebiotic lactulose, which stimulated the growth of endogenous *Lactobacillus* species. Continuous oral administration of *Lactobacillus plantarum* 299v to germ-free IL-10-deficient mice colonized with specific pathogen-free (SPF) flora could attenuate established colitis, whereas oral administration of *Lactobacillus rhamnosus* GG prevented relapse of colitis in SPF HLA-B27 transgenic rats after antibiotic treatment. A probiotic preparation including four species of *Lactobacillus*, three species of *Bifidobacterium*, and *Streptococcus salivarius* (VSL#3) maintained remission of refractory pouchitis in patients after transient antibiotic therapy. This probiotic cocktail (VSL#3) was also beneficial in the treatment of colitis in IL-10-deficient mice.

See Also the Following Articles

Bacterial Overgrowth • Candidiasis • *Helicobacter pylori*

Further Reading

Bullen, C. L., Tearle, P. V., and Willis, A. T. (1976). Bifidobacteria in the intestinal tract of infants: An *in vitro* study. *J. Med. Microbiol.* **9**, 325.

Dieleman, L. A., Goerres, M. S., Arends, A., Sprengers, D., Torrice, C., Hoentjen, F., Grenther, W. B., and Sartor, R. B (2003). *Lactobacillus* GG prevents recurrence of colitis in HLA-B27 transgenic rats after antibiotic treatment. *Gut* **52**, 370–376.

Falk, P. G., Hooper, L. V., Midtvedt, T., and Gordon, J. I. (1998). Creating and maintaining the gastrointestinal ecosystem: What we know and need to know from gnotobiology. *Microbiol. Mol. Biol. Rev.* **62**, 1157–1170.

Gionchetti, P., Gionchetti, P., Rizzello, F., Venturi, A., Brigidi, P., Matteuzzi, D., Bazzocchi, G., Poggioli, G., Miglioli, M., and Campieri, M. (2000). Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: A double-blind, placebo-controlled trial. *Gastroenterology* **119**, 305–309.

Hart, A. L., Stagg, A. J., Graffner, H., Glise, H., Falk, P., and Kamm, M. A. (eds.). (2002). "Gut Ecology." Martin Dunitz Ltd., London.

Hooper, L. V, Wong, M. H., Thelin, A., Hansson, L., Falk, P. G., and Gordon, J. I. (2001). Molecular analysis of commensal host–Microbial relationships in the intestine. *Science* **291**, 881–884.

Madsen, K. L. (2001). The use of probiotics in gastrointestinal disease. *Can. J. Gastroenterol.* **15**, 817–822.

Rotimi, V. O., and Duerden, B. I. (1981). The development of the bacterial flora in normal neonates. *J. Med. Microbiol.* **14**, 51–62.

Simon, G. L., and Gorbach, S. L. (1984). Intestinal flora in health and disease. *Gastroenterology* **86**, 174–193.

Tannock, G. W. (2001). Molecular assessment of intestinal microflora. *Am. J. Clin. Nutr.* **73**, 410S–414S.

Wilson, K. H., and Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Appl. Environ. Microbiol.* **62**, 2273–2278.

Migrating Motor Complex

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

enterohepatic circulation Movement of bile from the gallbladder to the small intestine and from the small intestine to the hepatic portal vein, then on to the liver, where it is again secreted for storage in the gallbladder.

interdigestive state Period of time between the end of the digestion and absorption of a meal and the ingestion of the next meal.

lithogenesis Formation of stones in the gallbladder or kidneys.

The motility detected in the stomach and small intestine during the interdigestive state consists of a migratory pattern of contractions called the migrating motor complex. This motility is present in most mammals, including humans, and is seen in conscious states and during sleep. It starts in the distal stomach as an increase in the strength of the regularly occurring propulsive contractions. From the stomach, the activity front progresses into the duodenum and on down the small intestine. In humans and other large mammals, it takes 80–120 minutes for the activity front to reach the ileum. As one activity front terminates in the ileum, another starts in the stomach.

INTRODUCTION

The interdigestive pattern of gastrointestinal motility begins from 2 to 3 hours after a meal and after digestion and absorption of the nutrients in the meal have taken place. The contractile behavior in the interdigestive state is detected either by placement of pressure sensors into the lumen or by attachment of electrodes to the gastrointestinal surface. At a single recording site, the migrating motor complex (MMC) pattern consists of three periods. The first is a silent period, called

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Bullen, C. L., Tearle, P. V., and Willis, A. T. (1976). Bifidobacteria in the intestinal tract of infants: An *in vitro* study. *J. Med. Microbiol.* **9**, 325.

Dieleman, L. A., Goerres, M. S., Arends, A., Sprengers, D., Torrice, C., Hoentjen, F., Grenther, W. B., and Sartor, R. B (2003). *Lactobacillus* GG prevents recurrence of colitis in HLA-B27 transgenic rats after antibiotic treatment. *Gut* **52**, 370–376.

Falk, P. G., Hooper, L. V., Midtvedt, T., and Gordon, J. I. (1998). Creating and maintaining the gastrointestinal ecosystem: What we know and need to know from gnotobiology. *Microbiol. Mol. Biol. Rev.* **62**, 1157–1170.

Gionchetti, P., Gionchetti, P., Rizzello, F., Venturi, A., Brigidi, P., Matteuzzi, D., Bazzocchi, G., Poggioli, G., Miglioli, M., and Campieri, M. (2000). Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: A double-blind, placebo-controlled trial. *Gastroenterology* **119**, 305–309.

Hart, A. L., Stagg, A. J., Graffner, H., Glise, H., Falk, P., and Kamm, M. A. (eds.). (2002). "Gut Ecology." Martin Dunitz Ltd., London.

Hooper, L. V, Wong, M. H., Thelin, A., Hansson, L., Falk, P. G., and Gordon, J. I. (2001). Molecular analysis of commensal host–Microbial relationships in the intestine. *Science* **291**, 881–884.

Madsen, K. L. (2001). The use of probiotics in gastrointestinal disease. *Can. J. Gastroenterol.* **15**, 817–822.

Rotimi, V. O., and Duerden, B. I. (1981). The development of the bacterial flora in normal neonates. *J. Med. Microbiol.* **14**, 51–62.

Simon, G. L., and Gorbach, S. L. (1984). Intestinal flora in health and disease. *Gastroenterology* **86**, 174–193.

Tannock, G. W. (2001). Molecular assessment of intestinal microflora. *Am. J. Clin. Nutr.* **73**, 410S–414S.

Wilson, K. H., and Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Appl. Environ. Microbiol.* **62**, 2273–2278.

Migrating Motor Complex

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

enterohepatic circulation Movement of bile from the gallbladder to the small intestine and from the small intestine to the hepatic portal vein, then on to the liver, where it is again secreted for storage in the gallbladder.

interdigestive state Period of time between the end of the digestion and absorption of a meal and the ingestion of the next meal.

lithogenesis Formation of stones in the gallbladder or kidneys.

The motility detected in the stomach and small intestine during the interdigestive state consists of a migratory pattern of contractions called the migrating motor complex. This motility is present in most mammals, including humans, and is seen in conscious states and during sleep. It starts in the distal stomach as an increase in the strength of the regularly occurring propulsive contractions. From the stomach, the activity front progresses into the duodenum and on down the small intestine. In humans and other large mammals, it takes 80–120 minutes for the activity front to reach the ileum. As one activity front terminates in the ileum, another starts in the stomach.

INTRODUCTION

The interdigestive pattern of gastrointestinal motility begins from 2 to 3 hours after a meal and after digestion and absorption of the nutrients in the meal have taken place. The contractile behavior in the interdigestive state is detected either by placement of pressure sensors into the lumen or by attachment of electrodes to the gastrointestinal surface. At a single recording site, the migrating motor complex (MMC) pattern consists of three periods. The first is a silent period, called

Copyright 2004, Elsevier (USA). All rights reserved.

Phase I, which has no contractile activity; the second period, Phase II, follows, consisting of irregularly occurring contractions. Phase II is followed by Phase III, which consists of regularly occurring contractions, and then Phase I begins again at the end of Phase III, to repeat the cycle. Multiple sensors positioned on the stomach and continuing at equally spaced intervals along the small intestine reveal that activity representing Phase II and Phase III starts in the stomach and propagates slowly down the intestine. The migratory nature of the contractions is the basis for describing the interdigestive pattern as the MMC.

Migrating Motor Complex Activity Front

The MMC, at any given time, occupies a limited length of intestine termed the "activity front." As such, the activity front has an upper boundary, where it begins, and a lower boundary, where it ends (Fig. 1). The activity front slowly advances (migrates) along the intestine at a rate that progressively slows as it approaches the ileum. Peristaltic propulsion of luminal contents in the anal direction occurs repeatedly within the activity front. Each successive peristaltic wave starts at the upper boundary of the activity front, propagates over the length of the activity front, and stops at the lower boundary. Successive peristaltic waves start a short distance further along the intestine than the previous one and propagate a short distance beyond the lower boundary where the last one stopped. This causes the activity front to migrate slowly down the intestine, sweeping the lumen clean as it progresses.

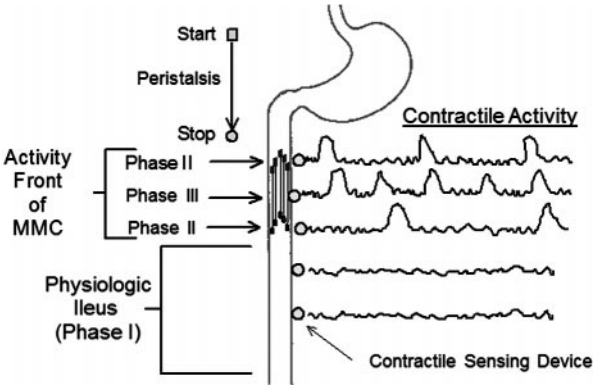


FIGURE 1 The activity front of the migrating motor complex consists of repetitive cycles of peristaltic propulsion as it slowly advances down the intestine. Peristaltic propulsion of luminal contents in the aboral direction occurs between the orad and aborad boundaries of the activity front. Successive waves of peristaltic propulsion start on average slightly further along in the aborad direction and propagate on average slightly beyond the boundary where the previous one stopped. This is responsible for the impression that the activity front migrates slowly in the aborad direction. Contractile activity described as Phase II or Phase III occurs because of irregularity of the orad boundary where each consecutive peristaltic wave starts and irregularity of the arrival of the contractile wave at the aborad boundary. As the leading and trailing edges of the activity front pass a contractile sensor, only the peristaltic waves that pass the sensor are recorded, and this gives the appearance of irregular contractions (i.e., Phase II). When the central region of the activity front passes a sensor, every recurrent peristaltic wave is recorded, giving the appearance of regular contractions at higher frequency (i.e., Phase III). Physiologic ileus is in effect in the regions of intestine outside the activity front (i.e., Phase I).

In humans, the time between cycles is longer during the day than at night. The activity front travels at 3–6 cm per minute in the duodenum and progressively slows to 1–2 cm per minute in the ileum. The speed of travel of the activity front is much slower that of the electrical slow waves, action potentials, and peristaltic contractions that make up the activity front. Slow waves with associated action potentials and contractions within the activity front travel 10 times faster than the activity front.

Phase III of the MMC reflects the propulsive contractions within the activity front as it slowly progresses along the intestine. Phase III lasts for 8–15 minutes as the activity front passes by a single recording site (Fig. 1). The shortest duration of Phase III is in the duodenum, with a progressive increase in duration occurring as the complex migrates toward the ileum. Each slow wave, with action potentials and associated contractions, reflects a single propulsive event occurring within the activity front. Phases I and II each occupy about half of the rest of the cycle as recorded at a single site. The irregular electrical and contractile activity of Phase II reflects the activity at the leading and trailing edges of the activity front as it moves along the intestine. At the leading edge, the irregularity reflects failure of some of the peristaltic waves within the front to propagate as far as others. At the trailing edge, irregular activity indicates failure of some slow-wave cycles to initiate the propulsive contractions as the end of the front passes the point on the bowel where the electrode or device for detecting contractions is placed.

A MEAL STOPS THE MIGRATING MOTOR COMPLEX

The MMC continues until it is terminated by the ingestion of food. Ingestion of a meal with sufficient caloric content halts the MMC. Termination requires contact of the meal with the upper digestive tract, because intravenous feeding does not end the fasting pattern. The

short time required for postprandial termination of the MMC suggests a neural or hormonal mechanism. Two hormones, gastrin and cholecystokinin, which are released on ingestion of a meal, terminate the MMC when injected intravenously. Nevertheless, this effect occurs only in the stomach and upper small intestine, not in the ileum.

The MMC is one of the motor programs in the program library of the enteric nervous system and requires that the enteric nervous system be fully functional. The MMC continues in the small intestine after interruption of all inputs from the central nervous system, but stops when it reaches a region of the intestine where the enteric nervous system has been interrupted experimentally. In experimental animal models, MMCs occur in autotransplanted stomachs that have no connections with the central nervous system. Presumably, command signals to the enteric neural circuits are necessary for initiation of the interdigestive pattern; however, whether the commands are neural or hormonal or both is unknown. Although levels of the hormone motilin increase in the blood at the onset of the interdigestive state, it is unclear whether motilin is a hormonal command to the enteric nervous system, to call up the MMC from its program library, or is released secondary to start up of the MMC program.

FUNCTIONAL SIGNIFICANCE

The adaptive significance of the MMC appears to be a mechanism for clearing indigestible debris, mucus secretions, and sloughed mucosal epithelial cells from the stomach and intestine during the fasting state. Large indigestible particles are emptied from the stomach only during the interdigestive state. Contraction of the gallbladder and delivery of bile into the duodenum occur as the activity front approaches the site where the bile duct enters the duodenum. Once in the duodenum, the activity front carries the bile downward to the ileum, where it is absorbed and transported to the liver to be secreted and again stored in the gallbladder. Contraction of the gallbladder and emptying of bile into the intestine are components of the neural program for the MMC and may be a mechanism that prevents bile from becoming overly concentrated in the gallbladder during the interdigestive state, when the digestive functions of bile are not required. Because gallstones form when bile becomes overly concentrated, the MMC may ensure movement of bile in the enterohepatic circulation during the interdigestive state and thereby protect against lithogenesis.

Bacterial overgrowth in the small intestine is associated with absence of the MMC in human patients. This suggests that the "housekeeper" function of the MMC may play a role in preventing the overgrowth of microorganisms that might occur in the small intestine if the contents were allowed to stagnate in the lumen.

See Also the Following Articles

Basic Electrical Rhythm • Duodental Motility • Enteric Nervous System • Gastric Motility • Pylorus • Small Intestinal Motility

Further Reading

Code, C. (1979). The interdigestive housekeeper of the gastrointestinal tract. *Perspect. Biol. Med.* **22**, S49–S55.

Code, C., and Marlett, J. (1975). The interdigestive myoelectric complex of the stomach and small bowel of dogs. *J. Physiol. (Lond.)* **246**, 289–309.

Szurszewski, J. (1969). A migrating myoelectric complex of the canine small intestine. *Am. J. Physiol.* **217**, 1757–1763.

Szurszewski, J. (1998). A 100-year perspective on gastrointestinal motility. *Am. J. Physiol.* **274**, G447–G453.

Wingate, D. (1981). Backwards and forward with the migrating complex. *Dig. Dis. Sci.* **26**, 641–666.

Wood, J., Alpers D., and Andrews, P. (2000). Fundamentals of neurogastroenterology: Basic science. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 31–90. Degnon Associates, McLean, Virginia.

Minimally Invasive Surgery

THOMAS E. CLANCY AND DAVID C. BROOKS
Brigham and Women's Hospital and Harvard University

fundoplication Fixation of the gastric fundus, usually around the distal esophagus, for the purpose of treating gastroesophageal reflux disease.
laparoscopic surgery The use of laparoscopy to perform diagnostic and therapeutic procedures in the abdominal cavity.
laparoscopy Visual inspection of the abdominal cavity via an endoscope or telescope, most often utilizing video monitoring via attachment of a camera to the endoscope.
pneumoperitoneum The presence of air in the peritoneal cavity; the condition is created to provide a working space in the abdomen.

With the introduction of minimally invasive surgery in the late 1980s, a revolutionary change has taken place in the way in which surgical problems are approached and the way in which operations are performed. Laparoscopic cholecystectomy was first performed in 1987 and subsequently was rapidly adopted by general surgeons as the standard of care of routine cholecystectomy. The ability to perform surgery with minimal trauma and the possibility for reduced postoperative pain, shorter hospitalization, and less disability led to the rapid development of a wide variety of laparoscopic applications for intra-abdominal operations. Minimally invasive approaches are now commonplace for the treatment of common bile duct stones, gastroesophageal reflux disease, preoperative staging of many intra-abdominal malignancies, splenectomy, and appendectomy in addition to cholecystectomy. The current article, although not comprehensive, will focus primarily on minimally invasive surgery for gastrointestinal disorders.

BACKGROUND

The majority of laparoscopic surgery utilizes carbon dioxide (CO_2) gas insufflation into the peritoneal cavity to a pressure of 10 to 15 mm Hg. The pneumoperitoneum allows for a working space within the abdomen. Carbon dioxide has the advantage of being noncombustible as well as rapidly eliminated via systemic absorption and excretion through the lungs. Pneumoperitoneum is established by either a closed or an open technique. With the closed method, CO_2 is insufflated via a Veress needle placed into the abdominal cavity and the initial trochar and sheath are placed blindly. In the open technique, a small incision is made and a sheath without the trochar (Hasson cannula) is inserted under direct vision. The laparoscope is inserted through this sheath and accessory sheaths of 5 or 10 mm in diameter are inserted under direct vision to allow access for laparoscopic instruments. The use of a video camera attached to the eyepiece of the laparoscope allows magnification as well as the ability for multiple viewers to see the operative field.

Many studies have demonstrated that laparoscopic surgery has a reduced effect on postoperative cardiac performance and perioperative pulmonary function. Stress responses, as measured by cortisol and catecholamine levels, are decreased compared to open surgery. Cellular immune function is preserved to a greater degree following laparoscopic procedures. Clinically, these benefits translate into decreased postoperative pain and ileus, more rapid return to full activity, shorter hospitalization, and often decreased costs.

Despite the advantages of minimally invasive procedures compared to standard "open" surgical approaches, several disadvantages exist. The two-dimensional video image limits perspective, there is somewhat limited freedom of movement of instruments, there is little tactile feedback when manipulating tissues, and potential adverse effects of carbon dioxide pneumoperitoneum have been described, including hypercarbia, acidosis, and arrhythmia. Clinically, complications with CO_2 pneumoperitoneum are quite rare. When minimally invasive procedures were introduced, several conditions were considered absolute contraindications to laparoscopic surgery, including pregnancy, bowel obstruction, obesity, cirrhosis, coagulopathy, and acute cholecystitis. With improvements in technique and increased experience, these are now considered only relative contraindications or are not considered contraindications at all (Table I).

Prior surgery was originally considered a contraindication due to the risk of bowel injury from trochar insertion with multiple adhesions. However, the formation and extent of adhesions are unpredictable

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Contraindications to Laparoscopic Surgery

Absolute	Relative
Unable to tolerate general anesthesia	Severe obstructive lung disease
Uncorrected coagulopathy	Prior abdominal surgery; adhesions
Generalized peritonitis	Morbid obesity
Other condition requiring laparotomy	Pregnancy
	Cirrhosis/portal hypertension
	Possible malignancy (other than staging procedures)

and laparoscopic adhesiolysis can be performed. Obese patients may present difficulties with trochar insertion and cannula displacement; however, with the use of longer trochars and slightly higher insufflation pressures, few patients with morbid obesity need be denied laparoscopy. With pregnancy, concerns about the effects of CO_2 pneumoperitoneum on the mother and fetus include increased intra-abdominal pressure on venous return, cardiac output, and uterine blood flow, with possible maternal and fetal acidosis. Risks include teratogenesis and miscarriage in the first trimester and preterm labor in the third trimester, with decreased risks of all complications in the second trimester. There have been multiple reports of safe laparoscopic procedures in all trimesters and pregnancy is therefore considered only a relative contraindication.

LAPAROSCOPIC CHOLECYSTECTOMY

First performed in the late 1980s, laparoscopic cholecystectomy was rapidly adopted as the method of choice for treating symptomatic cholelithiasis and some cholecystitis. The procedure was first popularized in the community setting and soon thereafter in academia. By 1993, it was estimated that greater than 85% of all cholecystectomies in the United States were performed laparoscopically. Now just over 10 years later, laparoscopic cholecystectomy is one of the most commonly performed general surgery operations.

The indications for laparoscopic cholecystectomy are the same as for open cholecystectomy and primarily include symptomatic cholelithiasis (biliary colic), cholecystitis, gallstone pancreatitis, and symptomatic biliary dyskinesia. Less than 20% of people with asymptomatic gallstones develop symptoms over many years, so generally cholecystectomy is not performed in these cases due to potential risks of surgery. With typical biliary colic, the only preoperative evaluation necessary is an ultrasound to document the presence and number of gallstones, gallbladder wall thickness, pericholecystic fluid, sludge, and the diameter of the common bile duct. A technetium-99-iminodiacetic acid scan may be useful in confirmation of acute cholecystitis. Biliary dyskinesia can be shown using cholecystokinin-stimulated gallbladder emptying of radionuclide. Endoscopic retrograde cholangiopancreatography (ERCP) is indicated for patients with gallstone pancreatitis without improvement or for those with abnormal liver function tests suggesting common bile duct (CBD) stones.

General anesthesia is utilized in all cases. The camera is inserted through a 10 mm periumbilical trochar; additional "working" trochar sites usually include two 5 mm subcostal trochars and a 10 mm epigastric trochar. A reverse trendelenberg position is used to increase visualization. The gallbladder is retracted superiorly over the liver and toward the diaphragm with a grasping instrument placed through the lateral subcostal port, and the neck of the gallbladder is retracted anteriorly and laterally using another grasping instrument placed in the medial subcostal trochar. Using blunt and sharp dissection with instruments placed through the 10 mm epigastric port, adhesions are dissected off the gallbladder, the cystic duct and artery are identified, clipped, and divided, and the gallbladder is dissected from the liver with electrocautery. The gallbladder is removed through the umbilical port after the laparoscope is moved to the epigastric cannula.

Cholangiography is usually performed selectively based on either history or anatomy. History of jaundice, pancreatitis, ultrasound evidence of choledocholithiasis, elevations of bilirubin, alkaline phosphatase, amylase, or lipase, and presence of a large cystic duct and small gallstones all have been considered indications. Before the cystic duct is clipped and divided, it may be incised anteriorly with introduction of a small catheter. Evaluation of the cholangiogram includes the size of the common bile duct, filling defects, location of the junction between cystic duct and CBD, free flow of contrast into the duodenum, and anatomy of the proximal biliary tree.

Conversion to open cholecystectomy occurs in 1.8 to 8.5% of cases and is usually greater earlier in a surgeon's career. Reasons for conversion include suspected bile duct or blood vessel injury, unclear anatomy, failure to progress in dissection, or unexpected pathology. Acute cholecystitis, if approached surgically more than 48 to 72 h after presentation, can provide unique challenges. The gallbladder is inflamed, thick-walled, and tensely distended with thickened and edematous surrounding tissue. Initiation of laparoscopic

cholecystectomy during this subacute phase will likely result in an increased chance of conversion to open cholecystectomy. With regards to other complications, most series report major complications in the 0.7 to 4.0% range, with bile duct injuries under 0.5–0.7%. Although low, this still exceeds the rate of bile duct injuries seen with open cholecystectomy. Mortality is extremely rare.

LAPAROSCOPIC CHOLEDOCHOLITHOTOMY

Choledocholithiasis is present in approximately 10% of patients who present with cholelithiasis. Definitive treatment of these patients includes clearance of the entire biliary ductal system. With the increased use of laparoscopic cholecystectomy, many surgeons have turned to methods such as ERCP. However, laparoscopic methods may be used for the management of CBD stones, including balloon catheter manipulation and choledochoscopy.

If CBD stones are suspected preoperatively, one option is to attempt ductal treatment preoperatively with ERCP and stone extraction with or without sphincterotomy. This is successful in clearing the duct in over 90% of cases, but is associated with slight morbidity (5–19%). Similarly, laparoscopic common bile duct exploration is successful in clearing the duct in over 90% of cases. Small CBD stones detected on intraoperative cholangiogram can usually be flushed out with a catheter threaded into the CBD after administration of IV glucagon, which relaxes the sphincter. If unsuccessful, a 4-French Fogarty balloon catheter may be used to retrieve stones and debris under fluoroscopic monitoring; similarly, a stone retrieval basket may be inserted into the cystic duct and followed with fluoroscopic guidance. If stones are large and the ducts are large (5–12 mm), extraction may be performed using a choledochoscope and stone basket or flexible electrohydraulic lithotriptor. Dilation of the cystic duct may be required to allow insertion of a 10-French choledochoscope. If stones are large or the cystic duct cannot be used to gain access to the CBD, the anterior wall of the CBD may be incised directly to allow insertion of a larger choledochoscope. The choledochotomy is traditionally closed over a T-tube, although primary closure of the choledochotomy is widely reported. Failure of the above techniques may require either postoperative ERCP and papillotomy or laparotomy and transduodenal sphincteroplasty. Generally, sphincterotomy is difficult to perform via the laparoscopic antegrade approach and is performed with the endoscopic retrograde approach.

LAPAROSCOPIC ANTIREFLUX SURGERY

Laparoscopic esophageal surgery is second only to biliary surgery in frequency of minimally invasive procedures in general surgery practice. Most common of these procedures on the esophagus is laparoscopic fundoplication for gastroesophageal reflux disease (GERD). Mechanisms thought to be important for the establishment of reflux control with fundoplication include (1) reconstruction of anatomical relations at the gastroesophageal junction, i.e., diaphragmatic crura, and (2) fixation of an adequate length of lower esophageal sphincter in the abdomen. Fundoplication is indicated for severe GERD, including patients with esophagitis, stricture, Barrett's esophagus, and patients dependent on proton pump inhibitors for symptomatic control. The most common indication is GERD in patients who have residual symptoms on maximal medical therapy or who are unwilling to accept lifelong medical therapy. Of note, fundoplication does not cause regression of Barrett's metaplasia and does not protect against malignant transformation and patients with Barrett's esophagus must therefore undergo long-term endoscopic surveillance for the presence of dysplasia. Whereas intestinal metaplasia and perhaps low-grade dysplasia are indications for an antireflux procedure, higher-grade dysplasia may be an indication for preemptive resection of the esophagus.

All patients under consideration for antireflux surgery should undergo a detailed diagnostic evaluation to corroborate the diagnosis and rule out other pathology. Upper endoscopy allows examination and/or biopsy of the esophagus, gastroesophageal junction (GEJ), stomach, and duodenum. Barrett's esophagus, as well as strictures or rings, may be identified. Distance to the GEJ is measured to rule out esophageal shortening. Twenty-four-hour pH monitoring, the definitive test for reflux, should be performed if the diagnosis is in question. A pH drop below 4 for more than 4% of a 24 h period is considered abnormal. Esophageal manometry must be performed to detect unsuspected functional motility disorders such as achalasia or scleroderma that would cause difficulty in propulsion of food into the stomach after a fundoplication. Barium swallow (esophagography) may demonstrate hiatal hernia, stricture, and esophageal shortening, but generally endoscopy is used for this anatomic information. Barium swallow is therefore not routinely utilized. Gastric emptying studies are employed selectively in patients with gastric symptoms, diabetes, peptic ulcer diseases, or history of bezoar. Delayed gastric emptying may require concurrent pyloroplasty or pyloromyotomy and may in fact indicate diffuse intestinal motility dysfunction.

One of the most commonly performed antireflux procedures is the Nissen fundoplication, which involves a 360° wrap of the fundus around the GEJ. The intra-abdominal esophagus and distal mediastinal esophagus are dissected and the short gastric vessels are divided. The right and left diaphragmatic crura are approximated with interrupted sutures and the fundus of the stomach is wrapped behind the esophagus, usually with a bougie in the esophagus. A fundoplication is created with interrupted sutures including two sides of the fundus and the esophagus. The Toupet fundoplication, similar in initial dissection but utilizing a 270° wrap, is indicated in patients with GERD but with poor esophageal motility. Such patients may otherwise experience postoperative dysphagia with a full 360° wrap. In patients with esophageal shortening (approximately 5% of patients presenting for surgery), a standard fundoplication may "slip" or herniate postoperatively. Such patients may be candidates for a Collis-Nissen technique to elongate the esophagus. This procedure uses stapling to create a tube of stomach as a continuation of the distal esophagus that easily lies beneath the diaphragm and may be followed by a standard fundoplication.

Multiple studies have shown that laparoscopic fundoplication can relieve typical reflux symptoms of regurgitation and dysphagia in more than 90% of patients at up to 3 years follow-up. Dysphagia is common postoperatively but commonly resolves by 3 months. Rare complications include gastric and esophageal perforation (0.5–0.8%), conversion to open technique (2–3%), pneumothorax or pneumomediastinum, and rarely bleeding.

OTHER APPLICATIONS—MINIMALLY INVASIVE ESOPHAGEAL SURGERY

Paraesophageal hernias, involving herniation of the gastric fundus into the thorax with or without cephalad displacement of the GEJ, may be associated with significant complications. Although the presence of an asymptomatic paraesophageal hernia is not universally recognized as an indication for surgery, given the risks of obstruction, perforation, and emergency surgery, most surgeons recommend elective repair. Laparoscopic repair of paraesophageal hernias is well described. However, volvulus of the stomach is associated with difficulty in identification of anatomy and the laparoscopic approach can be significantly more difficult than standard laparoscopic fundoplication. Surgery requires reduction of the hernia and removal of the entire peritoneal sac, followed by closure of the diaphragm. Given a high association with reflux and likely disruption of the normal antireflux mechanism during surgery, a fundoplication is performed after reduction of the hernia to maintain the GEJ in an intra-abdominal position. Complication rates are slightly higher than those for standard fundoplication.

Treatment of achalasia by eliminating the functional gastroesophageal obstruction includes pharmacologic methods, endoscopic injection of botulinum toxin, pneumatic dilation, and surgical cardiomyotomy. Surgery has been associated with some of the more durable symptomatic relief reported. Minimally invasive Heller myotomy can be performed through abdominal or thoracic approaches. Through the abdominal approach, the distal esophagus is mobilized and a myotomy is created with electrocautery for a distance of 4–6 cm. Taking care to preserve the anterior vagus nerve, the myotomy is carried 1 cm onto the anterior stomach. To minimize postoperative reflux after dissection of the hiatus, myotomy is followed by fundoplication, usually the Toupet procedure. Reported results are excellent, with approximately 90% of patients reporting good or excellent relief of dysphagia.

MINIMALLY INVASIVE GASTRIC SURGERY

Unlike many other commonly accepted laparoscopic applications, minimally invasive gastric surgery is largely in its infancy. Although applications exist for laparoscopic management of peptic ulcer disease, most minimally invasive gastric surgery consists of resection for benign and malignant masses. Critical to these procedures is the need to follow strict oncologic principles for resection of malignancy.

Rarely will optimal medical therapy for intractable peptic ulcer disease fail, leading to surgical intervention. Highly selective vagotomy (HSV) is associated with recurrence rates under 15% with a very low rate of complications. In the presence of a prepyloric ulcer (type II or III gastric ulcer), however, recurrence rates of greater than 30% are reported. Similarly, HSV is difficult with inflammation on the lesser curve of the stomach often present in type I gastric ulcers. Unlike duodenal ulcers, gastric ulcers may be managed best with either distal gastrectomy or truncal vagotomy with drainage procedure. Laparoscopic truncal vagotomy with gastrojejunostomy has been described but is not considered a standard of practice.

Laparoscopic gastric wedge resections may be performed for benign lesions such as leiomyomas. However, laparoscopic resection of gastric carcinoma

remains experimental. Although the role of staging laparoscopy in gastric carcinoma is well recognized (see below), pertinent questions include the sufficiency of lymphadenectomy with laparoscopy and the seeding of trochar sites with tumor cells.

MINIMALLY INVASIVE SURGERY FOR MORBID OBESITY

Severe obesity, defined as twice the ideal body weight or a body mass index (BMI) of more than 40 kg/m^2, is a source of serious health problems for millions people in the United States. However, conservative treatment is notoriously unsuccessful, with failure of more than 95% of treatment attempts. A number of surgical procedures have been developed to produce weight loss; two performed commonly today by laparotomy include the vertical banded gastroplasty (VBG) and the Roux-en-Y gastric bypass (RYGB). Due to the complexity of performing these procedures laparoscopically in morbidly obese patients, the minimally invasive variants of bariatric operations were slow to emerge but as experience has been gained they are more frequently performed. It is thought, however, that the well-documented benefits of laparoscopic surgery in nonobese patients may be more significant in the obese patient with associated higher risk for cardiopulmonary, infectious, and wound-related morbidity.

The VBG is primarily a gastric restrictive operation. Stapling techniques are utilized to create a narrow gastric pouch, thereby restricting enteral intake. In small series, patients undergoing laparoscopic VBG lost roughly 50% of excess weight at 2-year follow-up, results comparable to those with traditional VBG. A newer procedure, adjustable gastric banding, has been performed in some centers with encouraging results. The surgical approach is similar to the VBG but an inflatable band is placed around the 15 to 30 ml gastric pouch; band restriction is adjustable by manipulating a subcutaneous port on the abdominal wall. Weight loss is generally comparable to that obtained by VBG. Complications may be more common, with approximately one-third of patients requiring repeat surgery within 3 years, often for band erosion. The RYGB, which combines gastric restriction with intentional malabsorption, is now the most commonly performed bariatric operation in the United States. Long-term weight loss is reported to be far more durable than with the VBG, with 50–60% of excess body weight commonly reported at 5- to 15-year follow-up. The laparoscopic approach mimics the open procedure, which creates a 15 to 30 ml gastric pouch with an anastamosis to a 75 to 100 cm retrocolic, retrogastric Roux-en-Y limb of jejunum. Short (1- to 4-year) follow-up of patients after laparoscopic RYGB suggests weight loss similar to that obtained with the open technique, with 65–80% of extra weight lost. Complications are reported to be slightly higher with laparoscopic bariatric surgery, particularly leaks at the gastrojejunal anastamosis. Given the demonstrated benefits of gastric bypass for weight loss and the potential advantages of the minimally invasive approach, the incidence of laparoscopic RYGB will likely continue to increase.

MINIMALLY INVASIVE SURGERY FOR STAGING OF MALIGNANCY

Precise staging information is paramount in the management of patients with malignancy, for both prognostic and therapeutic considerations. Preoperative imaging modalities, including computerized tomography (CT), magnetic resonance imaging, ultrasound, and endoscopic procedures, although valuable, can often understage and lead to unnecessary surgery. The use of laparoscopy, particularly in combination with laparoscopic ultrasound, has been shown to improve the accuracy of staging, particularly in the recognition of metastases. Although surgery provides the only chance of cure with localized disease, it has no effect on outcome in most patients with metastatic disease. Identification of those patients with metastatic diseases through minimally invasive means has the potential to decrease nontherapeutic laparotomies and thus morbidity. Although laparoscopy is sometimes performed immediately prior to laparotomy to rule out unresectable disease or peritoneal metastases, staging laparoscopy should ideally be performed as a separate procedure to allow assessment of cytology or biopsy specimens.

Many different staging algorithms exist at different institutions and the role of laparoscopy is not universally agreed upon. However, studies have demonstrated excellent sensitivity of laparoscopy for detection of hepatic metastases and peritoneal dissemination. Evaluation generally consists of a systemic evaluation of the abdomen, inspection of the peritoneum and liver, and fluid sampling for cytology. The lesser omentum may be divided to inspect the celiac axis for lymphadenopathy, particularly in cases of esophageal, gastric, or pancreatic cancer. With pancreatic cancer, the pancreas may be inspected from a supragastric approach through the lesser omentum, from an infragastric approach through the gastrocolic ligament, or from an infracolic approach to the left of the middle colic

vessels. Laparoscopic staging of Hodgkin's lymphoma is reported in some centers, including sampling of portal, celiac, and splenic lymph nodes as well as liver biopsy and splenectomy. Non-Hodgkin's lymphomas are generally not staged laparoscopically.

Laparoscopic staging for colon cancer has not been as widely accepted as the above modalities. The minimally invasive approach prohibits the standard palpation of the liver and porta hepatis of conventional colon surgery and preoperative CT may detect only 70–80% of liver metastasis. Intraoperative ultrasound has been used by some clinicians to assess the liver for metastasis with good results. Preoperative laparoscopy to assess for metastasis from colorectal cancer does not have the same benefit as it does for other cancers since most patients will require resection even without liver or other distant metastases.

One potentially problematic issue with the use of minimally invasive surgery in cancer patients is implantation of tumor cells in the laparoscopic trochar sites (port site metastasis). Such metastases have been described with numerous types of tumors, including biliary, pancreatic, colonic, and gynecologic tumors. The risk of spread, however, is small and is not a contraindication to staging laparoscopy.

Of note, laparoscopic approaches to gastrointestinal malignancy—particularly colorectal cancer—have not been definitively proven to provide equivalent oncologic benefit to standard approaches. Minimally invasive surgery for cure of GI cancers should be performed primarily in the setting of prospective clinical trials.

MINIMALLY INVASIVE SURGERY FOR CROHN'S DISEASE

Complications requiring surgical intervention are common in patients with Crohn's disease, including strictures and obstruction, fistulas, bleeding, and intractability. Most of these, when requiring operation, have led to conservative enteric resection. Minimally invasive approaches to this disease are attractive in that Crohn's primarily affects young patients who often require multiple procedures. Some centers have recently described laparoscopic approaches to treating Crohn's disease with good outcomes. The minimally invasive approach can be quite technically challenging in Crohn's disease, however, given the presence of inflamed friable bowel and thickened mesentery with tendency for abscess and fistula formation. Chronic inflammation can obscure normal anatomical planes and may increase the risk of injury to other anatomic structures. The range of minimally invasive procedures for Crohn's disease includes the following: diagnostic laparoscopy, lysis of adhesions, fecal diversion for perineal sepsis or complex fistulas, closure of stomas, segmental bowel resection, and strictureplasty. Contraindications to laparoscopy for the above situations include the presence of dense adhesions (a "frozen abdomen"), critical illness or free perforation, peritonitis, and multiple complex fistulas.

LAPAROSCOPIC APPENDECTOMY

The issue of whether laparoscopic appendectomy provides benefit compared to standard open appendectomy remains to be shown. However, laparoscopic appendectomy may be particularly useful in several conditions. Reproductive-age women often have possible pelvic/gynecologic processes in the differential diagnosis, with a high rate of negative diagnostic laparotomy. Diagnostic laparoscopy may be performed with full view of intra-abdominal structures, followed by appendectomy if indicated. With no appendicitis on exploration, the appendix is often removed to avoid future confusion in care, unless significant regional pathology exists, such as severe pelvic inflammatory disease that could put the staple line at risk. Laparoscopic appendectomy is also particularly useful in obese patients who would otherwise require a larger incision. The laparoscopic approach is generally contraindicated in pregnancy or in patients with cardiorespiratory insufficiency. Patients with significant involvement of the cecum or base of the appendix may be at risk for stump leak in the case of laparoscopic appendectomy and an open procedure should be performed instead.

Whereas some authors have shown laparoscopic appendectomy to be a superior approach to appendicitis, other authors have reported the opposite result. In most cases, compared to open appendectomy, the lengths of hospital stay are similar or shorter, with similar or earlier return to normal activities and less postoperative pain. In addition, there is a clear advantage in cases of diagnostic uncertainty. Most authors, however, feel that there are not sufficiently clear data to consider it the gold standard for appendicitis.

SUMMARY

Minimally invasive surgery has dramatically altered the surgical approach to a host of intra-abdominal and gastrointestinal conditions and has provided, in most cases, comparable short- and long-term outcomes with minimal morbidity. This has led to increased patient

acceptance and satisfaction. Additionally, increased surgeon experience and technical advances have made performance of these procedures easier and safer.

See Also the Following Articles

Cholecystectomy • Crohn's Disease • Esophageal Surgery • Fast Track Surgery • Gastric Surgery • Gastroesophageal Reflux Disease (GERD) • Hernias • Laparoscopy • Obesity, Treatment of

Further Reading

Breen, E. M., and Ashley, S. W. (2000). Laparoscopic surgery for Crohn's disease?—A conditional yes. *Inflamm. Bowel Dis.* **6**, 43–45.

Brooks, D. C. (ed.) (1998). "Current Review of Minimally Invasive Surgery." Springer-Verlag, New York.

Rothschild, J. G. (1998). What alternatives has minimally invasive surgery provided the surgeon? *Arch. Surg.* **133**, 1156–1159.

Schauer, P. R., and Ikramuddin, S. (2001). Laparoscopic surgery for morbid obesity. *Surg. Clin. N. Am.* **81**, 1145–1179.

Scott-Conner, C. E. H. (ed.) (2000). Minimal access surgery. *Surg. Clin. N. Am.* **80**, 1093–1372.

Soper, N. J., Brunt, L. M., and Kerbl, K. (1994). Laparoscopic general surgery. *New Engl. J. Med.* **330**, 409–419.

Spivak, H., Lelcuk, S., and Hunter, J. G. (1999). Laparoscopic surgery of the gastroesophageal junction. *World J. Surg.* **23**, 356–367.

Wexner, S. D., and Hwang, Y. (2000). Clinical status of laparoscopic bowel surgery for GI malignancy. *Oncology* **14**, 1131–1143.

Motilin

PIERRE POITRAS
Université de Montréal, Canada

gastrointestinal motility Movements induced by contractions of the gastrointestinal wall muscles to transport the nutrient luminal content along the entire gastrointestinal tract, from the mouth to the anus.

hormone Term frequently used in a general sense as a synonym for regulatory peptide. A more precise definition relates to the endocrine action, i.e., mediated through variations in plasma levels, of a regulatory peptide (that could also be active via neurocrine, paracrine, or autocrine pathways).

interdigestive motility Two types of motility profiles can be seen in the stomach and small intestine: the intermittent and powerful cyclic contractions of the interdigestive fasting period and the constant mixing contraction seen during the digestive postprandial period.

migrating motor complex Organizational basic structure of the gastrointestinal tract motility during the interdigestive period; it is made of three successive patterns of contractions (Phases I, II, and III) that are cyclically recurring as long as fasting is maintained and that disappear after eating.

Motilin is one of the numerous gut-manufactured peptides that are involved in control of gut physiological activity. Intestinal wall contractions that transport nutrients along the gut are influenced by neural mechanisms as well as by circulating hormonal substances. Motilin is released periodically into the blood to help regulate motility of the upper gut during the interdigestive fasting period.

STRUCTURE AND LOCALIZATION

Motilin is a 22-amino-acid peptide synthesized in endocrine cells (M cells) of the duodenal and jejunal mucosae. Its existence was first suspected in the early 1970s by John Brown in Vancouver; working on physiological experiments, Brown found that antral contractions were stimulated by an alkaline solution instilled into the duodenum of dogs, even when the

Copyright 2004, Elsevier (USA). All rights reserved.

acceptance and satisfaction. Additionally, increased surgeon experience and technical advances have made performance of these procedures easier and safer.

See Also the Following Articles

Cholecystectomy • Crohn's Disease • Esophageal Surgery • Fast Track Surgery • Gastric Surgery • Gastroesophageal Reflux Disease (GERD) • Hernias • Laparoscopy • Obesity, Treatment of

Further Reading

Breen, E. M., and Ashley, S. W. (2000). Laparoscopic surgery for Crohn's disease?—A conditional yes. *Inflamm. Bowel Dis.* **6**, 43–45.

Brooks, D. C. (ed.) (1998). "Current Review of Minimally Invasive Surgery." Springer-Verlag, New York.

Rothschild, J. G. (1998). What alternatives has minimally invasive surgery provided the surgeon? *Arch. Surg.* **133**, 1156–1159.

Schauer, P. R., and Ikramuddin, S. (2001). Laparoscopic surgery for morbid obesity. *Surg. Clin. N. Am.* **81**, 1145–1179.

Scott-Conner, C. E. H. (ed.) (2000). Minimal access surgery. *Surg. Clin. N. Am.* **80**, 1093–1372.

Soper, N. J., Brunt, L. M., and Kerbl, K. (1994). Laparoscopic general surgery. *New Engl. J. Med.* **330**, 409–419.

Spivak, H., Lelcuk, S., and Hunter, J. G. (1999). Laparoscopic surgery of the gastroesophageal junction. *World J. Surg.* **23**, 356–367.

Wexner, S. D., and Hwang, Y. (2000). Clinical status of laparoscopic bowel surgery for GI malignancy. *Oncology* **14**, 1131–1143.

Motilin

PIERRE POITRAS
Université de Montréal, Canada

gastrointestinal motility Movements induced by contractions of the gastrointestinal wall muscles to transport the nutrient luminal content along the entire gastrointestinal tract, from the mouth to the anus.

hormone Term frequently used in a general sense as a synonym for regulatory peptide. A more precise definition relates to the endocrine action, i.e., mediated through variations in plasma levels, of a regulatory peptide (that could also be active via neurocrine, paracrine, or autocrine pathways).

interdigestive motility Two types of motility profiles can be seen in the stomach and small intestine: the intermittent and powerful cyclic contractions of the interdigestive fasting period and the constant mixing contraction seen during the digestive postprandial period.

migrating motor complex Organizational basic structure of the gastrointestinal tract motility during the interdigestive period; it is made of three successive patterns of contractions (Phases I, II, and III) that are cyclically recurring as long as fasting is maintained and that disappear after eating.

Motilin is one of the numerous gut-manufactured peptides that are involved in control of gut physiological activity. Intestinal wall contractions that transport nutrients along the gut are influenced by neural mechanisms as well as by circulating hormonal substances. Motilin is released periodically into the blood to help regulate motility of the upper gut during the interdigestive fasting period.

STRUCTURE AND LOCALIZATION

Motilin is a 22-amino-acid peptide synthesized in endocrine cells (M cells) of the duodenal and jejunal mucosae. Its existence was first suspected in the early 1970s by John Brown in Vancouver; working on physiological experiments, Brown found that antral contractions were stimulated by an alkaline solution instilled into the duodenum of dogs, even when the

Copyright 2004, Elsevier (USA). All rights reserved.

stomach had been totally denervated, leaving it under the sole control of circulating hormonal substances. Eventually, motilin was identified and the structure of motilin was deduced using an extract of porcine small intestine with a side fraction material prepared for the purification of cholecystokinin in the laboratory of Professor Victor Mutt at Karolinska Institute. The active peptide was purified by successive chromatographies (gel permeation and ion exchange), using the dog stomach *in vivo* as a bioassay for its recognition. Brown called the substance motilin because of its obvious action on gastrointestinal motility. Biochemical purification through successive chromatographies (gel permeation, ion exchange, and high-performance liquid chromatography) and using a radioimmunoassay (RIA) to recognize the immunoreactive material subsequently allowed the identification of dog motilin. Since 1980, molecular biology techniques have characterized motilin from numerous species, including humans, rabbits, chickens, horses, monkeys, and guinea pigs. Two points characterize the biochemistry of motilin, species heterogeneity and the marked structural heterogeneity of rodent motilin. Human motilin is identical to the originally purified peptide found in pigs, but the structure is much different in other species. This structural heterogeneity has led to difficulties in recognition of the molecule by RIA and may also be responsible for specific patterns of bioactivities. Rodent motilin has never been sequenced. The logical explanation is that the structural heterogeneity is so marked in rodents that tools (antisera or probes) developed from known motilin structure in other species cannot recognize rodent motilin. The practical consequence is that the most common and useful laboratory animals cannot be used in the field of motilin research.

STRUCTURE–ACTIVITY AND RECEPTOR

Using a goal-directed approach with a motilin probe for motilin receptor recognition, all attempts have failed to sequence the intestinal motilin receptor. In 1999, researchers in the Merck laboratories followed a different strategy by first cloning and sequencing a large variety of G-protein-coupled receptors, and then looking at the nature of ligands for these receptors. Starting from an orphan receptor found in the thyroid with close structural similarity to the receptor for the growth hormone secretagogue (GHS), mass screening of more than 500 peptide and nonpeptide molecules was performed to obtain positive results for motilin. Much of our knowledge on the motilin receptor has been derived up to now from structure–activity studies of peptide analogues and binding of radioactive ligands on various tissue membranes prepared *in vitro*.

Nuclear magnetic resonance shows that the motilin molecule (Fig. 1) is shaped like a golf club or shepherd's crook. The N-terminal portion is responsible for receptor binding and *in vitro* bioactivity. Peptide fragments containing the first 12 amino acids of the molecule have demonstrated full binding capacity *in vitro*, whereas removal of the first amino acid is enough to abolish motilin activity. Amino acids 4 and 7 also seem to play a major role in this N-terminal sequence. Interestingly, the C-terminal structure, not required for *in vitro* bioactivity, is essential for *in vivo* bioactivity. The N-terminal fragments of 15 amino acids, fully active *in vitro*, are inactive *in vivo*, whereas longer fragments with 19, 20, or 21 amino acids can mimic the motor activity of the 22-amino-acid peptide. In the field of peptide pharmacology, the apparent importance of the α-helix-shaped C-terminal structure in protecting the whole molecule, probably against circulating degrading enzymes is an interesting concept.

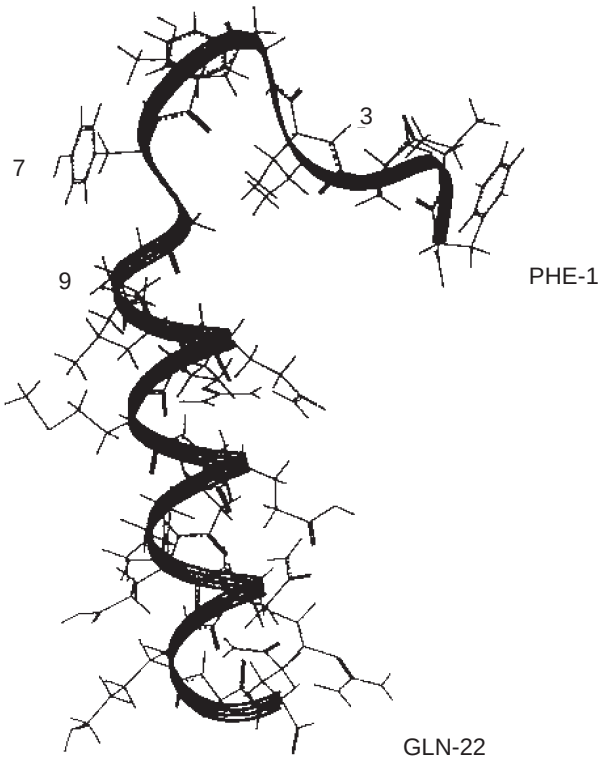


FIGURE 1 Structure of the 22-amino-acid motilin peptide revealed by nuclear magnetic resonance. The N-terminal curved sequence binds to the receptor whereas the α-helix C-terminal structure protects the molecule for *in vivo* bioactivity.

Receptors have been studied in membrane solutions prepared from various species (mostly humans or rabbits) and from neural or muscle elements of the intestinal wall. The binding of motilin analogues is different in membranes prepared from human or rabbit antrum, suggesting that motilin receptor structure is, as for motilin,, different among species. Solutions enriched in muscle or in neural elements of the intestinal wall have allowed the identification of specific and distinct responses to various motilin synthetic analogues, indicating the existence of structurally heterogeneous motilin receptor subtypes specific to muscle or neural elements (M and N receptor subtypes).

PHARMACOLOGICAL ACTION

As its name implies, motilin acts on digestive motility. Figure 2 shows the expected effect in humans after the administration of motilin or motilin receptor agonists. Stimulation of antroduodenal motility remains the dominant action of the peptide.

Initial *in vitro* studies with intestinal tissues from rabbits and humans have revealed that motilin can stimulate smooth muscle contraction by a direct effect on muscle cells; the peptide action was seen despite the addition of all neural blockers, including tetrodotoxin. Experiments with isolated muscle cells have confirmed the presence of motilin receptors on muscle cell membranes. However, *in vivo* studies in dog and human tissues clearly indicate that the motor action of motilin is mediated by muscarinic transmitters. As discussed previously, binding experiments support the existence of motilin receptor subtypes (M and N) that are functionally and structurally different. Currently, the most interesting hypothesis proposes that the interdigestive motor action of motilin (induction of Phase III contraction) is elicited at low doses through an action via neural cholinergic receptors, whereas postprandial stimulation of antral motility is evoked at higher doses via muscle receptors.

PHYSIOLOGICAL ROLE

The characteristic action of motilin is the induction of Phase III contractions of the migrating motor complex (MMC). The MMC is the basic organization of motor activity of the gut during the fasting interdigestive period. It lasts 80 to 120 min and comprises three successive phases: in Phase I, no significant contraction is seen for 20 to 60 min; in Phase II, intermittent and irregular contractions start to occur during 20 to 60 min before Phase III, when strong peristaltic contractions, lasting 3 to 10 min, start from the stomach and lower esophagus, migrating distally to the duodenum, jejunum, and ileum until reaching the colon. This Phase III peristaltic wave has been proposed to clean bacteria or nutrients from the gastrointestinal tract; accumulated microflora during the digestive period could have deleterious effects on the gut (bacterial overgrowth, etc.). Feeding interrupts the intermittent cyclical fasting motility and

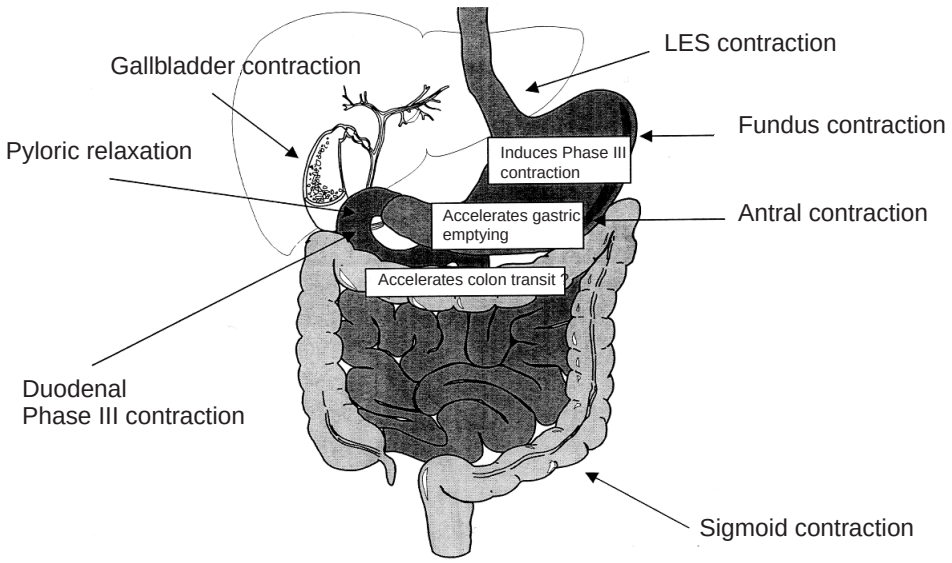


FIGURE 2 Activation of motilin receptors. Biological effects in the human gastrointestinal tract in response to injection of motilin or a motilin receptor agonist such as erythromycin. LES, Lower esophageal sphincter.

induces a more constant motor activity of moderate amplitude to allow the optimal absorption of ingested nutrients; accelerated transit during Phase III would impair nutrient absorption.

Regulatory peptides can exert their influence via different pathways: endocrine (hormonal), neurocrine, paracrine, or autocrine. Motilin is in the unique position, as a gastrointestinal (GI) regulatory peptide, of being an interdigestive hormone. This has been well established in the dog, for which all four of the criteria of Morton Grossman establishing the endocrine contributions of peptides have been fulfilled: (1) regulation of the cyclical pattern of the MMC is under the control of circulating factors, as shown in various experiments in which the MMC persisted in animals when the stomach had been completely denervated; (2) perfect correlation exists between circulating peak levels of motilin and the initiation of Phase III contractions from the stomach or proximal duodenum; (3) there is induction of Phase III contractions of the MMC by small doses of exogenous motilin, reproducing physiological plasma variations; and (4) inhibition of circulating motilin occurs following the administration of specific motilin antisera, which blocks Phase III contractions from the upper gut. Although the situation cannot be explored under such perfect experimental conditions in humans, most evidence suggests that circulating motilin also plays a key role in the regulation of Phase III initiated from the human antrum.

RELEASE MECHANISMS

Most gastrointestinal hormones are released after a meal to allow or facilitate the digestion and absorption of nutrients. Motilin is a unique hormone. It is released periodically during the interdigestive fasting period, and its cyclical release is abolished after a meal, as shown schematically in Fig. 3. Therefore, a "biological clock" somewhere in the organism periodically signals motilin cells to release the peptide into the circulation. *In vitro* preparations of intestinal mucosal cells enriched in motilin cells show that muscarinic receptors are present on the motilin cell membrane and that protein kinase C activators are the most potent second messengers eliciting motilin release. In the *ex vivo* perfused canine intestine, bombesin has been identified as a direct stimulant of motilin release, whereas the effect of opiates is mediated by acetylcholine. Phenylephrine and somatostatin appear to act on M cell membrane receptors to block release of the peptide. In humans, meal ingestion is followed by a very early and brief increase in plasma motilin before the interdigestive release cycle is interrupted. This early release can be mimicked by central stimulation with modified sham feeding and by distension of the fundus with an air-filled balloon. The contribution of this postprandial motilin release (not present in the dog) in the process of nutrient digestion remains to be characterized.

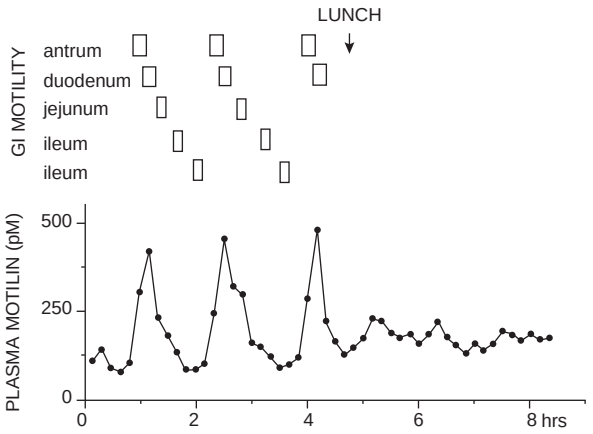


FIGURE 3 Schematic representation of canine plasma motilin variations. During the fasting interdigestive period, motilin is released cyclically every 80–120 min (lower panel) to induce Phase III contraction of the migrating motor complex from the stomach to the ileum (indicated by boxes in the upper panel). After eating, motilin cyclical peak increases are abolished for 2–8 hours (depending on the content and nature of the meal) while the fed pattern motility profile is taking place.

CONTRIBUTION TO CLINICAL MEDICINE

Some GI hormones, such as gastrin or vasoactive intestinal polypeptide (VIP), are important to clinicians because of the disease symptoms they can generate. Up to now, there has been no clinical phenotype attributed to motilin hypersecretion. High levels of circulating motilin have been documented in some patients with pancreatic tumors and Zollinger–Ellison syndrome, as well as in subjects with carcinoid tumors of the gut. Although it is tempting to speculate on the role of motilin in the diarrhea found in these patients, its contribution to the biological alterations remains unknown.

Because motilin hypersecretion could be expected to generate GI hypermotricity and hypersecretion with probable diarrhea, it is logical to expect that hypomotilinemia will induce GI hypomotility. Some investigators have indeed found low levels of plasma motilin in patients with idiopathic intestinal pseudo-obstruction and idiopathic or postoperative gastroparesis. Up to now, however, plasma motilin measurement has not proved useful for inclusion in the workup diagnosis of any clinical situations.

On the other hand, motilin has been recently of major interest to medical clinicians because of the capacity of motilin receptor agonists to act as powerful stimulants of GI motor activity in patients with hypokinetic disorders. Zen Itoh in Japan was the first to observe that erythromycin could mimic the motor effect of motilin when injected in dogs. It was soon established that erythromycin was in fact acting on motilin receptors, and Janssens *et al*. in Leuven made the capital observation that erythromycin was the most potent gastrokinetic ever tested to stimulate gastric emptying in diabetic patients with gastroparesis. Since then, intravenous or periorbital erythromycin has been used by many clinicians for the treatment of patients with gastroparesis or intestinal pseudo-obstruction. Motilides, i.e., motilin receptor agonist substances derived from the erythromycin macrolide and with improved gastrokinetic activity but devoid of antibiotic properties, have been developed by many pharmaceutical companies. At least three motilides have been tested in humans, but, for various reasons (rapid tachyphylaxis, no significant clinical benefit, or potential side effects), clinical trials with these newly derived molecules have failed to substantiate the impressive pharmacological potential seen with intravenous erythromycin. Whether the gastrokinetic capacity of motilin receptor agonists will be amenable to commercial development and clinical exploitation remains to be seen.

CENTRAL NERVOUS SYSTEM AND THE PEPTIDE FAMILY

Most GI peptides are found in the brain and/or act as neuropeptides, but the situation remains unclear for motilin. Motilin mRNA has been identified in brain tissues, but RIA determination of motilin content has provided ambiguous results. Motilin administration in brain tissue induces activities (suggesting that there are motilin receptors in the brain) that are quite unexpected, including appetite stimulation, growth hormone (GH) release, and anxiety suppression. Yet no valid data support the role of motilin as a neuropeptide.

Most peptides are members of a "peptide family" that includes structurally related compounds; gastrin/cholecystokinin and secretin/VIP are typical examples. A recent discovered peptide in the gastric mucosa shows 25% similarity to motilin. It has been called motilin related peptide (MTL-RP) by some, but the name "ghrelin," because of its effect on GH release, is more commonly used. Interestingly, ghrelin administration induces central actions that are similar to those described for motilin (such as GH release or appetite stimulation). Peripherally, ghrelin mimics motilin and appears to be the most potent gastrokinetic agent we have ever tested in the rodent. Future studies should tell us more about the importance of this new family of peptides.

See Also the Following Articles

Gastric Motility • Migrating Motor Complex

Further Reading

Brown, J. C., Cook, M. A., and Dryburgh, J. H. (1973). Motilin, a gastric motor activity stimulating polypeptide: The complete amino-acid sequence. *Can. J. Biochem.* **51**(5), 533–537.

Itoh, Z. (1997). Motilin and clinical application. *Peptides* **18**(4), 593–608.

Peeters, T. L. (1993). Erythromycin and other macrolides as prokinetic agents. *Gastroenterology* **105**(6), 1886–1899.

Poitras, P. (1984). Motilin is a digestive hormone in the dog. *Gastroenterology* **87**(4), 909–913.

Mucosa-Associated Lymphoid Tissue (MALT)

HYUN SIL KIM
University of California, San Diego

M cell Specialized type of epithelial microfold cell that is found in the epithelium overlying organized lymphoid follicles, the Peyer's patches. Endocytosis and transport of intestinal microbial organisms and large molecules from the lumen into the underlying lymphoid tissues are the main functions of M cells.

mucosa-associated lymphoid tissue Organized immune system found in mucosal sites. Gut-associated lymphoid tissue is a component of MALT; these are inductive sites where luminal antigens are processed.

Peyer's patch Lymphoid nodule located underneath M cells in the lamina propria; the sites of antigen recognition and processing, which induce activation of resident B and T cells.

The gastrointestinal tract represents a unique immunologic compartment that is actively involved in managing responses to luminal food antigens and protecting the host from infectious agents and environmental toxins. Given these exposures, one method the intestine has devised to defend itself from such antigenic challenges is by creating a physical barrier in the form of an epithelial single-cell layer that lines the mucosa from the esophagus to rectum. A mucosa-associated lymphoid tissue comprises the immune system that responds to disruption of the mucosal barrier.

INTRODUCTION

For intestinal luminal antigens, such as food particles and/or intestinal bacteria, to elicit a local and/or systemic immune response, an antigen must either cross an intact epithelium or secrete noxious mediators that disrupt the tight protective gastrointestinal barrier, allowing for entry into the lamina propria and interaction with immune-responsive cells. Many infectious diseases of the gastrointestinal tract involve the disruption of this protective epithelial barrier by bacteria or viruses and the invasion of these foreign pathogens into the host. Therefore, in addition to the physical barrier, the intestine has also devised an elaborate system of mucosal immunity to limit injury and regulate immune responses from potentially deleterious agents.

INNATE/ACQUIRED IMMUNE RESPONSES

The immune system is composed of innate (or natural) and adaptive (or acquired) immunity. The innate immune system consists of (a) cells, including monocytes, neutrophils, and intestinal epithelial cells (IECs); (b) toll-like receptors, which are cell surface and intracellular proteins that recognize bacterial components such as lipopolysaccharide (LPS), lipid/carbohydrate components of the bacterial wall, and bacterial DNA; and (c) soluble proteins, such as cytokines and complement products that can be released rapidly to initiate host immune response. The cells, receptors, and soluble mediators of the innate immune system provide immediate recognition of a general class of foreign antigens. In contrast, acquired immunity is mediated primarily by B and T lymphocytes. Such adaptive immunity develops over a longer period of time and provides recognition of specific antigens. The two types of immunity are closely dependent on each other for activation. In addition, the innate immune system play a key role in the selection of particular types of B and T cells needed to participate in an acquired immune response.

PEYER'S PATCHES AND M CELLS

As seen in Fig. 1, the gastrointestinal (GI) tract contains immune-responsive cells that reside both within the epithelial and subepithelial layers and that are important regulators of the mucosal immune system. These cellular elements function in direct phagocytosis/cytolysis of foreign intruders, in antibody and cytokine production, and in recruitment of additional immune-reactive cells to areas of infection and/or injury. The gut-associated lymphoid tissue (GALT) contains a loosely affiliated group of immune cells in the lamina propria and an organized component (Peyer's patches). Found predominantly in the small intestine, Peyer's patches are lymphoid nodules that are composed of a central B cell zone and a parafollicular T cell zone. Over the dome of Peyer's patches, a specialized epithelium containing microvillus fold cells (M cells)

 Copyright 2004, Elsevier (USA). All rights reserved.

functions in the transport of antigens from the lumen to the Peyer's patches. Macrophages and dendritic cells exist in close proximity to M cells within Peyer's patches. Known infectious agents that adhere to and are transported by M cells include human immunodeficiency virus, reovirus, poliovirus, *Vibrio cholerae*, mycobacteria, *Salmonella*, and *Shigella*. Within Peyer's patches antigen-specific B cells and T cells are activated. Some of these activated lymphocytes migrate from the lamina propria via lymphatics to mesenteric lymph nodes and eventually enter the systemic circulation, and may return to the lamina propria and other mucosal tissues. As such, the organized lymphoid tissue functions as an inductive or afferent site and the loosely affiliated cells in the lamina propria function as an efferent or effector site.

INTESTINAL EPITHELIAL CELLS

In addition to their function in nutrient absorption, IECs serve an important immune regulatory function in the GI tract. As a component of the innate immune system within the gut, IECs constitutively produce factors such as tumor growth factor-β (TGF-β), complement proteins (C3, C4, and factor B), and leukotrienes, all of which are involved in providing immune protection and controlling the recruitment and activation state of local T cells. TGF-β secreted by the IECs, for example, may regulate expression of an adhesive molecule on IECs that functions in intestinal intraepithelial lymphocyte (iIEL) and IEC interactions (see later). In addition, as shown in various *in vitro* models of bacterial infections, IECs produce a variety of chemokines and proinflammatory cytokines, including interleukin-8 (IL-8), granulocyte–macrophage colony-stimulating factor (GM-CSF), and IL-6. Similarly, pathogenic bacteria such as *Shigella dysenteriae*, *Yersinia enterocolitica*, enteroinvasive *Escherichia coli*, and *Salmonella dublin* induce IL-8, monocyte chemotactic protein, and tumor necrosis factor α (TNFα). IL-8 is a potent chemoattractant for neutrophils. Thus, the constitutive and induced expression of a specific set of cytokines by IECs demonstrates that IECs serve a vital role in initiating and amplifying the mucosal inflammatory response as an important aspect of gastrointestinal innate immunity. The paracrine and autocrine signals provided by IECs alert the local mucosal immune system to respond appropriately in an infectious process and initiate the processes of T cell activation and recruitment. Furthermore, IECs have been shown to process and present soluble antigens to activate mucosal T cells. Extracellular antigens that are internalized by IECs appear to enter an endolysosomal route and are presented on the IEC surface. Therefore, IECs have been shown to act as an important regulator of mucosal immune responses by responding to and secreting cytokines and potentially participating indirectly (through M cells) and directly in antigen presentation.

MUCOSAL B LYMPHOCYTES

Another crucial component of the mucosal immune system is B cells, which are abundantly present within the lamina propria. These lymphocytes mainly produce IgA, the principal immunoglobulin found in mucosal secretions. IgA in secretions recognizes antigens in the gut lumen while they are external and thus prevents pathogens from entering the host systemically. Therefore, the immune exclusion function of IgA represents one of the first lines of defense against many microbial organisms. In infections with noninvasive pathogens, as seen in *V. cholerae* and enterotoxigenic *E. coli*, secretory IgA is an important host defense mechanism along with other aspects of cell-mediated immunity. As seen in Fig. 1, B cells and plasma cells are located in the loose corrective tissue of the lamina propria. The basolateral surfaces of the epithelial cells lining the GI tract express a receptor, the polymeric Ig receptor, that functions in the transport of IgA and IgM. The dimeric IgA secreted by B cells and plasma cells in the lamina propria binds to the polymeric Ig receptor (pIgR) on the basolateral surface of IEC, after which it is internalized and transported through a vesicular pathway to the apical surface of the IEC facing the lumen. Prior to its release, the pIgR is cleaved at the transmembrane domain, and the extracellular portion of the receptor and IgA is released into the lumen as secretory IgA (sIgA). This form of IgA is resistant to proteolytic degradation by normal intestinal proteases, allowing it to bind to any pathogenic microorganisms and their associated toxins that are present in the lumen, thus preventing their attachment and/or absorption. The immune function of sIgA is not limited to binding luminal antigens; IgA can also neutralize infectious agents present within epithelial cells. It is thought that during the transport of IgA across epithelial cells, IgA can recognize and neutralize specific antigens within infected epithelial cells. In addition, IgA produced locally within the lamina propria has been shown to bind antigens that have gained access across the epithelial lining into the lamina propria. IgA can form immune complexes with these antigens and extrude viral pathogens back onto the lumen via the pIgR, thus containing infection locally and limiting local and/or systemic infection.

MUCOSAL T LYMPHOCYTES

The interactions between IECs and T cells must be tightly regulated in order to coordinate appropriate mucosal responses in normal and disease states. The careful balance between the two immunologic states is determined, in part, by specific IEC/T cell interactions that trigger differential expression of various cell surface molecules and secretion of cytokines. Depending on the recognition of a specific antigen in the context of major histocompatibility complex (MHC) or MHC-like molecules, T cells undergo activation either to maintain the normal hemostasis of the gut mucosa or to initiate inflammatory processes, as occurs in infections and idiopathic inflammatory diseases of the bowel.

Two types of T cells are observed throughout the GI tract: (1) intraepithelial T lymphocytes (iTEL) that are found between epithelial cells adjacent to the basolateral surfaces of IECs and (2) lamina propria T lymphocytes (LPLs) (see Fig. 1). The activation of T cells occurs when a T cell receptor (TCR) expressed on the surface of the T cell recognizes a specific antigen presented by an MHC molecule expressed on the surface of an antigen-presenting cell (APC) such as an IEC, macrophage, dendritic cell, or B cell. Along with the TCR, T cells also express a coreceptor, either the CD8 or CD4 molecule, on the cell surface. The processed antigen is usually a small peptide derived from a larger polypeptide, such as a viral or bacterial protein that entered the APC. Virtually all cells express MHC class I molecules on their surface; however, the expression of MHC class II molecules is primarily restricted to professional APCs, namely, B cells and mononuclear phagocytes, and nonprofessional APCs exposed to cytokines such as interferon γ (IFNγ). An antigen presented by an APC in the context of an MHC class I molecule and in the context of an MHC class II molecule activates CD8+ T cells and CD4+ T cells, respectively. CD8+ T cells are commonly referred to as cytolytic T cells because they can directly lyse and kill infected APCs expressing peptides derived from intracellular pathogens or tumors. CD4+ T cells are commonly known as helper T cells, which are activated by peptides derived from

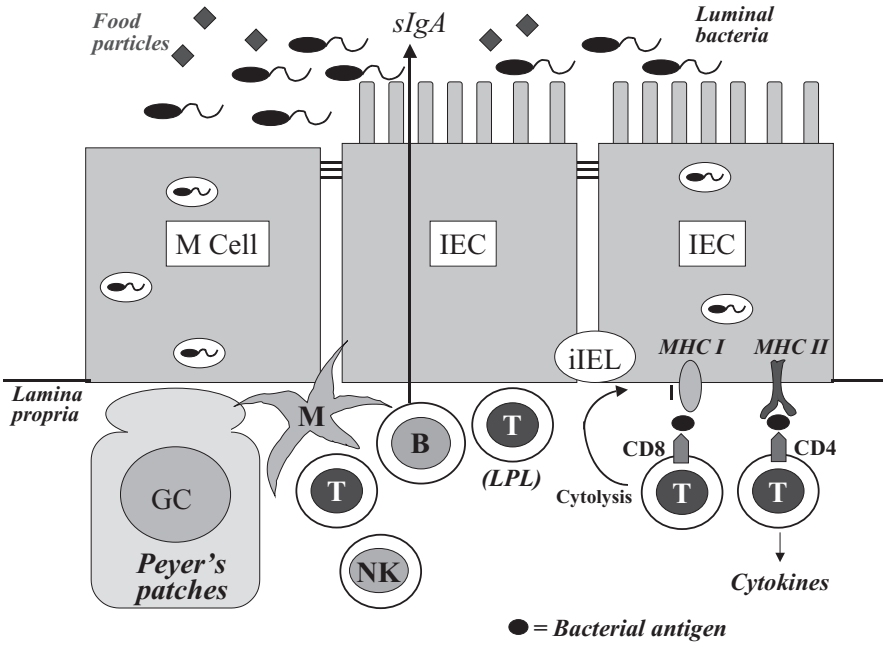


FIGURE 1 Mucosa-associated lymphoid tissue is an organized lymphoid tissue found in the intestine; it consists of M cells and intestinal epithelial cells (IECs) that transport and process luminal antigens (food particles and microbial organisms). The follicle-associated M cells are located above Peyer's patches, and B cells are found within germinal center (GC). The antigen-activated T cells—intraepithelial lymphocytes (IELs) and lamina propria T lymphocytes (LPLs)—are an important source of cytokines such as interferon γ (IFN-γ), interleukin-4 (IL-4), and tumor necrosis factor α (TNF-α) which are potent inflammatory cytokines that able to regulate local immune responses. A macrophage (M), a natural killer (NK) cell, and B and T lymphocytes are depicted. MHC, Major histocompatibility complex.

extracellular pathogens; on activation helper T cells release many important cytokines, such as IL-2, IL-4, IL-10, TGF-β, and interferon γ, which recruit other T cells, activate B cells to produce antibodies, and enhance the phagocytic function of macrophages and natural killer (NK) cells.

iIELs are located immediately adjacent to the basolateral surface of IECs in a unique compartment separated by a basement membrane from other immune cells found within the lamina propria. The proximity of the iIEL to the IEC suggests that iIELs may play a vital role in immunoregulation and immunosurveillance within the epithelium. Human iIELs are predominantly T cells and display a limited number of sets of TCRs on their cell surface, despite direct exposure of these cells to numerous antigens contained in the gut lumen. A majority of these cells are CD8+ memory T cells, which suggests that iIELs have already been activated to respond to a limited set of specific antigens in the context of an MHC class I or MHC class I-like molecule.

Lamina propria lymphocytes are most likely thymically derived lymphocytes that have previously responded to luminal gut antigens. These gut antigen-activated lymphocytes enter the systemic circulation after initial antigen encounter within the Peyer's patches and return to the lamina propria by a characteristic homing mechanism. The majority of LPLs are memory T cells. Similar to iIELs, many LPLs show a limited range of TCR expression on their surface. Some studies show overlap in TCR usage for iIELs and LPLs, suggesting that lymphocytes may be mobile between the epithelium and lamina propria compartments. Several features of LPLs indicate that they are a special subset of lymphocytes that have distinct modes of activation and function. First, LPLs exhibit an increased basal level of activation by expressing the IL-2 receptor, which is not seen in unactivated peripheral blood or splenic lymphocytes. LPLs also express other cell surface molecules that are markers of lymphocyte activation. Second, LPLs respond to antigens primarily with cytokine production, which may be responsible for the high levels of IgA production characteristic of this compartment. Finally, LPLs undergo programmed cell death, also known as apoptosis. Apoptosis may serve an important regulatory mechanism in which LPLs undergo cell death to limit inflammation and clonal expansion of antigen-activated LPLs, which may otherwise harm the host by initiating and/or perpetuating deleterious inflammatory processes.

SUMMARY

The gastrointestinal tract represents an important immunologic compartment in which active immune recognition and activation take place. The mucosa-activated lymphoid tissue is an organized immune system at mucosal sites and in the intestinal tract; it is composed of a variety of cell types, including M cells, dendritic cells, and macrophages and effector cells, represented by T cells, B cells, and NK cells. Challenged by luminal antigens such as food particles and microbial organisms, the MALT serves a vital role in protecting the host from foreign organisms and in the recognition of self from non-self antigens via activities of resident immune-responsive cells located within the intestine.

See Also the Following Articles

Lymph, Lymphatics, and Lymph Flow • Lymphocytes • Lymphomas • Mucosal Biopsies

Further Reading

Debard, N., Sierro, F., and Kraehenbuhl, J. P. (1999). Development of Peyer's patches, follicle-associated epithelium and M cell: Lessons from immunodeficient and knockout mice. *Semin. Immunol.* **11**, 183–191.

Farstad, I. N., Norstein, J., and Brandtzaeg, P. (1997). Phenotypes of B and T cells in human intestinal and mesenteric lymph. *Gastroenterology* **112**, 163.

Hein, W. R. (1998). Organization of mucosal lymphoid tissue. *Curr. Top. Microbiol. Immunol.* **236**, 1–15.

Neutra, M., Mantis, N. J., and Kraehenbuhl, J. P. (2001). Collaboration of epithelial cells with organized mucosal lymphoid tissues. *Nat. Immunol.* **2**(11), 1004–1009.

Ogra, P., Mastecky, R., and Lamm, M. E. (eds.) (1994). "Handbook of Mucosal Immunology." Academic Press, Orlando.

Phalipon, A., and Corthesy, B. (2003). Novel functions of the polymeric Ig receptor: Well beyond transport of immunoglobulins. *Trends Immunol.* **24**(2), 55–58.

Plevy, S. (2002). The immunology of inflammatory bowel disease. *Gastroenterol. Clin. North Am.* **31**(1), 77–92.

Mucosal Biopsies

HARVEY GOLDMAN
Beth Israel Deaconess Medical Center and Harvard Medical School

Barrett's esophagus Condition of chronic esophagitis, characterized by a change (metaplasia) of the normal squamous epithelium to a glandular mucosa.

collagenous colitis Condition of chronic colitis, characterized by thickening of the collagen layer beneath the surface epithelium. Typically seen in middle-aged and older women with watery diarrhea.

dysplasia Lesion of the epithelium (squamous or glandular), characterized by irregular and atypical proliferation and cytologic abnormalities. When seen in a patient with a chronic inflammatory disorder (Barrett's esophagus, chronic gastritis, ulcerative colitis, and Crohn's disease), the finding of dysplasia serves as a marker that the patient has or is especially apt to develop adenocarcinoma.

lymphocytic colitis Condition of chronic colitis, characterized by watery diarrhea, no gross radiographic or endoscopic abnormality, and biopsies showing prominent increase of lymphocytes within the surface and upper crypt epithelium.

metaplasia Change from the normal histology to a different mature differentiation. The finding of metaplasia serves as a marker that the patient has a chronic inflammatory lesion. Examples in the alimentary tract include the glandular metaplasia in Barrett's esophagus, intestinal metaplasia in chronic gastritis, gastric mucous cell metaplasia in chronic duodenitis, pyloric gland metaplasia in Crohn's enteritis, and Paneth cell metaplasia in ulcerative colitis.

mucosa-associated lymphoid tissue lymphoma Malignant lymphoma of the marginal-type B lymphocytes within the alimentary tract. This is most commonly seen in the stomach as a complication of chronic *Helicobacter pylori* gastritis.

pouchitis Inflammatory condition, either acute or chronic, involving the ileal pouch following a colectomy and ileal pouch–anal anastomosis.

pseudomembranous colitis Inflammatory condition, characterized by the presence of prominent inflammatory membranes on the colonic mucosal surface. Typically present in patients with *Clostridium difficile* colitis, but can be seen as well in cases of ischemic disease and other infections of the colon.

With the widespread use of flexible endoscopes, there has been a large and steady increase in the procurement of and dependence on mucosal biopsies of the gut. Lesions are being detected and evaluated at early stages, in some instances before they are grossly evident. Furthermore, with serial examinations, the spectrum of stages and the evolution of many disorders are being appreciated. Endoscopic biopsies are used not only to determine a specific diagnosis but also to assess the extent and severity of a lesion, to monitor the course, and to detect complications. It is imperative that relevant clinical information and particular reason for the examination be provided to ensure an appropriate and intelligent analysis of the biopsy, and the questions posed should be specifically answered in the pathologic report.

GENERAL FEATURES

This article concentrates on biopsy uses and interpretation in selected inflammatory disorders of the gut, with brief mention of the utility of mucosal biopsies in tumor detection. The jejunum, which is not ordinarily visualized by endoscopy, is not included.

Technical Aspects

Mucosal biopsies of the gut include those obtained directly at endoscopy and those obtained by aspiration to secure submucosal tissue. Orientation of the biopsy specimen is essential for analysis of the jejunal mucosa and is optimal when the endoscopist places the specimen luminal side upward on filter paper or other substrate that can be embedded and cut. For other biopsies, sufficient orientation is ordinarily achieved by obtaining multiple sections. Buffered formalin is an adequate fixative for most studies, provided that tissue and cellular shrinkage are considered. When finer nuclear detail is desired, as in the identification of lymphomas and other cellular tumors, a mercury- or picric acid-containing solution such as Bouin's or Hollande's is preferable. Because many mucosal lesions are tiny, two to three levels of any biopsy, with at least four sections at each level, are recommended. Ordinary hematoxylin and eosin (H&E) stain is usually sufficient, using special procedures when indicated. Cytologic smears are an important adjunct for identification

 Copyright 2004, Elsevier (USA). All rights reserved.

of tumor cells and organisms (viral inclusions, fungi, protozoa, and ova).

Effects of Preparation, Endoscopy, and Biopsy Procedures

Preparation, endoscopy, and biopsy procedures may cause variable degrees of trauma to the tissue, resulting in distortion and tears, flattening or denudation of surface epithelium, edema, congestion, and hemorrhage. These effects must be discounted before considering abnormalities. For evidence of injury, it is best to look for sure signs of epithelial damage coupled with the presence of neutrophils, eosinophils, or increased mononuclear inflammatory cells in the epithelial layer.

ESOPHAGUS

Normal Esophageal Mucosa

The esophagus is covered by a layer of nonkeratinizing, stratified squamous epithelium. The basal zone of proliferative cells is ordinarily only a few cells thick but may extend to up to 15% of the epithelial thickness; the papillae of the lamina propria reach up to 50% of the epithelial thickness. Squamous epithelium contains small number of mononuclear inflammatory cells (lymphocytes, Langerhans cells) but no neutrophils and only rare eosinophils. Endoscopic biopsies typically contain little or no lamina propria, resulting usually in less than optimal orientation of the epithelium. Aspiration-type biopsies contain substantial amounts of lamina propria and typically extend into the muscularis mucosae, which is normally very thick in the esophagus. The submucosa is usually not sampled.

Reflux Esophagitis

Features of milder acute (or active) lesions include basal zone hyperplasia exceeding 20–25% of total epithelial thickness, the presence of intraepithelial eosinophils or neutrophils, and elongation of papillae with dilated venules beyond 75% of the epithelial layer. Accurate estimates of basal zone hyperplasia and papillae length require well-oriented biopsies that are usually of the aspiration type, whereas intraepithelial granulocytes can be readily appreciated in the smaller endoscopic biopsies. More marked lesions show necrosis with ulceration and granulation tissue. Chronic lesions reveal the presence of glandular metaplasia (Barrett's esophagus), consisting of any admixture of gastric cardiac and intestinal mucous cell epithelia (Fig. 1). This must be differentiated from normal gastric fundic mucosa, which indicates a hiatal hernia; and from carditis, which is often due to *Helicobacter pylori*.

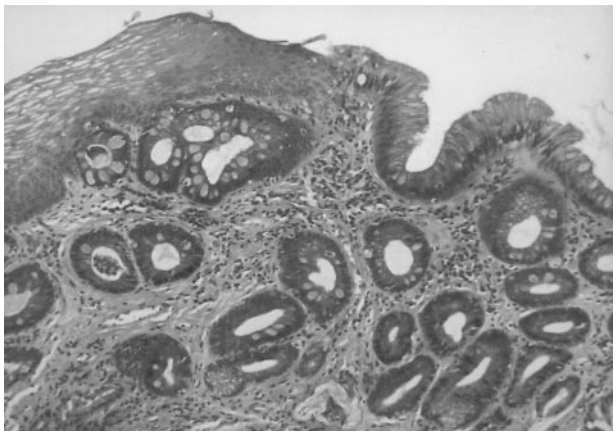


FIGURE 1 Esophageal mucosal biopsy, showing Barrett's esophagus. There is a small portion of normal epithelium in the upper left section, consisting of nonkeratinizing stratified squamous epithelium. The rest of the mucosa consists of glandular tissue, including gastric-type cells that are narrow and have mucus in the luminal end, and distended mucus-filled goblet-type cells of the intestinal type.

Other Forms of Esophagitis

Esophagitis may manifest in other forms:

1. Infections. Biopsy is used principally to demonstrate viral inclusions of herpes simplex in the epithelium, cytomegalovirus in ulcers, and fungi such as *Candida*. The lesions are often focal and may spare the distal esophagus, the typical site of reflux disease.
2. Radiation. Biopsy is done mainly to exclude tumor or opportunistic infections. Features of radiation at the mucosal level are relatively nonspecific; atypical mesenchymal cells and vascular ectasia may be seen.
3. Drug effects. Most cases are due to localized entrapment and pressure injury from a pill, resulting in a focal ulcer with nonspecific acute inflammation.
4. Neuromuscular disorders. Most common are systemic sclerosis and late effects of esophageal atresia. Injury is due to impaired motility (and acid clearance) and/or lower sphincter incompetence, which promote development of peptic esophagitis. Biopsy may also assist in surveillance for late tumors in cases of achalasia.
5. Immune disorders. Allergic (eosinophilic) gastroenteritis is more common in children, and the esophagus may also be involved. The features are similar to reflux injury, with a tendency to show large

aggregates of eosinophils; functional tests are needed to separate.

6. Rare disorders. Include pemphigus and other skin conditions, Crohn's disease, sarcoidosis, and Behçet's disease.

Tumors of the Esophagus

Squamous cell papilloma is a rare, small, wart-like lesion, and biopsy reveals a papillary hyperplasia of the squamous layer without dysplasia. Biopsy and cytology are used to detect and monitor lesions of squamous cell dysplasia and carcinoma. Adenocarcinomas arise from the gastric cardia, Barrett's metaplasia, heterotopic stomach, or esophageal glands. Cases developing in Barrett's mucosa are usually associated with glandular dysplasia, and this dysplasia should be sought as the precursor lesion. Adenocarcinomas located in the distal esophagus, esophagogastric junction, and gastric cardia have a common histology. Other tumors include lymphomas, endocrine and mesenchymal lesions, rare primary melanoma, and metastatic tumors.

STOMACH

Normal Gastric Mucosa

The antral mucosa has relatively long pits (foveolae) and prominent fibromuscular stroma in its lamina propria. Chronic inflammatory cells are commonly present in the lamina propria of the antrum and cardia, and this increases with age. The body and fundus have short pits, few inflammatory cells, and are composed mainly of compact glands containing parietal and chief cells. Also present are a variety of endocrine cells that are mainly located at the gland bases; the cells have central nuclei and clear cytoplasm and are best visualized with silver-binding stain (Grimelius) or specific immunocytochemical stains (chromogranin, synaptophysin, gastrin, or somatostatin).

Chemical and Drug Gastritis

Common causes are reflux of bile salts, ethanol, aspirin, and nonsteroidal antiinflammatory drugs (NSAIDS). The major histologic effects are edema and congestion, elongation and regeneration of the pits, with minimal or no inflammation; hemorrhage and erosion occur in severe cases (Fig. 2).

Chronic Gastritis

General features of chronic gastritis include an increase of mononuclear inflammatory cells with lymphoid nodules in the lamina propria, atrophy of specialized glands, hyperplasia of gastric pits, and the appearance of intestinal and pyloric metaplasia. Active disease shows neutrophils in the pits and lamina propria, with ulcers in severe cases.

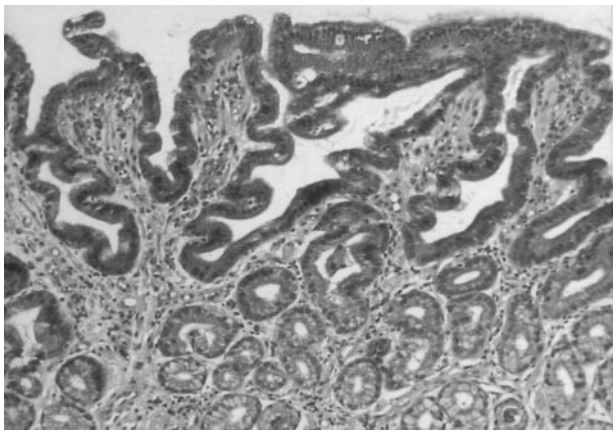


FIGURE 2 Gastric antral mucosal biopsy, showing the typical features of chemical- or drug- induced gastritis. The surface is at the top. In the bottom portion are seen the normal gastric antral mucous cells, which are distended with mucin. In the upper portion, there is marked regeneration of the gastric pits characterized by tortuous and elongated glands that contain cells with enlarged nuclei that occupy a large portion of the cell. There is also some edema and congestion of the lamina propria but minimal cellular inflammation.

Chronic antral gastritis occurs due to exogenous agents, and about two-thirds to three-quarters of cases of active chronic gastritis are associated with the presence of *H. pylori* on surface and in gastric foveolae. The lesion starts in the antrum but often progresses to involve the corpus/fundus and cardia. The organisms can be seen with H&E and are accented by Giemsa and silver (Warthin–Starry or Dieterle) stains. The bacteria may also be found wherever there are gastric surface mucous cells, including the duodenal mucosa at sites of gastric mucous cell metaplasia, Barrett's esophagus, and in heterotopic stomach.

Chronic fundic gastritis is an immunologic disorder with dominant effect in the corpus and fundus; it may progress to atrophic gastritis and primary pernicious anemia. The severe "atrophic" form shows complete loss of specialized fundic–corpus cells and prominent metaplasia of pyloric and intestinal glands. Biopsy is of limited use in overall diagnosis because of the potential of sampling error; biopsy is mainly done to exclude other diseases or as surveillance procedure to detect dysplasia and early carcinoma. The low acid may also lead to enhanced G cell stimulation in the antrum, resulting in proliferation of the enterochromaffin-like (ECL) cells in the body.

Biopsy helps to detect the several complications of chronic gastritis, including xanthoma, hyperplastic polyp, glandular dysplasia, adenocarcinoma, and endocrine cell hyperplasia/neoplasia.

Other Forms of Gastritis

There are several other causes and forms of gastritis:

1. Other infections. Biopsy can identify opportunistic agents, including cytomegalovirus, mycobacteria, and fungi.
2. Ischemic lesions. These are seen in stress and shock. Noted are hemorrhages, erosions, and ulcers, simulating changes seen in severe chemical gastritis.
3. Granulomatous gastritis. This is most marked in the antrum, and cases grossly mimic carcinoma. Causes include Crohn's disease, tuberculosis and other infections, foreign bodies, isolated (most common), and sarcoidosis.
4. Eosinophilic (allergic) gastroenteritis. Glandular necrosis and regeneration with many eosinophils are seen, most marked in the antrum. Associated lesions may be in the esophagus, small bowel, and/or colon. In cases with enteritis, gastric antral biopsy helps to separate an allergic disorder and other causes such as celiac disease or viral enteritis.

Chronic Peptic Ulcer

Diagnosis must distinguish between epithelial alterations due to degeneration or regeneration and true dysplasia; conservatism is the rule. The ulcers due to aspirin and other antiinflammatory drugs are suggested by noting normal mucosa at the edge (present in about one-half of cases). There is invariably inflammation in the chronic peptic ulcer cases. *Helicobacter pylori* are present in the inflamed antral mucosa in the large majority of cases of chronic peptic ulcers, both of the stomach and the duodenum.

Mucosal Hypertrophy and Hyperplasia of the Stomach

Biopsy can help to demonstrate the pure hyperplasia of the gastric pits in Menetrier's disease, and the increase of parietal cells in the Zollinger–Ellison syndrome. Other causes of rugal hypertrophy include normal variation, some inflammatory conditions, and tumors.

Gastric Polyps

Biopsy distinguishes the more common hyperplastic (inflammatory, regenerative) and cystic fundic gland polyps from the true neoplastic (adenomatous, villous) types. Rarer forms are juvenile, Peutz–Jegher, and inflammatory fibroid polyps. The major object of biopsy is to determine whether the polyp is an adenoma. If it is an adenoma, the polyp must be completely excised because of the strong chance of having or developing carcinoma. If the polyp is nonneoplastic, the particular type (and name) may depend on full clinical and distributional data.

Malignant Tumors of the Stomach

Adenocarcinomas, lymphomas, and other tumors must be differentiated. For optimal detection of adenocarcinomas, multiple biopsies are obtained from the edges and central ulceration of the lesion. The peripheral samples identify most tumors with gland formation, whereas the central biopsies are better to show lesions that are infiltrative under the mucosa, such as the signet ring cell form of adenocarcinoma and mural tumors. Biopsies are also useful in the detection of glandular dysplasia or early carcinoma in patients with promoting conditions, such as chronic gastritis or polyposis.

Patients with long-standing *H. pylori* gastritis may develop a malignant lymphoma involving the marginal B cells, commonly called mucosa-associated lymphoid tissue (MALT) lymphoma. Biopsy shows proliferation of immature lymphoid cells with invasion into the pit epithelium, termed lymphoepithelial lesion. Because the early lesions may be subtle and difficult to distinguish from simple chronic gastritis, other tests using fresh samples may be needed, including immunostains, flow cytometry, and polymerase chain reaction to determine monoclonality. Other gastric tumors include endocrine, mesenchymal, and metastatic lesions.

DUODENUM

Normal Duodenal Mucosa

Endoscopic biopsies are obtained to assess duodenitis and unusual ulcers, to screen for malabsorptive disorders, and to detect tumors of the duodenum and adjacent structures. In contrast to the jejunum, the mucosa of the proximal duodenum frequently reveals variation in villous height and more chronic inflammation, possibly reflecting constant physiologic damage from acid.

Duodenitis

In peptic duodenitis, there is variable villous shortening, epithelial cell degeneration, and active inflammation, in the form of increased intraepithelial lymphocytes or neutrophils. Chronic duodenitis causes

gastric mucous cell metaplasia and Brunner gland hyperplasia.

In infections, biopsy is done mainly to look for *Giardia* or for opportunistic infections in immunosuppressed patients. Common examples are viral inclusions of herpes and cytomegalovirus (CMV), *Mycobacterium avium* in acid-fast stains, fungi, cryptosporidia, and microsporidia in gram stains. Other causes of duodenitis include stress hemorrhages and ulcers, drug effects, radiation, and Crohn's disease.

Malabsorptive Disorders

Endoscopic biopsy of the duodenum is used to screen for a diffuse enteritis such as celiac disease. If biopsy is normal, it effectively excludes celiac disease, which would always involve the duodenum. Conversely, if biopsy reveals enteritis, it is not specific, because identical changes could be due to peptic injury. Features of celiac disease include shortening of villi, degeneration of surface absorptive cells, increased intraepithelial lymphocytes, crypt hyperplasia with more mitoses, and increased mononuclear inflammatory cells and eosinophils in the lamina propria. Biopsy may also help to detect focal lesions of Whipple's disease or lymphangiectasia, and a variety of other immunologic disorders.

Tumors of the Duodenum

Benign nodules include heterotopic stomach, Brunner gland hyperplasia, and lymphoid aggregates. Neoplasms may be detected in the duodendal mucosa and ampulla, showing the full range of adenoma, adenocarcinoma, lymphoma, and endocrine tumors.

ILEUM

Normal Ileal Mucosa

Compared to the jejunum, there is a greater number of goblet mucous cells in the villi and larger quantities of lymphoid tissue in the ileum. Furthermore, biopsies of ileum proximal to an anastomosis or stoma often reveal a variable degree of atrophy, characterized by villous shortening and crypt hyperplasia. For a diagnosis of active ileitis, it is necessary to see definite destruction of the surface or crypt epithelium, usually associated with neutrophils.

Ileitis

Biopsy is done in patients who have had ileocolic anastomosis, ileal pouch, or ileostomy and present with ileal dysfunction or suspected inflammation. The presence of pyloric gland metaplasia is especially helpful in detecting chronic ileitis.

Mechanical, vascular, or septic complications of a nonspecific nature may occur. Prestomal ileitis is characterized by nonspecific inflammation and ulceration, typically associated with a dilated intestinal segment. Pouch ileitis, or acute pouchitis, occurs in about 20–25% of cases and usually responds to antimicrobial drugs. Chronic changes of severe atrophy are noted in a small percentage of patients, and such patients have an increased risk for dysplasia and carcinoma in the pouch mucosa. Recurrent Crohn's disease is suspected by stenosis or sinuses away from anastomosis or stoma. There is a need to see both ulceration and granuloma in biopsy, because either alone may not signify specific or active disease.

Tumors of the Ileum

Exceptionally, lesions of the terminal ileum may be biopsied through an intact ileocecal valve. Endoscopy and biopsy are also used for the detection of dysplasia, adenomas, and carcinomas in an ileostomy segment and in pouches, in patients who have had colectomy for adenomatous polyposis or ulcerative colitis.

COLON AND RECTUM

Normal Colonic/Rectal Mucosa

Biopsies are employed for the diagnosis, extent, and severity, and complications of the various inflammatory disorders; and also for the detection and typing of abnormal mucosal growths—polyps, primary and metastatic cancers, dysplasia in chronic colitis, multiple and diffuse polyposis syndromes, and tumor-like conditions.

The crypts are generally arranged in a parallel fashion, are composed predominantly of goblet mucous cells, may show minor branching at the base, and extend from the surface to abut on the muscularis mucosae. The surface epithelial cells between the crypts are more variable and include nonmucous forms. There is a considerable variation in the quantity and types of chronic inflammatory cells (including lymphocytes, plasma cells, eosinophils, and macrophages) in the lamina propria, and also present are numerous lymphoid nodules covered by an attenuated epithelium of M cells. Because of this great variation, it is best not to make a diagnosis of inflammatory disease unless there is the presence of neutrophils and particularly the appearance of damage or repair of the epithelial cells. Paneth cells are normally present in the right portion of the colon. The lower part of the rectum may show shortened crypts and increased inflammation, simulating chronic inactive colitis.

Infections

The lesions are often focal in the early stage but then proceed to a more diffuse colitis, and granulomas may be present in some cases. For most viral and bacterial infections, specific cultures are required. Biopsy may be used to identify some specific agents—particularly in opportunistic infections affecting immunocompromised patients and others at high risk—including viral inclusions of herpes (more in anal squamous cells) and cytomegalovirus (in glandular epithelium and mesenchymal cells), fungi, protozoa, and helminthic ova.

Antibiotic-Associated Pseudomembranous Colitis

Early lesion is distinctive, showing discrete foci of dilated crypts and overlying marked mucus and neutrophils, in the shape of a mushroom. Later lesions are less specific, revealing more necrosis, inflammation, and hemorrhage and resembling ischemic disease or other acute colitis. Specific diagnosis requires identification of *Clostridium difficile* toxin.

Ischemic Disease

Early lesion of hemorrhage and crypt dropout with minimal acute inflammation are characteristic. Later lesions show ulceration and marked acute inflammation ("ischemic colitis") and mimic other causes of active colitis, especially infections due to *C. difficile* and enterohemorrhagic *E. coli*.

Ulcerative Colitis and Crohn's Disease

There are multiple reasons for the performance of endoscopy and biopsy in patients with idiopathic inflammatory bowel disease:

1. To detect or exclude, in selected cases, other specific causes of colitis or proctitis that have a distinctive histology (e.g., early ischemic disease, pseudomembranous colitis, protozoan infection).

2. To confirm, exclude, or document a case of suspected colitis or proctitis—subsequent biopsy may be able to distinguish between chronic disease in remission (by noting some alteration of crypts) and complete recovery of a reversible acute colitis (which would appear completely normal). But, this distinction may be confounded by the treatment, leading to a normal-appearing biopsy in patients with chronic inactive colitis.

3. To distinguish acute self-limited colitis (mostly, infectious) and chronic colitis (usually ulcerative colitis or Crohn's disease)—major features of chronic colitis include crypt irregularity in the form of complex branching, crypt atrophy, villiform surface, Paneth cell metaplasia, and the presence of lymphocytes and plasma cells in the base between the crypts and the muscularis mucosae (Fig. 3). Acute colitis is suggested by the presence of normal crypt architecture and a relatively superficial damage of the epithelium, with appearance of only neutrophils, but this is not absolute, because some cases of chronic colitis may initially show such features.

4. To assist in the diagnostic distinction between ulcerative colitis and Crohn's disease—accurate diagnosis of the particular type requires consideration of the gross features and distribution of the disease, as assessed by X ray and endoscopy as well as the mucosal biopsies. The patterns available on the superficial mucosal biopsy are limited. A normal biopsy as evidence of a spared rectum or other skip area, a focal colitis or proctitis, or the presence of granulomas would support or indicate CD. A diffuse colitis or proctitis would be compatible with either form, and in such cases the diagnosis would depend on other data.

5. To determine disease extent and severity—whatever the type, biopsies may be used to monitor the course and estimate activity in cases of apparent remission.

6. To distinguish polypoid dysplasia versus adenoma—sporadic adenoma- and colitis-associated dysplasia cannot be distinguished by histology. There is a tendency to favor adenoma in older patients, in those with disease of shorter duration, and the finding in an

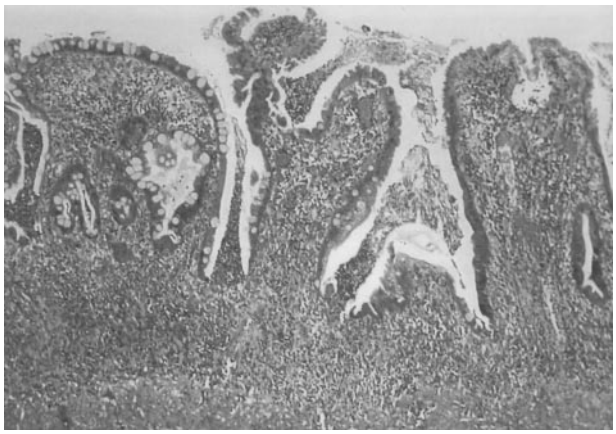


FIGURE 3 Colonic mucosal biopsy, showing chronic active colitis in a patient with ulcerative colitis. The luminal surface is at the top. There is diffuse irregularity of the crypts and marked infiltrate of mononuclear inflammatory cells and eosinophils in the base of the mucosa separating the crypts from the muscularis mucosae (at the bottom). Also seen is active disease in the form of crypt abscesses.

area without surrounding colitis, particularly if the lesion has a stalk. Conversely, dysplasia seems more likely in younger patients, in those with more extensive disease of longer duration, and in lesions that are sessile and present within an area of the colitis.

7. To detect dysplasia and carcinoma—biopsy is done of grossly evident mass or stenotic lesion and also in colonoscopic surveillance to detect dysplasia. Biopsies should be taken from any villous growth, from other atypical-appearing polyps, and from flat areas that seem the least inflamed. They must distinguish histologically between effects of degeneration and regeneration and true dysplastic epithelium. This may be difficult in an area of great inflammation, and experience is essential.

The following biopsy classification for dysplasia has been suggested:

a. Negative for dysplasia—includes normal mucosa and effects of inactive (or quiescent) colitis and active colitis.

b. Indefinite for dysplasia—specify particular technical or interpretative reason, such as excessive acute inflammation or appearance of unusual growth pattern. These are further rated as probably negative, unknown, or probably positive, because this assists in decisions about early rebiopsy.

c. Positive for dysplasia (Fig. 4)—rate by worse area as low-grade dysplasia or high-grade dysplasia. Low grade is roughly equivalent to mild dysplasia and shows architectural abnormalities that are not greater than those seen in chronic colitis but, in addition, palisading of nuclei that occupy up to half the cell height, not related to inflammation. High grade encompasses all abnormalities from moderate to severe to carcinoma *in situ*. Extension of atypical cells to the surface helps to distinguish dysplasia from effects of regeneration alone.

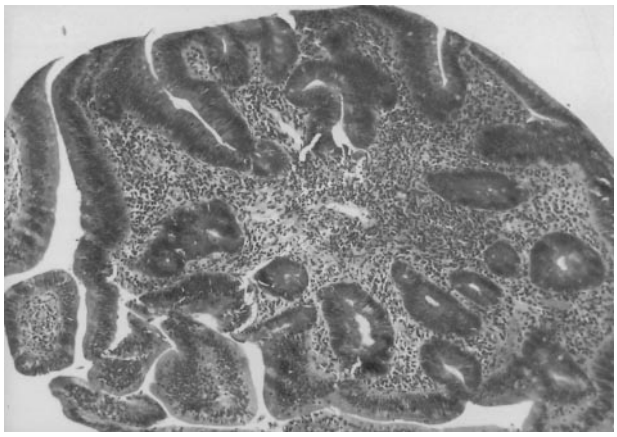


FIGURE 4 Colonic mucosal biopsy in a patient with chronic ulcerative colitis, revealing high-grade glandular dysplasia. The cells contain large and palisading nuclei that occupy more than half the height of the cell and show variable hyperchromatism. The altered cells extend to the surface at the top. Contrast with the nondysplastic epithelium seen in Fig. 3.

Other Inflammatory Disorders of the Colon and Rectum

Motor and mechanical conditions may require biopsy. Biopsy is used to look for ganglia in the submucosa and for associated colitis in cases of Hirschsprung's disease; and to evaluate cases of diverticular disease, diversion-related colitis, and obstructive colitis. Vascular disorders, i.e., ectasias and malformations, can be visualized by endoscopy and then cauterized. In vasculitis, the lesions tend to involve larger submucosal vessels and are missed by ordinary endoscopic biopsy, which usually contains only the mucosa; for these cases, deeper aspiration-type biopsy can be taken from the rectum.

Immunologic diseases can be assessed using biopsies. Allergic disease is either due to a single allergen (milk or soy protein) or is part of a generalized reaction (eosinophilic or allergic gastroenteritis). Patients with the form limited to the large intestine (allergic proctitis or colitis) are very young (average age, 2 months) and usually present with rectal bleeding alone or in combination with diarrhea. Biopsy reveals focal areas of crypt destruction with numerous eosinophils; the lesion may be very tiny, and multiple sections are recommended. There is often a more diffuse increase of eosinophils in the lamina propria and the musculus mucosae. An increase of apoptosis in the crypt epithelia is sought in cases of graft-versus-host disease.

Collagenous colitis occurs mostly in middle-age women, presenting with chronic watery diarrhea. Radiographic and gross endoscopic examinations are normal. The only constant abnormality is thickening (usually, greater than 10 μm) of the collagen layer beneath the surface epithelium. The lesion may be focal in early and in healing stages, and the rectum may be uninvolved. An increase in lymphocytes within the surface epithelial layer is often present. The collagen lesion is nonspecific; similar thickenings are noted uncommonly in other types of colitis, particularly due to radiation, and in hyperplastic polyps. In lymphocytic colitis, there are similar clinical features, but with a wider age range, and the disease also occurs in men. There is the presence of focal cryptitis and increased lymphocytes in the surface epithelial layer and lamina propria but no collagen thickening.

Biopsy may be useful in miscellaneous conditions that occur in many other disorders that can affect the large intestine, including radiation injury and foreign body reactions to barium and oil; chemical and drug effects from strong enema solutions, cathartics, NSAIDS, heavy metals, and chemotherapeutic agents; metabolic and infiltrative lesions seen in uremia, amyloidosis, and storage diseases; and tumorlike lesions such as endometriosis and pneumatosis.

Polyps of the Colon and Rectum

Polyps may be inflammatory, lymphoid, juvenile (retention), other hamartomatous, hyperplastic, and neoplastic (adenomatous or tubular, villous, mixed). If a polyp is neoplastic, look for the presence and extent of carcinoma—true *in situ* (limited to gland, without trespass of basement membrane), intramucosal (invasion into lamina propria but not through the muscularis mucosae), and submucosal. If there is submucosal invasion, it is important to separate those polyps with stalks from sessile lesions. It is also necessary to distinguish between invasion and effects of inflammation leading to "misplacement" of the benign epithelium into the stalk region. The misplaced tissue contains normal lamina propria, whereas carcinoma has an abnormal fibrovascular stroma.

Polypectomy is sufficient treatment for lesions that are adenomas or contain a carcinoma that is just *in situ* or intramucosal, whereas additional intestinal resection is usually recommended for sessile tumors that show invasion of carcinoma into the submucosa. Polypectomy may also be adequate for pedunculated adenomas with a moderately or well-differentiated carcinoma extending into the stalk but not at or within 1 mm of the resection margin (identified by cautery effect). If the carcinoma is at or very near the stalk margin, is poorly differentiated, or there is definite lymphatic invasion, colonic resection is ordinarily performed.

Malignant Tumors of the Colon and Rectum

Biopsy is done of polyps, masses, and ulcerated and stenotic lesions to detect primary carcinoma, mainly adenocarcinoma, carcinoid tumors, lymphomas, and other mural lesions tumors, and recurrent and metastatic tumors. The biopsy may also be helpful in the evaluation of the differentiation and degree of invasion of the tumor, the assessment of drug and radiation effects, the identification of opportunistic infections, and in surveillance for coexisting and future tumors.

See Also the Following Articles

Colorectal Cancer Screening • Endoscopy, Complications of • Esophageal Cancer Surveillance and Screening: Barrett's Esophagus and GERD • Gastrointestinal Tract Anatomy, Overview • Liver Biopsy

Acknowledgments

The author thanks Dr. Jeremy Ditelberg for providing the illustrations and Ms. Mariann Seymour for the preparation of the manuscript.

Further Reading

Goldman, H. (1996). "Gastrointestinal Mucosal Biopsy." Churchill-Livingstone, New York.

Goldman, H. (1994). Interpretation of large intestinal mucosal biopsy specimens. *Hum. Pathol.* **25**, 1150–1159.

Listron, M. B., and Fenoglio-Preiser, C. (1989). Short course. Gastrointestinal polyps. *Modern Pathol.* **2**, 161–181.

Ming, S.-C., and Goldman, H. (eds.) (1998). "Pathology of the Gastrointestinal Tract," 2nd Ed. Williams and Wilkins, Baltimore.

Odze, R. D. (1999). Adenomas and adenoma-like DALMs in chronic ulcerative colitis: A clinical, pathological and molecular review. *Am. J. Gastroenterol.* **94**, 1746–1750.

Riddell, R. H., Goldman, H., Ransohoff, D. F., *et al.* (1983). Dysplasia in inflammatory bowel disease: Standardized classification with provisional clinical applications. *Hum. Pathol.* **14**, 931–968.

Rotterdam, H., Sheahan, D. G, and Sommers, S. C. (1992). "Biopsy Diagnosis of the Digestive Tract," 2nd Ed. Raven Press, New York.

Whitehead, R. (1996). "Mucosal Biopsy of the Gastrointestinal Tract," 5th Ed. W. B Saunders, Philadelphia.

Multiple Endocrine Neoplasia (MEN)

JEFFREY C. BUEHLER AND GORDON C. HUNT
University of California, San Diego

catecholamines Group of physiologic substances, mainly epinephrine, norepinephrine, and dopamine; act as neurotransmitters in the functioning of the sympathetic nervous system.

gastrinoma Neuroendocrine tumor found mainly in the wall of the duodenum and in the pancreas; secretes excessive amounts of the growth hormone, gastrin.

insulinoma Insulin-producing tumor of the pancreatic beta cells in the islets of Langerhans.

kindred Family relationship by birth, or sometimes by marriage.

medullary thyroid carcinoma Malignant tumor of the thyroid gland; the most common tumor found in multiple endocrine neoplasia types 2A and 2B.

pancreatic islet cells Small groups of cells found throughout the pancreas; subdivided into alpha, beta, and delta islet cell types, which produce glucagon, insulin, and somatostatin, respectively.

prolactinoma Pituitary adenoma that secretes excessive amounts of prolactin, which may lead to the clinical symptoms of galactorrhea or amenorrhea.

VIPoma Endocrine tumor that produces vasoactive intestinal peptide, which belongs to the secretin family of hormones. The resulting syndrome causes a watery diarrhea–hypokalemia–achlorhydria syndrome.

The multiple endocrine neoplasia (MEN) syndromes are rare causes of hormone excess. The MEN syndromes involve hyperplasia, adenomas, and carcinomas of several different endocrine glands. MEN disorders are usually clustered in families and are inherited in an autosomal dominant pattern. Traditionally, MEN syndromes are diagnosed after patients present with clinical signs and symptoms related to elevated hormone levels, and evaluation reveals multiple endocrine abnormalities. Alternatively, MEN syndromes may be discovered when a single endocrine disorder is identified, but a family history of MEN leads to further evaluation and discovery of multiple endocrine abnormalities. With the recent advent of DNA testing for these disorders, people with a genetic predisposition for MEN are being diagnosed in an asymptomatic state.

MEN 1

MEN 1 is characterized by a predisposition to tumors of the parathyroid glands, the pancreatic islet cells, and the anterior pituitary gland. Parathyroid hormone (PTH) excess is the most common abnormality, occurring in excess of 90% of cases of MEN 1, with hypercalcemia being a frequent clinical presentation. Patients usually present in their second to fourth decade of life and are affected almost universally before age 50. Patients with excess PTH (and resultant hypercalcemia) may present with generalized symptoms of weakness and confusion or with gastrointestinal (GI) symptoms of constipation, nausea, or abdominal pain. More commonly, they present with kidney stones from excess calcium filtration into the urine. Elevated calcium levels may also be detected during routine lab testing.

Pancreatic islet cell tumors occur in 40–80% of MEN 1 patients. The tumors can often be asymptomatic and elusive to the physician despite a robust workup. Gastrinomas are the most common symptomatic islet cell tumors, producing excess gastrin and causing recalcitrant peptic ulcer disease. Symptomatic insulinomas occur less frequently. These tumors may present with hypoglycemic symptoms that can result clinically in an altered mental status and that can uncommonly manifest as seizures. Other less common islet tumors include vasoactive intestinal peptide tumors (VIPomas), glucagonomas, and nonfunctioning tumors.

Anterior pituitary tumors are present in a majority of patients with MEN 1 but are clinically apparent in only 15–20%. Presenting symptoms occur from tumor mass effect or from the hormone that is being produced in excess. Prolactinomas are most frequent, but adrenocorticotropic hormone (ACTH)-producing and growth hormone-producing tumors are also seen. Nonfunctional pituitary tumors also occur and frequently present with visual changes due to mass effect.

MEN 2

MEN 2 is subdivided into two types, MEN 2A and MEN 2B. Both disorders are autosomal dominant. MEN 2A is defined by a predisposition to medullary thyroid carcinoma, pheochromocytoma, and parathyroid hyperplasia. Medullary thyroid carcinoma is the most frequent manifestation, occurring in over 90% of cases.

 Copyright 2004, Elsevier (USA). All rights reserved.

The thyroid carcinoma of MEN 2A usually presents in the third and fourth decades of life, most often as a thyroid nodule or cervical adenopathy. Pheochromocytoma with excess catecholamine production (most often epinephrine) is present in about half of MEN 2 cases. Clinical presentation varies, and patients may develop anxiety, excess sweating, headache, and heart palpitations. Hypertension, often paroxysmal, is common and patients will occasionally present in hypertensive crisis. Pheochromocytomas rarely present before the medullary thyroid disease. The last component of MEN 2A is parathyroid excess, which is seen about 25% of the time.

Patients with MEN 2B, like those with MEN 2A, also have a predisposition to medullary thyroid carcinoma and pheochromocytoma, but do not have parathyroid abnormalities. Several other abnormalities are associated with the MEN 2B syndrome, including Hirschsprung's disease, mucosal neuromas, and a marfanoid habitus.

GENETIC TESTING

The MEN 1 gene was discovered in 1997 and mutations in the gene have been present in a majority of MEN 1 kindreds that have been tested. Testing for the MEN 1 gene defect is not yet standard, because there is no proved clinical benefit in determining a presymptomatic diagnosis.

Unlike MEN 1, for which early diagnosis through genetic testing confers an uncertain mortality benefit, the newly developed genetic test for the receptor tyrosine (RET) proto-oncogene defect as seen in MEN 2 patients is of profound benefit. The medullary thyroid carcinoma that is almost universally present in MEN 2 cases is curable with early surgery. Earlier diagnosis through genetic testing for the RET oncogene has helped identify patients needing prophylactic thyroidectomy. Genetic testing has also reduced the number of false positive tests that were generated by older biochemical stimulation testing. Earlier diagnosis is especially crucial in MEN 2B cases, in which the medullary thyroid carcinoma is aggressive, often presenting earlier in life.

GASTROINTESTINAL AND PANCREATIC ISLET CELL TUMORS

Pancreatic endocrine tumors, although often present in cases of MEN 1, are rarely seen in cases of MEN 2. The hyperparathyroidism of MEN 1 almost universally precedes the appearance of pancreatic endocrine tumors. The parathyroid and pituitary endocrinopathies of MEN 1 patients are often effectively managed with surgery and hormone replacement. Unlike pituitary and parathyroid surgery, in which involved organs are removed completely (pituitary) or almost completely (parathyroid), pancreatic tumors are usually enucleated, leaving the remaining pancreas behind. The remaining pancreatic tissue still has malignant potential and screening for recurrent tumor development is problematic and a subject of continued research.

Pancreatic endocrine tumors classically present in the fourth or fifth decade with symptoms of hormone excess. Approximately one-third of MEN 1 patients become symptomatic. More often these tumors are asymptomatic. Because asymptomatic tumors may go undetected, the true prevalence of pancreatic endocrine tumors is difficult to determine. Advanced biochemical testing and anatomical/pathological studies reveal these tumors in up to 80% of MEN 1 cases. Despite being asymptomatic, these pancreatic islet and gastrointestinal tumors may be malignant. In contrast, some symptomatic tumors can be difficult to find despite extensive imaging and evaluation.

Zollinger–Ellison (Gastrinoma) Syndrome

Approximately 60% of patients with MEN 1 have symptoms from excess gastrin or have asymptomatic elevations in serum gastrin. Conversely, about 25% of Zollinger–Ellison (ZE) cases are subsequently diagnosed with MEN 1. The number of peptic ulcers due to gastrin excess is estimated at 0.1% or less, but with improved medical therapy for peptic ulcer disease, ZE ulcer symptoms may be controlled and formal testing for a hypergastrinemic state not performed. This potential for misdiagnosing gastrinomas would result in a misrepresentation of the true prevalence.

Gastrinomas present most commonly after age 20 years as symptoms related to gastroesophageal reflux disease, abdominal pain, and peptic ulcers. The ulcer disease in MEN 1 gastrinomas is more severe than in common peptic ulcers. Gastrinoma-induced ulcers are more often multiple, frequently resistant to treatment, recurrent, and can be atypically located as distal as the jejunum. Diarrhea is also a common presenting symptom and can help distinguish gastrinoma-induced ulcers from other common ulcers. The diarrhea is believed to be the result of excess gastric acid production. Two mechanisms explain this diarrhea. First, excess acid output overwhelms the resorptive capacity of the small bowel and colon, causing a secretory diarrhea. Second, excess acid in the stomach lowers the pH in the duodenum, compromising pancreatic enzyme digestion of lipids, and steatorrhea results. As one would expect with a secretory diarrhea, fasting does not affect the diarrhea. Nasogastric suctioning may, however, alleviate the diarrhea.

TABLE I Causes of Hypergastrinemia

Pharmacologically (drug) induced
Helicobacter pylori infection
Hypochlorhydria/achlorhydria +/− pernicious anemia (i.e., gastric atrophy, chronic atrophic gastritis, gastric carcinoma)
Retained gastric antrum
Incomplete vagotomy
G cell hyperplasia
Renal insufficiency
Massive small bowel resection
Gastric outlet obstruction
Others (rheumatoid arthritis, vitiligo, diabetes, pheochromocytoma, gastroparesis)

A variety of tests are available for diagnosing gastrinomas. Secretin stimulation tests and fasting serum gastrin levels with pH testing of the stomach have been used. Blood samples for determination of fasting serum gastrin levels are drawn while documenting gastric acid hypersecretion through pH analysis of the stomach. Gastrin levels in excess of 1000 pg/ml are very specific for gastrinomas. Less profound levels of gastrin elevation are seen in other disorders, including pernicious anemia, atrophic gastritis, small bowel resection, gastric outlet obstruction, and renal insufficiency. Antisecretory therapy (i.e., proton pump inhibitor therapy) also causes hypergastrinemia by reducing stomach acid and causing gastrin levels to rise through the feedback inhibition loop. Common etiologies of hypergastrinemia are outlined in Table I.

In cases of mild elevation of gastrin, the secretin stimulation test is helpful. Secretin paradoxically causes a release of gastrin from gastrinoma cells, whereas gastrin release is inhibited in normal gastric G cells. This test, however, has been difficult to perform because of a nationwide shortage of secretin. Differentiating sporadic ZE cases from those associated with MEN 1 is important, because their management is different. For example, MEN 1-related hyperparathyroidism can exacerbate gastric acid secretion through hypercalcemia. Treatment of the hypercalcemia can improve gastrin-related symptoms (ulcer pain, diarrhea, reflux). The surgical approach and results are also different in MEN 1 gastrinomas in which tumors are often multiple, small, and often located outside of the pancreas, most commonly in the wall of the duodenum. Malignancy, defined by the presence of metastatic disease (rather than by histology), is present about 60% of the time in MEN 1-related cases, but hepatic metastases are more frequently seen in sporadic gastrinoma cases. Largely due to the aforementioned differences between sporadic ZE gastrinomas and MEN 1 gastrinomas, surgical attempts to cure MEN 1 patients have been less successful. However, some surgeons advocate surgical intervention in patients with MEN 1 only when tumors exceed 2.5 cm in size. In contrast to surgical therapy, medical therapy for gastrinomas is very effective. Proton pump inhibitor (PPI) medications, which reduce gastric acid output and control symptoms, are the mainstay of therapy.

Insulinoma

Insulinomas are the next most common functional pancreatic endocrine tumor, occurring in approximately 20% of MEN 1 patients (Fig. 1). The average age of presentation in most series is between 40 and 50 years of age, with 60% of the insulinomas occurring in females. Most insulinomas present with hypoglycemic (neuroglycopenic) symptoms, including confusion, lightheadedness, blurry vision, behavioral changes, drowsiness, and, rarely, seizure. Hypoglycemic symptoms often occur during periods of increased demand (exercise) or decreased supply (fasting). Sweating, tremor, and palpitations from catecholamine excess occur less often. Because of the general nature of these symptoms, there is often a delay in diagnosis on the order of years, and not infrequently, patients are misdiagnosed with neurologic or psychiatric disorders. Taking a careful history may reveal a pattern of frequent meals to suppress symptoms; consequently, obesity is present in approximately 40% of patients at diagnosis. The diagnosis of insulinoma is made by documenting hypoglycemia in the setting of elevated blood insulin

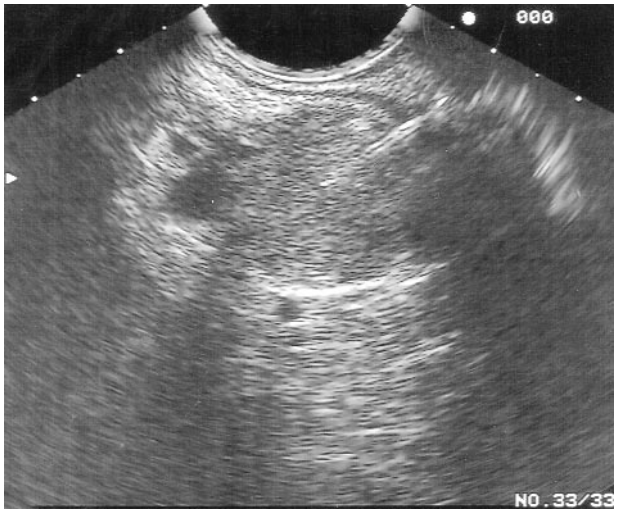


FIGURE 1 EUS image (7.5-MHz linear array echoendoscope) of a 1.7-cm hypoechoic insulinoma as seen in the tail of the pancreas. Fine needle aspiration needle is seen entering the tumor from the 1 o'clock position. This 55-year-old male patient presented with episodes of hypoglycemia and was ultimately diagnosed as having MEN 1 syndrome.

levels with classic hypoglycemic symptoms that reverse with the administration of glucose. A 72-hour fast with serial blood tests for insulin and glucose is the most reliable test to make the diagnosis. Serum samples should also be obtained for C peptide levels and for proinsulin, which will help differentiate from other causes of inappropriately elevated insulin levels.

Unlike gastrinomas, which are frequently found outside of the pancreas, insulinomas are almost invariably present within the pancreas. Insulinomas are characteristically small, measuring less than 1 cm almost half of the time. The insulinomas found in MEN 1 patients are more often multiple and are often accompanied by other types of islet cell tumors. Malignancy is found in a minority of cases, approximately 10%. Recurrence of insulinomas at 10 years occurs approximately 20% of the time in the setting of MEN 1, and 5% of the time in sporadic cases. Therapy is largely surgical, with tumor resection showing the best results. Medical therapy is used in patients who are poor surgical candidates, for whom surgery fails to eradicate the hypoglycemia, or who have metastatic disease. The somatostatin analogue octreotide controls hypoglycemia in roughly 40% of cases.

Other Pancreatic Endocrine Tumors

Other endocrine tumors seen in MEN 1 include pancreatic polypeptide tumors (PPomas), VIPomas, glucagonomas, and growth hormone-releasing factor tumors (GRFomas), the latter three occurring in less than 5% of MEN 1 cases. Pancreatic polypeptide (PP)-secreting PPomas are probably the most common tumors of the endocrine pancreas, but because they are nonfunctional tumors, their true prevalence is difficult to determine. PP is elevated in many of the other functional pancreatic endocrine tumors and thus is not diagnostic of PPoma. Without clinical warning signs, these tumors may appear late in the course of disease, with signs of local enlargement (back pain) or metastatic disease (hepatomegaly). The rare endocrine vasoactive intestinal peptide-secreting VIPomas result in VIP syndrome, characterized by a severe, watery, secretory diarrhea with hypokalemia and hypochlorhydria. Most VIPomas are sporadic, with less than 1% occurring as part of the MEN 1 syndrome. Glucagonomas are rare glucagon-secreting tumors that cause a syndrome of glucose intolerance, weight loss, and anemia. Migratory necrolytic erythema, a rash specific to glucagonomas, develops in the majority of cases. The rash begins with erythema that then blisters and heals with scarring and hyperpigmentation. Growth hormone-releasing factor secreting GRFomas cause acromegaly. These rare tumors are associated with MEN 1 in about one-third of cases.

TUMOR LOCALIZATION

Once clinical symptoms and/or biochemical testing identify a pancreatic endocrine tumor, imaging is necessary in order to determine the location of the primary tumor and whether metastatic disease is present. Traditional imaging with ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) is often used initially because these tests are widely available and noninvasive (Fig. 2). Unfortunately, the sensitivity of the tests is directly dependent on tumor size and many of the pancreatic endocrine tumors, specifically insulinomas, are small (<1 cm) and therefore difficult to detect. The sensitivity of these tests for recognizing primary pancreatic tumors is generally below 50%. The sensitivities for detecting liver metastases are significantly better but usually not above 80%. Imaging is further compromised by the variable location of tumors. For example, gastrinomas, which frequently occur in the duodenal wall, would be missed on imaging studies dedicated to viewing the pancreas.

Two other modalities for localizing endocrine tumors are being used with increasing frequency, endoscopic ultrasound (EUS) (Fig. 3) and somatostatin

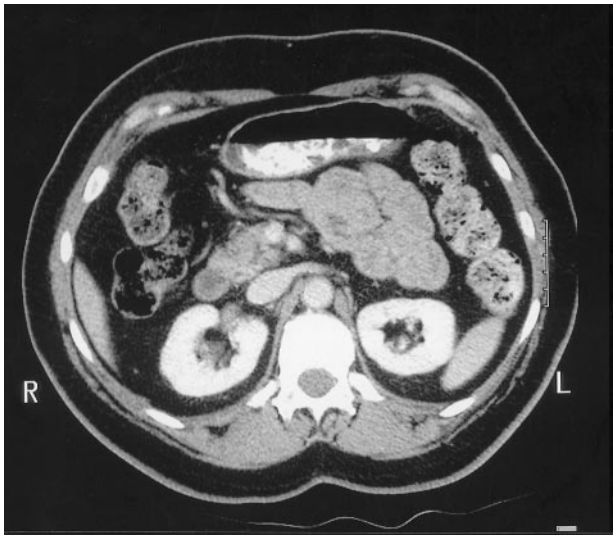


FIGURE 2 Computed tomography scan image of a nonfunctioning (asymptomatic) neuroendocrine tumor in the tail of the pancreas. The tomography scan was ordered to follow up an abnormal ultrasound of the liver, which was performed due to mildly elevated liver function tests. A pancreatic tumor is seen just above the left kidney and below the stomach. No hepatic metastases were present.

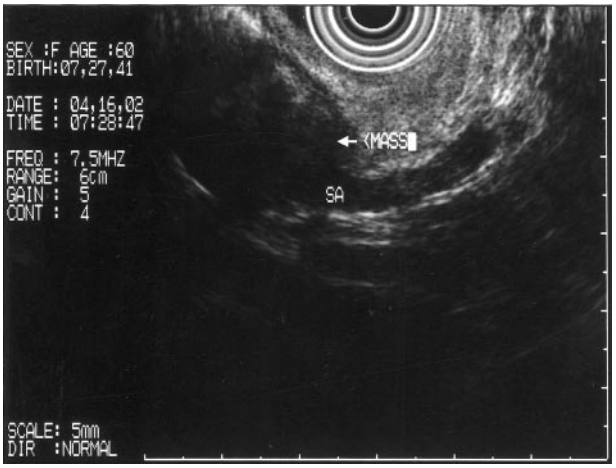


FIGURE 3 Endoscopic ultrasound image (7.5-MHz radial echoendoscope) of a nonfunctioning (asymptomatic) neuroendocrine tumor in the tail of the pancreas, as seen in Fig. 2. The mass is invading the splenic artery (SA) on this image, and also was invading the splenic vein.

receptor scintigraphy (SRS). The proximity of the EUS probe to the pancreas allows for detection of small structures, even those less than 5 mm. The reported sensitivity of EUS for pancreatic tumors is in excess of 80%. EUS is also able to evaluate extrapancreatic tumors in the duodenal wall and local lymph nodes with a sensitivity of over 50%. EUS requires specialized endoscopists and on occasion is unable to visualize the pancreatic tail adequately. Further advances in EUS are being made with the introduction of ultrathin probes, which actually cannulate the main inner duct of the pancreas. Because of their proximity to the pancreas, ultrathin probes provide detailed imaging and accurate detection of pancreatic endocrine tumors.

SRS is based on the high concentration of somatostatin receptors found in most pancreatic endocrine tumors. In SRS, a radiolabeled somatostatin analogue is injected into the patient's bloodstream and binds to tumors with somatostatin receptors. Subsequent imaging reveals the location of these tumors. The primary advantage of SRS is the ability to scan the entire body, including the pancreas, liver, surrounding lymph nodes, and duodenum. Although SRS detects PPomas, gastrinomas, VIPomas, and glucagonomas from 77 to 100% of the time, it is less successful in detecting insulinomas, with an approximate sensitivity of only 50%.

All of these techniques are not widely available, thus the ideal imaging technique is institution dependent. Newer innovations are continuously being refined and developed. The type of tumor dictates which imaging studies are most useful. Broadly speaking, the evaluation of insulinomas starts with one of the three traditional modalities, and if no tumor is found, EUS is often used as the next step. Gastrinomas and other endocrine tumors are well detected by SRS imaging; because of its superior detection rate for both primary and metastatic disease, SRS is often used in addition to traditional imaging. When further localization is needed, more invasive techniques of portal venous sampling and angiography are conducted.

Once a diagnosis is made and the location of the tumor is determined accurately, therapy is possible. Surgical resection is the primary treatment of nonmetastatic disease, and in metastatic disease, surgical debulking may play a role. MEN 1-associated pancreatic tumors, often being multiple and small, are more difficult to cure compared to sporadic cases. In the setting of metastatic disease, different chemotherapy regimens have been tried with variable success. Liver transplantation is increasingly being used in patients with metastatic neuroendocrine tumors.

See Also the Following Articles

Gastrinoma • Hyperparathyroidism • Vipoma

Further Reading

Arnold, A. (1998). Pathogenesis of endocrine tumors. *In* "Williams Textbook of Endocrinology" (J. D. Wilson *et al.*, eds.), 9th Ed., pp. 150–151. W. B. Saunders, Philadelphia.

Del Valle, J., and Scheiman, J. M. (1999). Zollinger–Ellison syndrome. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 1, pp. 1145–1162. Lippincott Williams & Wilkins, Philadelphia.

Jensen, R. T., and Norton, J. A. (1999). Endocrine neoplasms of the pancreas. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 2, pp. 2193–2228. Lippincott Williams & Wilkins, Philadelphia.

Li, M. L., and Norton, J. A. (2001). Gastrinoma. *Curr. Treat. Options Oncol.* **2**, 337–346.

Lips, C. J., Landsvater, R. M., Hoppener, J. W., *et al.* (1994). Clinical screening as compared with DNA analysis in families with multiple endocrine neoplasia type 2A. *N. Engl. J. Med.* **331**, 828–835.

McGuigan, J. (1998). Zollinger–Ellison syndrome and other hypersecretory states. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathology/Diagnosis/Management" (M. Feldman, ed.), 6th Ed., pp. 679–691. W. B. Saunders, Philadelphia.

Metz, D. C., and Jensen, R. T. (1995). Endocrine tumors of the pancreas. *In* "Bockus Gastroenterology" (W. S. Haubrich and F. Schaffner, eds.), 5th Ed., pp. 3007–3034. W. B. Saunders, Philadelphia.

Modlin, I. M., and Tang, L. H. (1997). Approaches to the diagnosis of gut neuroendocrine tumors: The last word (today). *Gastroenterology* **112**, 583–590.

Pipleers-Marichal, M., Somers, G., Willems, G., *et al.* (1990). Gastrinomas in the duodenums of patients with multiple endocrine neoplasia type 1 and the Zollinger–Ellison syndrome. *N. Engl. J. Med.* **322**, 723–727.

Münchausen's and Münchausen by Proxy Syndromes

HILARY K. KLEIN
St. Louis University, Missouri

Münchausen by proxy syndrome Illness intentionally produced in a child by a parent who wants to be seen as concerned and caring.
Münchausen's syndrome Self-inflicted illness in a conscious attempt to gain attention.

Münchausen's syndrome (MS) is a psychiatric disorder in which patients self-inflict symptoms of illness in a conscious attempt to gain attention. Münchausen by proxy syndrome is a type of child abuse in which serious illness is intentionally produced in a child by a parent who wants to be seen as caring and concerned.

MÜNCHAUSEN'S SYNDROME

Münchausen's syndrome (MS) is a chronic psychiatric disorder in which patients consciously simulate or self-inflict symptoms of illness in repeated attempts to gain hospital admission. These patients seem to be motivated by the desire to play the sick role and to be the center of attention.

Clinical Features

Clinical features include the following: pathological lying with presentation of the history in a dramatic, vague, and inconsistent manner; evidence of prior treatment at various hospitals; multiple scars; medical sophistication; disruptive behavior while hospitalized; symptoms that shift from one organ system to another; tolerance for painful and invasive procedures; frequent demands for analgesic medications without signs of withdrawal when they are discontinued; and absence of visitors.

Symptoms of gastrointestinal cases include chronic diarrhea, abdominal pain, nausea and vomiting, ingestion of a foreign body, and patient-reported intestinal obstruction or inflammatory bowel disease.

Diagnosis

The diagnosis of MS is usually made by exclusion after diagnostic procedures have failed to elucidate a physiologic derangement. All means of noninvasive testing should be used to confirm a diagnosis of MS. For example, in cases of chronic diarrhea, stool testing for phenolphthalein, sulfate, magnesium, or anthracene derivatives can identify laxative use. Finally, psychiatric consultation should be sought.

Evidence gathering should be limited to situations in which significant morbidity or mortality is imminent. When imminent danger is not clear, legal and ethical consultation should be sought prior to initiating any action that violates the patient's rights to privacy and confidentiality. Obtaining the patient's consent for videotape surveillance has been successfully used in some cases because it provides these patients with the attention that may be a primary motivating factor in MS. None of these actions should be undertaken without first discussing the situation with legal personnel.

Treatment

Managing the patient with MS is difficult, with many patients continuing to deny their complicity in their illness and fleeing the hospital to avoid further confrontation. The primary goal of treatment is to prevent secondary complications from unnecessary tests and procedures. Staff that feels angry may make premature confrontations, resulting in discharge of the patient against medical advice. If confrontation of the patient is advisable, a psychiatrist can help to plan a conference where the disparities between the patient's presentation and findings are discussed with the patient. Confrontation, however, is not a cure. An offer for referral for psychiatric treatment can be made but these offers may be rebuffed. Co-morbid psychiatric disorders should be diagnosed and treated. Finally, the medical chart should be flagged in order to communicate the

 Copyright 2004, Elsevier (USA). All rights reserved.

necessity for definitive diagnostic testing prior to any invasive treatment.

MÜNCHAUSEN BY PROXY SYNDROME

Münchausen by proxy syndrome (MBPS) is a chronic form of child abuse in which the parent, usually the mother, produces serious illness in her child. She is usually motivated by a desire to be publicly recognized for her apparent devotion to the child. In approximately 25% of cases, the abusing parent also has a history of overt MS.

Clinical Features

Clinical features include the following: dramatic presentation of the history by a parent who claims to have saved the child's life; occurrence of illness events when the parent is present and absence of episodes of illness when the parent is separated from the child; and the absence of any independent observer who could corroborate the onset of the illness symptoms.

Diagnosis

Covert surreptitious videotaping in order to catch the parent in the act of abusing the child may frequently reveal a parent who ignores the child for long periods of time when no one else is present. It may also provide direct legal evidence of criminal behavior since the child may be unable to reveal the abuse. The necessity to protect a child from harm is a legal and ethical principle that takes precedence over the principle of confidentiality. A mortality rate of 9% and incalculable physical and psychological morbidity make intervention critical.

Treatment

The safety and protection of the child are the primary concerns. Other siblings may need protection as they have as much as a 25% risk that they are also being victimized. In addition, children who are victimized in this way have an increased risk of developing MS as adults. Suspicions must be reported to the appropriate child protective services agency immediately so that a formal investigation can proceed as quickly as possible.

See Also the Following Articles

Emesis • Foreign Bodies • Nausea

Further Reading

Folks, D., and Freeman, A. (1985). Münchausen's syndrome and other factitious illness. *Psychiatric Clin. North Am.* **8**, 263–278.

Hyler, S., and Sussman, N. (1981). Chronic factitious disorder with physical symptoms (the Münchausen syndrome). *Psychiatric Clin. North Am.* **4**, 365–377.

Schreier, H., and Libow, J. (1993). "Hurting for Love: Münchausen by Proxy Syndrome." The Guilford Press, New York.

Mycobacterial Infection

HYUN SIL KIM
University of California, San Diego

***Mycobacterium avium* complex** A member of the *Mycobacterium* family, it causes diffuse systemic infection in immunocompromised patients and can lead to cachexia, fever, and weight loss.

Mycobacterium tuberculosis An acid-fast intracellular bacterium that causes tuberculosis. The primary site of infection is the lung but infection can involve the gastrointestinal tract and can spread throughout the body, where it is termed miliary tuberculosis.

Mycobacterial infection of the gastrointestinal tract is relatively uncommon in the United States, although the number of reported cases has increased since the late 1980s due to the epidemic of acquired immunodeficiency syndrome (AIDS). *Mycobacterium tuberculosis* is the most common mycobacterial infection and accounts for the majority of mortality associated with mycobacterial infection. *Mycobacterium avium* complex is an opportunistic infection since it affects mostly immunocompromised patients, such as those who have AIDS, those on immunosuppressive medications, the severely ill, and the elderly.

INTRODUCTION

Mycobacterium is an acid-fast intracellular bacterium found throughout the world. In the United States, 90% of mycobacterial infections are caused by *Mycobacterium tuberculosis* (Mtb) and $<5\%$ are due to *Mycobacterium avium* complex (MAC). The influx of new immigrants from indigent areas with a high prevalence of Mtb and the epidemic of acquired immunodeficiency syndrome (AIDS) have presented a major challenge to patients and health care professionals in recent years, and the resurgence of Mtb and MAC infection in the United States has generated a renewed focus on the epidemiology and various treatment options for patients infected with mycobacterial organisms.

MYCOBACTERIUM TUBERCULOSIS

Historically, tuberculosis infection has been a major culprit of a significant mortality worldwide and over one-third of the world's population is currently infected with Mtb. Prior to the AIDS epidemic, Mtb infections were predominantly seen in immunocompromised patients and those who live in endemic areas, such as Southeast Asia, Africa, and Latin America. The disease is transmitted from person to person through respiratory aerosolization. In rare cases, tuberculosis infection of the gastrointestinal tract can occur when unpasteurized milk from *Mycobacterium bovis*-infected cattle is consumed.

Mtb infection of the intestine is relatively uncommon and it is usually seen in the setting of miliary tuberculosis. Gastrointestinal involvement by Mtb is thought to occur either by swallowing Mtb-infected respiratory secretions or by ingesting the organism directly. The most common sites of involvement are the terminal ileum, cecum, and appendix. Patients may present with abdominal pain, diarrhea, fever, weight loss, anorexia, and, in patients with severe pulmonary Mtb infection, cough and night sweats. Mtb lesions of the intestine initially may show mucosal ulcerations with necrosis or inflammation, but such complications as perforation, hemorrhage, fistula formation, and obstruction due to severe fibrosis can be seen in chronic infection. Often, enlargement of mesenteric lymph nodes accompanies intestinal infection and the lymph nodes can become caseous and granulomatous. It is thought that similar to pulmonary lesions, the pathogenesis of intestinal Mtb is mediated by lamina propria macrophages that ingest Mtb, which then release various pro-inflammatory cytokines and induce the host mucosal immune response, leading to caseating necrosis and inflammatory changes with granuloma formation, the pathologic hallmark of tuberculosis infection.

Diagnosis of Mtb infection is made by culture, histology, and most recently, polymerase chain reaction amplification techniques using tissue biopsies and stool specimens with Mtb organism-specific sequences. Histologic examination may reveal acid-fast bacilli with large foamy macrophages and granulomas. The treatment of Mtb requires multidrug regimens due to the high incidence of antibiotic resistance but is effective in most cases. A standard initial therapy includes isoniazid,

 Copyright 2004, Elsevier (USA). All rights reserved.

rifampin, and pyrazinamide, and other anti-mycobacterial agents, such as ethambutol and streptomycin, can be added according to the culture and susceptibility results. Occasionally, due to strictures and obstruction, surgical resection may be indicated.

MYCOBACTERIUM AVIUM COMPLEX

Although MAC infection is usually subclinical and self-limited in healthy individuals, it is a common gastrointestinal bacterial infection in AIDS patients whose CD4 counts are $<50-100/mm^3$ and in patients who receive immunosuppressive treatment. MAC can be found in many environmental sources, including water, soil, and dust. The pathogenesis of MAC infection is similar to that of Mtb, which involves the phagocytosis of the MAC organism by macrophages, leading to intense inflammatory responses resulting in observed clinical characteristics of diffuse systemic lymphadenopathy, organomegaly, and thickening of the intestine. In addition, MAC infection of the liver and biliary tree is common in AIDS patients. Infected patients typically present with persistent fever, diarrhea, abdominal pain, weight loss, and progressive wasting. MAC infiltration of the small intestine can lead to malabsorption and a severe wasting syndrome. Major intestinal bleeding and colitis with ulceration can be found in some patients with MAC infection. Diagnosis can be made by blood or bone marrow cultures or histology (e.g., intestinal or liver biopsy), which may demonstrate acid-fast organisms within enlarged macrophages.

The current recommendation for patients infected with human immunodeficiency virus and patients with AIDS whose CD4 counts fall below $50/mm^3$ is a prophylactic treatment against MAC infection with azithromycin, clarithromycin, or rifabutin. For those patients with documented MAC infection, treatment regimens include clarithromycin, ethambutol, rifabutin, ciprofloxacin, ofloxacin, or amikacin. Despite multidrug treatment, however, the response is slow, the eradication of MAC remains difficult, and many patients have a recurrence of symptoms from persistent MAC infection.

SUMMARY

Mycobacterium is the cause of significant mortality and morbidity throughout the world, and since the AIDS epidemic, two major mycobacterial organisms, *M. tuberculosis* and *M. avium* complex, have emerged as important agents that infect the intestine and cause diarrhea, inflammation, bleeding, and malabsorption. Careful workups that include necessary microbial studies to identify the organism are essential for the proper treatment and follow-up of patients.

See Also the Following Articles

AIDS, Gastrointestinal Manifestations of • Gastric Infection (non-*H. pylori*)

Further Reading

Bermudez, L. E., and Sangari, F. J. (2001). Cellular and molecular mechanisms of internalization of mycobacteria by host cells. *Microbes Infect.* **3**, 37–42.

Jakubowski, A., Elwood, R. K., and Enarson, D. A. (1988). Clinical features of abdominal tuberculosis. *J. Infect. Dis.* **158**, 687–692.

Marshall, J. B. (1993). Tuberculosis of the gastrointestinal tract and peritoneum. *Am. J. Gastroenterol.* **88**, 989–999.

Sangari, F. J., Parker, A., and Bermudez, L. E. (1999). *Mycobacterium avium* interaction with macrophages and intestinal epithelial cells. *Front. Biosci.* **4**, D582–D588.

Shah, S., Thomas, V., Mathan, M., *et al.* (1992). Colonoscopic study of 50 patients with colonic tuberculosis. *Gut* **33**, 347–351.

Nausea

ROBERT M. STERN
The Pennsylvania State University

area postrema An area of the medulla where toxins can be detected and vomiting is initiated.
gastric tachyarrhythmia Abnormal dysrhythmic gastric activity that occurs at a frequency of 4–9 cycles per minute and that is often accompanied by reports of nausea.
optokinetic drum A device with alternating black and white vertical stripes that is rotated around a subject to provoke motion sickness.
vasopressin An antidiuretic hormone that is secreted by the posterior pituitary and that increases in the blood when subjects report nausea.
vection The illusion that a stationary subject is rotating caused by the rotation of the surround.

Nausea is a difficult-to-describe sick or queasy sensation usually attributed to the abdominal area. Nausea is similar to pain and fatigue in that it is an unpleasant sensation, but it may be crucial for survival. Very little is known, however, about the complex pathophysiology or biopsychosocial mechanisms of nausea.

INTRODUCTION

Nausea is often assumed to be the conscious awareness of unusual activity in the chemoreceptor trigger zone, the so-called vomiting center in the medulla of the brainstem, but despite numerous references to a vomiting center in the literature, the existence of such an anatomically well-localized structure remains controversial, as does its relationship, if any, to nausea. The subjective nature of nausea, as opposed to vomiting, with its easily observable and quantifiable characteristics, may partly account for the paucity of research and knowledge about nausea.

Nausea, in addition to being a sensation, is a complex control mechanism with multiple detectors that inhibits food intake in certain situations, such as when the available food is perceived as being disgusting, when the available food has been previously associated with nausea (conditioned taste aversion), or when one's stomach is not functioning normally.

Nausea is frequently followed by vomiting and sometimes the symptoms of nausea abate following vomiting. This sequence of events has led some to assume that nausea and vomiting are on a continuum. However, there are situations in which nausea is present to varying degrees, but the individual does not vomit, such as in some individuals receiving chemotherapy for cancer, and other situations in which individuals vomit without any sensations of nausea, such as astronauts in space. Furthermore, there are drugs, such as ondansetron, that act as anti-emetics, effectively reducing vomiting following chemotherapy, but that do little to relieve nausea. The relationship of nausea to emesis is obviously not simple or clear.

THE ROLE OF NAUSEA IN THE SURVIVAL OF THE ORGANISM

It is generally thought that in cases in which a toxin has been ingested, the biological basis for emesis is to remove the toxin, whereas the role of nausea is to create a conditioned aversion to that substance so that it will be avoided in the future. Such life-saving aversion might not develop in the absence of nausea, since uncomplicated vomiting may not be particularly noxious; e.g., a person might drink 10 beers, vomit, and drink 10 more beers.

Several investigators have discussed the role of nausea in each of four levels of toxin defense: (1) nausea and avoidance evoked by smell and taste; (2) detection of ingested toxins by gastric receptors followed by a central reflex that evokes nausea and vomiting; (3) detection of circulating toxins by chemoreceptors in the area postrema followed by a central reflex that evokes nausea and vomiting; and (4) detection of disturbances in normal sensory input and/or motor control. Disturbance of this system is often suggested to be the cause of the nausea and vomiting that are seen as symptoms of motion sickness.

As indicated above, nausea is a control mechanism and, as such, inhibits ingestion either when there may be something wrong with food being approached or eaten or when there may be something wrong with one's stomach; i.e., it is not functioning normally because of abnormal local activity or because of unusual signals

 Copyright 2004, Elsevier (USA). All rights reserved.

from the central nervous system. The survival value of avoiding foods that have made one nauseated in the past, conditioned taste avoidance, is obvious. Several authors have pointed out that the most potent internal event for producing taste aversion is nausea.

THE THRESHOLD FOR NAUSEA

There are large individual differences in the threshold for the complex control mechanism that is labeled "nausea" that depend on the interaction of inherent and situational factors and the threshold may change from moment to moment. What follows is a brief description of some of the factors that have been studied, but the list is not exhaustive nor are the factors included here thought to be independent.

Inherent Factors

1. Age: Younger people tend to report more nausea than older individuals.
2. Gender: Women report more nausea than men.
3. Race: Asian American people report more nausea than European Americans or African Americans.

Situational Factors

There is great overlap among the following situational factors and, indeed, what one investigator refers to as, for example, "anticipation," another may label "expectation."

1. Anxiety: Reports of nausea are common in patients with various anxiety disorders, such as generalized anxiety disorder, and it has been found that high anxiety contributes to greater side effects, including more nausea and vomiting associated with chemotherapy.

2. Expectation: Several studies have found that expectations about nausea prior to one's first chemotherapy treatment can lower that individual's nausea threshold.

3. Anticipation: Anticipatory nausea prior to chemotherapy is thought to be a learned response; it is seldom reported by patients who have not experienced nausea during previous treatments. Data tend to support a classical conditioning model with the chemotherapy nurse or the sight or smell of the clinic acting as the conditioned stimulus, the chemotherapy drugs acting as the unconditioned stimulus, and nausea and vomiting being the unconditioned response.

4. Adaptation: Adaptation occurs in most individuals exposed to provocative motion who are reexposed to the nauseogenic stimulus within 48 h.

5. Pregnancy: Nausea and vomiting during the first trimester of pregnancy are normal and common, but little is known about the cause, treatment, or role of nausea during the first trimester of pregnancy.

THE SUBJECTIVE EXPERIENCE OF NAUSEA

Nausea is a subjective experience and, as such, its occurrence and characteristics are best described by individuals experiencing the feeling. Health care professionals and caregivers tend to underestimate significantly the frequency, duration, and severity of nausea experienced by patients. A variety of self-report questionnaires have been developed to describe the experience of nausea. Visual analogue scales have been used frequently to describe its occurrence and severity. Morrow developed a questionnaire to assess the nausea and vomiting of cancer patients receiving chemotherapy based on the frequency, severity, and duration of nausea and vomiting.

In a recent study designed to explore the possibility that the experience of nausea may be very different in different situations and for different individuals, descriptors were obtained for the sensation of nausea from a large group of subjects voluntarily undergoing vection-induced motion sickness inside of a rotating optokinetic drum. The descriptors were factor-analyzed and Muth *et al*. developed a nausea profile. The results indicated that there are three dimensions of nausea that can be compared across subjects or patients in a variety of situations. Using this new questionnaire, a total nausea score is calculated, and a nausea profile is determined for each subject or patient that is composed of a score on each of three dimensions: bodily distress, gastrointestinal distress, and emotional distress.

PHYSIOLOGICAL CHANGES THAT ACCOMPANY THE DEVELOPMENT OF NAUSEA

Subjects who experience nausea when exposed to a rotating optokinetic drum show an increase in sympathetic nervous system activity and a decrease in parasympathetic nervous system activity, followed by a change in gastric myoelectrical activity from a regular 3 cycles per minute (cpm), the normal gastric frequency of humans, to dysrhythmic 4–9 cpm activity, or gastric tachyarrhythmia. The disruption in normal gastric myoelectrical activity is usually followed by reports of nausea and an increase in vasopressin levels in the blood.

Similar changes in gastric activity have been found in patients reporting nausea following chemotherapy, suggesting a similar role for the autonomic nervous system in the expression of nausea in that context.

Several studies have demonstrated that nausea is accompanied by abnormal dysrhythmic electrical activity in the stomach and little or no normal muscular activity. In short, the stomach is functionally shut down and may not be able to empty. If one were to eat while one's stomach was in this state, the food would remain in the stomach longer than usual, possibly leading to a feeling of fullness, epigastric discomfort, and more severe nausea.

Vasopressin, an antidiuretic hormone, is released by the posterior pituitary and increases in the blood of individuals who report nausea after injection of apomorphine, after cancer chemotherapy agents, and after the stimulation of sitting in a rotating chair. It is not clear in these situations whether plasma vasopressin increases immediately before or immediately after the experience of nausea, although several authors have stated that nausea causes an increase in vasopressin release. However, in those few experiments in which several vasopressin measurements were made over time rather than just one prestimulus and one poststimulus measurement, reports of nausea did not usually precede increases in vasopressin levels but rather the two phenomena covaried. In a more recent study by Xu *et al.*, gastric dysrhythmias preceded the onset of nausea and vasopressin release. Vasopressin levels in the blood of symptomatic subjects increased along with reports of nausea and decreased as nausea subsided. In contrast, levels of stress hormones, such as epinephrine, increased with reports of nausea, but did not decrease until long after nausea subsided. Asymptomatic subjects developed neither gastric dysrhythmias nor increased vasopressin release during exposure to drum rotation.

To summarize, the physiological measures that correlate highly with the development of nausea are increased sympathetic nervous system activity, decreased parasympathetic nervous system activity, increased gastric tachyarrhythmia, and increased plasma vasopressin levels.

See Also the Following Articles

Emesis • Parasympathetic Innervation • Rumination Syndrome • Sympathetic Innervation • Taste and Smell

Further Reading

Blum, R. H., Heinrichs, W. L., and Herxheimer, A. (2000). "Nausea and Vomiting." Whurr, London.

Davis, C. J., Harding, R. R., Leslie, R. A., and Andrews, P. L. R. (1986). The organization of vomiting as a protective reflex. *In* "Nausea and Vomiting: Mechanisms and Treatment" (C. J. Davis, G. V. Lake-Bakaar, and D. G. Grahame-Smith, eds.), pp. 65–77. Springer-Verlag, Berlin.

Hu, S., Grant, W., Stern, R. M., and Koch, K. L. (1991). Motion sickness severity and physiological correlates of repeated exposure to a rotating optokinetic drum. *Aviation Space Environ. Med.* **62**, 308–314.

Koch, K. L. (2001). Nausea: An approach to a symptom. *Clin. Perspect. Gastroenterol.* Sept./Oct., 285–297.

Koch, K. L., Stern, R. M., Vasey, M. W., Seaton, J. F., Demers, L. M., and Harrison, T. S. (1990). Neuroendocrine and gastric myoelectric responses to illusory self-motion in men. *Am. J. Physiol.* **258**, E304–E310.

Kucharczyk, J., Stewart, D. J., and Miller, A. D. (eds.) (1991). "Nausea and Vomiting: Recent Research and Clinical Advances." CRC Press, Boca Raton, FL.

Morrow, G. R. (1992). A patient report measure for the quantification of chemotherapy-induced nausea and emesis: Psychometric properties of the Morrow Assessment of Nausea and Emesis (MANE). *Br. J. Cancer* **66**, S72–S74.

Muth, E. R., Stern, R. M., Thayer, J. F., and Koch, K. L. (1996). Assessment of the multiple dimensions of nausea: The nausea profile (NP). *J. Psychosomatic Res.* **40**, 511–520.

Sleisenger, M. H. (ed.) (1993). "Handbook of Nausea and Vomiting." Parthenon, New York.

Xu, L. H., Koch, K. I., Summy-Long, J., Stern, R. M., Seaton, J. F., Harrison, T. S., Demers, L. M., and Bingaman, S. (1993). Hypothalamic and gastric myoelectric responses during vection-induced nausea in healthy Chinese subjects. *Am. J. Physiol.* **265**, E578–E584.

Necrotizing Enterocolitis

ESTHER JACOBOWITZ ISRAEL* AND CLAUDIO MORERA†

Massachusetts General Hospital and Harvard University and †Hospital General del Oeste, Venezuela

preterm newborn Human born before 37 weeks of gestation.
term newborn Human born with or after 37 weeks of gestation.

Necrotizing enterocolitis is a disease of focal or diffuse ulceration and necrosis of the gastrointestinal tract, primarily the distal small bowel and colon, occurring in the newborn period.

EPIDEMIOLOGY

Necrotizing enterocolitis (NEC) is an almost exclusively newborn disease. The incidence is 1 to 3 in 1000 live births and the prevalence is 1 to 5% of all infants in neonatal intensive care unit, especially premature infants. There is no seasonal, geographic, or gender disparity. The age at onset of NEC is inversely related to birth weight and gestational age. Immature infants, particularly low-birth-weight (LBW) babies, are at prolonged risk as they can develop NEC as late as 10 weeks of age. Infants weighing less than 1000 g are particularly vulnerable. Advances in neonatal intensive care have improved survival of premature infants beyond the first days of life. This situation has been accompanied by a further increase in NEC-associated infant mortality, from 11.5 deaths per 100,000 live births in the presurfactant era to 12.3 in the postsurfactant era. Nevertheless, 10 to 35% of affected neonates are full-term infants with different risk factors than those found in premature infants (Table I).

PATHOLOGIC FINDINGS

The distal ileum and ascending colon are the areas mainly affected, although lesions have been noted throughout both organs. The earliest pathologic lesions are superficial mucosal ulceration and submucosal edema and hemorrhage with little infiltration of acute inflammatory cells, evidence against a primary infectious process. The mucosal process progresses to coagulative transmural necrosis, leading to perforation and, in some cases, intramural gas in the submucosa and subserosa. If there is no progression to complete transmural disease and perforation, healing occurs by epithelialization and fibroblast proliferation, and granulation tissue is laid down, sometimes leading to stenosis.

Changes in the neural elements within the intestinal wall have also been described in NEC. The pathological features resemble those of acquired hypoganglionosis and may result in dysfunctional intestinal motility.

PATHOGENIC MECHANISMS

NEC has a multifactorial pathogenesis with no evidence that there is one universal cause, although host prematurity is clearly the most important predisposing factor. Enteral feeding, ischemia, and infectious agents may contribute to the process (Fig. 1).

Enteral Feeding

Milk feeding is a nearly universal observation among patients with NEC. Oral alimentation may provide a substrate for bacteria, and during digestion and absorption, increased oxygen demand increases the risk of intestinal tissue hypoxia. Hyperosmolar feedings can directly damage the intestinal mucosa and may alter intestinal perfusion. Maldigestion of nutrients may be a contributory factor, particularly lactose maldigestion in the very premature infant because of a relative deficiency of the lactase enzyme. The lipid component of

TABLE I Risk Factors for Necrotizing Enterocolitis

Premature	Full-term
Lower gestational age	Cyanotic: congenital heart disease
Feeding	Polycythemia
	Exchange transfusions
	Perinatal asphyxia
	Small for gestational age
	Umbilical catheters
	Maternal preeclampsia
	Antenatal cocaine

 Copyright 2004, Elsevier (USA). All rights reserved.

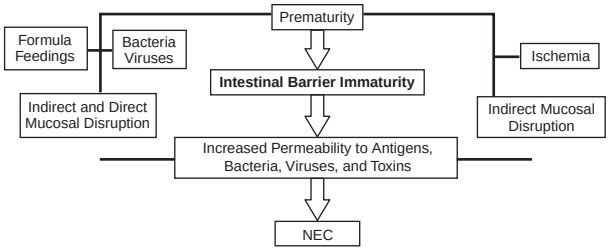


FIGURE 1 A proposed pathogenic framework for necrotizing enterocolitis.

formula may increase mucosal permeability, also predisposing to NEC. On the other hand, minimal enteral intake early after birth may facilitate gastrointestinal maturation and has not been shown to cause an increase in the incidence of NEC.

Artificial formula feeding compared with the feeding of human milk seems to predispose to NEC. Maternal milk remains an optimal source of nutrition, although for certain infants supplementation of feeding with special formulas is necessary. Breast milk contains elements that participate in classic host defense, such as immunoglobulin A (IgA), lymphocytes, and macrophages, which secrete anti-inflammatory substances, such as platelet activating factor-acetyl hydroxylase (PAF-AH). Other nonspecific host defense factors in human milk include complement, lysozyme, lactoferrin, and antioxidant compounds.

Ischemia

The pathologic findings in NEC are consistent with ischemic events, but case-control studies have identified few risk factors that might cause ischemia in the premature infant. However, the fact that NEC most commonly occurs in the distal ileum and proximal colon, which make up the watershed area of the superior and inferior mesenteric arteries, suggests that derangement of the circulatory system is involved. The occurrence of NEC in full-term infants and the predisposing risk factors (see Table I) also indicate that ischemia is an important etiologic factor in these infants.

In animals, the intestinal circulation in neonates has a decreased vascular resistance, an increased blood flow, and increased oxygen delivery when compared with older animals.

The vascular endothelium can help regulate vascular tone by producing vasoactive agents such as nitric oxide (NO) and PAF. PAF antagonists attenuate hypoxic ischemia and increased levels of PAF have been demonstrated in human preterm infants who have been fed. Other inflammatory mediators that may be involved in ischemia in NEC are tumor necrosis factor α, endothelin-1, and cachectin, a potent vasoconstrictor peptide. Lumenal nutrients, especially lipids, seem to intensify tissue injury from ischemia. Cocaine compromises the uterine blood flow and fetal oxygenation. In LBW or premature infants, the risk is two to three times higher in babies who have *in utero* exposure to cocaine than in controls.

Bacteria

The histopathology of the disease suggests that bacterial proliferation and invasion result from the already compromised intestinal mucosal barrier. The role of milk leukocytes in the protection against experimental NEC as well as the efficacy of feedings of an IgA–IgG supplement to premature infants in preventing NEC supports the implication of bacteria in the pathogenesis of NEC. No specific pathogen has been implicated but colonization with potential pathogens is thought to be a prerequisite to develop NEC.

Host Factors

Prematurity seems to be the most important factor in all of the epidemiologic studies of NEC. Up to 90% of the infants with NEC are of LBW and the disease is more frequent and more severe in those infants with the earliest postconceptual age.

Intestinal host defense components mature through the prenatal and perinatal period. Basal acid output, protease levels, IgA concentration in saliva, stool, and serum, and the number of intestinal T lymphocytes, and levels of other components, are low at birth. Intestinal motility is slower in the prenatal and perinatal period, potentially contributing to the overgrowth of bacteria.

The regulation of the immune response mediators also develops postnatally. In the human newborn, PAF-AH activity is decreased and PAF synthesis pathways are increased. This imbalance places the newborn at special risk of an increased PAF response before adequate immune stimuli are developed. The expression of defensins and epidermal growth factor receptor ("lumenal surveillance" peptides available to stimulate repair with a key role in intestinal maturation) is decreased in the preterm infant, as well.

CLINICAL FEATURES

The clinical course of NEC may vary from a slow, indolent process to a fulminant one. The classic triad of symptoms includes abdominal distension, bilious vomiting, and blood in the stools. Few patients have such a classic sequence of symptoms, but instead show less

specific signs indicative of generalized sepsis. Signs vary in the time of their appearance according to the severity of the disease process. The diagnosis is confirmed radiologically on the basis of pneumatosis intestinalis or air in the porta hepatis. Intestinal distension with multiple dilated loops of small bowel is most commonly seen, with separation of the loops suggesting mural edema. Pneumoperitoneum and intra-abdominal fluid are ominous signs, indicating the need for immediate surgical intervention. Levels of IL-6 in umbilical cord blood and reducing substances, α1-anti-trypsin, endotoxin, and short-chain fatty acids in the stool may be early signs of the disease. Portal vein ultrasonography is useful especially when the abdominal radiograph gives normal or nonspecific results. Although these tests may be helpful, they have not replaced the clinical presentation and plain abdominal radiograph in the diagnostic evaluation. A set of diagnostic criteria have been designed as a helpful tool in effecting some uniformity in therapeutic decision-making (Table II). Signs of clinical decompensation accompany the advancing course of NEC. Bowel necrosis generally accompanies this clinical deterioration.

TREATMENT

The management of the infant with NEC is largely supportive. As soon as the diagnosis is suspected, certain measures are undertaken to prevent the progression of the disease. All oral feedings are withheld and a nasogastric tube is placed either to gravity or very gentle suction. Intravenous access is ensured to provide fluids, electrolytes, and nutrition while enteral feedings are withheld. The duration of restriction of oral intake depends on the confirmation of the diagnosis of NEC, which is dictated by the radiographs and the clinical status. The reinstitution of feedings generally is done in a slow and cautious manner, using an elemental formulation to allow for optimal absorption of all nutrients and to avoid further potential injury to the intestinal mucosa.

Cultures of blood, stool, urine, and cerebrospinal fluid are obtained and broad-spectrum parenteral antibiotics are given. It is important to know the prevailing flora of the infants in the neonatal intensive care unit (NICU) and the current NEC bacteriology as NEC cases cluster, as does the predominant organism. Fluid and electrolyte monitoring is essential because large third-space losses may occur with edema and severe metabolic acidosis. Perfusion and blood pressure must be maintained. An abdominal radiograph is obtained to look for the pathognomonic signs of NEC and is repeated periodically to assess the progression of the disease as well as to determine the existence of pneumoperitoneum, indicating perforation.

Different experimental therapies have been studied, mostly in animal models, to prevent the onset of the disease or to diminish its severity. L-Arginine, to increase the NO production, vitamin E, to function as an antioxidant agent, and recombinant PAF-AH have been proposed as preventive or therapeutic agents. None of these has gained universal acceptance.

The indications for surgery include pneumoperitoneum and cellulitis of the anterior abdominal wall, suggesting peritonitis and gangrenous bowel. The classic operative approach consists of resection

TABLE II Staging Criteria for the Management of Necrotizing Enterocolitis

Stage I (Suspect)
- a. Any one or more historical factors producing perinatal stress
- b. Systemic manifestations: temperature instability, lethargy, apnea, bradycardia
- c. Gastrointestinal manifestations: poor feeding, increasing pregavage residuals, emesis (may be bilious or test positive for occult blood)
- d. Abdominal radiographs show distension with mild ileus

State II (Definite)
- a. Any one or more historical factors
- b. Above-mentioned signs and symptoms plus persistent occult or gross gastrointestinal bleeding; marked abdominal distension
- c. Abdominal radiographs show significant intestinal distension with ileus, small bowel separation (edema in bowel wall or peritoneal fluid), unchanging or persistent "rigid" bowel loops, pneumatosis intestinalis, portal vein gas

Stage III (Advanced)
- a. Any one or more historical factors
- b. Above-mentioned signs and symptoms plus deterioration of vital signs, evidence of septic shock or marked gastrointestinal hemorrhage
- c. Abdominal radiographs may show pneumoperitoneum in addition to others listed under IIC

Reprinted from Bell, M. I., *et al.* (1987). Neonatal necrotizing enterocolitis: Therapeutic decisions based upon clinical staging. *Ann. Surg.* **187**, 1, with permission.

of the grossly gangrenous bowel and proximal enterostomy. A conservative approach to surgery when a significant amount of the small bowel is affected involves a proximal enterostomy without resection and a return to the operating room usually within 1 week to assess the damage to the intestine. This approach diminishes the incidence of short gut syndrome. In clinically deteriorated patients and in extremely premature infants, short-duration procedures have been explored. In premature infants with intestinal perforation, drainage alone may allow acute improvement with subsequent laparotomy for definitive treatment.

PROGNOSIS

The decrease in the mortality rate over the years (from 78 to 20%) is most likely due to earlier diagnosis combined with intensive medical and surgical therapy. Mortality is directly related to the presence of bacteremia and low birth weight as well as lower gestational age. Morbidity related to long-term complications is seen in 10 to 30% of affected children.

Strictures of the intestines, particularly in the colon, are the most common sequelae of NEC, occurring in 10 to 36% of survivors. Some of these spontaneously resolve, whereas others require more aggressive treatment, such as surgery.

The short gut syndrome is seen in a small number of the surgically treated infants and survival after extensive resection can occur. If more than 70% of the intestine is removed, serious nutritional-consequences prevail. Preservation of the terminal ileum and of the ileocecal valve is an important survival factor.

Overall, the prognosis for infants surviving the acute stage of NEC is good. Patients discharged from their primary hospitalization have a greater than 80–95% chance of long-term survival with a good quality of life in 75% of patients.

PREVENTION

Since there is no specific treatment for the disease, research has been directed at prevention of the disease. NEC might be ameliorated if the factors important in its pathogenesis, such as bowel ischemia, bacterial proliferation, and the effects of oral feeding or gut immaturity, were eliminated or decreased.

Feeding

A very slow regimen for increases in feeding with maternal milk is optimal. Alternatively, a formula containing low-osmolar protein hydrolysate, easily digested glucose polymers, and MCT oil seems to be most prudent for initiating artificial-formula feedings in sick, LBW infants. Omega-3 fatty acids and egg phospholipid-containing diets have been used. These lipids seem to have an anti-inflammatory effect and a significant decrease in the incidence of stage III NEC in premature infants born prior to 32 weeks of gestation.

Bacterial Proliferation

The use of prophylactic antibiotics in treating bacterial proliferation has been shown to be ineffective in preventing NEC. Supplementing formula with an IgA–IgG preparation in LBW babies (in whom mucosal sIgA is suboptimal) was shown in some studies to prevent the development of NEC although further studies have failed to prove the advantage of oral IgG as a preventive measure.

Host Maturation

Since immaturity is the most significant predisposing factor, strategies directed toward making the immature intestinal barrier more potent have been investigated. The effectiveness of host defense enhancement has been demonstrated for breast milk macrophages in the rodent model of NEC. Corticosteroids, other hormones, and growth factors stimulate the maturation of the intestinal barrier. Furthermore, in the suckling rat, these agents accentuate the sequence of glycoconjugate development on the intestinal surface necessary for bacterial attachment and colonization to occur, stimulate the metabolism, and inhibit the synthesis of PAF. Corticosteroids have also been shown to be effective in preventing the occurrence of NEC in some studies in humans.

See Also the Following Articles

Colitis, Radiation, Chemical, and Drug-Induced • Colonic Ischemia • Growth Factors • Pneumatosis, Benign and Serious

Further Reading

Ford, H., *et al*. (1997). The role of inflammatory cytokines and nitric oxide in the pathogenesis of necrotizing enterocolitis. *J. Pediatr. Surg.* **32**, 275.

Furlano, R. I., and Walker, W. A. (1998). Immaturity of gastrointestinal host defense in newborns and gastrointestinal disease states. *Adv. Pediatr.* **45**, 201.

Holman, R. C., *et al*. (1997). The epidemiology of necrotizing enterocolitis infant mortality in the United States. *Am. J. Public Health* **87**, 2026.

Horwitz, J. R., *et al*. (1995). Complications after surgical intervention for necrotizing enterocolitis: A multicenter review. *J. Pediatr. Surg.* **30**, 994–998.

Israel, E., and Walker, W. (1987). Development of intestinal mucosal barrier function to antigens and bacterial toxins. *In* "Recent Advances in Mucosal Immunology" (I. Mestecky and I. McGhee, eds.), p. 673. Plenum Press, New York.
LaGamma, E., and Browne, L. (1994). Feeding practices for infants less than 1500 grams at birth and the pathogenesis of necrotizing enterocolitis. *Clin. Perinatol.* **21**, 271.
Morrison, S., and Jacobson, J. (1994). The radiology of necrotizing enterocolitis. *Pediatr. Clin. North Am.* **21**, 347.
Norwicki, P., and Naukervis, C. (1994). The role of the circulation in the pathogenesis of necrotizing enterocolitis. *Clin. Perinatol.* **21**, 219.
Peter, C. S., *et al.* (1999). Necrotising enterocolitis: Is there a relationship to specific pathogens? *Eur. J. Pediatr.* **158**, 67.
Rayyis, S. F., *et al.* (1999). Randomized trial of "slow" versus "fast" feed advancements on the incidence of necrotizing enterocolitis in very low birth weight infants. *J. Pediatr.* **134**, 293.

Nematodes

BRIAN F. JONES AND MICHAEL CAPPELLO
Yale University School of Medicine

autoinfection Replication of a nematode parasite within the primary host.
larva Immature form of a nematode that requires molting to develop to adult stage.
molting Process of developmental transformation toward the adult worm stage.

In terms of global health and disease, nematodes (roundworms) stand out as a most important class of human parasite. Billions of people worldwide are infected with one or more species of nematode. The adult stage of many parasitic nematodes commonly resides in the intestine, although some are localized to the bloodstream, lymphatics, or somatic tissues. Those nematode parasites that reside in the intestinal tract cause a variety of gastrointestinal symptoms, including pain, vomiting, and diarrhea, in addition to physiologic, neurologic, immunologic, and systemic symptoms.

GASTROINTESTINAL NEMATODES

Ascaris lumbricoides

Ascaris lumbricoides is the largest intestinal nematode that infects humans; adult worms reach up to 30 cm in length. Ascariasis has a worldwide distribution, with more than 1 billion people infected in both rural and urban environments. Infection occurs when eggs shed in the feces of the host are ingested. The eggs embryonate, hatch, and release first-stage larvae in the small intestine. These larvae penetrate the mucosa, enter the venous circulation, and are carried to the lungs, where they access the alveolar space. The larvae then migrate up the trachea and are swallowed, ultimately returning to the small intestine. During this migration the worms undergo a series of developmental molts and eventually become adults. Light to moderate infections with *Ascaris* may be asymptomatic, whereas heavy infections may cause fatal obstruction of the intestine. Additionally, adult worms may aberrantly migrate from the intestinal lumen to the appendix, liver, or pancreas. In these cases, infection with even a single worm may cause severe complications.

Ascaris infection is routinely diagnosed by identifying eggs in the feces under light microscopy. The eggs are ovoid in shape and range from 45 to 70 μm long by 30 to 50 μm wide (Fig. 1). The recommended treatments for ascariasis include albendazole (400 mg orally, single dose), mebendazole (100 mg/day orally for 3 days), or pyrantel pamoate (11 mg/kg/day for 3 days), each of which is highly effective at eradicating the infection. The initiation of therapy may trigger active movement of the adult worms, leading to obstruction or extraintestinal migration. Strategies for prevention of initial infection include sanitary disposal of solid human waste to avoid contamination of food sources with eggs, as well as targeted chemotherapy directed at high-risk populations.

Copyright 2004, Elsevier (USA). All rights reserved.

Israel, E., and Walker, W. (1987). Development of intestinal mucosal barrier function to antigens and bacterial toxins. *In* "Recent Advances in Mucosal Immunology" (I. Mestecky and I. McGhee, eds.), p. 673. Plenum Press, New York.
LaGamma, E., and Browne, L. (1994). Feeding practices for infants less than 1500 grams at birth and the pathogenesis of necrotizing enterocolitis. *Clin. Perinatol.* **21**, 271.
Morrison, S., and Jacobson, J. (1994). The radiology of necrotizing enterocolitis. *Pediatr. Clin. North Am.* **21**, 347.
Norwicki, P., and Naukervis, C. (1994). The role of the circulation in the pathogenesis of necrotizing enterocolitis. *Clin. Perinatol.* **21**, 219.
Peter, C. S., *et al.* (1999). Necrotising enterocolitis: Is there a relationship to specific pathogens? *Eur. J. Pediatr.* **158**, 67.
Rayyis, S. F., *et al.* (1999). Randomized trial of "slow" versus "fast" feed advancements on the incidence of necrotizing enterocolitis in very low birth weight infants. *J. Pediatr.* **134**, 293.

Nematodes

BRIAN F. JONES AND MICHAEL CAPPELLO
Yale University School of Medicine

autoinfection Replication of a nematode parasite within the primary host.
larva Immature form of a nematode that requires molting to develop to adult stage.
molting Process of developmental transformation toward the adult worm stage.

In terms of global health and disease, nematodes (roundworms) stand out as a most important class of human parasite. Billions of people worldwide are infected with one or more species of nematode. The adult stage of many parasitic nematodes commonly resides in the intestine, although some are localized to the bloodstream, lymphatics, or somatic tissues. Those nematode parasites that reside in the intestinal tract cause a variety of gastrointestinal symptoms, including pain, vomiting, and diarrhea, in addition to physiologic, neurologic, immunologic, and systemic symptoms.

GASTROINTESTINAL NEMATODES

Ascaris lumbricoides

Ascaris lumbricoides is the largest intestinal nematode that infects humans; adult worms reach up to 30 cm in length. Ascariasis has a worldwide distribution, with more than 1 billion people infected in both rural and urban environments. Infection occurs when eggs shed in the feces of the host are ingested. The eggs embryonate, hatch, and release first-stage larvae in the small intestine. These larvae penetrate the mucosa, enter the venous circulation, and are carried to the lungs, where they access the alveolar space. The larvae then migrate up the trachea and are swallowed, ultimately returning to the small intestine. During this migration the worms undergo a series of developmental molts and eventually become adults. Light to moderate infections with *Ascaris* may be asymptomatic, whereas heavy infections may cause fatal obstruction of the intestine. Additionally, adult worms may aberrantly migrate from the intestinal lumen to the appendix, liver, or pancreas. In these cases, infection with even a single worm may cause severe complications.

Ascaris infection is routinely diagnosed by identifying eggs in the feces under light microscopy. The eggs are ovoid in shape and range from 45 to 70 μm long by 30 to 50 μm wide (Fig. 1). The recommended treatments for ascariasis include albendazole (400 mg orally, single dose), mebendazole (100 mg/day orally for 3 days), or pyrantel pamoate (11 mg/kg/day for 3 days), each of which is highly effective at eradicating the infection. The initiation of therapy may trigger active movement of the adult worms, leading to obstruction or extraintestinal migration. Strategies for prevention of initial infection include sanitary disposal of solid human waste to avoid contamination of food sources with eggs, as well as targeted chemotherapy directed at high-risk populations.

Copyright 2004, Elsevier (USA). All rights reserved.

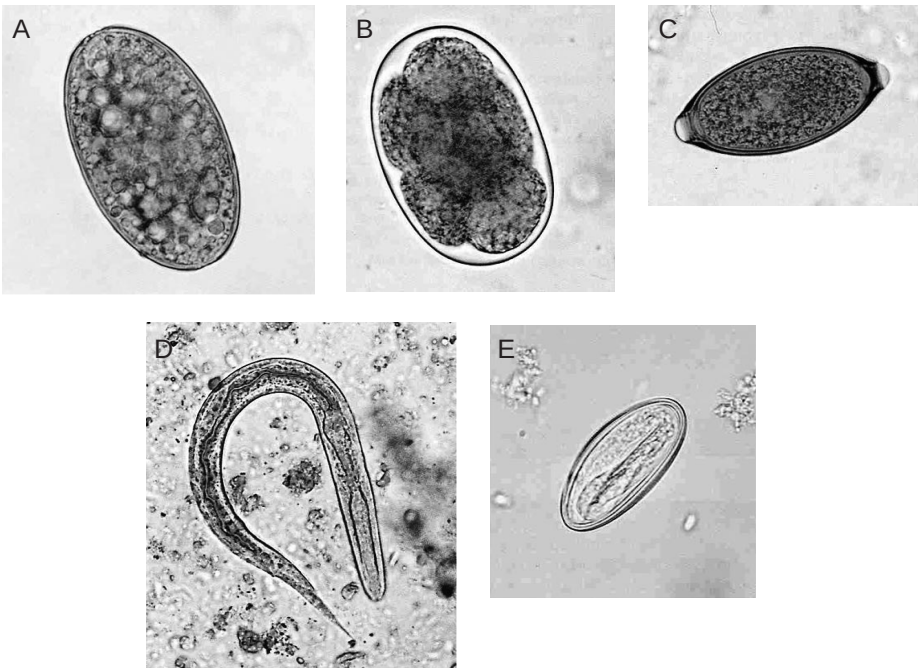


FIGURE 1 Diagnostic stages of major intestinal nematodes. (Top row) *Ascaris lumbricoides* (A), hookworm (B), and *Trichuris trichiura* (C) are diagnosed by fecal egg examination using light microscopy. (Bottom row) *Strongyloides stercoralis* (D) releases rhabditiform larvae, not eggs, in the feces. *Enterobius vermicularis* (E) eggs are deposited on the perianal skin and can be removed with adhesive tape and viewed under light microscopy. Reproduced with permission from Ash and Orihel (1997). Copyright © 1997 by the American Society of Clinical Pathologists. Reprinted with permission.

Hookworms: *Necator americanus* and *Ancylostoma* spp.

Infections with the human hookworms, *Necator americanus* and *Ancylostoma duodenale*, are a leading cause of iron-deficiency anemia in the developing world. Although eventually eradicated due to the efforts of the Rockefeller Sanitary Commission, hookworm was a major public health problem in the southeastern United States during the first half of the twentieth century. Presently, hookworms remain endemic throughout much of Central and South America, Africa, and Asia.

Infection with hookworm occurs when a mammalian host contacts free-living infective (filariform) larvae. These larvae quickly penetrate the host skin and travel via the venous bloodstream to the lungs, where they exit blood vessels and enter the airway. The larvae then migrate up the trachea, are swallowed to gain access to the small intestine and molt to the adult stage. There the adult hookworms attach to the small intestine and lacerate mucosal vessels using cutting plates or teeth in their buccal capsule. Adult hookworms mate and release eggs in the feces of the infected host.

The pathogenesis of hookworm anemia is a direct result of intestinal blood loss that occurs during adult worm attachment to the mammalian small intestine. Each adult hookworm may feed on up to 0.2 ml of blood per day, with some infected individuals harboring hundreds to thousands of parasites. Iron-deficiency anemia and protein malnutrition are the primary clinical illnesses associated with hookworm infection. In children, especially those with concomitant dietary iron deficiency, the chronic effects of hookworm may ultimately lead to growth and cognitive delays. In Australia, human infection with the dog hookworm, *Ancylostoma caninum*, has been implicated as a cause of eosinophilic enteritis, often associated with the presence of a single adult blood-feeding worm.

The diagnosis of hookworm infection is made by identifying eggs in the feces by light microscopy (Fig. 1). Laboratory evaluation may reveal iron-deficiency anemia, hypoalbuminemia, and mild eosinophilia. Recommended treatments for hookworm include albendazole (400 mg orally, single dose), mebendazole (100 mg/day orally for 3 days), or pyrantel pamoate (11 mg/kg/day for 3 days). Prevention of infection can be accomplished by taking preventive measures to limit exposure to the

infective larvae. Proper waste disposal, including avoidance of the use of human feces as fertilizer (night soil), is the primary method employed to remove infective larvae from the environment.

Trichuris trichiura (Whipworm)

Trichuris trichiura, or whipworm, occurs in most parts of the world where other soil-transmitted helminths, e.g., *Ascaris* and hookworm, are found. Infection occurs when ingested eggs hatch in the intestine and develop into adults when they reach the large intestine. Although most individuals with light infection are asymptomatic, heavily infected individuals may experience a severe inflammatory colitis associated with significant hemorrhage (*Trichuris* dysentery). Chronic infection in children has been associated with rectal prolapse, as well as growth delay and cognitive deficits. The diagnosis of *T. trichiura* infection requires identification of the characteristic oval, operculated eggs in the feces (Fig. 1). The recommended treatments for whipworm include albendazole (400 mg orally, single dose) or mebendazole (100 mg/day orally for 3 days), although *Trichuris* is resistant to pyrantel pamoate. Prevention of infection is accomplished by the sanitary disposal of contaminated feces.

Enterobius vermicularis (Pinworm)

Human infection by the parasitic nematode *Enterobius vermicularis*, or pinworm, is the most common nematode infection in developed countries. Pinworm infections are often found in preschool-aged children, particularly those who are not toilet trained. Adult female worms, which may grow to 1 cm in length, reside in the large intestine. At night, the females migrate out of the body and release eggs on the perianal skin. The eggs become infective within several hours, and the life cycle is completed when the eggs are ingested and hatch, to release the larval form of the worm in the host small intestine. The pinworm larvae undergo several molts, eventually developing into adults when they reach the colon. Pinworm eggs are light and may be dispersed in fomites, potentially spreading the infection in institutions and daycare centers because of contamination of environmental surfaces.

The clinical sequelae of pinworm infection are generally mild, and include pruritus ani and superficial cellulitis of the perianal skin. However, aberrant migration of adults has also been recognized to cause urinary tract infection and salpingitis in girls, as well as appendicitis. Local colonic inflammation may also give rise to abscess formation characterized by a profound eosinophilic infiltrate. An association between pinworm and infection with the protozoan parasite *Dientamoeba fragilis* has also been reported. The diagnosis can be made using the "Scotch tape test," in which adhesive tape that has been applied to the perianal skin is examined under a microscope for the presence of eggs. The eggs are ovoid and measure 50–60 μm long by 20–30 μm wide (Fig. 1). The recommended treatment for *E. vermicularis* infection is albendazole (400 mg orally) in a single dose, or mebendazole (100 mg orally) or pyrantel pamoate (11 mg/kg orally), in one dose, with a second dose 14 days later to treat the recently developed adult worms.

Strongyloidiasis

Strongyloidiasis is the disease caused by the nematode parasites *Strongyloides stercoralis* and *Strongyloides fuelleborni*. *Strongyloides stercoralis* is endemic throughout the tropics and subtropics, including parts of the southeastern United States. In contrast, *S. fuelleborni* is found primarily in the South Pacific, particularly in Papua New Guinea, as well as in parts of Africa. Infection occurs when free-living larvae penetrate the host skin and travel to the lungs via the venous circulation. In the lungs, the larvae enter the airways and migrate up the trachea and are swallowed. In the small intestine, the worms develop to the adult stage and embed within the intestinal epithelium. Eggs are shed by adult female worms into the gut and develop into larvae to be released in the feces. The life cycle of *Strongyloides* is unique in two respects. First, unlike nearly all other parasitic nematodes, the worm can complete a free-living life cycle outside the host, with larvae developing in the soil to mature adults. Second, larvae may develop to the infectious filariform stage during transit in the gut, ultimately penetrating the colonic mucosa or perianal skin to complete their life cycle within a single host. This process, termed autoinfection, may have severe consequences in certain hosts, particularly those who are receiving glucocorticoids for the treatment of malignant or autoimmune diseases. In addition, autoinfection explains how certain individuals remain infected with *S. stercoralis* for decades after leaving endemic areas.

Most patients with chronic strongyloidiasis are asymptomatic, although diarrhea and nonspecific abdominal pain may occur. Penetration of the skin by infectious larvae may be associated with a pruritic and erythematous rash, referred to as larva currens (racing larva), which, in cases of autoinfection, may be localized to the perineum and buttocks. Children with *S. fuelleborni* infection may present with a constellation of findings known as "swollen belly syndrome," characterized by severe abdominal distension, malnutrition,

and lethargy. Disseminated strongyloidiasis occurs in immunocompromised hosts, often following treatment with cytotoxic chemotherapy or glucocorticoids. In this setting, increased numbers of larvae undergo the autoinfective cycle, leading to amplification of the infection and overwhelming larval migration. Such patients may develop a hemorrhagic pneumonitis from massive larval transit through the lungs, or a secondary meningitis caused by bacteria carried with larvae that penetrate the central nervous system.

Laboratory test results generally reveal mild to moderate eosinophilia along with increased levels of circulating immunoglobulin E (IgE). Serologic assays are available to detect serum antibodies to *S. stercoralis*, although cross-reactivity with other parasitic nematode infections can occur. The definitive diagnosis of *Strongyloides* infection is accomplished by identification of the first-stage (rhabditiform) larvae in the feces of the host. Because egg production is low, it may be necessary to employ stool-concentrating techniques or to test multiple samples in order to confirm the diagnosis. Disseminated strongyloidiasis should be suspected in patients receiving glucocorticoid therapy who present with bacterial meningitis caused by enteric pathogens. Larvae may also be detected in sputum or lung biopsy specimens. Interestingly, most patients with disseminated infection do not have peripheral eosinophilia. The recommended treatment for chronic strongyloidiasis is ivermectin (200 μg/kg/day for 2 days). An alternative therapy is thiabendazole (50 mg/kg twice daily for 3 days), although this agent is associated with more significant side effects. Treatment of disseminated disease with ivermectin (200 μg/kg/day) should be continued for atleast 1 week. Because the mortality from disseminated strongyloidiasis is high (at least 50%), screening of individuals from endemic areas prior to initiation of immunosuppressive therapy is essential, because even those who are asymptomatically infected should be treated.

Trichinella

Infection with *Trichinella spiralis* occurs throughout most of the world, perhaps due to its lack of an environmental life cycle stage. Infection occurs after ingestion of raw or undercooked meat containing encysted *T. spiralis* larvae. The encysted larvae are released, migrate to the small intestine, penetrate the columnar epithelium, and develop into adult worms. Adult females release motile larvae that enter the host circulatory system and migrate to somatic tissues, including striated muscle. The larvae remain within a specialized cell termed a nurse cell, where they await consumption by a new host.

Acute symptoms associated with ingestion of contaminated meat include nausea, vomiting, diarrhea, and fever. A second phase of illness, characterized by the migration of larvae to somatic tissues, is characterized by edema of the face and eyelids, muscle pain, fever, and conjunctivitis. Both myocarditis and respiratory symptoms may also occur secondary to the host inflammatory response directed at larvae invading the heart and diaphragm, respectively. The diagnosis of *Trichinella* infection requires a high index of suspicion. Eosinophilia is a nearly universal laboratory finding, and levels may reach 50% or greater of the total leukocyte count. A definitive diagnosis is made by identifying encysted larvae in a muscle biopsy. Serologic testing is available through specialty laboratories and the United States Centers for Disease Control and Prevention. A history of raw or undercooked meat consumption, especially pork or horsemeat, is suggestive of the diagnosis.

The role of antihelminthic therapy in the management of trichinellosis is controversial, because the activity of most available agents against tissue-encysted larvae is variable. Both mebendazole (200 mg twice a day for 10 days) and thiabendazole (25 mg/kg twice a day for 10 days) have been reported to improve symptoms of trichinellosis in adults compared to placebo, although thiabendazole was associated with significant side effects. Albendazole (400 mg twice a day) is also recommended. Oral prednisone may be helpful in relieving inflammatory symptoms associated with *Trichinella* infection.

See Also the Following Articles

Cestodes • Helminth Infections • Parasitic Diseases, Overview • Trematodes • *Trichinella*

Further Reading

Ash, L., and Orihel, T. (1997). "Atlas of Human Parasitology," 4th Ed. ACSP Press, Chicago.

Crompton, D. W. T. (2000). The public health importance of hookworm disease. *Parasitology* **121**, S39–S50.

Dupouy-Camet, J., Kociecka, W., Bruschi, F., Bolas-Fernandez, F., and Pozio, E. (2002). Opinion on the diagnosis and treatment of human trichinellosis. *Expert Opin. Pharmacother.* **3**, 1117–1130.

Gilles, H. M. (1996). Soil-transmitted helminths. *In* "Manson's Tropical Diseases" (G. C. Cook, ed.), 20th Ed., pp. 1403–1408. W. B. Saunders, London.

Moorhead, A., Grunenwald, P. E., Dietz, V. J., and Schantz, P. M. (1999). Trichinellosis in the United States, 1991–1996: Declining but not gone. *Am. J. Trop. Med. Hyg.* **60**, 66–69.

Schmidt, G. D., and Roberts, L. S. (1989). "Foundations of Parasitology," 4th Ed. Times Mirror/Mosby College Publishing, Boston.

Watt, G., Saisorn, S., Jongsakul, K., Sakolvaree, Y., and Chaicumpa, W. (2000). Blinded, placebo-controlled trial of antiparasitic drugs for trichinosis myositis. *J. Infect. Dis.* **182**, 371–374.

Neonatal Cholestasis and Biliary Atresia

JUDITH A. O'CONNOR AND RONALD J. SOKOL
University of Colorado Health Sciences Center and Children's Hospital, Denver

biliary atresia Progressive fibro inflammatory process resulting in obliteration of the extra hepatic biliary duct in infants.
cholestasis Condition in which bile excretion is blocked.
***FIC1* gene** Familial intrahepatic cholestasis gene involved in inherited disorders of neonatal cholestasis.
progressive familial intrahepatic cholestasis Inherited disorders of neonatal cholestasis.

Neonatal cholestasis is defined as an impairment of bile flow in the first 3 months of life. It is generally associated with a measured conjugated (direct-acting) bilirubin fraction of greater than 2 mg/dl or more than 20% of the total bilirubin. Prolonged conjugated hyperbilirubinemia in the newborn period is never physiologic and always requires an expeditious evaluation. The incidence of neonatal cholestasis is estimated to be about 1 in 2500 live births. There are multiple causes of cholestasis in the first 3 months of life, of which the majority fall into a few discrete yet overlapping categories. The importance of prompt recognition and definitive diagnosis of neonatal cholestasis cannot be overemphasized. Many causes of cholestasis, such as sepsis, endocrinopathies, galactosemia, fructosemia, tryosinemia, inborn errors of bile acid synthesis, biliary atresia, and choledochal cyst, are either reversible or treatable.

INTRODUCTION

Although the list of disorders causing neonatal cholestasis seems exhaustive (Table I), idiopathic neonatal hepatitis and biliary atresia account for over 60% of all cases. The overlap of these entities has been well recognized, and in 1974, Benjamin Landing, a pediatric pathologist, proposed that biliary atresia, neonatal hepatitis, and choledochal cyst represent a spectrum of disease, resulting from a unifying process. He observed that histologically all three disorders share common findings that include syncytial giant cell transformation of hepatocytes, extramedullary hematopoiesis, lobular disarray, and cellular and canalicular cholestasis with variable degrees of portal and lobular inflammation. He proposed that a viral infection primarily affecting either the hepatocyte or cholangiocyte was responsible for these entities and coined the term "infantile obstructive cholangiopathies."

In general, disorders of extrahepatic obstruction present in an asymptomatic, thriving infant who has clinical jaundice. Often the infant is seen in the clinic because neighbors, family, or friends comment on the infant's yellow skin color. Further investigation will reveal an otherwise healthy infant with light or acholic stools, dark urine, and the presence of hepatomegaly without splenomegaly. Initial laboratory analysis will reveal a moderate direct hyperbilirubinemia, elevated canalicular enzymes (alkaline phosphatase and γ-glutamyl transferase), and normal to mild elevation of amino-transferases. Early in the disease process, growth velocity is normal and hepatic synthetic function, clotting factors, and albumin are preserved.

Infants with cholestasis as a result of an infectious origin usually have a male predominance and failure to thrive or growth failure. An infectious risk may or may not be identified. These infants usually have a marked elevation of serum aminotransferases, variable elevation of canalicular enzymes, and a modest coagulopathy that is commonly a result of mild disseminated intravascular coagulation (DIC). Skin lesions or intracranial calcifications may be pathognomonic.

Infants suffering from a metabolic defect often have mild growth failure and present with a sepsislike picture manifested as hepatic synthetic dysfunction. Metabolic disease may present with ascites and a coagulopathy out of proportion to elevated aminotransferase or canalicular enzymes. These infants have the most varied presentation in terms of age at onset, associated symptoms, and rate of progression to hepatic failure. The clinical course depends on the underlying metabolic disorder.

EXTRAHEPATIC NEONATAL CHOLESTASIS

Biliary atresia is the most common cause of obstructive cholestasis in infants younger than 3 months of age. This disorder occurs in 1 in 8000–15,000 live births and is

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Classification of Cholestatic Disorders[a]

Classification	Example
I. Bile duct obstruction	
Extrahepatic disorders	Biliary atresia, choledochal cysts, spontaneous perforation of common bile duct, Caroli's disease, bile duct stenosis, cholelithiasis, tumors/masses (intrinsic and extrinsic)
Intrahepatic disorders	Alagille's syndrome, nonsyndromic paucity of intrahepatic ducts, neonatal sclerosing cholangitis, congenital hepatic fibrosis, inspissated bile/mucous plug
II. Neonatal Hepatitis	
Viral	Cytomegalovirus, herpes (simplex, zoster, human type 6), rubella, reovirus type 3, adenovirus, enteroviruses, parvovirus B19, hepatitis B, human immunodeficiency virus, bacterial and parasitic, bacterial sepsis, listeriosis, tuberculosis, toxoplasmosis, malaria
Idiopathic	
III. Cholestatic syndromes	
Progressive familial intrahepatic cholestatis caused by transport defects	Type 1 (Byler's disease, defect in FIC1, a P-type ATPase), type 2 (defect in BSEP, canalicular bile acid pump), type 3 (defect in MDR-3, a canalicular phospholipid transporter)
Hereditary cholestasis with lymphedema (Aagenaes syndrome)	
Cholestasis in North American Indians	
Nielsen syndrome (Greenland Eskimos)	
Benign recurrent cholestasis (defect in same gene as PFIC type 1)	
IV. Metabolic disorders	
α1-Antitrypsin deficiency	
Cystic fibrosis	
Neonatal iron storage disease	
Endocrinopathies	Hypopituitarism (septooptic dysplasia), hypothyroidism
Amino acid disorders	Tyrosinemia, hypermethionemia, mevalonate kinase deficiency
Lipid disorders	Niemann–Pick disease, Gaucher's disease, Wolman's disease, cholesterol ester storage disease
Urea cycle disorders	Arginase deficiency
Carbohydrate disorders	Galactosemia, fructosemia, glycogen storage disease type IV
Mitochondrial disorders (respiratory chain)	
Peroxisomal disorders	Zellweger syndrome, infantile Refsum's disease, other enzymopathies
Bile acid synthetic defects	3β-Hydroxy-Δ^5-C_{27}-steroid dehydrogenase isomerase, Δ^4-3-oxosteroid 5β-reductase, oxysterol 7α-hyroxylase
V. Toxic	
Drugs, parenteral nutrition, aluminum	
VI. Miscellaneous Associations	Shock/hypoperfusion, histiocytosis X, neonatal lupus erythematosus, Indian childhood cirrhosis, autosomal trisomies, graft-versus-host disease, erythrophagocytic lymphohistiocytosis, extracorporeal membrane-oxygenation, venoocclusive disease, Donahue syndrome (leprechaunism), arthrogryposis, cholestatic pigmentary-disease, renal dysfunction syndrome, erythro-blastosis fetalis

[a] Modified with permission from Suchy F. (2001), Approach to the infant with Cholestasis. *In* "Liver Disease in Children" (Suchy, Sokol and Balistreri, eds.). Copyright Lippincott, Williams & Wilkins, 2001.

slightly more common in females and infants of Asian or African American descent. Formerly called extrahepatic biliary atresia, it is now clear that this disorder results from a progressive inflammatory sclerosing fibroobliterative process involving both the intra- and extrahepatic ducts in the majority of infants. The result is biliary ductular luminal obliteration and the development of biliary cirrhosis. If untreated, biliary atresia is uniformly fatal by 2 years of age. Death results from end-stage liver disease, portal hypertension, and variceal

bleeding. Biliary atresia is the leading indication for liver transplantation in children, accounting for 50% of all pediatric liver transplants.

Biliary atresia is a clinicopathologic entity and as proposed by Landing may represent a spectrum of disorders. There has been no familial clustering, and reports of discordance in monozygotic twins do not support a classic genetic basis for the disease. There appear to be two distinct phenotypes, an embryonic or fetal type and a perinatal type. Although the clinical consequences of the embryonic or fetal type of biliary atresia are similar to the classic or perinatal form, the etiologies may be very different.

Etiology of Biliary Atresia

The embryonic or fetal type accounts for approximately 20% of all cases and is associated with other congenital anomalies (e.g., polysplenia syndrome). Associated congenital anomalies include malrotation of abdominal viscera, interrupted inferior vena cava, midline liver, preduodenal portal vein, polysplenia, situs inversus, and congenital heart anomalies. Early-onset cholestasis and acholic stools characterize this form of biliary atresia, which may be caused by a defect in morphogenesis of the biliary tree. It is proposed that the constellation of anomalies is caused by abnormalities of laterality. A disorder similar to biliary atresia—polysplenia syndrome exists in the *inv* mouse. This transgenic mouse has a recessive deletion of the inversion (*inv*) gene that results in situs inversus and obstructive jaundice. Studies suggest that abnormalities in the human orthologue of the inversion gene might be involved in the fetal form of biliary atresia. However, no mutations in the *INV* human gene have been found in human patients with biliary atresia.

The perinatal form of biliary atresia is far more common and accounts for almost 80% of all cases. It is believed to occur at or following birth with progressive postnatal destruction of a normally formed biliary tree. Clinically, the infant presents with later-onset jaundice following a jaundice-free period. The jaundice-free period may be obscured by breast-feeding jaundice. Although the pathogenesis of this form of biliary atresia is not understood, several factors have been proposed and are undergoing investigation. Most data suggest that there is an infectious, probably viral, insult with a secondary immunologic pathogenesis that is specific to the neonate less than 3 months of age. The most promising viral etiologies currently under investigation include reovirus type 3, and rotavirus group C.

Observations of the similarity of weanling mice with hepatic histologic lesions due to infection with reovirus type 3 and infants with biliary atresia raised interest in reovirus as a candidate pathogen for biliary atresia. Two large series of patients with neonatal cholestasis detected serologic evidence of reovirus infection in 60% of infants with biliary atresia or neonatal cholestasis, versus 12% of controls or patients with other cholestatic disorders. Reovirus antigen and viral particles have been found in the bile duct of an infant rhesus monkey that spontaneously developed a disorder similar to biliary atresia. Reovirus antigen and viral-like particles have also been identified in bile duct remnants resected from infants with biliary atresia. Although other investigators have not shown an association, this may reflect the age of the patient when studied, tissue preparation, and viral isolation techniques. Utilizing a reverse transcriptase and polymerase chain reaction (RT-PCR) assay specific for Reovirus RNA, virus was detected in over 50% of patients with perinatal biliary atresia (11/20), in none of the fetal type of biliary atresia (0/3), and in 80% (7/9) of infants with a choledochal cyst. Reovirus virus was detected in only 8–15% of autopsy controls or infants less than 1 year of age with other liver disease.

Rotavirus C, another virus of the Reoviridae family, has recently attracted interest as a candidate pathogen for biliary atresia. Studies in which rotavirus was administered orally or by intraperitoneal injection to newborn mice resulted in an intrahepatic lesion strikingly similar to human biliary atresia. Further studies have shown that this model mimics human biliary atresia, resulting in extrahepatic biliary atresia and intrahepatic cholangitis with bile duct proliferation. Studies using interferon α (IFNα), an antiviral that disrupts viral replication, prior to rotavirus infection aborted this disease progression. There was no protective effect if the antiviral was given after infection. Investigation of patients with biliary atresia for evidence of rotavirus has been inconclusive and probably reflects intrinsic problems with study design and patient selection. The similarities between the infant mouse models of reovirus and rotavirus infection and the association of at least one study demonstrating a 50% frequency of each virus in patients with biliary atresia warrant further investigation. Additionally, human papillomavirus has recently been proposed as another candidate virus associated with neonatal cholestasis.

The multihit hypothesis of the etiology of perinatal biliary atresia is gaining popularity. This hypothesis proposes that there is a genetic/immune predisposition, which requires a viral infection at the appropriate time, resulting in the exposure of neoantigens on biliary epithelium, simulating an inflammatory process. The importance of biliary atresia as a significant disorder of

infancy that is the primary indication for pediatric liver transplantation has resulted in the National Institutes of Health (NIH)-supported Biliary Atresia Clinical Research Consortium. This consortium of nine large pediatric liver centers will embark on an effort to define the etiology and to improve current treatment outcomes.

Despite these recent advances, the etiology of biliary atresia remains elusive. Prognosis is dependent on an accurate and timely diagnosis and subsequent surgical intervention to restore bile flow.

Diagnosis of Biliary Atresia

In the diagnosis of biliary atresia, patient history should be compatible with an obstructive presentation. Infants are usually healthy appearing, thriving, with clinical jaundice, hepatomegaly, and rarely splenomegaly. Stools are consistently acholic. Many first-time parents may not recognize the stools as unusual, thus direct examination of the stool is often beneficial. There is no single laboratory or radiology test to diagnosis biliary atresia definitively. Abdominal ultrasound often will note the absence of the gallbladder, but the presence of a gallbladder on ultrasound does not rule out biliary atresia. Importantly, obstructive cholestasis in the neonatal liver results in ductular neoproliferation and not ductular dilation as in the adult. Radionucleotide imaging has been used to determine biliary obstruction. Early in the disease, patients with biliary atresia will have good uptake of radiolabel from the circulating pool, but no excretion into the bowel after 24 hours. Many other disorders, however, will mimic this pattern, and this test will not eliminate the need for histologic analysis. Novel approaches to define the biliary tree include endoscopic retrograde cholangiopancreatography (ERCP) and magnetic resonance cholangiogram (MRC). These tests currently remain investigational. Evaluation of hepatic tissue reveals periportal inflammation, bile ductular proliferation, bile plugs, and occasional giant cell transformation, with relative preservation of hepatic lobular architecture. Intraoperative cholangiogram and laparotomy demonstrating nonpatency of the biliary tree confirm the diagnosis.

Prognosis and Management of Biliary Atresia

The treatment of choice is a hepatoportoenterostomy with a Roux-en-Y enteroanastomosis (Kasai procedure). This procedure involves the excision of the obliterated extrahepatic ducts and apposition of the resected surface of the porta hepatis to the bowel mucosa. Through the years, many variations of the Kasai procedure have been proposed. Despite these modifications, the outcome remains dependent on the experience of the center where the surgical correction is performed and the age of the infant at the time of the procedure. Of infants who are less than 2 months of age at the time of surgical intervention, 80% will reestablish bile flow. Successful establishment of bile flow decreases to less than 20% if the infant is older than 90 days at the time of surgery. Complications following the Kasai procedure include recurrent cholangitis, pruritus, growth failure, hypersplenism, and portal hypertension with variceal bleeding. Medical treatment options are summarized in Table II. Despite successful establishment of bile flow, up to 80% of children with biliary atresia will ultimately require liver transplantation to obtain long-term survival.

INTRAHEPATIC NEONATAL CHOLESTASIS

Paucity of Intralobular Biliary Ducts

Syndromic Paucity of Interlobular Bile Ducts—Alagille's Syndrome

Alagille's syndrome (AGS) was first described in 1969 in a group of children with idiopathic bile duct paucity and a constellation of anomalies. This disorder has been termed arteriohepatic dysplasia, intrahepatic biliary hypoplasia, Watson–Alagille syndrome, and syndromic paucity of interlobular bile ducts. It was recognized that many family members would display similar characteristics and that the disorder was inherited as an autosomal dominant with variable expressivity. Associated findings include cardiac, musculoskeletal, ocular, facial, renal, pancreatic, and neurodevelopmental abnormalities. The prevalence has been reported to be about 1 per 100,000 live births, although this probably underestimates the true frequency, because some affected patients have subclinical expression.

Etiology of AGS In 1997, mutations in the *JAG1* (*Jagged 1*) gene were shown to be associated with AGS. *JAG1* is located on chromosome 20p and is a cell surface protein that functions as a ligand of the Notch receptor. This pathway is a part of an evolutionarily conserved signaling pathway for cell fate. The Notch pathway has been primarily studied in the fruit fly, *Drosophila melanogaster*. Notch and *JAG1* derived their names from the phenotypic observation of a notched or jagged wing found in heterozygote mutations in these flies. Homozygote mutations are uniformly fatal. The finding that mutations in *JAG1* cause AGS indicates that Notch signaling is important in the development of the involved organ systems, i.e., the liver, heart, skeleton,

TABLE II Treatment of Complications of Chronic Cholestatic Liver Disease[a]

Indication	Treatment	Dose	Toxicity/comments
Intrahepatic cholestasis	Phenobarbital	3–10 mg/kg/day	Drowsiness, irritability, interference with vitamin D metabolism
	Cholestyramine/colestipol hydrochloride	250–500 mg/kg/day	Constipation, acidosis, binding of drugs, increased steatorrhea
	Ursodeoxycholic acid	15–20 mg/kg/day	Transient increase in pruritus
Pruritus	Phenobarbital or cholestyramine/colestipol (or both)	15–20 mg/kg/day	—
	Antihistamines		
	Diphenhydramine	5–10 mg/kg/day	Drowsiness
	Hydrochloride hydroxyzine	2–5 mg/kg/day	Drowsiness
	Ultraviolet B light	Exposure as needed	Skin burn
	Carbamazepine	20–40 mg/kg/day	Hepatotoxicity, marrow suppression, fluid retention
	Rifampin	10 mg/kg/day	Hepatotoxicity, marrow suppression
	Ursodeoxycholic acid	15–20 mg/kg/day	Transient increase in pruritus
Steatorrhea	Formula containing medium-chain triglycerides (e.g., Pregestimil)	120–150 calories/kg/day for infants	Expensive
	Oil supplement containing medium-chain triglycerides	1–2 ml/kg/day	Diarrhea, aspiration
Malabsorption of fat-soluble vitamins	Vitamin A	10,000–25,000 units/day	Hepatitis, pseudotumor cerebri, bone lesions
	Vitamin D	800–5000 units/day	Hypercalcemia, hypercalciuria
	25-Hydroxycholecalciferol (vitamin D)	3–5 mcg/kg/day	Hypercalcemia, hypercalciuria
	1,25-Dihydroxycholecalciferol (vitamin D)	0.05–0.2 mcg/kg/day	Hypercalcemia, hypercalciuria
	Vitamin E (oral)	25–200 IU/kg/day	Potentiation of vitamin K deficiency
	Vitamin E (oral, TPGS)[b]	15–25 IU/kg/day	Potentiation of vitamin K deficiency
	Vitamin E (intramuscular)	1–2 mg/kg/day	Muscle calcifications
	Vitamin K (oral)	2.5 mg twice per week to 5 mg/day	—
	Vitamin K (intramuscular)	2–5 mg every 4 weeks	—
Malabsorption of other nutrients	Multiple vitamins	Up to twice the standard dose	
	Calcium	25–100 mg/kg/day	Hypercalcemia, hypercalciuria
	Phosphorus	25–50 mg/kg/day	Gastrointestinal intolerance
	Zinc	1 mg/kg/day	Interference with copper and iron absorption

[a] Modified with permission of The McGraw-Hill companies from Sokol and Narkewicz (2003) in "Current Pediatric Diagnosis" (Hay, Hayward *et al*. eds.), 16th Ed.
[b] D-α-Tocopheryl polyethylene glycol-1000 succinate.

eye, face, and kidney. The timing and specificity of the Notch receptor–ligand interactions are incompletely understood at this time. Over 230 AGS probands have been studied molecularly and *JAG1* mutations have been demonstrated in approximately 70%. Mutations are distributed over the entire coding region of the *JAG1* gene and 50–70% of mutations occur *de novo*.

Diagnosis of AGS Other than the identification of a mutation in *JAG1*, there is no specific test to identify patients with AGS. The diagnosis is based on recognition of associated anomalies and hepatic histology with a paucity of intralobular bile ducts (<0.5 bile ducts per portal tract). Both sexes are equally affected and infants usually demonstrate mild growth failure. Patients with AGS present with mild to severe cholestasis although serum alkaline phosphatase, γ-glutamyl transferase (GGT), and cholesterol are markedly elevated, especially early in life. Serum bile acids are always elevated. Aminotransferases are mildly increased, but clotting factors and other liver proteins are usually normal.

A characteristic facies is often described at an early stage, although it becomes more obvious with age. During childhood, the forehead, ears and nasal bridge are prominent, eyes are deep set and widely spaced, and the chin is small, slightly pointed and projected forward. Cardiac murmurs are present in 95% of patients and cardiac anomalies range from the more common peripheral and valvular pulmonary stenosis to complex lesions such as tetralogy of Fallot. Butterfly vertebrae (incomplete fusion of the vertebral body or anterior arch) occur in up to 50% of affected individuals. Ocular anomalies are varied; however, the most common is posterior embryotoxon, which is identified on ophthalmic evaluation in up to 90% of patients. Posterior embryotoxon is a prominent Schwalbe's ring or line at the point where corneal endothelium and uveal trabecular meshwork join. Renal anomalies occur in about 50% of patients and include solitary, bifid, ectopic, or horseshoe kidney; renal tubular acidosis; or interstitial nephropathy.

AGS must be differentiated from other causes of cholestasis, and most importantly from biliary atresia. Abdominal ultrasound will exclude choledochal cyst. Radionuclide imaging may be abnormal but is not specific; up to 60% of AGS patients will not show excretion of label into the bowel. Liver histology is the most useful preoperative study for the discrimination of AGS from biliary atresia. Unfortunately, in very young infants, bile ducts may be present, confounding the diagnosis. Intraoperative cholangiogram may reveal an anatomically normal but very narrow extrahepatic biliary tree. The diagnosis is made with appropriate hepatic pathology and at least two of the above-mentioned extrahepatic organ system anomalies, or by finding a *JAG1* mutation.

Prognosis and management of AGS Treatment is aimed at medical management of chronic cholestasis (Table II). Growth retardation with normal to increased levels of growth hormone is common. Pruritus is the most severe of any chronic liver disease. Although rarely present before 3–5 months of age, it is observed in most children by the third year of life, even in those who are anicteric. Xanthomas are common in patients with a serum cholesterol >500 mg/dl and cholesterol may reach levels up to 2000 mg/dl. Total serum bile salt elevation up to 100 times the upper limit of normal is the most dramatic laboratory abnormality.

Approximately 30–40% of patients develop complications and progress to cirrhosis, portal hypertension, or synthetic liver failure. Liver transplantation is eventually necessary in 20–50% of patients who have hepatic symptoms in infancy. Indications for transplantation include synthetic failure, complications of portal hypertension, bone fractures, intractable pruritus, disfiguring xanthomata, and growth failure. Posttransplant survival is approximately 90% in patients without significant cardiac or renal anomalies.

Nonsyndromic Paucity of Interlobular Biliary Ducts

Nonsyndromic bile duct paucity is the term used to describe all instances of paucity except those occurring in AGS. This term includes cases either with or without an associated "primary" disease. The overall prognosis for nonsyndromic paucity is poor, with over 70% of cases progressing to cirrhosis. The prognosis is associated with the prognosis of the primary disease, which can include trisomy 21, hypopituitarism, cystic fibrosis, α1-antitrypsin deficiency, cytomegalovirus infection, congenital rubella, chromosomal abnormalities, graft-versus-host disease, primary sclerosing cholangitis (idiopathic), Zellweger syndrome, and some cases of progressive familial intrahepatic cholestasis or inborn errors of bile acid synthesis. Nonsyndromic paucity without an identifiable primary disorder may be a result of a yet unknown metabolic defect.

Progressive Familial Intrahepatic Cholestasis

Progressive familial intrahepatic cholestasis (PFIC) manifests as inherited disorders that present as cholestasis in the neonatal period up to the first year of life. There are several distinct phenotypes. The pattern of affected children within families is suggestive of autosomal recessive inheritance. Recent molecular and genetic studies have allowed the identification of

genes responsible for three types of PFIC. Additionally, recognition of abnormal bile acid patterns has allowed for identification of several disorders of bile synthesis.

PFIC Type 1 (Byler's Disease)

Progressive familial cholestasis type 1 was first described in the Amish, as descendants of Jacob Byler. This disorder consists of at least three clinical entities previously considered to be unrelated: PFIC-1, benign recurrent intrahepatic cholestasis (BRIC), and Greenland familial cholestasis (GFC). Patients usually present with severe pruritus, diarrhea, jaundice, and low or normal GGT and cholesterol levels, and may develop sensorineural hearing loss. In both PFIC-1 and GFC, the disease is progressive and leads to cirrhosis and death from liver failure in the first decade of life.

Etiology of PFIC-1 A locus for PFIC-1 has been mapped to 18q21−q22 in the original Byler pedigree. Genetic studies have revealed locus heterogeneity in the *ATP8B1* (*FIC1*) gene locus. The gene product is a P-type ATPase that may be involved in transport of aminophospholipids. The gene is expressed on the canalicular membrane in the liver, small intestine, pancreas, and other organs. The exact function is not known, but it is postulated that mutations in *FIC1* indirectly disturb biliary secretion of bile acids.

Diagnosis and management of PFIC-1 Diagnosis is based on the exclusion of other cholestatic disorders, especially biliary atresia. There is no specific diagnostic test, but patients will have varying degrees of cholestasis and normal or low serum GGT and cholesterol levels. Serum aminotransferases may be variably elevated and clotting factors are normal. Genetic tests may reveal mutations of *FIC1*. Liver histology reveals fibrosis, canalicular cholestasis, and absence of ductular proliferation. Electron microscopy reveals characteristic coarsely granular bile within the canaliculus.

No specific treatment is available for PFIC-1. Medical management is aimed at the treatment of chronic cholestasis (Table II). Ursodeoxycholic acid and partial external biliary diversion may present alternative options to liver transplantation. Liver tests normalize in 40% of patients treated with ursodeoxycholic acid. Partial external biliary diversion can provide effective relief from pruritus, improve growth, and appears to stabilize or even reverse liver fibrosis.

PFIC Type 2

Progressive familial intrahepatic cholestasis type 2 is clinically indistinguishable from PFIC type 1 except that diarrhea may be less common. Patients are unrelated to the original Byler family and have previously been considered to have "Byler syndrome." Patients present with cholestasis, pruritus, and low or normal GGT and cholesterol levels. Serum bile acids are elevated and biliary bile acids are low.

Etiology of PFIC-2 Genetic studies of patients with phenotypic findings suggestive of PFIC-1 but unrelated to the Byler kindred have revealed mutations in the *ABCB11* locus, which has been mapped to 2q24. The gene product is the ATP-dependent bile salt export pump (BSEP), which is localized to the canalicular membrane. Multiple mutations have been identified. It is proposed that impaired bile secretion causes accumulation of intracellular bile acids and ongoing hepatocellular damage.

Diagnosis, prognosis, and management of PFIC-2 Diagnosis is the same as for PFIC-1 with the exception of liver histology. Liver histology reveals fibrosis, canalicular cholestasis, and absence of biliary ductular proliferation as in PFIC-1, but there is an increase in lobular disarray, fibrosis, inflammation, and presence of giant cell transformation. Electron micrographs reveal amorphous bile in the biliary canaliculus.

Patients with PFIC-2 have a more rapid progression to cirrhosis compared to patients with PFIC-1 and usually suffer end-stage liver disease in the first few years of life. Medical management is similar and, because most patients ultimately undergo liver transplantation, early treatment with ursodeoxycholic acid and/or partial external biliary diversion should be considered.

PFIC Type 3

Progressive familial intrahepatic cholestasis type-3 can be distinguished from the other forms of PFIC by a high serum GGT level. In general, patients with PFIC-3 usually present somewhat later in childhood, with only 30% of cases presenting during the first year of life. Gastrointestinal bleeding due to portal hypertension is often the presenting symptom in older children.

Insight into the etiology of PFIC-3 has resulted from the identification of a similar nonsuppurative cholangitis in mice with a homozygous mutation of the *mdr*2 gene. The mouse *mdr*2 and the human multidrug resistance 3 gene, (*MDR*-3) are P-glycoproteins that function as phospholipid translocators involved in biliary phosphatidylcholine secretion. The gene product is predominantly expressed in the canalicular membrane of hepatocytes. It is proposed that the liver pathology is a result of the toxic effects of bile acids on bile duct epithelia in the absence of biliary phospholipids.

Diagnosis, prognosis, and management of PFIC-3 Patients present with mild and variable pruritus,

moderately elevated serum bile acids, and normal biliary bile acids. Cholesterol is normal but GGT is markedly elevated, distinguishing type 3 from types 1 and 2. Serum lipoprotein X is absent. Liver histology of PFIC-3 also differs from PFIC types 1 and 2. There is a prominence of portal tract inflammatory infiltrate and profound bile ductular proliferation. Care must be given to distinguish PFIC-3 from biliary atresia when the patient is a neonate.

Patients with PFIC-3 ultimately progress to end-stage liver disease, portal hypertension, and variceal hemorrhage. Due to the progressive nature of the disorder, children often require rescue liver transplantation. Partial biliary diversion has not been uniformly helpful. There is a varied response to treatment with ursodeoxycholic acid and up to 45% of patients will have some improvement in liver tests. Responders may represent a partial or missense mutation, and have some residual phospholipid transport.

Idiopathic Neonatal Hepatitis

Neonatal hepatitis is defined as a clinicopathologic disorder in the absence of an identifiable etiology and is among the most common causes of neonatal cholestasis. Previously, neonatal hepatitis accounted for up to 50% of all cases. Over time, the relative percentage of cases has decreased to 25–30% due to recognition of specific disorders mimicking neonatal hepatitis. Two distinct patterns have been identified, familial and sporadic. The sporadic form occurs more frequently and has a better prognosis; over 80% of infants will recover or have residual nonprogressive liver disease. Familial neonatal hepatitis has a poorer prognosis, with up to 80% of patients developing end-stage liver disease. Most investigators believe that the familial form represents variants of PFIC and other undiscovered genetic disorders. Other diseases that originally were classified as idiopathic neonatal hepatitis include identified infections, α1-antitrypsin deficiency, Alagille's syndrome, Niemann–Pick type C, PFIC, bile acid synthesis defects, and other inherited metabolic defects.

Diagnosis, Treatment, and Management of Neonatal Hepatitis

Neonatal hepatitis has a higher prevalence in males and there is often an associated history of intrauterine growth retardation, prematurity, poor feeding, emesis, growth failure, and partially or intermittently acholic stools. The liver may be enlarged or small but is often firm, and splenomegaly may be present. Laboratory studies are varied; however, serum aminotransferases are usually markedly elevated. Alkaline phosphatase and GGT levels are variable. The prothrombin time may be abnormal due to vitamin K deficiency. Abdominal ultrasound may reveal a heterogeneous hyperechoic liver with a normal gallbladder. Radionuclide studies often are nonexcreting and do not reliably distinguish neonatal hepatitis from biliary atresia. Liver histology reveals marked lobular disarray, hepatocellular ballooning, focal hepatic necrosis, multinuclear giant cell transformation, hepatocellular cholestasis, and extramedullary hematopoiesis. Varying degrees of fibrosis may be present. Portal tracts may be expanded with an inflammatory infiltrate but there is an absence of bile ductular proliferation. Intraoperative cholangiogram will reveal an intact extra- and intrahepatic biliary tree, although it may appear hypoplastic.

There is no specific treatment for neonatal hepatitis. As with other neonatal cholestatic disorders, the treatment is aimed at prevention and ultimately treatment of complications of cholestasis (Table II). Prognosis for sporadic neonatal hepatitis is generally good, with most infants having either a complete or partial recovery without progression to end-stage liver disease.

Metabolic or Genetic Disorders Associated with Neonatal Cholestasis

Metabolic disorders present with a wide variety of clinical manifestations depending on the nature of underlying defect. Table III provides a summary of these disorders, their metabolic defect, characteristic hepatic pathology, and definitive diagnostic test.

Disorders of Bile Acid Synthesis

In 1984, infants with suspected neonatal cholestasis were screened for defects in bile acid synthesis using the then-novel technique of fast-atom bombardment ionization mass spectrometry (FAB-MS). The presumption was that infants with idiopathic neonatal cholestasis had an unrecognized infection or genetic defect as the primary cause of their disorder. Using this approach, six defects in bile acid synthesis have been identified to date and account for 2–3% of cholestatic disorders of infancy and childhood. In neonates, the presentation is similar to PFIC. Laboratory tests are variable, but all infants have atypical urinary bile acids identified by FAB-MS.

3β-Hydroxy-C_{27}-steroid dehydrogenase/isomerase deficiency This defect was first identified in a Saudi Arabian patient who was the third of five children to be affected with a progressive familial cholestasis. All of the children were products of a consanguineous marriage and the older two affected siblings had died. All had elevated levels of aminotransferases and a normal

TABLE III Metabolic and Genetic Causes of Neonatal Cholestasis[a]

Disease	Inborn error	Hepatic pathology[b]	Diagnostic studies
Galactosemia	Galactose-1-phosphate uridylyltransferase deficiency	Cholestasis, steatosis, necrosis, pseudoacini, fibrosis	Galactose-1-phosphate uridylyltransferase assay of red blood cells
Fructose intolerance	Fructose-1-phosphate aldolase deficiency	Steatosis, necrosis, pseudoacini, fibrosis	Liver fructose-1-phosphate aldolase assay or leukocyte DNA analysis
Tyrosinemia	Fumarylacetoacetase deficiency	Necrosis, steatosis, pseudoacini, portal fibrosis	Urinary succinylacetone, fumarylacetoacetase assay of red blood cells
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator gene mutations	Cholestasis, neoductular proliferation, excess bile duct mucus, portal fibrosis	Sweat test and leukocyte DNA analysis
Hypopituitarism	Deficient production of pituitary hormones	Cholestasis, giant cells	Thyroxine, thyroid-stimulating hormone, cortisol levels
α1-Antitrypsin deficiency	α1-Antitrypsin gene mutations (PiZZ phenotype)	Giant cells, cholestasis, steatosis, neoductular proliferation, fibrosis, PAS-diastase-resistant cytoplasmic granules	Serum α1-antitrypsin phenotype
Gaucher's disease	β-Glucosidase deficiency	Cholestasis, cytoplasmic inclusions in Kupffer cells (foam cells)	β-Glucosidase assay in leukocytes
Niemann–Pick disease	Lysosomal sphingomyelinase deficiency	Cholestasis, cytoplasmic inclusions in Kupffer cells	Sphingomyelinase assay of leukocytes or liver or fibroblasts (type C); leukocyte DNA analysis
Glycogen storage disease type IV	Branching enzyme deficiency	Fibrosis, cirrhosis, PAS-diastase-resistant cytoplasmic inclusions	Branching enzyme analysis of leukocytes or liver
Neonatal iron storage disease	Unknown	Giant cells, portal fibrosis, hemosiderosis, cirrhosis	Histology, iron stains, buccal mucosa biopsy
Peroxisomal disorders (e.g., Zellweger syndrome)	Deficient peroxisomal enzymes or assembly	Cholestasis, necrosis, fibrosis, cirrhosis, hemosiderosis	Plasma very-long-chain fatty acids, qualitative bile acids, plasmalogen, pipecolic acid, liver electron microscopy
Abnormalities in bile acid metabolism	Several enzyme deficiencies defined	Cholestasis, necrosis, giant cells	Urine, serum, duodenal fluid analyzed for bile acids by fast-atom bombardment mass spectroscopy
PFIC types 1 and 2	*FIC1* and BSEP gene mutations	Cholestasis, necrosis, giant cells, fibrosis, (pericentral)	Histology, family history, normal cholesterol, low or normal γ-glutamyltranspeptidase, DNA analysis
PFIC type 3	MDR-3 gene mutations	Cholestasis, bile duct proliferation, portal fibrosis	Bile phospholipid level, DNA analysis
Alagille's syndrome (syndromic paucity of interlobular bile ducts)	*Jagged* 1 gene mutations	Cholestasis, paucity of interlobular bile ducts, increased copper levels	Three or more clinical features, liver histology, DNA analysis

[a] Modified from Sokol and Narkewicz (2003). "Current Pediatric Diagnosis and Treatment." With permission of The MacGraw-Hill Companies.
[b] PAS, periodic acid–Schiff.

GGT level. Total serum bile acid levels were normal, despite the degree of cholestasis. The defect identified was in the second step in primary bile acid synthesis, disrupting the conversion of 7α-hydroxycholesterol to 7α-hydroxy-4-cholesten-3-one. FAB-MS of urine identified a qualitative absence of normal glyco- and tauro-conjugates of primary bile acids. Treatment with oral cholic acid, 250 mg/day, normalized biochemical parameters and reversed liver disease in all cases.

Δ^4-3-Oxosteroid 5β-reductase deficiency The defect was first identified in monochorionic twin boys whose older brother had died of neonatal hepatitis. Unlike 3β-hydroxy-C_{27}-steroid dehydrogenase/isomerase deficiency, serum GGT was elevated. In this condition, serum bile acids may be elevated, depending on the assay methodology. FAB-MS identified that over 90% of urine bile acids were atypical. If treatment with cholic and ursodeoxycholic acid is initiated prior to the onset of significant liver injury, resolution of jaundice and normalization of biochemical parameters can be achieved. The dose of oral bile acids is based on monitoring urine Δ^4-3-oxo bile acids.

Oxysterol 7α-hydroxylase deficiency Described in only one 10-week-old male infant, this disorder is manifested by progressive cholestasis, cirrhosis, and liver synthetic failure. This infant had hepatosplenomegaly, elevated serum aminotransferases, and a low GGT level. FAB-MS of urine revealed an absence of primary bile acids and an abundance of monohydroxy bile acids. Treatment with ursodeoxycholic acid led to deterioration of liver tests, and cholic acid therapy proved ineffective.

α1-Antitrypsin Deficiency

α1-Antitrypsin (A1AT) deficiency was first described in 1969 as a distinct entity causing neonatal cholestasis. A1AT deficiency accounts for up to 10% of all cases of neonatal cholestasis. There are three major alleles, M, S, and Z. A1AT deficiency is an autosomal recessive disorder associated with an 85–90% reduction of serum A1AT levels. Homozygous ZZ is relatively common, affecting about 1 in 2000 live births. Significant liver disease occurs in up to 20% of affected individuals. Compound heterozygotes (SZ) are at similar risk for liver disease. The gene product is a serine protease inhibitor that is active against proteases released from activated neutrophils. The gene defect reduces the stability of the monomeric form, causing polymerization, misfolding, and retention of the gene product in the endoplasmic reticulum. The mechanism of how the abnormally retained product causes liver disease has yet to be elucidated. Homozygous ZZ adults suffer from emphysema, most likely as a result of decreased circulating levels of the protease inhibitor, and pulmonary damage by neutrophil elastase.

Affected infants presenting with neonatal cholestasis often have mild growth failure. Serum aminotransferase, alkaline phosphatase, and GGT levels are often elevated. About 10% of affected infants will have hepatosplenomegaly, ascites, and liver synthetic dysfunction. This disorder must be distinguished from biliary atresia. There is a report of A1AT deficiency and biliary atresia occurring in a single patient. Liver ultrasound is usually normal. Often patients exhibiting ZZ or SZ phenotypes have no radionuclide biliary excretion. Serum A1AT level will be markedly reduced; serum A1AT level in conjunction with a Pi phenotype should be measured. Liver biopsy will reveal characteristic periodic acid–Schiff (PAS)-positive diastase-resistant globules in the periportal hepatocytes. There is no specific treatment for this disorder. Liver transplantation is curative for end-stage liver disease and prevents the development or worsening of lung disease.

Hereditary Tyrosinemia Type 1

Tyrosinemia was first described in 1932. Almost all infants died of liver disease or hepatocellular carcinoma in infancy or early childhood. In 1977, the defect was identified as a deficiency of fumaryl acetoacetate hydrolase, which is the last enzyme in the tyrosine degradation pathway. The result is an accumulation of succinylacetone, which can be measured in urine. Tyrosinemia is inherited as an autosomal recessive trait. Overall, this is a relatively rare disorder with a frequency of 1 in 100,000 live births. It is most common in French Canadians, particularly the Lac St.-Jean area of northern Quebec, where the incidence is 1 in 1846 live births.

Infants usually present in the first 2 months of life with cirrhosis or decreased hepatic synthetic function manifested as a coagulopathy and ascites for which the cause is not evident. The diagnosis is made by identification of succinylacetone in the urine. Previously, all infants underwent urgent liver transplantation until the discovery of 2(-2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) as a therapeutic agent. NTBC is an herbicide that inhibits the second enzyme in the tyrosine degradation pathway. The initial dose is 1 mg/kg/day, and blood tyrosine levels are monitored. In a recently published study of over 300 patients worldwide, over 95% responded to treatment without apparent side effects. Concern remains regarding the development of hepatocellular carcinoma despite effective treatment. In addition to NTBC, patients

continue a phenylalanine- and tyrosine-restricted diet and are monitored for blood tyrosine and α-fetoprotein levels.

SUMMARY

Cholestatic disorders can be classified into several large categories. These include intra- and extrahepatic biliary obstruction, metabolic disorders, congenital or acquired infections, and infiltrative disease. Despite the wide variety of causes, the clinical presentation of the different types of neonatal cholestasis is similar, reflecting the response of a physiologically immature liver to a variety of insults. There are, however, some unique clinical clues to allow for a directed and cost-effective approach to diagnosis and treatment of neonatal cholestasis. Further research will undoubtedly identify other genetic defects causing neonatal cholestasis.

See Also the Following Articles

Alagille Syndrome • Alpha-1-Antitrypsin (α1AT) Deficiency, Pediatric • Biliary Tract, Developmental Anomalies of the • Bilirubin and Jaundice • Carbohydrate and Lactose Malabsorption • Cholestatic Diseases, Chronic • Cystic Fibrosis • Galactosemia • Glycogen Storage Disease • Malabsorption • Neonatal Hyperbilirubinemia • Tyrosinemia

Further Reading

Dellert, S. F., and Balistreri, W. F. (2000). Neonatal cholestasis. *In* "Pediatric Gastrointestinal Disease" (W. A. Walker, P. R. Durie, J. R. Hamilton, J. A. Walker-Smith, and J. B. Watkins, eds.), 3rd Ed., pp. 890–894. B. C. Decker, Hamilton, Ontario.

Grompe, M. (2001). The pathophysiology and treatment of hereditary tryosinemia type 1. *Semin. Liver Dis.* **21**(4), 563–571.

Jacquemin, E., De Vree, J. M. L., Cresteil, D., Sokal, E. M., Sturm, E., Dumont, M., Scheffer, G. L., Paul, M., Burdelski, M., Bosma, P. J., Bernard, O., Hadchouel, M., and Oude Elferink, R. P. J. (2001). The wide spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. *Gastroenterology* **120**, 1448–1458.

Piccoli, D. A., and Spinner, N. B. (2001). Alagille syndrome and the Jagged1 gene. *Semin. Liver Dis.* **21**(4), 525–534.

Setchell, K. D. R., and O'Connell N. C. (2001). Disorders of bile acid synthesis and metabolism: A metabolic basis for liver disease. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), 2nd Ed., pp. 701–733. Lippincott Williams & Wilkins, Philadelphia.

Sokol, R. J., and Mack, C. (2001). Etiopathogenesis of biliary atresia. *Semin. Liver Dis.* **21**(4), 517–524.

Sokol, R. J., and Narkewicz, M. R. (2003). Liver and pancreas. *In* "Current Pediatric Diagnosis and Treatment" (W. W. Hay, A. R. Hayward, M. J. Levin, and J. M. Sondheimer, eds.), 16th Ed., pp. 647–692. Lange Medical Books/McGraw-Hill.

Suchy, F. J. (2001). Approach to the infant with cholestasis. *In* "Liver Disease in Children" (F. J. Suchy, R. J. Sokol, and W. F. Balistreri, eds.), 2nd Ed., pp. 187–194. Lippincott Williams & Wilkins, Philadelphia.

Thompson, R., and Strautnieks, S. (2001). BSEP: Function and role in progressive familial intrahepatic cholestasis. *Semin. Liver Dis.* **21**(4), 545–550.

van Mil, S. W. C., Klomp, L. W. J., Bull, L. N., and Houwen, R. H. J. (2001). FIC1 disease: A spectrum of intrahepatic cholestatic disorders. *Semin. Liver Dis.* **21**(4), 535–544.

Neonatal Hemochromatosis

KAREN F. MURRAY AND KRIS V. KOWDLEY
University of Washington, Seattle

β2-microglobulin A 12 kDa nonpolymorphic immunoglobulin-like polypeptide that is homologous to the C_3 domain of IgG and is one subunit of the major histocompatibility class I antigens.
hereditary hemochromatosis A common autosomal recessive disorder of excessive intestinal absorption of iron, resulting in iron deposits on the liver and other organs, affecting adults.
oligohydramnios Insufficient amniotic fluid.
polyhydramnios Excessive amniotic fluid.
Sjögren's syndrome An immunological disorder causing mucosal and conjunctival dryness from progressive destruction of exocrine glands and associated with other autoimmune phenomena.

Hereditary hemochromatosis is a common autosomal recessive disorder, resulting in excessive iron deposition in the liver as well as in extrahepatic sites. Symptoms related to hepatic insufficiency, cardiac dysfunction, or endocrinological abnormalities usually present between 40 and 60 years of age. Involvement in adults under the age of 30 years is uncommon and involvement in children has only rarely been reported. Neonatal hemochromatosis (NH), also known as neonatal iron storage disease, is a phenotypically similar disorder; however, its extremely early onset of liver failure makes it notably unique. Originally described in 1957, there have now been over 100 cases reported. Liver failure in the first 30 days of life is uncommon, with neonates representing less than 2% of children listed for liver transplantation, but of these few patients, NH may be one of the most common causes of liver failure. Despite a lack of clear etiology and pathogenesis and the possibility that it is not a primary disease state, the syndrome of NH is now well recognized and universally found to have an aggressive course and carry a poor prognosis. This article discusses the most common clinical presentation as well as less common clinical associations. Genetic inheritance and theories that pertain to abnormal iron metabolism versus NH as the final pathway of multiple possible *in utero* insults are also reviewed. Methods of diagnosis, liver histology, treatment, and prognosis are discussed.

CLINICAL PRESENTATION

When not stillborn, infants with neonatal hemochromatosis (NH) are frequently premature or are small for gestational age. The pregnancy may be complicated by intrauterine growth retardation, oligohydramnios, placental edema, or sometimes polyhydramnios. Illness is usually evident within hours of birth, although there have been individuals diagnosed at a few weeks of age. Patients have features of liver failure with hypoalbuminemia, hypoglycemia, coagulopathy, low fibrinogen, and, frequently, thrombocytopenia and anemia. If not present at birth, ascites develops shortly thereafter, as does hyperbilirubinemia. Transaminase levels are characteristically low.

Clearly the onset of the liver disease is *in utero* with end-stage liver disease already established even in the prematurely born infant. Hepatocellular synthetic insufficiency can explain the coagulopathy, low fibrinogen, and hypoalbuminemia. The hypoalbuminemia in turn contributes to a low intravascular oncotic pressure, edema, contracted blood volume, oliguria, and resultant oligohydramnios. The edema and ascites may be due to anemia, heart failure, or portal hypertension. The characteristically low transaminases and hypoglycemia, which is secondary to poor glycogen stores, are evidence of overall hepatocellular loss. The frequently seen anemia is probably multifactorial, caused by severely limited hepatic erythropoiesis, acquired defects in erythrocyte membranes due to liver disease, and hypotransferrinemia and low erythropoietin levels due to hepatocellular synthetic insufficiency.

OTHER CLINICAL ASSOCIATIONS

The NH phenotype, with respect to clinical presentation, laboratory abnormalities, liver histology, and parenchymal iron deposition, has also been seen in children ultimately diagnosed with Δ^4-3-oxosteroid 5β-reductase deficiency, a defect in bile acid synthesis. Clayton points out, however, that this enzyme is labile and hence its activity may be substantially reduced with hepatocyte damage. Consequently, Δ^4-3-oxosteroid 5β-reductase deficiency may or may not be a primary diagnosis in a subset of NH patients.

A report of an infant with NH born to a mother with Sjögren's syndrome and high levels of anti-Ro/SS-A and

 Copyright 2004, Elsevier (USA). All rights reserved.

anti-La/SS-B antibodies raises the possibility of an autoimmune pathophysiology to this condition.

Most patients with NH do not have other significant congenital abnormalities; however, patients with two separate syndromes have been described with some frequency. Four children, two pairs of siblings, have been reported with trichomalacia, diarrhea, facial dysmorphology, and some degree of congenital heart disease; Verloes *et al*. coined the name "tricho-hepato-enteric syndrome" to describe the condition. Additionally, four patients have been described with NH and renal tubular dysgenesis, also an autosomal recessive condition. These findings support the concept that NH is the result of an *in utero* insult to the fetus.

GENETICS

There have been family analyses to determine whether there is a genetic basis for NH and whether there is a genetic link to hereditary hemochromatosis. Transmission of the disorder has been described as autosomal recessive, codominant, and autosomal dominant with variable penetrance. There are now multiple reports of recurrence of NH in siblings and half-siblings. In most analyses of family occurrences, an autosomal recessive pattern of inheritance has emerged. Another observation, however, cannot be ignored. Several women have had infants with NH fathered by different men, yet no man has fathered children with NH with different women. This suggests possible mitochondrial inheritance, an acquired and persistent maternal factor, or gonadal mosaicism for a dominant mutation lethal in spermatogenesis but not in oogenesis.

A genetic relationship with hereditary hemochromatosis is an attractive idea since the distribution of parenchymal iron and the liver histology in the two conditions are similar. The HFE (hereditary hemochromatosis) gene is mutated in the majority of patients with hereditary hemochromatosis. This gene is located on the short arm of chromosome 6 and there is an association with the human leukocyte antigen (HLA)-A3 alloantigen. The HFE gene mutation has been found in parents of children with NH and parents and affected children may share the HLA-A3 alloantigen. Other studies have disputed this link and association, however. Furthermore, infants with the HFE gene mutation do not necessarily have NH.

IRON METABOLISM AND THE PHYSIOLOGIC BASIS OF DISEASE

The bulk of the body's iron is found as essential iron, specifically heme compounds: enzymes with iron–sulfur complexes. The largest store of essential iron is erythron in red blood cells, the next largest fraction is found in myoglobin in muscle, and, the third largest fraction is present in mitochondrial enzymes. These iron fractions are essential for bodily functioning. Iron in these forms is relatively stable, limited by the life span of the cells in which it is contained.

Tissue iron is stored primarily in ferritin. Ferritin is a polypeptide sphere that encloses iron atoms and it is found in cellular cytoplasm and lysosomal membranes. Glycosylated, iron-free ferritin is secreted into the serum from all ferritin-producing cells and therefore reflects tissue iron stores. Ferritin secretion into serum increases, however, under conditions of inflammation, particularly when the inflammation involves ferritin-rich tissues. In lysosomes, the ferritin molecules aggregate and the degraded polypeptides and iron coalesce into hemosiderin, which can be seen with Prussian blue staining under light microscopy.

Transferrin, the main molecule that mediates the transfer of iron between tissues, is a single polypeptide with two iron-binding sites. Occupation of none (apoferric), one (monoferric), or both (diferric) of the iron-binding sites determines the degree of iron saturation. Transfer of the iron to tissues occurs by the interaction of the transferrin–iron complex with specific membrane receptors and subsequent internalization of the transferrin–iron complex. Fibroblasts from patients with NH synthesize ferritin and transferrin receptors and iron regulates the cellular expression of ferritin and transferrin receptors to the same degree as do fibroblasts from individuals without NH. Furthermore, the intracellular processes by which iron is metabolized are unaffected in NH fibroblasts. It is now known that the transferrin receptor interacts with both HFE protein and β2-microglobulin (β2M). Proper interaction of these molecules allows for regulated internalization and subsequent cycling back to the cell surface of the transferrin–iron–HFE protein–β2M complex, therefore maintaining correct tissue-iron concentrations. In hereditary hemochromatosis, mutation of the HFE gene results in an abnormal HFE protein that does not associate with β2M and after internalization is not transported back to the cell surface. HFE protein is expressed in the crypt cells of the small intestine. It is believed that in individuals with homozygosity for specific HFE gene mutations (C282Y and H63D), the mutant HFE is trapped extracellularly, hence impeding the transport of iron via the transferrin–transferrin receptor interaction. Parkkila *et al*. found that the transferrin receptor and HFE protein–β2M complex are also located in the apical plasma membrane of the syncytotrophoblast cells of the fetal placenta, thus

playing a role in iron transport across the placenta. Placental iron levels, in turn, may regulate the expression of transferrin receptors on the syncytotrophoblast apical plasma membrane and therefore maintain steady iron uptake into the placenta.

Non-transferrin-bound iron (NTBI) ligands may also play a role in hemochromatosis. In hereditary hemochromatosis, a chelator-inaccessible fraction of NTBI has been observed. The identity of this presumably high-affinity ligand has not yet been elucidated. In the case of hemochromatosis, it is possible that this NTBI-ligand fraction is extruded from the cytoplasm of damaged cells. Cytoferrin (host-associated iron transfer factor), an iron-binding compound of lower molecular weight than transferrin, is found mainly in the cellular cytoplasm but also in low concentrations in serum. Cytoferrin concentrations are known to increase with liver disease or in the setting of iron loading. Knisely *et al*. compared the serum concentration of cytoferrin among adult nonselected women, parturient women, normal infants (cord serum), and an infant with NH. The difference between parturient and nonparturient women was not significant. Cord sera from normal neonates, however, had a cytoferrin concentration twice that of the women and sera from the NH infant had a concentration 270 times that of the women. The transplacental gradient of cytoferrin suggests that this compound may be involved in maternofetal transport of iron. The dramatic increase in cytoferrin in the NH neonate may simply be secondary to the liver disease, but its possible role in this disorder warrants further investigation.

DIAGNOSIS

Diagnosis is generally made after other causes of neonatal liver failure are excluded (Table I) and confirmatory tests have been performed. Serum concentrations of ferritin are elevated in patients with NH; thus, an elevated ferritin will help to confirm the diagnosis. However, body-iron stores are normally high in infancy; furthermore, elevated ferritin is nonspecific and may simply represent total body overload, nonspecific liver disease, or inflammation. When measured, the iron-binding capacity (transferrin) is low, reflecting the impaired synthetic ability of the liver. The percentage iron saturation, however, has been reported to markedly exceed the normally high saturation seen in the cord blood of newborns without liver disease (80%), with saturations of 95 to 100% being common.

Siderosis of the liver is normal in the third trimester but can be distinguished from NH hepatic siderosis by the extrahepatic siderosis with reticuloendothelial sparing. Specifically, although the iron deposition in NH is most notable in the liver, it is also present in the heart, pancreas, exocrine and endocrine organs, intestines, and gastric and salivary glands. The siderosis explicitly spares the reticuloendothelial elements, however, as it does in hereditary hemochromatosis. Siderosis of extrahepatic sites including the salivary glands makes it possible to verify the diagnosis histologically without the need for a liver biopsy. Knisely *et al*. showed that siderosis of acinar epithelial cells was found in the minor salivary glands of all patients proven to have NH and in none of 30 patients without liver disease. Salivary gland siderosis was also seen in patients with tyrosinemia, parvovirus B19, rubella, α-thalassemia, and renal–hepatic–pancreatic cystic dysplasia of Ivemark, but these could generally be distinguished on other clinical and laboratory grounds.

Due to the paramagnetic influence of ferric ions (Fe^{3+}) on the image signal during magnetic resonance imaging (MRI), this radiological tool can be useful in the diagnostic evaluation of an infant with possible NH. The signal alterations caused by the ferric ions cause shortening of the T1 and, more impressively, the T2 relaxation times. Any tissue containing iron will therefore have low signal intensity. In NH with T2-weighted images, the signal intensity of the liver and pancreas will be lower than that of the normal-intensity spleen. Furthermore, the more rapid imaging possible with T2 gradient-recalled echo can be used for morphologic and functional assessment of myocardial siderosis. In situations where a fetus is at risk of having NH, MRI can also be used to evaluate the infant during the third trimester of pregnancy. In this case, fetal liver intensity can be compared to other fetal tissues and to the mother's liver.

LIVER HISTOLOGY

Nodular cirrhosis with severe cholestasis is typically found at the time of biopsy or autopsy. It is not unusual to see regenerating nodules separated by wide bands of bland fibrosis containing proliferating bile ducts and multinucleated giant cells. Central vein sclerosis with thickened venular walls and lumens either stenosed or occluded by fibrous tissue is also common. The siderosis is always, by definition, significant. Within tubular hepatocytes, hemosiderin occupies the apical cytoplasm adjacent to the bile canaliculi. In hepatic giant cells, the iron and bile are more diffusely distributed. Extramedullary erythropoiesis, usually obviously apparent in the neonate, is found only in the sinusoids of liver nodules in NH and probably contributes to the anemia in these patients. Although less common, hepatocellular carcinoma has been reported in cases of NH at the time

TABLE I Differential Diagnosis in Neonatal Liver Failure

Disease category	Disease	Diagnostic tests
Metabolic	Galactosemia	Urine reducing substances
		Red blood cell galactose-1-phosphate uridyl transferase
	Tyrosinemia	Plasma quantitative amino acids
		Urine succinylacetone
		α-Fetoprotein
	Neonatal hemochromatosis	Salivary gland or liver biopsy
		Hypersaturation of transferrin with decreased absolute transferrin concentrations
		Appropriate pattern of hepatic and extrahepatic siderosis on MRI
	Δ^4-3-Oxosteroid 5β-reductase deficiency	Urinary bile acid analysis by fast-atom-bombardment mass spectroscopy
	Hereditary fructose intolerance	Liver aldolase activity
		Allele-specific oligonucleotide analysis of peripheral genomic DNA
Storage diseases	α_1-Antitrypsin deficiency	Quantitative serum α1-antitrypsin level
		Protease inhibitor typing
	Zellweger syndrome	Plasma very long plasma fatty acids
	Niemann–Pick disease type C	Skin biopsy for fibroblast culture
		Studies of cholesterol esterification and accumulation
	Glycogen storage disease type IV	Skin biopsy for branching enzyme assay in cultured fibroblasts
		Liver biopsy
Infectious	Parvovirus B19	IgM
		Blood PCR
	Echovirus 9	Culture from pharynx or blood
		PCR
	Coxsackie virus	Culture from urine or blood
		PCR
	Adenovirus	IgM
		Culture from urine or blood
	Cytomegalovirus	Buffy coat antigen
		IgM
	Herpes simplex virus	Culture from nasopharynx or rectal swab
		Buffy coat PCR
	Hepatitis A virus	IgM
	Hepatitis B virus	HBsAg
		PCR
	Hepatitis C virus	PCR
	Rubella	IgM
		Culture from urine or nasopharynx
	Syphilis	VDRL or RPR
	Toxoplasmosis	DS-IgM EIA or ISAGA
	Human herpes virus 6	Blood PCR
Other	Hemophagocytic lymphohistiocytosis	Hemophagocytic histiocytosis in liver and bone marrow
	Mitochondrial cytopathy	Lactic acidosis
		Muscle or liver biopsy for analysis of respiratory chain enzyme function
	Neonatal lupus erythematosus	Maternal anti-Ro/SS-A and anti-La/SS-B antibodies
	Hepatic neuroblastoma	Ultrasound or computed tomography
		Biopsy of mass
	Budd–Chiari syndrome	Doppler ultrasound

Note. DS-IgM EIA, Double-sandwich IgM enzyme immunoassay; HBsAg, Hepatitis B surface antigen; IgM, Immunoglobulin M; ISAGA, IgM Immunosorbent agglutination assay; PCR, Polymerase chain reaction; RPR, Rapid plasma reagin; VRDL, Venereal disease research laboratory.

TABLE II Disorders of the Neonate with High Iron Deposition in the Liver and Other Organs with Reticuloendothelial Organ Sparing

Diagnosis	Hepatic iron content
Neonatal hemochromatosis	+++
Δ^4-3-Oxosteroid 5β-reductase deficiency	+++
Echovirus 9	+++
Cytomegalovirus	+++
Neonatal lupus	+++
Herpes simplex virus	++
Tyrosinemia	++
Mitochondrial cytopathy	++
Zellweger syndrome	+
Rubella	+
Parvovirus B19	+
Congenital hepatic fibrosis	+

of transplantation or autopsy, reinforcing the prenatal onset of this condition.

Whereas the average liver weight is lower in NH patients than in normal individuals (reflecting loss of hepatic parenchyma) even when adjustments are made for gestational age and body size, the mean hepatic iron concentration is higher in NH than in comparable controls. Knisely found that the hepatic iron concentration in patients with NH ranged from 240 to 38,200 μg/g of dry weight, whereas that of a healthy neonate was 250 μg/g of dry weight. It must be noted, however, that despite the clearly elevated liver iron in most cases, the range overlaps with normal perinatal hepatic iron concentrations in some subjects. Furthermore, there are conditions other than NH in which the iron content is elevated in the liver and other nonreticuloendothelial organs (Table II). This elevated iron content in other organs is particularly likely with cirrhosis and may indicate a shunt siderosis similar to that frequently seen in adult liver disease.

TREATMENT AND PROGNOSIS

NH is nearly universally fatal and experience with treatment is limited because many patients die before diagnosis. Deferoxamine therapy is not efficacious and it has been suggested that its use may potentiate bacterial growth. Additionally, despite early preliminary reports to the contrary, deferoxamine combined with an antioxidant cocktail has not been proven to be universally successful.

Although experience is very limited, liver transplantation has been successful in treating NH. In one study of 14 infants with NH who were treated with an antioxidant cocktail, 5 patients survived to transplantation and 3 were alive 1 year after transplantation. In most cases, the iron overload resolves slowly after transplantation. In 1 patient, however, iron accumulated in the allograft as early as 7 days posttransplant and the patient died of cardiac arrhythmia 2 months posttransplant. The iron accumulation in the allograft liver was thought to be secondary to redistribution of the preexisting body iron and the authors speculated that deferoxamine and antioxidant therapy may be beneficial after transplant.

SUMMARY

NH is a dramatic condition that, though rare, may be one of the most common causes of liver failure in the newborn. Its cause is not clear and there is even controversy over whether it is a unique disease state. It is known, however, that NH can recur in families and has shown inheritance patterns and it may or may not be related to hereditary hemochromatosis. Furthermore, although medical treatments have not been successful, liver transplantation provides some hope for the families and individuals affected. Ongoing research into its cause and pathogenesis will eventually uncover some of the mysteries related to this disease state.

Acknowledgment

The article was previously published in *Pediatrics*, Murray, K. (2001). Neonatal hemochromatosis. **108**(4), pp. 960–964. Reproduced by permission of *Pediatrics*.

See Also the Following Articles

Hereditary Hemochromatosis • Iron Absorption • Liver Failure, Pediatric • Liver Transplantation

Further Reading

Bacon, B. B., Powell, L. W., Adams, P. C., Kresina, T. F., and Hoofnagle, J. H. (1999). Molecular medicine and hemochromatosis: At the crossroads. *Gastroenterology* **116**, 193–207.

Bierings, M. B., Baert, M. R., van Eijk, H. G., and van Dijk, J. P. (1991). Transferrin receptor expression and the regulation of placental iron uptake. *Mol. Cell. Biochem.* **100**, 31–38.

Clayton, P. T. (1994). Δ^4-3-Oxosteroid 5β-reductase deficiency and neonatal hemochromatosis. *J. Pediatr.* **125**, 845–846.

Hardy, L., Hansen, J. L., Kushner, J. P., and Knisely, A. S. (1990). Neonatal hemochromatosis: Genetic analysis of transferrin-receptor, H-apoferritin, and L-apoferritin loci, and of the human leukocyte antigen class 1 region. *Am. J. Pathol.* **137**, 149–153.

Kaikov, Y., Wadsworth, L. D., Hassell, E., Dimmick, J. E., and Rogers, P. C. J. (1992). Primary hemochromatosis in children: Report of three newly diagnosed cases and review of the pediatric literature. *Pediatrics* **90**, 37–42.

Knisely, A. S. (1992). Neonatal hemochromatosis. *Adv. Pediatr.* **39**, 383–403.

Lund, D. P., Lillehei, C. W., Kevy, S., *et al.* (1993). Liver transplantation in newborn liver failure: Treatment for neonatal hemochromatosis. *Transplant. Proc.* **25**, 1068–1071.

Merryweather-Clarke, A., Simonsen, H., Shearman, J. D., Pointon, J. J., Nørgaard-Pedersen, B., and Robson, K. J. H. (1999). A retrospective anonymous pilot study in screening newborns for HFE mutations in Scandinavian populations. *Hum. Mutat.* **13**, 154–159.

Oddone, M., Bellini, C., Bonacci, W., Toma, P., and Serra, G. (1999). Diagnosis of neonatal hemochromatosis with MR imaging and duplex Doppler sonography. *Eur. Radiol.* **9**, 1882–1885.

Parkkila, S., Waheed, A., Britton, R. S., *et al.* (1997). Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary hemochromatosis. *Proc. Natl. Acad. Sci. USA* **94**, 13198–13202.

Rand, E. B., McClenathan, D. T., and Whitington, P. F. (1992). Neonatal hemochromatosis: Report of successful orthotopic liver transplantation. *J. Pediatr. Gastroenterol. Nutr.* **15**, 325–329.

Schneider, B. L. (1996). Neonatal liver failure. *Curr. Opin. Pediatr.* **8**, 495–501.

Shneider, B. L., Setchell, K. D. R., Whitington, P. F., Neilson, K. A., and Suchy, F. J. (1994). Δ^4-3-oxosteroid 5β-reductase deficiency causing neonatal liver failure and hemochromatosis. *J. Pediatr.* **124**, 234–238.

Schoenlebe, J., Buyon, J. P., Zitelli, B. J., Friedman, D., Greco, M. A., and Knisely, A. S. (1993). Neonatal hemochromatosis associated with maternal autoantibodies against Ro/SS-A and La/SS-B ribonucleoproteins. *Am. J. Dis. Child.* **147**, 1072–1075.

Siafakas, C. G., Jonas, M. M., and Perez-Atayde, A. R. (1997). Abnormal bile acid metabolism and neonatal hemochromatosis: A subset with poor prognosis. *J. Pediatr. Gastrenterol. Nutr.* **25**, 321–326.

Sigurdsson, L., Reyes, J., Kocoshis, S. A., Hansen, T. W. R., Rosh, J., and Knisely, A. S. (1998). Neonatal hemochromatosis: Outcomes of pharmacologic and surgical therapies. *J. Pediatr. Gastroenterol. Nutr.* **26**, 85–89.

Silver, M. M., Valberg, L. S., Cutz, E., Lines, L. D., and Phillips, M. J. (1993). Hepatic morphology and iron quantitation in perinatal hemochromatosis. *Am. J. Pathol.* **143**, 1312–1325.

Verloes, A., Lombet, J., Lambert, Y., *et al.* (1997). Tricho-hepato-enteric syndrome: Further delineation of a distinct syndrome with neonatal hemochromatosis phenotype, intractable diarrhea, and hair anomalies. *Am. J. Med. Genet.* **68**, 391–395.

Verloes, A., Temple, I. K., Hubert, A.-F., *et al.* (1996). Recurrence of neonatal haemochromatosis in half-sibs born of unaffected mothers. *J. Med. Genet.* **33**, 444–449.

Waheed, A., Parkkila, S., Zhou, X. Y., *et al.* (1997). Hereditary hemochromatosis: Effects of C282Y and H63D mutations on association with β2-microglobulin, intracellular processing, and cell surface expression of the HFE protein in COS-7 cells. *Proc. Natl. Acad. Sci. USA* **94**, 12384–12389.

Neonatal Hyperbilirubinemia

MICHAEL KAPLAN*,† AND CATHY HAMMERMAN*,‡

**Shaare Zedek Medical Center, Jerusalem, †Hebrew University, Jerusalem, and ‡Ben Gurion University of the Negev, Israel*

choreoathetosis Abnormal movements of the body of combined choreic and athetoid pattern.

gene promoter A noncoding DNA sequence of a gene at which RNA polymerase and transcription factors bind and regulate transcription.

heme The oxygen-carrying portion of hemoglobin.

hemolysis Destruction of red blood cells with concomitant release of hemoglobin.

hyperbilirubinemia A condition in which an abnormally large amount of bilirubin circulates in the blood, resulting in serum total bilirubin concentrations greater than the 95th percentile for hour of life.

kernicterus A severe neurologic condition associated with yellow staining and degenerative lesions of the basal ganglia of the brain of severely jaundiced term newborns or moderately hyperbilirubinemic premature neonates.

Jaundice is a common condition of the newborn and refers to a yellow discoloration of the skin and sclerae, the result of bilirubin deposition in these visible tissues. Hyperbilirubinemia occurs when the serum total bilirubin concentration exceeds the 95th percentile for the neonate's age in hours. Kernicterus, a permanent neurological condition associated with severe hyperbilirubinemia, occurs when bilirubin becomes deposited in the basal ganglia of the brain.

INTRODUCTION

Neonatal jaundice is one of the most common occurrences encountered by the neonatologist or pediatrician, with visible yellowing of the skin and/or sclerae,

 Copyright 2004, Elsevier (USA). All rights reserved.

Lund, D. P., Lillehei, C. W., Kevy, S., *et al.* (1993). Liver transplantation in newborn liver failure: Treatment for neonatal hemochromatosis. *Transplant. Proc.* **25**, 1068–1071.

Merryweather-Clarke, A., Simonsen, H., Shearman, J. D., Pointon, J. J., Nørgaard-Pedersen, B., and Robson, K. J. H. (1999). A retrospective anonymous pilot study in screening newborns for HFE mutations in Scandinavian populations. *Hum. Mutat.* **13**, 154–159.

Oddone, M., Bellini, C., Bonacci, W., Toma, P., and Serra, G. (1999). Diagnosis of neonatal hemochromatosis with MR imaging and duplex Doppler sonography. *Eur. Radiol.* **9**, 1882–1885.

Parkkila, S., Waheed, A., Britton, R. S., *et al.* (1997). Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary hemochromatosis. *Proc. Natl. Acad. Sci. USA* **94**, 13198–13202.

Rand, E. B., McClenathan, D. T., and Whitington, P. F. (1992). Neonatal hemochromatosis: Report of successful orthotopic liver transplantation. *J. Pediatr. Gastroenterol. Nutr.* **15**, 325–329.

Schneider, B. L. (1996). Neonatal liver failure. *Curr. Opin. Pediatr.* **8**, 495–501.

Shneider, B. L., Setchell, K. D. R., Whitington, P. F., Neilson, K. A., and Suchy, F. J. (1994). Δ^4-3-oxosteroid 5β-reductase deficiency causing neonatal liver failure and hemochromatosis. *J. Pediatr.* **124**, 234–238.

Schoenlebe, J., Buyon, J. P., Zitelli, B. J., Friedman, D., Greco, M. A., and Knisely, A. S. (1993). Neonatal hemochromatosis associated with maternal autoantibodies against Ro/SS-A and La/SS-B ribonucleoproteins. *Am. J. Dis. Child.* **147**, 1072–1075.

Siafakas, C. G., Jonas, M. M., and Perez-Atayde, A. R. (1997). Abnormal bile acid metabolism and neonatal hemochromatosis: A subset with poor prognosis. *J. Pediatr. Gastrenterol. Nutr.* **25**, 321–326.

Sigurdsson, L., Reyes, J., Kocoshis, S. A., Hansen, T. W. R., Rosh, J., and Knisely, A. S. (1998). Neonatal hemochromatosis: Outcomes of pharmacologic and surgical therapies. *J. Pediatr. Gastroenterol. Nutr.* **26**, 85–89.

Silver, M. M., Valberg, L. S., Cutz, E., Lines, L. D., and Phillips, M. J. (1993). Hepatic morphology and iron quantitation in perinatal hemochromatosis. *Am. J. Pathol.* **143**, 1312–1325.

Verloes, A., Lombet, J., Lambert, Y., *et al.* (1997). Tricho-hepato-enteric syndrome: Further delineation of a distinct syndrome with neonatal hemochromatosis phenotype, intractable diarrhea, and hair anomalies. *Am. J. Med. Genet.* **68**, 391–395.

Verloes, A., Temple, I. K., Hubert, A.-F., *et al.* (1996). Recurrence of neonatal haemochromatosis in half-sibs born of unaffected mothers. *J. Med. Genet.* **33**, 444–449.

Waheed, A., Parkkila, S., Zhou, X. Y., *et al.* (1997). Hereditary hemochromatosis: Effects of C282Y and H63D mutations on association with β2-microglobulin, intracellular processing, and cell surface expression of the HFE protein in COS-7 cells. *Proc. Natl. Acad. Sci. USA* **94**, 12384–12389.

Neonatal Hyperbilirubinemia

MICHAEL KAPLAN*,† AND CATHY HAMMERMAN*,‡

**Shaare Zedek Medical Center, Jerusalem, †Hebrew University, Jerusalem, and ‡Ben Gurion University of the Negev, Israel*

choreoathetosis Abnormal movements of the body of combined choreic and athetoid pattern.

gene promoter A noncoding DNA sequence of a gene at which RNA polymerase and transcription factors bind and regulate transcription.

heme The oxygen-carrying portion of hemoglobin.

hemolysis Destruction of red blood cells with concomitant release of hemoglobin.

hyperbilirubinemia A condition in which an abnormally large amount of bilirubin circulates in the blood, resulting in serum total bilirubin concentrations greater than the 95th percentile for hour of life.

kernicterus A severe neurologic condition associated with yellow staining and degenerative lesions of the basal ganglia of the brain of severely jaundiced term newborns or moderately hyperbilirubinemic premature neonates.

Jaundice is a common condition of the newborn and refers to a yellow discoloration of the skin and sclerae, the result of bilirubin deposition in these visible tissues. Hyperbilirubinemia occurs when the serum total bilirubin concentration exceeds the 95th percentile for the neonate's age in hours. Kernicterus, a permanent neurological condition associated with severe hyperbilirubinemia, occurs when bilirubin becomes deposited in the basal ganglia of the brain.

INTRODUCTION

Neonatal jaundice is one of the most common occurrences encountered by the neonatologist or pediatrician, with visible yellowing of the skin and/or sclerae,

 Copyright 2004, Elsevier (USA). All rights reserved.

indicating a serum total bilirubin (STB) concentration of 5–6 mg/dl, appearing in approximately 60% of otherwise healthy, term newborns during the first days of life. In most cases, the STB concentrations do not exceed the normal range, the upper limit for which is currently defined as the 95th percentile for hour of life. This point corresponds to STB concentrations of 15.0 to 17.0 mg/dl, with a maximum STB level noted on the third to fourth day of life. An individual neonate's peak may be affected by prematurity, being breast-fed, and East Asian parentage. Not every jaundiced newborn is necessarily hyperbilirubinemic; only in a fraction of cases do the STB concentrations exceed the 95th percentile and enter the dangerous range.

Rarely, dangerously high bilirubin levels may be reached and bilirubin encephalopathy, or permanent, irreversible brain damage called kernicterus, may ensue. Prior to the 1990s, kernicterus was seldom encountered, as STB concentrations were for the most part prevented from exceeding 20.0 mg/dl. This concept was derived from the outcome of rhesus factor (Rh) isoimmunized neonates, who rarely developed kernicterus provided STB values were kept below this concentration. However, there is evidence of a recent resurgence of kernicterus not only in North America, where over 125 cases have been recorded in an informal kernicterus registry, but in other countries as well. A more liberal attitude toward neonatal hyperbilirubinemia, the (subsequently disproved) belief that kernicterus will not occur in the absence of hemolysis, and a growing trend to mandatory early discharge prior to STB concentrations peaking, have all contributed to this tragedy. In most cases, with the possible exception of severe, sudden hemolysis due to glucose-6-phosphate dehydrogenase (G-6-PD) deficiency, kernicterus should be a preventable condition provided management guidelines of the American Academy of Pediatrics (AAP) are adhered to. Unfortunately, this was not always the case.

In the majority of cases of neonatal jaundice, the serum bilirubin is of the unconjugated, or indirect-reacting, type. As conjugated, or direct-reacting, bilirubin does not cross the blood–brain barrier and therefore does not cause bilirubin encephalopathy, in the remainder of this discussion only indirect hyperbilirubinemia will be addressed.

BILIRUBIN FORMATION AND ELIMINATION

More than 80% of the bilirubin produced in the human body derives from heme catabolism, liberated from senescent red cells. Heme is degraded by the enzyme heme oxygenase to biliverdin and subsequently to bilirubin (see Fig. 1). For each molecule of bilirubin produced by this mechanism, equimolar quantities of carbon monoxide (CO) are produced. The latter binds to hemoglobin to form carboxyhemoglobin, in which form it is transported to the lungs and exhaled. Bilirubin itself enters the bloodstream and binds to albumin, on which it is transported. As long as this unconjugated bilirubin remains bound to albumin, it will be prevented from entering the basal ganglia. To begin its excretion process, bilirubin enters the hepatocyte, where it is rendered water-soluble by conjugation with glucuronic acid, a process catalyzed by the enzyme UDP-glucuronosyltransferase 1A1 (UGT). The resulting conjugated bilirubin fractions are excreted via the bile into the bowel, where much is converted to nonreabsorbable products and excreted in the stool. However, in the postnatal period, especially in breast-fed neonates, the glucuronide molecules can be cleaved off by the mucosal and bacterial enzyme β-glucuronidase. The resulting unconjugated bilirubin can now be reabsorbed, reentering the bilirubin pool. This mechanism is known as the enterohepatic circulation of bilirubin.

PHYSIOLOGICAL JAUNDICE

Jaundice occurs when the rate of bilirubin production exceeds the body's ability to eliminate it. *In utero*, the placenta is responsible for removal of bilirubin, and, except for instances of severe intrauterine hemolysis, the fetus does not usually become jaundiced. However, within a few days of delivery, STB concentrations may increase rapidly from approximately 1.5 mg/dl in umbilical cord blood to an average of 5–6 mg/dl. Mild imbalance between the bilirubin production and elimination processes should not result in dangerously high STB concentrations. Physiological jaundice peaks at approximately 3 to 4 days and diminishes toward the end of the first week of life.

Conditions resulting in physiological jaundice include those increasing bilirubin production, those diminishing conjugation or excretion, increased enterohepatic circulation, or a combination of all mechanisms. Increased heme production may be the result of high hemoglobin concentrations at the time of delivery, shortened neonatal RBC life span, bruising, or cephalhematoma. Maturation of UGT enzyme activity is a function of progressing gestational age and prematurity is therefore a major cause of diminished bilirubin conjugation. Diminished activity of UGT may also result from hypoxia, hypothermia, and hypothyroidism. The

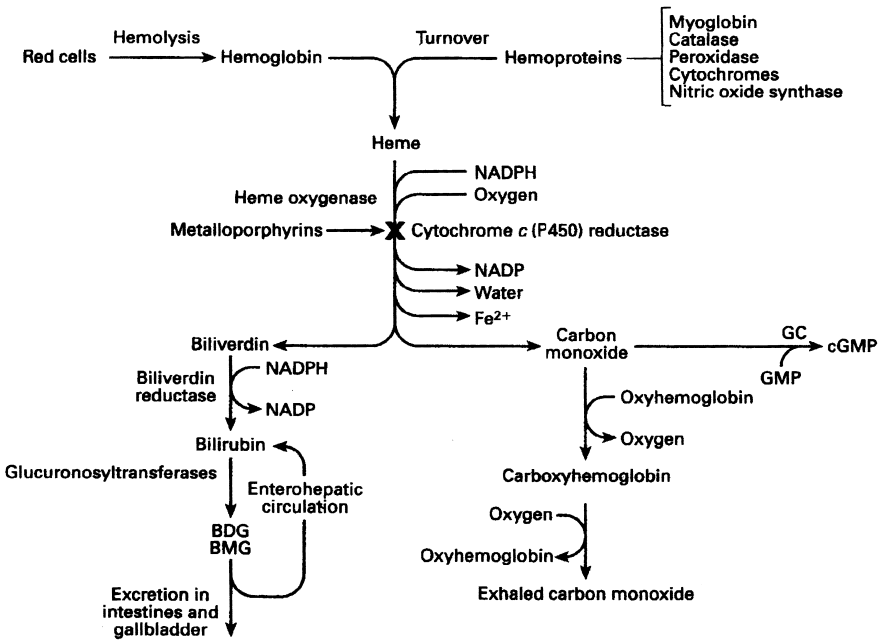


FIGURE 1 Metabolic pathway of the degradation of heme and the formation and excretion of bilirubin. BDG, bilirubin diglucuronide; BMG, bilirubin monoglucuronide. Reprinted from Dennery, P. A., Seidman, D. S., and Stevenson, D. K. (2001). Drug therapy: Neonatal hyperbilirubinemia. *N. Engl. J. Med.* **344**, 581–590, with permission. Copyright 2001, Massachusetts Medical Society. All rights reserved.

enterohepatic bilirubin circulation has been shown to increase under fasting conditions and may result in augmented bilirubin reabsorption in poorly feeding nursing infants.

Beneficial Properties of Bilirubin

Despite its potential for encephalopathy, bilirubin may have beneficial, physiologic properties. *In vitro* studies have demonstrated antioxidative properties associated with bilirubin and it is possible that *in vivo* some degree of bilirubinemia may help prevent reactive oxygen species-mediated diseases.

PATHOLOGICAL JAUNDICE

As in physiological jaundice, conditions resulting in pathological hyperbilirubinemia can be divided into those causing increased bilirubin production, on the one hand, and those causing diminished bilirubin conjugation or elimination, on the other.

Increased Bilirubin Production

Increased production of bilirubin, or hemolysis, may be due to immune causes, such as blood group incompatibilities. Rh isoimmunization was, in the past, a frequent cause of hyperbilirubinemia and kernicterus. However, prophylaxis of this condition by antenatal and immediate postnatal administration of Rhogam (anti-D antibody) have made this condition rare. In the remainder of cases, severity of disease has been ameliorated by vigilant antenatal surveillance and intrauterine blood transfusion where indicated. Undoubtedly, the most common blood group incompatibility currently causing immune hemolysis is ABO blood group heterospecificity, in which situation the mother has blood group O and her anti-A or anti-B antibodies cross the placenta and hemolyze the fetal A or B red blood cells (RBCs). Isoimmunization may also be the result of other antibodies, such as anti-E and anti-C. A positive direct Coomb's test, indicating maternal antibody deposited on the neonate's RBCs, will frequently be found, but is a poor predictor of hemolysis and hyperbilirubinemia.

Nonimmune causes of hemolysis include G-6-PD deficiency, a common enzyme deficiency occurring in some Mediterranean, Asian, and African populations. Severe hemolysis, often the result of an identifiable chemical trigger, may occur. Structural abnormalities of the RBC membrane (hereditary spherocytosis, elliptocytosis), propensity to increased

hemolysis in certain racial groups (e.g., Asian), pyruvate kinase deficiency, and extravasated blood (cephalhematoma, bruising) are other causes of nonimmune, increased hemolysis.

Diminished Bilirubin Elimination

Of the conditions resulting in diminished bilirubin conjugation, Crigler-Najjar syndrome is the most severe (also known as Arias type 1). This is a rare condition in which coding area mutations of the *UGT1A1* gene result in an abnormally structured UGT enzyme. The type 1 variant frequently leads to kernicterus early in neonatal life. Type 2 is less severe and typically responds to phenobarbital therapy, an inducer of UGT activity. Gilbert's syndrome is a genetic condition associated in Western populations with additional TA sequences in the promoter region of the *UGT1A1* gene. Jaundice in this condition may be the result of diminished bilirubin conjugation due to decreased UGT enzyme expression and also of increased heme catabolism combined with decreased bilirubin uptake into the hepatocyte. Gilbert's syndrome may combine with factors increasing bilirubin production to potentiate severe hyperbilirubinemia. A gene interaction has been proposed between Gilbert's syndrome and G-6-PD deficiency; neither Gilbert's syndrome nor G-6-PD deficiency, individually, results in an increase in the incidence of STB concentration >15.0 mg/dl over that of baseline. However, when the two conditions coexist, the incidence of hyperbilirubinemia increases significantly. Gilbert's syndrome also potentiates hyperbilirubinemia in hereditary spherocytosis and ABO blood group incompatibility and is associated with prolonged jaundice. In Asian populations, Gilbert's syndrome may be caused by a *UGT* gene-coding area mutation, Gly71Arg, which is also associated with hyperbilirubinemia.

Miscellaneous Causes

Sepsis, macrosomia associated with maternal diabetes, hypothyroidism, and being breast-fed are other conditions known to increase the severity of neonatal hyperbilirubinemia.

JAUNDICE IN BREAST-FED NEONATES

Some degree of jaundice is very common in breast-fed neonates. Not only do they have higher STB concentrations than formula-fed counterparts, but a higher incidence of hyperbilirubinemia as well. Almost all neonates readmitted to the hospital for hyperbilirubinemia are breast-fed. The jaundice may be due to unsuccessful breast-feeding, associated with chronic dehydration and failure to gain weight, which can be attributed to decreased milk or caloric intake. This condition usually becomes manifest in the first week of life and is frequently associated with weight loss >7% of birth weight, few wet diapers, and a small number of stools daily. The problem can be resolved by frequent nursing, including at night, and counseling regarding breast-feeding techniques. This condition has been distinguished from true breast milk jaundice, which develops gradually toward the second week of life. The cause is unclear, but exaggeration of the enterohepatic bilirubin circulation or the presence of β-glucuronidase or other substances in the milk may be associated. Increased hemolysis has not been demonstrated in these neonates. The diagnosis is one of exclusion, by ruling out other causes of indirect hyperbilirubinemia in a breast-fed infant. Usually the jaundice resolves, but STB concentrations may occasionally reach dangerous levels and kernicterus has been reported in breast-fed neonates with no additional cause of their hyperbilirubinemia. Temporary cessation of breast-feeding or a short period of phototherapy is usually effective in preventing further increases in STB.

ASSESSMENT OF JAUNDICE AND PREDICTION OF HYPERBILIRUBINEMIA

Age-Specific Serum Total Bilirubin Concentrations

It is important to assess the degree of jaundice according to an infant's age in hours. A study of the nomogram in Fig. 2 will reveal that there is a rapid increase in STB concentrations over the first few days of life. Thus, a neonate with a STB concentration of 12.0 mg/dl at age 72 h will be below the 40th percentile for hour of life, whereas a similar infant with the same STB concentration at 36 h will be above the 95th percentile. This concept is of extreme importance because of the current trend toward early discharge; babies sent home at age 48 h or less will frequently not yet have developed visible jaundice and bilirubin concentrations can be expected to increase during the next few days. In a study of universal predischarge bilirubin screening, babies whose STB concentrations were >95th percentile for hour of life at 18 to 72 h had a 40% risk of developing subsequent hyperbilirubinemia, whereas, in contrast, those whose concentrations were <40th percentile had zero risk. This method of

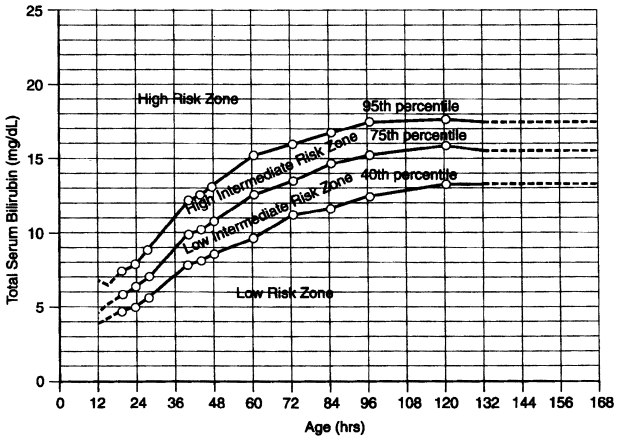


FIGURE 2 Nomogram for hour-specific serum total bilirubin values and the risk category for hyperbilirubinemia, in term and near-term neonates. Reprinted from Bhutani, V. K., Johnson, L., and Sivieri, E. M. (1999). Predictive ability of a predischarge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near-term newborns. Reproduced with permission from *Pediatrics* vol. **103**, 6–14, Fig. 2. Copyright 1999, American Academy of Pediatrics.

prediction is not infallible but does serve to indicate which neonates can be sent home with a wide margin of safety.

Assessment of Hemolysis

The AAP has recommended determining whether hemolysis, a risk factor for kernicterus, is present when evaluating a jaundiced neonate. Commonly used indices, such as falling hemoglobin or hematocrit count, reticulocytosis, or haptoglobin, are unreliable indicators of hemolysis in newborns. Assessment of the endogenous rate of CO production affords an accurate index of the rate of bilirubin production (Fig. 1). Commercially available, noninvasive devices are currently available for determining the end-tidal CO (ETCO) concentration in the neonates' exhaled breath. ETCO concentrations >95th percentile accurately detect increased hemolysis but, because the rates of bilirubin elimination vary, are not predictive of the development of hyperbilirubinemia.

KERNICTERUS

Kernicterus occurs when free bilirubin (unbound to serum albumin) crosses the blood–brain barrier and is deposited in the basal ganglia of the brain. Bilirubin may also enter the brain after the blood–brain barrier has been damaged by conditions such as asphyxia or hyperosmolarity. Pathological findings of kernicterus include staining and necrosis of the basal ganglia, hippocampus, and subthalamic brain nuclei. It is thought that bilirubin-induced damage is the result of interference with oxygen utilization by the cerebral tissue, possibly by causing damage to the cell membrane.

Current thinking suggests that it is unlikely that kernicterus will develop in an otherwise healthy, term infant, with no evidence of hemolysis, if STB concentrations do not exceed 25 mg/dl. This level of bilirubin may not be safe, however, in actively hemolyzing neonates or in prematures. Hypoalbuminemia, conditions such as acidosis, sepsis, and hypothermia, or drugs including sulfisoxazole and benzyl alcohol may affect the ratio of free bilirubin : albumin bound bilirubin and facilitate the entry of free bilirubin into the brain.

Clinically, bilirubin encephalopathy is characterized by acute and chronic phases. Classically, the first stage of kernicterus is characterized by poor sucking and hypotonia in an obviously severely jaundiced neonate. A typical high-pitched cry, seizures, and stupor ensue. Several days later (second stage), hypertonia of the extensor muscles, especially arching of the back, opisthotonus, retrocollis, and fever develop. The third stage, following the first week, is characterized by hypotonia, diminishing yet persisting stupor, and poor sucking and swallowing movements. Survivors of the acute stage demonstrate hypotonia and delayed motor skills in the first year, followed by extrapyramidal abnormalities including movement disorders, choreoathetosis, and involuntary muscle spasms. Upward gaze, diminished cognitive function, and high-frequency sensorineural hearing loss frequently occur. Some children may be more mildly affected by bilirubin encephalopathy and display moderate neuromuscular incoordination, varying degrees of hearing loss, and learning problems, which become apparent only on entering school.

TREATMENT OF HYPERBILIRUBINEMIA

The primary reason for treating hyperbilirubinemia is to prevent bilirubin encephalopathy. The mainstay of therapy is phototherapy, with exchange transfusion held in reserve for the few cases nowadays that do not respond to phototherapy. Additional therapeutic modalities include the use of intravenous immune

globulin (IVIG), drug induction of UGT, and synthetic metalloporphyrins for the inhibition of bilirubin production.

The AAP has formulated guidelines, currently in the process of revision, for the management of jaundice in healthy, term neonates. It should be realized that these guidelines have not been based on controlled, therapeutic trials, but rather on experience and clinical practice. Treatment is instituted at lower levels of STB in situations in which hyperbilirubinemia is evident on the first day of life and in septic, otherwise ill or premature infants. Newborns with evidence of hemolysis should be treated with great vigilance and STB concentrations should not be allowed to exceed 20 mg/dl in these cases. Neonates with very high STB levels (>25 mg/dl, or >20 mg/dl in the presence of hemolysis) who do not respond to a trial of approximately 4 h of intensive phototherapy or whose STB levels continue to rise despite treatment should have exchange transfusion performed or a trial of IVIG therapy where indicated (see below).

Phototherapy

Phototherapy has been in clinical use for over 30 years. It is safe and effective and has been instrumental in preventing kernicterus and avoiding the need for exchange transfusion in neonates worldwide. Phototherapy should be instituted in order to prevent STB concentrations from reaching levels that are associated with neurotoxic damage.

Bilirubin is one of the few body substances that absorb light energy, especially in the blue range of 450 nm. Light causes isomerization of the bilirubin molecule, forming lumirubin, and this water-soluble substance can be more readily excreted via the liver than regular bilirubin. Photooxidation of bilirubin accounts for a small fraction of the bilirubin eliminated via phototherapy.

The rate of lumirubin production, and therefore the effectiveness of phototherapy, is dependent on the total intensity of irradiance delivered. The light energy can be increased by replacing white light bulbs with blue when using fluorescents and decreasing the distance from the bulbs to the infant or increasing the body surface area exposed to light. The intensity of light irradiance can be measured at the bedside in units of microwatts per square centimeter per nanometer and phototherapy units should be regularly assessed, as light output of the bulbs within the given spectrum may weaken with time. Fluorescent bulbs are frequently used, but halogen lamps or fiberoptic blankets or mattresses are also available. Phototherapy is generally regarded as safe to use, although the eyes must be well covered prophylactically.

Exchange Transfusion

Exchange transfusion definitively corrects hyperbilirubinemia by physically removing bilirubin, as well as antibodies when present, from the body. The procedure is usually performed via an umbilical catheter. Aliquots of blood are removed from the infants and replaced by donor RBCs mixed with plasma. One double volume exchange transfusion is usually effective, but sometimes additional exchange procedures are necessary depending on the degree of postexchange STB rebound.

Complications of the procedure include apnea, bradycardia, cardiorespiratory arrest, necrotizing enterocolitis, and a small risk of death. Exchange transfusion also carries the risk of complications associated with blood transfusion. Prior to embarking on an exchange transfusion, the risks should clearly be weighed against the benefits.

Pharmacologic Therapy

Phenobarbital

Phenobarbital, administered either to the pregnant mother or to the already delivered baby, is instrumental in inducing UGT enzyme activity and also in enhancing bilirubin excretion. However, it may take several days for it to be effective and may itself have neurotoxic properties.

Heme Oxygenase Inhibitors

Heme oxygenase is the enzyme responsible for converting heme to biliverdin, from which bilirubin is subsequently formed. Some metalloporphyrins have been shown to be powerful inhibitors of heme oxygenase, thereby preventing the formation of bilirubin from heme. Although not yet approved for routine clinical use, studies in Argentina and Greece have demonstrated the efficacy of tin-protoporphyrin in preventing hyperbilirubinemia. Toxicity appears to be minimal, in the form of transient erythema.

Intravenous Immune Globulin

IVIG has been shown to reduce the need for exchange transfusion in Rh isoimmunization and also to decrease the degree of hemolysis in Coombs' positive ABO blood group heterospecificity. It is thought that IVIG acts by blocking the Fc receptor in the reticuloendothelial system, thereby inhibiting hemolysis.

CONCLUSIONS

Neonatal jaundice is extremely common and in most cases resolves with no ill effects. However, the resurgence of kernicterus is a frightening reminder of the tragic and devastating potential danger of hyperbilirubinemia. All newborn babies should be closely monitored for icterus. If a serum or transcutaneous bilirubin test is performed, the risk of hyperbilirubinemia should be assessed by plotting the result on the hour-by-hour nomogram. Parents must be interviewed about hyperbilirubinemia in previous siblings or even in themselves as infants. They should be closely questioned about their family's geographic area of origin in order to exclude the possibility of G-6-PD deficiency. The AAP has recommended that all neonates discharged at 48 h or less should be examined 2 to 3 days later for the development of jaundice. Vigilant management of all neonates and careful explanations to parents are necessary if future cases of kernicterus are to be prevented.

See Also the Following Articles

Bilirubin and Jaundice • Kernicterus • Neonatal Cholestasis and Biliary Atresia

Further Reading

American Academy of Pediatrics, Provisional Committee for Quality Improvement and Subcommittee on Hyperbilirubinemia (1994). Practice parameter: Management of hyperbilirubinemia in the healthy term newborn. *Pediatrics* **94**, 558–562.

Dennery, P. A., Seidman, D. S., and Stevenson, D. K. (2001). Neonatal hyperbilirubinemia. *N. Engl. J. Med.* **344**, 581–590.

Gourley, G. R. (2002). Breast-feeding, neonatal jaundice and kernicterus. *Semin. Neonatol.* **7**, 135–141.

Hansen, T. W. (2002). Mechanisms of bilirubin toxicity: Clinical implications. *Clin. Perinatol.* **29**, 765–778.

Johnson, L. H., Bhutani, V. K., and Brown, A. K. (2002). System-based approach to management of neonatal jaundice and prevention of kernicterus. *J. Pediatr.* **140**, 396–403.

Kaplan, M., Hammerman, C., and Maisels, M. J. (2003). Bilirubin genetics for the non-geneticist: Hereditary defects of neonatal bilirubin conjugation. *Pediatrics* **111**, 886–893.

Kaplan, M., Renbaum, P., Levy-Lahad, E., Hammerman, C., Lahad, A., and Beutler, E. (1997). Gilbert syndrome and glucose-6-phosphate dehydrogenase deficiency: A dose-dependent genetic interaction crucial to neonatal hyperbilirubinemia. *Proc. Natl. Acad. Sci. USA* **94**, 12128–12132.

Maisels, M. J., and Watchko, J. F. (eds.) (2000). "Neonatal Jaundice." Harwood Academic, Amsterdam.

Stevenson, D. K., Fanaroff, A. A., Maisels, M. J., *et al.* (2001). Prediction of hyperbilirubinemia in near-term and term infants. *Pediatrics* **108**, 31–39.

Neonatal Intestinal Obstruction

AMIR KAVIANI, DARIO FAUZA, AND MORITZ M. ZIEGLER
Children's Hospital Boston and Harvard Medical School

atresia Absence or closure of a natural passage or channel of the body.
meconium Dark greenish mass of desquamated cells, mucus, and bile that accumulates in the fetal bowel and is typically discharged shortly after birth.
polyhydramnios Excessive accumulation of amniotic fluid.

Intestinal obstruction is the most common surgical emergency of the newborn. The incidence of neonatal intestinal obstruction is approximately 1 in every 500–1000 live births. Approximately 50% of these cases are associated with intestinal atresia or stenosis.

ETIOLOGY

Intestinal obstruction in the neonate may be caused by the comprehensive list of conditions depicted in Table I. Practically, neonatal intestinal obstruction is most commonly due to intestinal atresia, malrotation with or without volvulus, meconium ileus, and Hirschsprung's disease. Table II lists a practical age-related etiology for the most commonly encountered cases of generic pediatric intestinal obstruction.

CLINICAL PRESENTATION AND DIAGNOSIS

In most cases, intestinal obstruction in the newborn becomes apparent shortly after birth. In some cases, proximal obstructing lesions may result in proximal bowel dilatation and hyperperistalsis, demonstrated by prenatal ultrasonography. Polyhydramnios may be present when a "high" intestinal obstruction is present. The classic "double-bubble" sign associated with duodenal atresia can also be identified by prenatal ultrasonography.

Findings suggesting intestinal obstruction in the neonate include (1) maternal polyhydramnios, (2) excessive gastric aspiration at birth accompanied by the inability to tolerate initial feeds, (3) abdominal distension, (4) bilious emesis, and (5) obstipation. The presence and severity of each of these findings depend largely on the level of obstruction along the gastrointestinal tract.

Maternal Polyhydraminos

Amniotic fluid is normally swallowed by the growing fetus and is absorbed by the fetal gastrointestinal tract. The fetal kidneys excrete the fluid back into the amniotic sac, it crosses into the maternal circulation through the placenta, and is ultimately excreted by the mother's kidneys. Proximal gastrointestinal obstruction in the fetus results in interruption of this normal process and leads to accumulation of excess amniotic fluid. Distal intestinal obstruction usually does not result in excessive amniotic fluid accumulation, because most of the absorption takes place in the proximal small bowel.

Excessive Gastric Aspiration

Passage of a nasogastric tube or orogastric tube is commonly performed in the neonatal care unit in premature infants or in infants with a history of maternal polyhydramnios. Intestinal obstruction should be suspected if the initial aspirate is large (50–100 ml), or bilious in nature.

TABLE I Causes of Neonatal Intestinal Obstruction

Common
Malrotation (duodenal obstruction, volvulus, internal hernia)
Duodenal atresia, stenosis, annular pancreas
Jejunal atresia or stenosis
Ileal atresia or stenosis
Simple meconium ileus
Meconium ileus with perforation
Hirschsprung's disease
Hypertrophic pyloric stenosis
Uncommon
Pyloric atresia or web
Tumors
Intussusception
Segmental intestinal dilatation
Small left colon syndrome
Milk bolus obstruction
Colonic atresia
Functional intestinal obstruction
Intestinal pseudo-obstruction

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Age-Based Classification of the Most Common Etiologies of Intestinal Obstruction in the Pediatric Population

Age	Etiology
Newborn	Atresia
	Malrotation ± volvulus
	Meconium ileus
	Hirschsprung's disease
2 Months to 2 years	Incarcerated inguinal hernia
	Intussusception
Over 2 years	Adhesions

Bilious Emesis

Although preterm infants may display bilious emesis as a result of a poorly developed pyloric sphincter, bilious emesis in a term infant is distinctly abnormal. In cases of intestinal obstruction, vomiting begins usually shortly after birth in a proximal, complete lesion but may be delayed several hours in more distal or incomplete obstruction. Bilious emesis in the newborn remains a particularly urgent finding that may herald the diagnosis of malrotation with volvulus.

Failure to Pass Meconium

A normal newborn typically passes copious amounts of thick, dark green meconium within the first 8–24 hours after birth. Failure to do so may be normal, but it may also suggest obstruction. Passage of meconium does not, by itself, exclude the possibility of obstruction. Preterm infants may have delayed passage of meconium, and this must be taken into consideration when evaluating the premature infant for a bowel obstruction.

Intestinal patterning and peristalsis that is visible through the abdominal wall are suggestive of obstruction, and distended loops of small bowel may be palpated through the abdominal wall as tubular masses. Hypotonia and lethargy develop late in the obstructed infant and are usually a result of sepsis. Ecchymosis and discoloration of the abdominal wall may suggest free perforation or frank intestinal ischemic infarction.

DIAGNOSIS

The diagnosis of neonatal intestinal obstruction is based on clinical findings supplemented with radiographs. Flat, upright, and decubitus abdominal radiographs are helpful in elucidating the diagnosis. Air swallowed by the newborn shortly after birth normally reaches the proximal small intestine in 30 minutes and it passes to the level of the colon by 3–4 hours. In a complete bowel obstruction, the air pattern stops abruptly just proximal to the site of obstruction. Multiple dilated small bowel loops with the characteristic "step-ladder" pattern may be seen on an upright radiograph in cases of distal obstruction. Peritoneal or scrotal calcifications may signify intrauterine perforation with meconium peritonitis.

Contrast barium enema radiographs are useful when attempting to delineate the cause of a distal intestinal obstruction; for example, in cases of meconium ileus, meconium plug syndrome, distal bowel atresia, and Hirschsprung's disease, specific findings may be present. Upper gastrointestinal (GI) contrast studies are most useful in cases in which abnormal bowel rotation is suspected.

PATHOPHYSIOLOGY

Proximal bowel obstruction in the neonate leads to a loss of fluid that has high concentrations of hydrogen, potassium, and chloride ions, leading to a hypochloremic, hypokalemic metabolic alkalosis. More distal obstruction leads to significant intravascular volume depletion due to the osmotic fluid shifts resulting from the fluid sequestered in the intestinal lumen. When pronounced, this may lead to oliguria and shock. Ultimately, severe dehydration may lead to poor end organ perfusion, and when coupled with bowel distension, perforation, and necrosis, metabolic acidosis may result. Significant abdominal distension may also impair diaphragmatic excursion and lead to respiratory acidosis.

TREATMENT

When neonatal intestinal obstruction is suspected, the early management includes gastric decompression with a nasogastric tube to prevent vomiting and aspiration. Resuscitation with appropriate fluids and electrolytes should begin immediately. Surgical intervention, when indicated, should be undertaken immediately once adequate resuscitation has been achieved.

See Also the Following Articles

Cystic Fibrosis • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Atresia • Malrotation • Volvulus

Further Reading

Filston, H. C. (1999). Other causes of intestinal obstruction. *In* "Pediatric Surgery" (J. A. O'Neill, Jr., *et al.,* eds.), 5th Ed., pp. 941–967. Mosby, St. Louis.

Haller, Jr., J. A., and Talbert, J. L. (1972). Gastrointestinal emergencies. *In* "Surgical Emergencies in the Newborn" (J. A. Haller, Jr., and J. L. Talbert, eds.), pp. 86–111. Lea & Febiger, Philadelphia.

Raffensperger, J. G., Seeler, R. A., and Moncada, R. (1970). Intestinal obstruction in the newborn. *In* "The Acute Abdomen in Infancy and Childhood" (J. G. Raffensperger *et al.,* eds.), pp. 1–19. J. B. Lippincott, Philadelphia.

Neonatal Tracheoesophageal Anomalies

AMIR KAVIANI, DARIO FAUZA, AND MORITZ M. ZIEGLER
Children's Hospital Boston and Harvard Medical School

atresia Absence or closure of a natural passage or channel of the body.

fistula Abnormal passage or communication, usually between two internal organs or leading from an internal organ to the surface of the body.

VACTERL Acronym for abnormalities of vertebrae, anus, cardiovascular tree, trachea, esophagus, renal system, and limb buds; condition associated with administration of sex steroids during early pregnancy.

Tracheoesophageal malformations in the neonate refer to a spectrum of developmental anomalies in which an abnormal anatomic arrangement exists between the esophagus and the tracheobronchial tree.

INTRODUCTION AND HISTORICAL BACKGROUND

The first description of esophageal atresia (EA) dates back to 1670 in *A Narrative of a Monstrous Birth in Plymouth*, in which William Durston described a blind-ending esophageal pouch in one infant of a set of conjoined twins. Thomas Gibson, in 1697, was the first author to describe EA with the typical tracheoesophageal fistula (TEF), in *The Anatomy of the Human Body Epitomized*. Numerous case reports appeared in the literature in the eighteenth and nineteenth centuries regarding these conditions, meticulously detailing some of the associated anomalies, including spina bifida, horseshoe kidney, and imperforate anus.

The first successful report of primary repair of EA with TEF was described by Cameron Haight in 1941. After Haight's initial success, numerous other reports appeared in the literature, detailing successful operative repair and various surgical approaches to these conditions.

EPIDEMIOLOGY

The incidence of congenital tracheoesophageal malformations in the United States is approximately 1 in every 4500 live births. In some regions of the world (i.e., Finland), the incidence may be as high as 1 in every 2400 births. There is a slight male preponderance for EA, with a male : female ratio of 1.2 : 1. There is a reported increased incidence of EA and TEF in first pregnancies and a slight correlation with increasing maternal age.

Although most cases of EA and TEF are sporadic, familial patterns of inheritance have been reported. Parents with one affected child have a 1–2% chance of a tracheoesophageal malformation affecting subsequent offspring. If more than one offspring is affected, the risk is 20%. The rate of twinning among infants with EA and TEF is higher than that of the general population (6 vs. 1%), as is the incidence of chromosomal anomalies. The most common chromosomal anomalies include trisomies 13 and 18, and the well-described VACTERL (involving abnormalities of the vertebrae, anus, cardiovascular tree, trachea, esophagus, renal system, and limb buds) association. Environmental teratogens associated with these conditions include prolonged exposure to contraceptive pills and exposure to estrogen and progesterone during pregnancy. In addition, a higher incidence has been described in infants of diabetic mothers and in those with intrauterine exposure to thalidomide.

 Copyright 2004, Elsevier (USA). All rights reserved.

Haller, Jr., J. A., and Talbert, J. L. (1972). Gastrointestinal emergencies. *In* "Surgical Emergencies in the Newborn" (J. A. Haller, Jr., and J. L. Talbert, eds.), pp. 86–111. Lea & Febiger, Philadelphia.

Raffensperger, J. G., Seeler, R. A., and Moncada, R. (1970). Intestinal obstruction in the newborn. *In* "The Acute Abdomen in Infancy and Childhood" (J. G. Raffensperger *et al.,* eds.), pp. 1–19. J. B. Lippincott, Philadelphia.

Neonatal Tracheoesophageal Anomalies

AMIR KAVIANI, DARIO FAUZA, AND MORITZ M. ZIEGLER
Children's Hospital Boston and Harvard Medical School

atresia Absence or closure of a natural passage or channel of the body.

fistula Abnormal passage or communication, usually between two internal organs or leading from an internal organ to the surface of the body.

VACTERL Acronym for abnormalities of vertebrae, anus, cardiovascular tree, trachea, esophagus, renal system, and limb buds; condition associated with administration of sex steroids during early pregnancy.

Tracheoesophageal malformations in the neonate refer to a spectrum of developmental anomalies in which an abnormal anatomic arrangement exists between the esophagus and the tracheobronchial tree.

INTRODUCTION AND HISTORICAL BACKGROUND

The first description of esophageal atresia (EA) dates back to 1670 in *A Narrative of a Monstrous Birth in Plymouth*, in which William Durston described a blind-ending esophageal pouch in one infant of a set of conjoined twins. Thomas Gibson, in 1697, was the first author to describe EA with the typical tracheoesophageal fistula (TEF), in *The Anatomy of the Human Body Epitomized*. Numerous case reports appeared in the literature in the eighteenth and nineteenth centuries regarding these conditions, meticulously detailing some of the associated anomalies, including spina bifida, horseshoe kidney, and imperforate anus.

The first successful report of primary repair of EA with TEF was described by Cameron Haight in 1941. After Haight's initial success, numerous other reports appeared in the literature, detailing successful operative repair and various surgical approaches to these conditions.

EPIDEMIOLOGY

The incidence of congenital tracheoesophageal malformations in the United States is approximately 1 in every 4500 live births. In some regions of the world (i.e., Finland), the incidence may be as high as 1 in every 2400 births. There is a slight male preponderance for EA, with a male : female ratio of 1.2 : 1. There is a reported increased incidence of EA and TEF in first pregnancies and a slight correlation with increasing maternal age.

Although most cases of EA and TEF are sporadic, familial patterns of inheritance have been reported. Parents with one affected child have a 1–2% chance of a tracheoesophageal malformation affecting subsequent offspring. If more than one offspring is affected, the risk is 20%. The rate of twinning among infants with EA and TEF is higher than that of the general population (6 vs. 1%), as is the incidence of chromosomal anomalies. The most common chromosomal anomalies include trisomies 13 and 18, and the well-described VACTERL (involving abnormalities of the vertebrae, anus, cardiovascular tree, trachea, esophagus, renal system, and limb buds) association. Environmental teratogens associated with these conditions include prolonged exposure to contraceptive pills and exposure to estrogen and progesterone during pregnancy. In addition, a higher incidence has been described in infants of diabetic mothers and in those with intrauterine exposure to thalidomide.

 Copyright 2004, Elsevier (USA). All rights reserved.

ANATOMY AND EMBRYOLOGY

Normal development of the esophagus and trachea remains poorly understood, and therefore the pathogenesis of EA and TEF is uncertain. Nevertheless, numerous theories have been put forth, mostly based on the original work of Wilhelm His, who described the invagination of the lateral longitudinal ridges as they give rise to the dorsal digestive tract and the ventral respiratory system. Most theories on the origin of EA and TEF conclude that an aberrant migration of the longitudinal ridges accounts for the spectrum of anomalies seen in EA and TEF.

Based on their studies of human embryos, O'Rahilly and Muller have proposed an alternate theory for the pathogenesis of EA and TEF. Their hypothesis suggests that the lung bud separates from the digestive tract by day 28 of gestation, descending caudally into the ventral mesenchyme of the foregut. Rather than proposing that the failure of regression of a normal connection between the two structures accounts for TEF, they suggest that a fistula is created by an abnormal epithelial-lined communication between the two originally separate tubes. They postulate that this event likely occurs on day 33 of gestation, when the esophagus and the tracheal bifurcation are in very close proximity.

Candidate genes responsible for EA and TEF have been proposed in chromosomal segments believed to be responsible for limb and foregut patterning. These genes may be responsible for normal migration and cell–matrix interactions during normal organ development.

CLASSIFICATION

Numerous classification systems exist for describing the spectrum of anomalies encountered in EA and TEF. In 1944, Ladd introduced a numeric form of classification that consisted of specific anomalies. Gross altered this numeric system in 1953, and introduced an alphabetical system that is commonly referenced today. Type A lesions are isolated esophageal atresia without a tracheoesophageal fistula and are frequently associated with a "long gap" between the proximal and distal esophageal segments. Type B lesions are esophageal atresias in association with a proximal tracheoesophageal fistula and are very rare, accounting for less than 1% of lesions. Type C lesions constitute the most common anomaly, and include a blind-ending proximal esophageal pouch with a distal tracheoesophageal fistula. In type D anomalies, there are two tracheoesophageal fistulas, one each from the proximal and distal esophageal segments. In type E anomalies, a tracheoesophageal fistula is present without an atresia (H-type fistula). Type F anomalies consist of congenital esophageal stenosis and are exceedingly rare.

ASSOCIATED ANOMALIES

Numerous congenital anomalies are associated with EA and TEF (Table I). Some reports have suggested that between 50–70% of infants with EA have other associated anomalies. These anomalies are most common in cases of EA without TEF, and least common in cases of H-type TEF. The most common anomalies involve the

TABLE I Anomalies Associated with Tracheoesophageal Fistula and Esophageal Atresia

Anomaly	Defect	Incidence (%)
Cardiovascular	Ventricular septal defect, tetrology of Fallot, atrial septal defect, patent ductus arteriosus, coarctation of aorta	49
Genitourinary	Hypospadias, cryptorchdism, renal agenesis, renal hypoplasia, cystic renal diseases, hydronephrosis, vesicoureteral reflux, ureteric duplication, urachal anomalies, cloaca or bladder exstrophy, posterior urethral valves	24
Gastrointestinal	Anorectal atresia, duodenal atresia, ileal atresia, malrotation, annular pancreas, pyloric stenosis	24
Neurologic	Hydrocephalus, neural tube defects, holoprosencephaly, microcephaly	23
Skeletal	Vertebral anomalies, radial limb deformities	13
VACTERL	Vertebral, anal, cardiovascular tree, tracheal, esophageal, renal, limb bud defects	25
CHARGE	Coloboma, heart, atresia (choanal), retardation, genital and ear defects	3
Choanal atresia	—	5.2
Facial cleft	—	7.2
Abdominal wall defect	—	4.3
Diaphragmatic hernia	—	2.9
Tracheobronchial	—	50

cardiovascular system (49%), followed by the genitourinary (24%), gastrointestinal (24%), and skeletal (13%) systems. Associated gastrointestinal anomalies include anorectal atresia, duodenal atresia, ileal atresia, malrotation, pyloric stenosis, and annular pancreas. Genitourinary defects include hypospadias, undescended testes, renal agenesis, cystic renal disease, hydronephrosis, and vesicoureteral reflux. Complex cardiac anomalies account for most of the deaths associated with EA. The most common single cardiac anomaly is ventricular septal defect, followed by tetrology of Fallot, patent ductus arteriosus, and atrial septal defects.

Quan and Smith proposed a broad spectrum of associated malformations in newborns. They described the acronym VATER (vertebral defects, anal atresia, tracheoesophageal fistula, esophageal atresia, and renal defects.) The affected individuals had no family history of any given disorder, no chromosomal abnormalities, and no known exposure to a teratogen *in utero*. Later, the acronym was expanded to VACTERL, incorporating the frequently seen cardiac and limb anomalies. Infants with esophageal atresia who have features of VACTERL have a high mortality and poor overall prognosis.

CLINICAL PRESENTATION

Infants with EA and TEF typically become symptomatic in the first few hours of life. The most common symptom shortly after birth is the presence of excessive salivation resulting from pooling of secretions in the proximal blind-ending esophageal pouch and oropharynx. Feeding often results in regurgitation, choking, and cyanosis. Atelectasis, tachypnea, and respiratory distress occur as a result of reflux of gastric contents through the TEF into the airway, or aspiration of saliva or formula from the proximal blind pouch. Inspired air passing through the TEF into the stomach further distends the stomach and elevates the diaphragm, leading to more respiratory compromise. These symptoms may be less apparent in children with tracheoesophageal fistula without esophageal atresia (H-type fistula), for which the clinical presentation may be delayed until indolent or recurrent pulmonary symptoms require further evaluation.

DIAGNOSIS

The prenatal diagnosis of EA is suggested by the finding of a small or absent gastric bubble in conjunction with maternal polyhydramnios. Although suggestive, this finding occurs in only 42% of cases of EA and has a positive predictive value of 56%. *In utero* ultrasonographic visualization of the blind upper esophageal pouch has been reported, but this relies on advanced ultrasonographic skills and is an inconsistent finding.

Pooling of secretions in the posterior oropharynx may be one of the earliest findings, and typically the first feeding is followed by regurgitation and coughing. Cyanosis with or without feeding may be present, and passage of a gastric decompression tube will be impossible. Resistance is typically encountered when the tube is passed to about 11−12 cm. A plain chest radiograph typically demonstrates the nasogastric tube coiling in the proximal esophageal pouch. The abdominal portion of the X ray may reveal bowel gas or even a "double-bubble" sign if an associated obstruction is present. The distance between the distal most margin of the coiled tube and the carina is used to estimate the gap between the blind-ending proximal esophageal pouch and the distal esophageal segment. A distance of less than two vertebral bodies is associated with a higher likelihood of being able to effect a primary repair. If diagnostic uncertainty exists, a small volume of water-soluble contrast (1−2 ml) can be carefully injected through the catheter under radiographic guidance to outline the blind-ending pouch. A proximal coiled tube associated with a gasless abdomen suggests that the correct diagnosis is an isolated esophageal atresia without a TEF. Once the diagnosis is confirmed, a thorough search for other recognizable congenital defects should begin.

TREATMENT

Initial treatment of infants with tracheoesophageal malformations is aimed at prevention of aspiration and resultant pulmonary soiling. A suction catheter is placed in the proximal esophageal pouch to continuously aspirate saliva and limit aspiration. The infant is positioned in an upright position to minimize gastroesophageal reflux and prevent aspiration through the TEF. Broad-spectrum antibiotics and a histamine-2 (H2) blocker are empirically started. Routine endotracheal intubation is avoided if at all possible, because ventilated air entering through the tracheoesophageal fistula can cause worsening respiratory distress resulting from abdominal distension, and in a worst-case scenario, gastric perforation may occur.

The surgical approach depends on the specific anomaly present. For type C lesions, division of the fistula and the primary anastomosis of the esophagus is the procedure of choice. The usual approach is a right-sided posterolateral extrapleural thoracotomy through the 4th interspace, unless preoperative evaluation reveals a right-sided aortic arch. Recently, minimally

TABLE II Spitz Classification and Survival

Group	Description	Total representation (%)	Survival (%)
I	Birth weight >1500 g without major congenital cardiac defect	79	97
II	Birth weight <1500 g or major congenital cardiac defect	19	59
III	Birth weight <1500 g and major congenital cardiac defect	2	22

invasive thoracoscopic techniques have been applied to this operation. In cases of isolated tracheoesophageal fistulas (E type), a cervical approach is usually possible, and the fistula is divided through a right-sided incision. In cases of "long-gap" esophageal atresia, other modalities of delayed or staged therapy may be necessary, including colonic interposition and gastric pull-up in order to reestablish a conduit.

OUTCOMES

Overall survival of neonates with tracheoesophageal malformations in modern series is approximately 90%. The original Waterston classification incorporated risk factors associated with EA and TEF and was based on birth weight, presence of pneumonia, and other congenital anomalies. These factors predicted those infants with a poor risk for survival and they helped guide operative treatment and neonatal critical care. Infants in the "good" risk category (A) were typically treated with immediate operative repair, "moderate-risk" infants (B) were managed with delayed repair, and "high-risk" infants (C) were treated by staged repair. Given the numerous advances in neonatal intensive care, more options have become available for the infant with multiple congenital anomalies, leading to development of newer classification schemes. In 1992, Spitz and colleagues published their results of a large review demonstrating that birth weight and presence of major cardiac disease remained important predictors of survival (Table II).

Complications after repair of tracheoesophageal anomalies include anastomotic leak (14–20%), recurrent tracheoesophageal fistula (3–14%), esophageal stricture (20–40%), gastroesophageal reflux (40–70%), and tracheomalacia and/or obstruction (10–20%). Esophageal dysmotility in the esophageal segment below the anastomosis is common and is part of the original lesion, not secondary to surgical transection of the esophagus.

See Also the Following Articles

Esophagus, Development • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Webs

Further Reading

Haight, C., and Towsley, H. (1943). Congenital atresia of the esophagus with tracheoesophageal fistula: Extrapleural ligation of fistula and end-to-end anastomosis of esophageal segments. *Surg. Gyn. Obstet.* **76**, 672.

Harmon, C. M., and Coran, A. G. (1999). Congenital anomalies of the esophagus. *In* "Pediatric Surgery" (J. A. O'Neill, Jr., *et al.*, eds.), 5th Ed., pp. 941–967. Mosby, St. Louis.

Spitz, L. (1996). Esophageal atresia: Past, present, and future. *J. Pediatr. Surg.* **31**, 19.

Spitz, L., Kiely, E., and Brereton, R. J. (1987). Esophageal atresia: Five year experience with 148 cases. *J. Pediatr. Surg.* **22**, 103.

Waterston, D. J., Bonham-Carter, R. E., and Aberdeen, E. (1962). Esophageal atresia: Tracheoesophageal fistula. A study of survival in 218 infants. *Lancet* **1**, 819.

Neurogastroenterology

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

functional gastrointestinal disorders Associated with symptoms that are present in the absence of any observable organic or physical abnormalities; symptoms include abdominal pain, bloating, early satiety, and urgency of defecation.

Neurogastroenterology is the name given to a relatively new scientific and clinical subspecialty that developed in the last quarter of the twentieth century. Neurogastroenterological research is a logical progression of the neurologically oriented research in basic medical sciences and of the advances in the understanding of functional gastrointestinal disorders in clinical gastroenterology.

NEUROGASTROENTEROLOGY DEFINED

In clinical medicine, neurogastroenterology subspecializes further the subspecialty of gastroenterology. Neurogastroenterology encompasses the investigative sciences dealing with functions, malfunctions, and malformations in the brain and spinal cord and the sympathetic, parasympathetic, and enteric divisions of the autonomic innervation of the digestive tract. Somatomotor systems are included insofar as pharyngeal phases of swallowing and pelvic floor involvement in defecation and continence are concerned. Basic physiology of smooth muscles, as it relates to enteric neural control of motor movements, is a part of neurogastroenterology. Psychologic and psychiatric relations to gastrointestinal disorders are significant components of the neurogastroenterologic domain, especially in relation to projections of discomfort and pain to the digestive tract.

Neurogastroenterology as the descriptor for a specialized sector of gastroenterology has emerged from the parallel fertilization of research in the basic medical sciences and of diagnosis and treatment of functional gastrointestinal disorders in clinical medicine. Basic neurogastroenterologic research integrates physiology, biochemistry, neurobiology, anatomy/histology, endocrinology, microbiology, immunology, and pharmacology into a unified subspecialty focused on the innervation of the digestive system. Clinical gastroenterology applies the emergent knowledge in the diagnosis and treatment of functional gastrointestinal disorders, examples of which are irritable bowel syndrome, functional dyspepsia, and brain–gut relations in physical and emotional stress.

See Also the Following Articles

Brain–Gut Axis • Defecation • Functional (Non-Ulcer) Dyspepsia • Irritable Bowel Syndrome

Further Reading

Grundy, D. (1993). The changing face of gastrointestinal motility. *J. Gastrointest. Mot.* **5**, 231–232.

Wood, J. (1994). Progress in neurogastroenterology. *In* "The Growth of Gastroenterologic Knowledge during the Twentieth Century" (J. Kirsner, ed.), pp. 159–180. Lea & Febiger, New York.

Wood, J., Alpers, D., and Andrews, P. (1999). Fundamentals of neurogastroenterology. *Gut* **45**, 1–44.

Wood, J., Alpers, D., and Andrews, P. (2000). Fundamentals of neurogastroenterology: Basic science. *In* "The Functional Gastrointestinal Disorders: Diagnosis, Pathophysiology and Treatment: A Multinational Consensus" (D. Drossman, N. Talley, W. Thompson, E. Corazziari, and W. Whitehead, eds.), 2nd Ed., pp. 31–90. Degnon Associates, McLean, Virginia.

 Copyright 2004, Elsevier (USA). All rights reserved.

Neurotensin

ARTHUR SHULKES
University of Melbourne

autocrine action Activity elicited from a nonneuronal cell in response to factors released by that same cell.
enterogastrone Hormonal substance released in response to the presence of intestinal fats; causes physiological inhibition of gastric acid secretion.
neurocrine action Activity elicited from a neuron in response to factors released by an adjacent cell or neurone.
neuromodulator Substance that can alter the release or effect of a neurotransmitter.
paracrine action Activity elicited from a nonneuronal cell in response to factors released by an adjacent cell.

Neurotensin is a 13-amino-acid peptide originally isolated from the hypothalamus; similar peptides have been subsequently found in other parts of the brain, the pituitary, and the intestine. The gut, predominantly the ileum, contains more than 90% of immunoreactive neurotensin. Although circulating neurotensin concentration is increased by a meal, the mechanisms of action are predominantly neurocrine, paracrine, and neuromodulator rather than endocrine. The most important functions in the gut are vasodilation, stimulation of the exocrine pancreas, growth of normal and cancerous tissues, and slowing of gastrointestinal transit.

CHEMISTRY, ISOLATION, AND DISTRIBUTION

During the isolation of substance P from bovine hypothalami, Leeman and colleagues in 1973 isolated a 13-amino-acid peptide; because of the vasodilatory and cyanotic effects of the peptide, it was named neurotensin. Neurotensin (NT) was subsequently shown to have a widespread distribution in the central nervous system, pituitary, and intestine. Furthermore, neurotensin and neurotensin-related peptides are represented across the phylogenetic tree (Table I). The C terminus, which is the bioactive core, is conserved, whereas the N-terminal peptide, NT(1–8), may be an inactive but stable degradation product generated following cleavage at the dibasic Arg^8–Arg^8 residues. Some of the early literature is contradictory and confusing, often because the specificity of the antisera is not stated and the extraction technique could generate artifacts or degradation products. The ileum contains the highest concentration of neurotensin, with lesser amounts in the jejunum and duodenum. Neurotensin is present in specific endocrine cells, the N cells, and in nerves, predominantly in the myenteric plexus. Overall, the intestine contains more than 90% of neurotensin. Most of the remaining NT is in the brain, particularly the hypothalamus and pituitary.

GENE AND PRECURSOR STRUCTURE

The complementary deoxyribonucleic acid (cDNA) for neurotensin was first obtained in 1987 using cultured canine ileal mucosal cells, which would have been

TABLE I Comparison of the Amino Acid Sequence for Neurotensin and Related Peptides from Different Species[a]

Neurotensin family species	Sequence												
Bovine/canine/human/porcine/rat	pE	L	Y	E	N	K	P	R	R	P	Y	I	L
Guinea pig	-	-	-	-	-	-	S	-	-	-	-	-	-
Chicken	-	-	H	V	-	-	A	-	-	-	-	-	-
Possum	-	-	H	V	-	-	A	-	-	-	-	-	-
Amphibian xenopsin						pE	G	K	-	-	W	-	-
Canine/turkey/rat xenopsin					F	H	-	K	-	-	W	-	-
Bovine/canine/human/porcine/rat/neuromedin N								K	I	-	-	-	-
Chicken LANT-6								K	N	-	-	-	-

[a] The entire bovine/canine/human/porcine/rat neurotensin sequence is shown. Only those amino acids that differ from this sequence are shown.

 Copyright 2004, Elsevier (USA). All rights reserved.

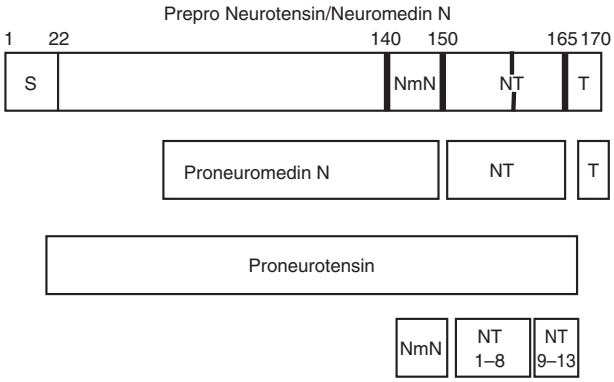


FIGURE 1 Schematic representation of human prepro neurotensin/neuromedin (170 amino acids). The thick vertical lines mark the positions of the dibasic residues (cleavage sites). The different peptide products generated from the precursor are shown (not to scale). S, Signal peptide; NmN, neuromedin N; NT, neurotensin; T, tail peptide.

enriched for both neurotensin peptide and mRNA. Neurotensin is synthesized within a 170-amino-acid precursor that also encodes the related peptide neuromedin N (Fig. 1). These two peptides are located near the C terminus and are separated by basic residue pairs. The rat neurotensin/neuromedin N gene was isolated by screening a genomic library with the canine cDNA. The rat gene has 10.2 kb and is divided into four exons, with exon 4 containing the coding region for both neurotensin and neuromedin N. Two species of mRNA have been identified (1.0 and 1.5 kb); the differences reside in the lengths of the 3′ untranslated tail. A similar pattern is seen in the human gene. The 1.0-kb form predominates in the intestine and pituitary, with equivalent amounts of the 1.0- and 1.5-kb forms in the brain and in tumors. The nucleotide and amino acid sequences are highly conserved (>80%) between humans, dogs, cows, and rats. Not all the precursor molecules are processed and proneurotensin and proneuromedin are found in the gastrointestinal tract and in tumor cell lines. In the gut, proneurotensin accounts for 2–5% of neurotensin immunoreactivity, but proneuromedin N is the major form of gut neuromedin. Proneurotensin is an order of magnitude less potent than neurotensin, but because of increased stability may contribute to the actions of neurotensin. The prohormone convertases, PC1 and PC, have the major roles in the processing of gut proneurotensin.

SECRETION

Ingestion of fat is the most potent stimulant of gastrointestinal neurotensin release; glucose and amino acids cause insignificant increases. The majority of neurotensin is rapidly metabolized into biologically inactive but stable N-terminal fragments, predominantly neurotensin(1–8), with the bioactive C-terminal fragment neurotensin(9–13) being very rapidly metabolized. Thus, the increase in circulating bioactive neurotensin following a meal is quite modest and less than that required for a biological effect, such as inhibition of acid secretion or gastric emptying. Nevertheless, neurotensin may function in concert with other regulatory peptides, such as cholecystokinin, to influence gastrointestinal functions. Infusion of fat into the jejunum but not the ileum of humans and dogs increases NT release. However, ileal resection abolishes the increase in NT. Together, these data suggest that neurotensin is released from the ileum via a local humoral or neural stimulus emanating from the proximal intestine. Two facts—that the initial release of neurotensin following a meal occurs within 10 minutes, which is before chyme would reach the ileum, and that it is abolished by atropine—support this reflex hypothesis. Neurally mediated release appears to involve a nonvagal cholinergic mechanism, because truncal vagotomy does not abolish the postprandial increase and stimulants of vagal tone, such as sham feeding and insulin hypoglycemia, have no effect on neurotensin. Cholinergic independent mechanisms may also be involved, because bombesin-stimulated release of neurotensin is not atropine sensitive.

RECEPTORS

Neurotensin binding sites were first demonstrated in rat brain synaptic terminals and subsequently detected in jejunum and ileum, especially in the submucosal and myenteric plexuses. Three receptor subtypes that recognize the C-terminal, biologically active part of neurotensin have been cloned and characterized. Neurotensin receptor type 1 (NTR1) and NTR2 are both G protein-coupled receptors, whereas NTR3 is identical to the previously cloned human gp95/sortilin receptor.

Neurotensin Receptor Type 1

The cDNA clone for NTR1 was obtained from a rat brain cDNA library using an electrophysiological assay in *Xenopus* oocytes as the end point. The cDNA encodes a 424-amino-acid protein (similar in size to that reported earlier by chemical isolation) that, as suggested by pharmacological studies, is a member of the G protein-coupled receptor family. The human receptor has been cloned from the colonic adenocarcinoma

cell line HT29. The receptor has a high affinity (0.2 n*M*), is insensitive to levocabastine, and is expressed in the intestine and brain. Studies with NTR1 knockout mice have demonstrated a major role for this receptor in mediating central temperature control and gastric and colonic motility. Signal transduction mechanisms are through coupling with phospholipase C and subsequent increases in intracellular Ca and inositol 1,4,5-trisphosphate. Mitogen-activated protein kinases are also activated by neurotensin. Both pathways are inhibited by nonpeptide neurotensin receptor antagonists such as SR 48692 and SR 142948. Based largely on studies with these antagonists, NTR1 is the probable mediator of the growth effects of NT on pancreas, and gastrointestinal tract and on pancreatic and colorectal cancer cell lines, although NTR3 may also play a contributory role (see later). NTR1 mRNA has been detected in many colorectal and pancreatic carcinoma cell lines. However. receptor autoradiography of resected colorectal tumors was positive for NTR1 in only one of 25 tumors.

Neurotensin Receptor Type 2

NTR2 is a low-affinity (5−7 n*M*) receptor that shares about 60% homology with NTR1 and is sensitive to levocabastine. Although it has the characteristics of a G protein-coupled receptor, the transduction mechanisms have not been completely characterized and appear to be system and species specific. Furthermore, neurotensin antagonists have paradoxical agonist functions in some species. The receptor is expressed mostly in the brain but in a distribution pattern different to NTR1, so there is rarely coexpression. The NTR2 receptor appears to mediate neurotensin-induced analgesia. No specific gastrointestinal role has been demonstrated for NTR2 and no NTR2 binding sites could be detected in the human colon and no NTR2 mRNA in human pancreatic and colon cancer cell lines.

Neurotensin Receptor Type 3

A third neurotensin binding site was isolated by affinity cross-linking of brain extracts; following cloning, the site was shown to have 100% homology with the previously cloned human gp95/sortilin receptor. This protein has a large luminal domain, a single-transmembrane domain, and a short cytoplasmic tail. The majority of NTR3 is located in intracellular vesicles, but some is also found on the cell membrane. NTR3 mRNA and intracellular binding sites are detected in pancreatic and colonic cancer cell lines. NTR3 mediates NT-stimulated cell growth in cancer cell lines either alone or as a heterodimer with NTR1.

BIOLOGICAL EFFECTS

Neurotensin has a variety of biological effects; more than 30 different *in vitro* and *in vivo* effects have been documented. Peripheral effects include vasodilation, cyanosis, increased histamine release, stimulation of the endocrine and exocrine pancreas, effects on gastrointestinal tract smooth muscle activity and motility, stimulation of intestinal secretion, and inhibition of blood flow to adipose tissue. Neurotensin promotes the growth of the gastrointestinal tract as well as gut- and pancreatic-derived cancer cell lines. Centrally, neurotensin has hypothermic and antinociceptive effects and modulates brain dopamine systems, luteinizing hormone, and prolactin release. Only the gastrointestinal effects will be considered further.

Neurotensin secretion is stimulated by fat and is thought to be the mediator of many of the effects influenced by fat ingestion. These include inhibition of gastrointestinal transit and gastric acid secretion and stimulation of pancreatic secretion and intestinal blood flow. However, infusion studies mimicking the postprandial release of neurotensin indicate that the amount secreted is insufficient to mediate these actions. Neurotensin is therefore thought to function by interacting with other hormones, such as cholecystokinin, secretin, and the cholinergic and noncholinergic components of the enteric nervous system. Finally, neurotensin functions as a locally acting paracrine agent, and its location in specific enteric neurones indicates a role as a neurotransmitter.

Gastrointestinal Secretions

Neurotensin administered either intracerebroventricularly or peripherally inhibits gastric acid secretion. The peripheral effect is dependent on an intact vagus. The concentration of neurotensin required to inhibit gastric acid secretion is higher than that achieved following a high-fat meal and it is likely that neurotensin acts in concert with other meal-stimulated enterogastrones, such as secretin.

A similar interaction is observed with exocrine pancreatic secretion. Neurotensin alone increases exocrine pancreatic secretions whereas the combination of neurotensin with secretin potentiates pancreatic protein output and the combination of neurotensin with cholecystokinin potentiates pancreatic bicarbonate secretion. Cholinergic pathways also contribute to the pancreatic stimulatory effect of neurotensin. Early studies reporting that neurotensin stimulated insulin secretion used pharmacological doses. Neurotensin is now not thought to have a role in regulating the

endocrine pancreas. Intestinal secretion is also stimulated by relatively low doses of neurotensin.

Motility

Many of the motility effects of neurotensin mimic the motility changes that occur following the conversion from the fed to the fasted state. Neurotensin or ingestion of a fat meal reduces lower esophageal sphincter pressure, slows gastric emptying and intestinal transit, and inhibits the interdigestive migratory complex. As with other effects of neurotensin, an interaction with vagal cholinergic components is required. Once fat reaches the ileum, there is a further inhibition of gastric and intestinal motility, and this appears to be mediated by the distal small intestine peptide YY (PYY) together with neurotensin. *In vitro* studies performed to determine the mechanism of the motility effects of neurotensin have been difficult to interpret because the responses are dose, region, and species dependent. Neurotensin *in vitro* has both a direct relaxant and an indirect contractile action, with the relaxant effect being tetrodotoxin resistant and the contractile effect dependent on cholinergic and substance P-mediated mechanisms. Studies with both neurotensin antiserum and antagonists suggest that neurotensin is one of the noncholinergic, nonadrenergic mediators of relaxation in the small intestine. Overall, the consequence of these motility changes is that the chyme from a meal rich in fat is retained longer in the stomach and intestine, which should aid digestion and absorption of nutrients.

Growth

Many of the peptides produced by the intestine are growth factors. These include epidermal growth factor, transforming growth factor-α (TGF-α), insulin-like growth factor-1 (IGF-1), and glucagon-like peptide-2 (GLP-2). Neurotensin also has a trophic effect on the gastrointestinal tract and pancreas, but the specific contribution of neurotensin among the constellation of growth factors has not been resolved.

Animal studies have shown that neurotensin infusion is trophic for the pancreas, stomach, and intestine. However, relatively high doses are required. Neurotensin enhances the trophic effect of GLP-2, whereas the trophic effects of IGF-1 and TGF-α are mediated in part by increases in neurotensin production. The hypoplasia induced by feeding rats an elemental diet is reversed by neurotensin infusion. Similarly, the adaptive response following either small or large bowel resection is augmented by neurotensin administration.

PATHOLOGY

Neurotensin-producing pancreatic endocrine tumors are quite rare and are often associated with the overproduction of other bioactive peptides, such as vasoactive intestinal peptide. Because vasoactive intestinal peptide and neurotensin both stimulate intestinal secretion, it has been difficult to determine whether neurotensin is responsible for the diarrhea often associated with pancreatic endocrine tumors. Studies with neurotensin antagonists in patients with the watery diarrhea hypokalemia achlorhydria syndrome have not been reported. Hepatic fibrolamellar tumors and some prostatic tumors also overproduce neurotensin, often resulting in elevated blood neurotensin concentrations. Again, no specific symptoms have been associated with neurotensin overproduction in these tumors. The nature of the neurotensin produced is quite variable, ranging from intact neurotensin, proneurotensin, inactive neurotensin(1–8), and bioactive C-terminal fragments 6–13 or 9–13.

Neurotensin gene expression has been detected in 10 of 40 colorectal cancers tested. This expression was not detected in adjacent normal colonic mucosa. Interestingly, neurotensin gene and peptide are present in fetal colon, suggesting a recapitulation of the fetal phenotype under some circumstances. Using receptor autoradiography, only a small proportion of human colorectal cancers express NTR1, although the surrounding muscle and nerve cells are labeled. Whether the NTR3 subtype is present is unknown.

Neurotensin administration to rats increases the number and size but not the incidence of colon tumors induced by azoxymethane. The mechanism has been ascribed to an increase in proliferation, although the doses of neurotensin and azoxymethane were high.

Extensive studies with colon cancer cells either *in vitro* or *in vivo* after xenografting have confirmed a proliferative effect of neurotensin that is inhibited by NTR1 antagonists. However, NTR1 and NTR3 are both detected on colon cancer cell lines and interact in a cooperative manner. About half of colorectal cancer cell lines tested contain neurotensin receptors, determined either by binding studies or expression of receptor mRNA. The NTR1 receptor antagonist alone inhibits the endogenous growth of xenografted colon cancer cells, raising the possibility that neurotensin antagonists could be used therapeutically. The neurotensin receptor antagonist may function by inhibiting an autocrine regulatory loop, because many colon cancer cell lines contain neurotensin mRNA and peptide. Whether endogenous, meal-stimulated neurotensin has a stimulatory role on tumor growth has not been determined.

Similar to other peptide growth factors detected in tumors, processing efficiency is diminished and there is a preponderance of neurotensin precursor detected in colon cancer lines. However, proneurotensin is more stable than the fully processed neurotensin(1–13) and is able to activate NTR1 receptors in the nanomolar range.

Neurotensin receptors (both mRNA and binding) have been detected in nearly all pancreatic carcinomas. However, unlike normal colon, neurotensin receptors are also detected in normal pancreas, albeit at lower levels than in the tumors. In pancreatic cancer cell lines, the neurotensin antagonist SR 48692 inhibits both basal and neurotensin-stimulated proliferation, consistent with an endocrine or autocrine regulation of growth.

Patients with the dumping syndrome have exaggerated meal-stimulated release of neurotensin, probably as the result of the rapid passage of nutrients from the stomach to the small intestine. Neurotensin is not the responsible agent, because intravenous infusion does not mimic the symptoms. Another instance of exaggerated neurotensin release occurs in patients with severe chronic pancreatitis. The excess neurotensin response is probably the result of undigested fat products in the small intestine, because it is reversed by enzyme replacement therapy.

In terms of therapeutic value, the trophic effect of neurotensin could be of theoretical benefit in patients who have atrophy of the bowel associated with long-term parenteral feeding, or in patients who have had small intestine resection. However, a long-lasting analogue will be required for this to be tested in human clinical trials. Finally, the efficacy of the NTR1 subtype antagonist SR 48692 in moderating experimental pancreatic and colorectal carcinogenesis indicates a potential role for neurotensin antagonists.

Acknowledgment

A. S. is a National Health and Medical Research Council of Australia Senior Principal Research Fellow.

See Also the Following Articles

Fat Digestion and Absorption • Substance P

Further Reading

Carraway, R. E., Mitra, S. P., and Spaulding, G. (1992). Posttranslational processing of the neurotensin/neuromedin-N precursor. *Ann. N. Y. Acad. Sci.* **668**, 1–16.

Evers, B. M. (2002). Endocrine gene neurotensin: Molecular mechanisms and a model of intestinal differentiation. *World J. Surg.* **26**, 799–805.

Ferris, C. (1989). Neurotensin. *In* "Handbook of Physiology, Section 6: The Gastrointestinal System. Volume II, Neural and Endocrine Biology" (S. G. Schultz, G. M. Makhlouf, and B. B. Rauner, eds.), pp. 559–586. American Physiological Society, Baltimore, Maryland.

Martin, S., Navarro, V., Vincent, J. P., and Mazella, J. (2002). Neurotensin receptor-1 and -3 complex modulates the cellular signaling of neurotensin in the HT29 cell line. *Gastroenterology* **123**, 1135–1143.

Pettibone, D. J., Hess, J. F., Hey, P. J., Jacobson, M. A., Leviten, M., Lis, E. V., Mallorga, P. J., Pascarella, D. M., Snyder, M. A., Williams, J. B., and Zeng, Z. (2002). The effects of deleting the mouse neurotensin receptor NTR1 on central and peripheral responses to neurotensin. *J. Pharmacol. Exp. Ther.* **300**, 305–313.

Rostene, W. H., and Alexander, M. J. (1997). Neurotensin and neuroendocrine regulation. *Front. Neuroendocrinol.* **18**, 115–173.

Shulkes, A. (1994). Neurotensin. *In* "Gut Peptides: Biochemistry and Physiology" (J. H. Walsh and G. J. Dockray, eds.), pp. 371–388. Raven Press, New York.

Vincent, J. P., Mazella, J., and Kitabgi, P (1999). Neurotensin and neurotensin receptors. *Trends Pharmacol. Sci.* **20**, 302–309.

Nitric Oxide

PAUL KUBES
University of Calgary, Canada

inflammation Body's response to infection or injury, including increased swelling, redness, heat, pain, and white cell infiltration (puss formation).
inflammatory bowel disease Chronic inflammation of the intestines.
nitric oxide Gas produced by cells of the brain, blood vessels, and immune system.
nitric oxide synthase-2 Protein that produces abundant bactericidal nitric oxide in response to inflammation.
nitric oxide synthase-3 Protein that continuously produces small amounts of nitric oxide to maintain normal blood pressure.

Nitric oxide is synthesized from a guanidino group of L-arginine and can be produced by virtually all mammalian cells, including endothelium lining the vasculature, epithelial cells, neurons of the central and enteric nervous systems, and cells of the immune system. Nitric oxide is constitutively produced by an enzyme normally present primarily in endothelium lining the vasculature (nitric oxide synthase-3; NOS3) and a neurally associated nitric oxide synthase found in neurons of the brain and enteric nervous system (NOS1). The constitutive forms of nitric oxide synthase, including neuronal NOS (NOS1) and endothelial NOS (NOS3), are critical to normal physiology. Nitric oxide produced from NOS3 maintains adequate perfusion, regulates microvascular permeability, modulates platelet homotypic aggregation as well as platelet adhesion to vessel walls, and regulates leukocyte–endothelial cell interactions. In the gastrointestinal tract, NOS3 appears to also regulate epithelial permeability. NOS1 is associated with the enteric nervous system and presently there is very limited information about its potential inflammatory/antiinflammatory function. Although NOS2, an inducible isoform of nitric oxide synthase, is not considered constitutive, small amounts of NOS2 message and protein have certainly been reported under normal conditions in the intestinal tract, and so a role for this isoform cannot be excluded in intestinal homeostatic function.

INTESTINAL INFLAMMATION

Acute and chronic inhibition of constitutive nitric oxide causes many of the hallmark features of intestinal inflammation, including increased neutrophil recruitment, increased oxidative stress, mast cell degranulation, and increased microvascular and epithelial permeability. Many studies of acute inflammatory models have documented that exogenous nitric oxide protects the gastrointestinal mucosa against noxious stimuli. Although improvement in blood flow associated with increased NO delivery would be an important factor in countering the compromised blood flow due to damaging effects of lumen-based insults such as HCl, ethanol, and indomethacin, NO can also protect at subdilatory levels.

NITRIC OXIDE SYNTHASE-2

Nitric oxide synthase-2 (NOS2) is an enzyme that requires protein synthesis for significant expression in endothelium, epithelium, and inflammatory cells; it is induced by cytokines and lipopolysaccharide (LPS) and produces large amounts of nitric oxide for extended periods of time. Three key observations underlie the hypothesis that NOS2 contributes to prolonged inflammation in a negative manner: (1) large quantities of nitric oxide are produced by NOS2, (2) expression patterns of NOS2 have correlated nicely with prolonged inflammation, and (3) many studies have shown that NO synthase inhibition reduces inflammation; this third observation is the most compelling evidence to suggest that NOS2 does contribute to intestinal inflammation. NO synthase inhibitors provide extremely beneficial results in a model of chemically induced (trinitrobenzenesulfonic acid; TNBS) guinea pig ileitis. Increased nitric oxide production from NOS2 has also been documented in a spontaneous, idiopathic, rhesus monkey colitis model. In that model, NOS2 inhibitor reduced the diarrhea but did not alter the morphological features of the disease. The TNBS-induced colitis was greatly improved in mice lacking NOS2 (NOS2−/− mice) and resulted in 90% mortality in wild-type mice and 38% mortality in the NOS2−/− counterparts.

It is important to note that not all studies support a detrimental role for NOS2. In a model of mucosal injury and repair not involving a chronic phase of inflammation,

 Copyright 2004, Elsevier (USA). All rights reserved.

the inducible NOS (iNOS)-deficient mice healed less effectively than did their wild-type counterparts, an observation recently noted in other tissues. In another study using TNBS-induced colitis in iNOS−/− mice, within the first 72 hours of TNBS-induced colitis the iNOS−/− mice had an approximately 50% greater damage score and increased neutrophilic infiltrate. By day 7, as the model entered the chronic phase of inflammation (mast cell hyperplasia and macrophage and lymphocyte infiltration), there was no difference in the various parameters of injury measured between iNOS−/− and iNOS+/+ mice. NOS2 deficiency is not important in interleukin-10 (IL-10)-deficient mice spontaneously developing chronic intestinal inflammation.

PEROXYNITRITE

Peroxynitrite, a potent nitrosating agent, may be the product that makes NOS2 detrimental. NO and superoxide react in essentially a diffusion-limited fashion to produce peroxynitrite. This product has a half-life of 1.9 seconds at pH 7.4 and exists in equilibrium with peroxynitrous acid. Peroxynitrous acid generates an excited isomer that acts as a hydroxyl-like oxidizing species, or, by direct homolysis of peroxynitrous acid, produces hydroxyl. In this fashion, peroxynitrite oxidizes a variety of molecules (sulfhydryls, thiols, and ascorbate) and triggers cytotoxic processes, including lipid peroxidation and DNA damage. Initial work suggests that peroxynitrite but not nitric oxide can also nitrosate tyrosine, to produce 3-nitrotyrosine. Because peroxynitrite is presently impossible to measure *in vivo* due to its high reactivity, 3-nitrotyrosine formation is used as a fingerprint of peroxynitrite formation. Intense immunostaining of nitrotyrosine has been detected in rodent and human colitis. It should be noted that that 3-nitrotyrosine could be formed from nitrite (NO_2^-) in the presence of hypochlorous acid or human polymorphonuclear leukocytes (a source of hypochlorous acid), suggesting this is not a specific marker of peroxynitrite. Although current work does question the specificity of nitrotyrosine, it does not necessarily refute the existence of peroxynitrite *in vivo*. A criticism levied against the formation of peroxynitrite per se has been the requirement for the simultaneous generation of equivalent amounts of nitric oxide and superoxide to generate optimal levels of peroxynitrite. Because the relative cellular concentration of superoxide is in the order of 1000 times lower than nitric oxide, the amount of peroxynitrite produced is limited, in a site-specific fashion, to areas of high superoxide anion generation or, alternatively, to low nitric oxide levels. Surprisingly, peroxynitrite under some conditions may function to prevent injury, an event that requires further investigation.

See Also the Following Articles

Enteric Nervous System • Gastric Reservoirs

Further Reading

Eiserich, J. P., Hristova, M., Cross, C. E., Jones, A. D., Freeman, B. A., Halliwell, B., and van der Vliet, A. (1998). Formation of nitric oxide-derived inflammatory oxidants by myeloperoxidase in neutrophils. *Nature* **391**, 393−397.

Grisham, M. B., Jourd'Heuil, D., and Wink, D. A. (1999). Nitric oxide. I. Physiological chemistry of nitric oxide and its metabolites: Implications in inflammation. *Am. J. Physiol.* **276**, G315−G321.

McCafferty, D.-M., Sihota, E., Muscara, M. N., Wallace, J. L., Sharkey, K. A., and Kubes, P. (2000). Spontaneously developing chronic colitis in IL-10/iNOS-double deficient mice. *Am. J. Physiol.* **279**, G90−G99.

Moncada, S. (1992). The L-arginine: Nitric oxide pathway. *Acta Physiol. Scand.* **145**, 201−227.

Nathan, C., and Xie, Q. (1994). Nitric oxide synthases: Roles, tolls and controls. *Cell* **78**, 915−918.

Pryor, W. A., and Squadrito, G. L. (1995). The chemistry of peroxynitrite: A product from the reaction of nitric oxide with superoxide [see comments]. *Am. J. Physiol.* **268**, L699−L722.

Ribbons, K. A., Zhang, X.-J., Thompson, J. H., Greenberg, S. S., Moore, W. M., Kornmeier, C. M., Currie, M. G., Lerche, N., Blanchard, J., Clark, D. A., and Miller, M. J. S. (1995). Potential role of nitric oxide in a chronic model of colitis in rhesus macaques. *Gastroenterology* **108**, 705−711.

Wink, D. A., and Mitchell, J. B. (1998). Chemical biology of nitric oxide: Insights into regulatory, cytotoxic, and cytoprotective mechanisms of nitric oxide. *Free Radic. Biol. Med.* **25**, 434−456.

Non-Steroidal Anti-Inflammatory Drugs (NSAIDs)

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

cyclooxygenases Family of enzymes, of which at least two isoforms exist, cyclooxygenase-1 and cyclooxygenase-2. They act on arachidonic acid to produce a number of compounds, including prostaglandins and thromboxane.

prostaglandins One form of eicosanoids (biologically active lipids formed by the oxidation of 20-carbon fatty acids); produced by the cyclooxygenase pathway, they are responsible for a variety of physiologic and inflammatory reactions.

Non-steroidal Anti-inflammatory drugs (NSAIDs) are a heterogeneous group of chemical compounds that share common mechanisms of action, therapeutic effects, and toxicities. The prototypical drug is aspirin. There are over 20 NSAIDs currently available in the United States. Nearly 20 million persons in the United States use these drugs on a daily basis, making this group of drugs one of the most widely prescribed agents. Systematic reviews have not found important differences in efficacy between different NSAIDs, but differences in toxicity have been demonstrated.

HISTORY

Salicylic acid is the active ingredient in willow bark, the medicinal effects of which have been known for several centuries. Sodium salicylate was first used in 1875 for the treatment of rheumatic fever. Hoffman, a chemist working for the pharmaceutical company Bayer, synthesized acetylsalicylic acid in 1897. This compound was introduced into medicine in 1899 under the name of aspirin. Since the introduction of indomethacin in 1963, a number of similar agents have been synthesized. In 1971, John Vane established that aspirin inhibits the synthesis of prostaglandins. Within the past decade, it has become clear that inhibition of two enzymes is involved. With the elucidation of these two enzymes, cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2), increasingly sophisticated and specific compounds have been produced in attempts to improve efficacy and lower toxicity. Cyclooxygenase-3 (COX-3) is a recently described variant of COX-1.

NOMENCLATURE AND CLASSIFICATION

The NSAIDs are classified based on COX specificity and chemical structure. The nonselective COX inhibitors include salicylate derivatives such as aspirin, sodium salicylate, sulfasalazine, and diflunisal. The *para*-aminophenol derivatives include acetaminophen, and the indole acetic acid derivatives include indomethacin and etodolac. The heteroarylacetic acid derivatives are typified by diclofenac. Arylpropionic acids include ibuprofen and naproxen. The fenamates include mefenamic acid, and the enolic acids include piroxicam and meloxicam. Nabumetone is an alkalone. Some of the nonselective COX inhibitors (meloxicam, etodolac, and diclofenac) display COX-2 selectivity, especially at low doses. COX-2-selective inhibitors include nimuselide, rofecoxib, celecoxib, valdecoxib, etoricoxib, and lumiracoxib. Of these, rofecoxib, valdecoxib, and celecoxib are currently licensed in the United States.

MECHANISM OF ACTION

Non-steroidal Anti-inflammatory drugs act principally by inhibiting the synthesis of prostaglandins. This appears to be the basis of both the Anti-inflammatory and analgesic effects. This inhibition is mediated by competitively inhibiting the enzyme, cyclooxygenase, responsible for the synthesis of prostaglandins. Additionally, the NSAIDs also inhibit the formation of prostacyclin and thromboxane, both of which are involved in vascular permeability and platelet aggregation. This is the basis of the anticoagulant effect seen with the NSAIDs, especially that of aspirin, which irreversibly acetylates the COX enzymes.

As already noted, there appear to be two well-defined isoforms of the COX enzymes that differ both in their tissue distribution and their inducibility. COX-1 is widely expressed in most tissues, with a constitutive level of expression, being especially important in the upper gastrointestinal tract, where it helps maintain the integrity of the mucosa. COX-2 is considered an

 Copyright 2004, Elsevier (USA). All rights reserved.

inducible enzyme and is typically undetectable in most tissues. During states of inflammation, the levels of COX-2 are greatly increased. The targeting of COX-2 with concomitant sparing of COX-1 has been the driving force behind the development of COX-2-selective NSAIDs. It is increasingly recognized that the dichotomy of physiologic significance of COX-1 and COX-2 isoforms is not mutually exclusive, with COX-1 likely involved in the prevention of gastrointestinal ulcers and central perception of pain and COX-2 being critical to renal blood flow and electrolyte balance.

Both isoforms of the COX enzyme are inhibited to some degree by all currently available NSAIDs. Several traditional NSAIDs have been shown to be more selective for COX-2 than for COX-1, especially at low doses. This is typical of etodolac and meloxicam. At higher doses, this relative specificity is typically lost. The COX-2-specific inhibitors are much more likely to retain this selectivity even at higher doses. Large, prospective trials of the COX-2-specific inhibitors have suggested a decreased incidence of clinically important upper gastrointestinal events, although the influence on renal function and effects on the coagulation cascade remain to be fully elucidated. Prospective trials to understand the effects of COX-2-specific inhibitors on the coagulation pathway are underway.

An exciting new development has been the demonstration of a COX-1 variant, called COX-3, suggesting that multiple COX isoenzymes could be derived from just two distinct genes providing a COX continuum of enzymes and products. Other proposed mechanisms of action for NSAIDs are based on their lipophilic properties. Direct interruption of protein−protein interactions may underlie some of the effects attributed to NSAIDs.

PHARMACOLOGY

NSAIDs are well absorbed following oral administration. Once absorbed they are typically highly bound to serum albumin because they are weak organic acids. Most NSAIDs are metabolized predominantly in the liver and are subsequently excreted in the urine. This has important implications in patients with hepatic and/or renal dysfunction. Some NSAIDs (indomethacin, sulindac, and piroxicam) have prominent enterohepatic circulation. This may result in a prolonged half-life. NSAIDs with short half-lives achieve maximum plasma concentrations quickly whereas those with longer half-lives achieve this more slowly. Sulindac and nabumetone require conversion to an active compound after first-pass metabolism through the liver. Once steady-state pharmacokinetics have been achieved, tissue concentrations of NSAIDs do not vary greatly. Highly lipid-soluble NSAIDs can penetrate the central nervous system and may cause adverse effects such as headache; this is most frequently noted with indomethacin.

CLINICAL USES

Pain

NSAIDs have been widely used as single-dose therapy or as short-term intermittent therapy, to relieve mild to moderate pain. Most episodes of acute musculoskeletal pain, muscular headaches, dysmenorrhea, and postoperative pain respond well to NSAIDs, although there is a paucity of randomized controlled trials demonstrating improved efficacy over acetaminophen in acute musculoskeletal syndromes. NSAIDs are occasionally effective as monotherapy in chronic pain, but are typically used as an adjunct to opioids in the management of severe pain. They may be particularly effective in bony pain of malignant origin. NSAIDs are increasingly used for the management of postoperative pain, especially following day-surgery, because of the lack of sedative effects.

Fever

NSAIDs are effective antipyretic agents, comparable to acetaminophen. Aspirin is generally avoided in children with fever because of a possible link between its use and the development of Reye's syndrome. Ibuprofen is widely utilized in children with febrile illnesses.

Osteoarthritis

NSAIDs, because of their anti-inflammatory effects, are particularly attractive analgesics in patients with osteoarthritis. Systematic reviews have found that NSAIDs are effective in reducing the pain of osteoarthritis. No particular NSAIDs has been found more effective than the others have in this regard. Comparative trials of NSAIDs versus simple analgesics such as acetaminophen (which may act via COX-3) have not shown good evidence that NSAIDs are superior to simple analgesics in the treatment of osteoarthritis.

Rheumatoid Arthritis

NSAIDs are widely prescribed in the management of rheumatoid arthritis, especially in early disease, for which they are utilized as "bridge" therapy, because traditional disease-modifying antirheumatic drugs achieve benefits only after several weeks of therapy.

Other Uses

NSAIDs have been utilized for the management of biliary and renal colic, acute gout, the spondyloarthropathies, and as topical agents for the management of inflammatory ocular disorders. High doses of celecoxib have been utilized in the management of familial adenomatous polyposis (FAP).

ADVERSE EFFECTS

Gastrointestinal effects include an increased incidence of dyspepsia, peptic ulceration, intestinal ulceration, and blood loss. Renal effects must be considered with patients whose renal function is dependent on prostaglandin synthesis; such patients, who are especially vulnerable to renal side effects from NSAID therapy, include the elderly and patients with diabetes, cirrhosis, hypertension, congestive heart failure, and chronic renal insufficiency from any cause. Such patients on NSAID therapy have predictable increases in blood pressure and in peripheral edema and decreases in glomerular filtration rate. The COX-2-selective agents are as likely to cause nephrotoxicity (and therefore offer no renal advantage) when compared with nonselective NSAIDs.

Other adverse effects are much less common and include hepatotoxicity, photosensitivity eruptions, headache with or without meningitis, tinnitus (especially with aspirin), worsening of asthma and allergic rhinitis, and worsening of osteoarthritis (specifically with indomethacin). The antiplatelet effects of aspirin may manifest as easy bruisability. It remains to be determined definitively whether COX-2-selective NSAIDs have prothrombotic effects.

See Also the Following Articles

NSAID-Induced Injury • Pharmacology, Overview

Further Reading

Bombardier, C., Laine, L., Reicin, A., *et al.* (2000). Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis: VIGOR Study Group. *N. Engl. J. Med.* **343**, 1520–1528.

Chandrasekharan, N. V., Dai, H., Roos, K. L., *et al.* (2002). COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: Cloning, structure, and expression. *Proc. Natl. Acad. Sci. U.S.A* **99**, 13926–13931.

Roberts II, L. J., and Morrow, J. D. (2001). Analgesic–antipyretic and anti-inflammatory agents and drugs employed in the treatment of gout. *In* "Goodman and Gilman's: The Pharmaceutical Basis of Therapeutics," 10th Ed., pp. 687–713. McGraw-Hill, New York.

Simon, L. S. (2001). Non-steroidal anti-inflammatory drugs. *In* "Primer on the Rheumatic Diseases," 12th Ed., pp. 583–591. Arthritis Foundation, Atlanta.

Vane, J. R. (1971). Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nat. New Biol.* **23**, 232–235.

Nonalcoholic Fatty Liver Disease

PAUL ANGULO
Mayo Medical School, Clinic, and Foundation

acanthosis nigricans Disorder involving localized skin hyperpigmentation; appears as velvety hyperpigmentation, primarily in flexural areas. It is usually associated with obesity, type 2 diabetes mellitus, Cushing's syndrome, acromegaly, and the Stein–Leventhal syndrome.

insulin resistance Decreased ability of insulin to act effectively on peripheral target tissues, especially muscle and liver; resistance is relative, because supernormal levels of circulating insulin will normalize the plasma glucose.

oxidative stress Refers to increased generation of reactive oxygen species or free radicals, including superoxide anions, hydrogen peroxide, and hydroxyl radicals; occurs as a consequence of accumulation of fat within hepatocytes. The reactive oxygen species or free radicals seem to induce chemotaxis and accumulation of inflammatory cells in the liver.

truncal obesity Denotes waist/hip ratio (waist circumference divided by hip circumference) equal to or greater than 0.85 and 0.90 in women and men, respectively.

Nonalcoholic fatty liver disease is an increasingly recognized clinicopathological condition that may progress to end-stage liver disease. The pathological picture resembles alcohol-induced liver injury, but occurs in patients who do not abuse alcohol. This disease comprises a wide spectrum of liver damage, ranging from simple steatosis, to steatohepatitis, to advanced fibrosis and cirrhosis. Steatohepatitis (nonalcoholic) represents only a stage within the spectrum of nonalcoholic fatty liver disease. Nonalcoholic fatty liver disease and steatosis with or without hepatitis resulting from secondary causes should be differentiated (Table I) because they have distinctly different pathogeneses and outcomes.

RISK FACTORS

Obesity, type 2 (non-insulin-dependent) diabetes mellitus, and hyperlipidemia are comorbid conditions frequently associated with nonalcoholic fatty liver disease (NAFLD). The prevalence of obesity reported in several series of NAFLD studies varies between 30 and 100%; the prevalence of type 2 diabetes varies between 10 and 75% and the prevalence of hyperlipidemia varies

TABLE I Causes of Fatty Liver Disease

Nutritional	Drugs[a]	Metabolic/genetic	Other
Protein-calorie malnutrition	Glucocorticoids	Lipodystrophy	Inflammatory bowel disease
Starvation	Synthetic estrogens	Dysbetalipoproteinemia	Small bowel diverticulosis with bacterial overgrowth
Total parenteral nutrition	Aspirin	Weber–Christian disease	HIV infection
Rapid weight loss	Calcium channel blockers	Wolman's disease	Enviromental hepatotoxins
Gastrointestinal surgery for obesity	Amiodarone	Cholesterol ester storage	Phosphorus poisoning
	Tamoxifen	Acute fatty liver of pregnancy	Petrochemical exposure
	Tetracycline		Toxic mushrooms
	Methotrexate		Organic solvents
	Perhexiline maleate		*Bacillus cereus* toxins
	Valproic acid		
	Cocaine		
	Antiviral agents		
	Zidovudine		
	Didanosine		
	Fialuridine		

[a] Partial list of agents that produce fatty liver in humans. Some drugs produce inflammation as well. The association of fatty liver with calcium channel blockers and valproic acid is weak, whereas the association with amiodarone is strong. Drug-induced fatty liver may have no sequelae (i.e., corticosteroids), or can eventuate in cirrhosis (i.e., methotrexate and amiodarone).

 Copyright 2004, Elsevier (USA). All rights reserved.

between 20 and 92%. The prevalence of NAFLD increases by 4.6-fold in obese people. Regardless of body mass index, type 2 diabetes mellitus significantly increases the prevalence and severity of NAFLD. Truncal obesity seems to be an important risk factor for NAFLD, even in subjects with normal body mass index. About half of patients with hyperlipidemia were found to have NAFLD on ultrasound examination in one study. Hypertriglyceridemia rather than hypercholesterolemia may increase the risk of NAFLD. NAFLD may affect any age group and has been described in most racial groups.

PREVALENCE

NAFLD affects 10−24% of the general population worldwide. The prevalence of NAFLD, however, increases to 57.5−74% in obese people. NAFLD affects 2.6% of children and this figure increases to 22.5−52.8% in the obese pediatric population. NAFLD represents a common explanation for abnormal liver tests in blood donors, and it is the cause of asymptomatic elevation of aminotransferases in up to 90% of cases once other causes of liver disease are excluded.

The prevalence of NAFLD in the United States is unknown, although a good estimate can be derived from the known prevalence of obesity and type 2 diabetes mellitus in the general population. Obesity as defined by a body mass index (BMI) $\geq 30\ kg/m^2$ affects 22.5% of people aged 20 years or older. Steatosis, regardless of diabetic status, is found in over two-thirds of the obese population and in more than 90% of the severely obese (BMI $\geq 35\ kg/m^2$) individuals. Steatohepatitis affects about 3% of the lean population, 19% of the obese individuals, and almost half of severely obese people. Hence, when extrapolated to the 2002 United States population, an estimated 30.1 million obese adults in this country may have steatosis, and about 8.6 million may have steatohepatitis. Diabetes mellitus affects 7.8% of the United States adult population whereas 50% (range 21−78%) of diabetic patients have NAFLD.

CLINICAL FEATURES

Most patients with NAFLD have no symptoms or signs of liver disease at the time of diagnosis, although many patients complaint of fatigue or malaise and a sensation of fullness or discomfort in the right upper abdomen. Hepatomegaly is the only physical finding in most patients. Acanthosis nigricans may be found in children with NAFLD. Stigmata of chronic liver disease suggest advanced disease with cirrhosis. A high proportion of patients with cryptogenic cirrhosis shares many of the clinical and demographic features of patients with NAFLD, suggesting that cryptogenic cirrhosis was previously unrecognized NAFLD.

Mild to moderate elevation of serum aminotransferases is the most common and often the only laboratory abnormality found in patients with NAFLD. The aspartate aminotransferase/alanine transaminase (AST/ALT) ratio is usually less than 1, but this ratio increases as fibrosis advances, leading to a loss of its diagnostic accuracy in patients with cirrhotic stage. Serum alkaline phosphatase and/or γ-glutamyl transpeptidase are above the normal range in many patients. Other abnormalities, including hypoalbuminemia, prolonged prothrombin time, and hyperbilirubinemia, may be found in patients with cirrhotic stage NAFLD. Elevated serum ferritin levels are found in a half of patients and increased transferrin saturation is found in 6−11% of patients. Hepatic iron index and hepatic iron concentration, however, are usually in the normal range.

Fatty infiltration of the liver produces a diffusely increased echogenicity when compared to the kidney texture on ultrasonography. Cirrhosis regardless of etiology is associated with a similar appearance. Ultrasonography has a sensitivity of 89% and a specificity of 93% in detecting steatosis, and a sensitivity and specificity of 77 and 89%, respectively, in detecting increased fibrosis. Fatty infiltration of the liver produces a low-density hepatic parenchyma on a computer tomography (CT) scan. Magnetic resonance (MR) imaging can distinguish space-occupying lesions from focal fatty infiltration or focal fatty sparing.

LIVER HISTOLOGY

NAFLD is histologically indistinguishable from the liver damage resulting from alcohol abuse. Liver biopsy features include steatosis, mixed inflammatory cell infiltration, hepatocyte ballooning and necrosis, glycogen nuclei, Mallory's hyaline, and fibrosis. The presence of these features, alone or in combination, accounts for the wide spectrum of NAFLD. Fatty infiltration, necroinflammation, and hepatocyte ballooning are mostly seen in acinar zone 3, but their distribution is more diffuse in more severe cases of NAFLD. Mallory's hyaline is usually located in ballooned hepatocytes, but it is neither unique nor specific for NAFLD. Portal tracts are relatively spared from inflammation, although children with NAFLD may show a predominance of portal inflammation as opposed to a lobular infiltrate. Mallory's hyaline is notably sparse or absent in children with NAFLD.

Fibrosis in NAFLD suggests more advanced and severe liver injury. Some degree of fibrosis is found in up to 66% of patients at the time of diagnosis. Collagen is first laid down in the pericellular space around the central vein and in the perisinusoidal region in zone 3. This pattern of fibrosis helps to distinguish NAFLD and alcoholic liver disease from other forms of liver disease in which fibrosis shows an initial portal distribution. The combination of (a) steatosis; (b) mixed mononuclear and/or polymorphonuclear cell infiltration; and (c) hepatocyte ballooning and spotty necrosis is known as nonalcoholic steatohepatitis (NASH).

PATHOGENESIS

The pathogenesis of NAFLD has remained poorly understood since the earliest description of the disease. Much current thinking remains hypothetical and the mechanisms are still being worked out. It is not yet understood why some patients develop simple steatosis whereas others develop steatohepatitis and progressive disease; differences in body fat distribution or antioxidant systems, possibly in the context of a genetic predisposition, may be among the explanations. A net retention of lipids within hepatocytes, mostly in the form of triglycerides, is a prerequisite for the development of NAFLD. The primary metabolic abnormality leading to lipid accumulation, however, is not well understood, but it could potentially result from insulin resistance and alterations in the uptake, synthesis, degradation, or secretory pathways of hepatic lipid metabolism. Insulin resistance represents the most reproducible factor for the development of NAFLD. Increased intrahepatic concentration of fatty acids provides a source of oxidative stress, which may be in large part responsible for the progression from steatosis to steatohepatitis to cirrhosis. Mitochondria are the main cellular source of reactive oxygen species (ROS), which may trigger steatohepatitis and fibrosis by two main mechanisms: lipid peroxidation and cytokine induction.

Humans with steatohepatitis exhibit ultrastructural mitochondrial lesions, including linear crystalline inclusions in megamitochondria. This mitochondrial injury is absent in most patients with simple steatosis or in healthy subjects. After a fructose challenge, which causes acute hepatic ATP depletion, patients with steatohepatitis slowly resynthesize ATP *in vivo*. This impaired ATP recovery may reflect the mitochondrial injury found in patients with steatohepatitis. Thus, although obese, diabetic, or hyperlipidemic patients with fatty liver rarely develop symptoms of liver disease, the steatotic liver may be vulnerable to further injury when challenged by additional insults. This has led some to presume that progression from simple steatosis to steatohepatitis to advanced fibrosis results from two operating "hits." First, insulin resistance leads to accumulation of fat within hepatocytes, and second, mitochondrial ROS lead to lipid peroxidation and cytokine induction.

DIAGNOSIS

NAFLD is usually suspected in individuals with asymptomatic elevation of aminotransferases, radiological findings of fatty liver, or unexplained persistent hepatomegaly. Imaging studies help in determining the presence and amount of fatty infiltration of the liver, but the clinical suspicion of NAFLD and its severity can be confirmed only with a liver biopsy.

The diagnosis of NAFLD requires exclusion of alcohol abuse as the cause of liver disease; abuse involves a daily alcohol intake as low as 20 g in females and 30 g in males. Other etiologies, such as viral, autoimmune, metabolic/hereditary, and drug/toxin-induced liver disease, should be ruled out. Specific laboratory tests along with a number of histologic findings on liver biopsy make the diagnosis of these other liver diseases straightforward in most cases.

Given the lack of effective medical therapy for all patients with NAFLD, a liver biopsy may not be necessary for the simple purpose of making the diagnosis in a patient with clear risk factors (obesity, diabetes, or hyperlipidemia). However, a liver biopsy may provide the most useful diagnostic information for those patients who do not have any risk factors for NAFLD, as well as for those patients whose liver test results do not improve after weight loss and appropriate metabolic control. Some factors can help to identify the NAFLD patient for whom the liver biopsy may provide the most prognostic information. Age older than 45 years, presence of obesity or type 2 diabetes mellitus, and an AST/ALT ratio greater than 1 are noteworthy indicators of advanced liver fibrosis.

NATURAL HISTORY

The natural history of NAFLD is being defined, but it seems to be determined by the severity of histological damage. Out of 257 patients with NAFLD reported in five different series, 54 had subsequent liver biopsies performed during an average followup of 3.5–11 years. Of these patients, 28% had progression of liver damage, 59% remained essentially unchanged, and 13% showed improvement or resolution of liver injury. Thus, in many patients, NAFLD follows a

relatively benign course, whereas, in other patients, the disease progresses to cirrhosis and its complications. Patients with pure steatosis on liver biopsy seem to have the best prognosis within the spectrum of NAFLD, whereas features of steatohepatitis or more advanced fibrosis are associated with a worse prognosis.

MANAGEMENT

In patients with diabetes mellitus or hyperlipidemia, good metabolic control is always recommended, but is not always effective in reversing NAFLD. Improvement in liver tests is almost universal in obese adults and children after weight reduction. The degree of fatty infiltration usually improves with weight loss in most patients, although the degree of necroinflammation and fibrosis may worsen. How fast weight loss is achieved is important and may play a critical role in determining whether liver histology will improve or, in fact, worsen. A weight loss of about 500 g/week in children and 1600 g/week in adults has been advocated. The use of medications that can directly reduce or reverse liver damage independent of weight loss is a desired alternative. However, only small pilot studies lasting 1 year or less have been reported to date. Medications that have been tested in patients with NAFLD are summarized in Table II. These medications deserve further evaluation in carefully controlled powerful clinical trials that include clinically relevant end points.

TABLE II Medications Evaluated in the Treatment of Nonalcoholic Fatty Liver Disease

Medication type	Example
Insulin sensitizing	Metformin
	Troglitazone[a]
	Rosiglitazone
	Pioglitazone
Antioxidant	Vitamin E
	Vitamin C
	Betaine
	N-Acetylcysteine
Antiobesity	Orlistat
Hepatoprotective	Ursodeoxycholic Acid
Lipid lowering	Gemfibrozil
	Clofibrate
	Bezafibrate
	Atorvastain

[a] Troglitazone has been removed from the market because of its potential hepatotoxicity.

General Recommendations

An attempt at gradual weight loss along with appropriate control of serum glucose and lipid levels is a useful first step. Perhaps this should be the only treatment recommendation for patients with NAFLD with pure steatosis and no evidence of necroinflammation or fibrosis. Because most patients who develop problems from NAFLD have steatohepatitis, treatment is more likely to be aimed at those patients. Patients with steatohepatitis, particularly those with fibrosis on liver biopsy, should be monitored closely, make a greater effort for adequate metabolic control, and be offered enrollment in clinical trials. For the NAFLD patient with decompensated cirrhosis, liver transplantation is a potential therapeutic alternative. NAFLD, however, may recur in the allograft or develop after liver transplantation for cryptogenic cirrhosis.

See Also the Following Articles

Diabetes Mellitus • Fibrogenesis • Hyperlipidemia • Obesity, Treatment of

Further Reading

Angulo, P. (2002). Nonalcoholic fatty liver disease. *N. Engl. J. Med.* **346**, 1221–1231.

Angulo, P., Keach, J. C., Batts, K. P., and Lindor, K. D. (1999). Independent predictors of liver fibrosis in patients with nonalcoholic steatohepatitis. *Hepatology* **30**, 1356–1362.

Baldridge, A. D., Peres-Atayde, A. R., Graeme-Cook, F., Higgins, L., and Lavine, J. E. (1995). Idiopathic steatohepatitis in childhood: A multicenter retrospective study. *J. Pediatr.* **127**, 700–704.

Ballentani, S., Saccoccio, G., Masutti, F., Croce, L. S., Brandi, G., Sasso, F, Cristanini, G, and Tiribelli, C. (2000). Prevalence of and risk factors for hepatic steatosis in northern Italy. *Ann. Intern. Med.* **132**, 112–117.

Chitturi, S., and Farrell, G. C. (2001). Etiopathogenesis of nonalcoholic steatohepatitis. *Semin. Liver Dis.* **21**, 27–41.

Day, C. P., and James, O. F. W. (1998). Steatohepatitis: A tale of two "hits"? *Gastroenterology* **114**, 842–845.

Matteoni, C. A., Younossi, Z. M., Gramlich, T., Boparai, N., Liu, Y. C., and McCullough, A. J. (1999). Nonalcoholic fatty liver disease. A spectrum of clinical and pathological severity. *Gastroenterology* **116**, 1413–1419.

Ratziu, V., Giral, P., Charlotte, F., Bruckert, E., Thibault, V., Theodorou, I., Khalil, L., Turpin, G., Opolon, P., and Poynard, T. (2000). Liver fibrosis in overweight patients. *Gastroenterology* **118**, 1117–1123.

Silverman, J. F., O'Brien, K. F., Long, S., Leggett, N., Khazanie, P. G., Pories, W. J., Norris, H. T., and Caro, J. F. (1990). Liver pathology in morbidly obese patients with and without diabetes. *Am. J. Gastroenterol.* **85**, 1349–1355.

Nosocomial Infections

ARMINE MARGARET SEFTON
Barts and the London, Queen Mary's School of Medicine and Dentistry, London

coliforms Gram-negative aerobic rods, which are part of the family *Enterobacteriacae* and which are found in normal bowel flora.
nosocomial Hospital acquired.

Nosocomial infections of the gastrointestinal (GI) tract include bacterial and viral diarrhea, sepsis relating to surgical and invasive procedures on the GI tract, hospital-acquired infective hepatitis, and nosocomial infections resulting from the colonization and/or infection of other areas of the body with organisms arising from the GI tract.

INTRODUCTION

A nosocomial infection is one that is hospital acquired. These infections can have significant morbidity and mortality and have a large financial impact on hospital resources. They lead to increased stay length of infected patients, resulting in decreased total throughput of patients. They also lead to increased utilization of antimicrobials and high-intensity beds for patients developing nosocomial infections. In addition, they lay the hospital open to the potential risk of litigation.

The three most common types of nosocomial infection are urinary tract infections, wound and soft tissue infections [including those resulting after procedures on the gastrointestinal (GI) tract], and nosocomial pneumonia, which frequently occurs due to aspiration of bacteria colonizing the oropharynx or upper gastrointestinal tract. However, nosocomial infections, per se, of the gastrointestinal tract are also a significant cause of morbidity and sometimes mortality in hospitals and nosocomial gastrointestinal infections can lead to multiple ward closures.

NOSOCOMIAL GASTROINTESTINAL INFECTIONS

Diarrhea is the most common nosocomial infection of the gastrointestinal tract. It can be bacterial or viral in origin and can sometimes result in multiple ward closures within a hospital.

Bacterial Infections

Antibiotic-Associated Diarrhea and Pseudomembranous Colitis

Giving a patient broad-spectrum antimicrobials changes his or her bowel flora and can lead to antibiotic-associated diarrhea. This is usually mild, is not associated with a particular organism, and resolves when the antibiotics are stopped. However, approximately 20% of cases of antibiotic-associated diarrhea are due to a specific bacterium, *Clostridium difficile*, that has a spectrum of disease ranging from an asymptomatic carrier state to fulminant colitis. It affects mainly the large bowel, and in severe disease, when the bowel is examined using a sigmoidoscope, sometimes pseudomembrane formation can be seen; hence the name pseudomembranous colitis.

C. difficile is a spore-forming, Gram-positive, anaerobic rod. It was given its name as it was initially found to be difficult to grow in culture and isolate. The spores are very hardy and can survive well in the ward environment. Hence, it can easily be transmitted by person-to-person contact within the hospital environment unless strict hand-washing and enteric precautions are used when dealing with infected patients. It is primarily a nosocomial infection and is thought to cause approximately three million cases of diarrhea or colitis each year in the United States.

Treatment of symptomatic *C. difficile* infection should begin with supportive measures and stopping the offending antimicrobial therapy if at all possible; if this is not possible, therapy should be changed to an antimicrobial not commonly implicated in causing *C. difficile* diarrhea. Specific therapy includes oral metronidazole or oral vancomycin. Metronidazole is usually given as first-line therapy because of the increasing incidence of vancomycin-resistant enterococci in hospitals and the worry that nonessential use of oral vancomycin might increase this incidence further. Very occasionally, the colitis caused by *C. difficile* does not respond to therapy and the patient must undergo a colectomy. Preventive measures to minimize the spread and acquisition of *C. difficile* diarrhea include the isolation of symptomatic patients to prevent

 Copyright 2004, Elsevier (USA). All rights reserved.

cross-infection and the judicious use of antibiotics such as broad-spectrum cephalosporins.

Other

In the past, there have been outbreaks of diarrhea in hospitals due to organisms such as *Salmonella* spp. and *Clostridium perfringens* resulting from the inadequate cooking or reheating of food. Salmonella infections may happen if inadequately cooked chicken or egg dishes are given to patients; if enteric precautions are not taken with infected patients, patient-to-patient transmission can then occur as well. Diarrheal illness due to *C. perfringens* is associated with poorly reheated meat dishes. These infections should not happen if food is properly prepared and stored.

Viral Infections

Small Round Structured Viruses

Small round structured viruses (SRSVs) are the most common cause of outbreaks of gastroenteritis in hospitals as well as in nursing homes, schools, hotels, etc. The norovirus (known as Norwalk virus until recently) is one of the most commonly implicated causes. Nosocomial outbreaks often lead to ward closures and can sometimes affect several wards in one hospital. This in turn leads to major disruption in the normal functioning of the hospital. Ward outbreaks tend to affect both patients and staff and sometimes have an attack rate in excess of 50%. This results in significant staffing problems and often staff aware of this situation try to return to work too early and while still infectious; this further compounds the problem. SRSVs may spread by a variety of routes including fecal–oral, vomit/aerosols, food, and water. Once the SRSVs have been introduced into the ward by any one of these routes, person-to-person spread then occurs. By the time an outbreak has been identified, most susceptible patients on the ward will have already been exposed to the virus. Hence, stringent attempts should be made to try and prevent spread to other wards and clinical areas by restricting patient and staff movements to other areas, enforcing good hand hygiene, and implementing effective environmental decontamination. It is important to prevent patients from becoming dehydrated but no specific therapy is available.

Rotavirus

Whereas SRSVs are the major cause of ward outbreaks of viral gastroenteritis in adults, rotaviruses are the most common cause of these outbreaks in pediatric wards. Outbreaks tend to occur in the winter months and the virus can spread within the ward environment very quickly. The virus is named for its characteristic wheel-like appearance when viewed under an electron microscope (see Fig. 1). The incubation period of rotavirus is approximately 2 days and the disease is characterized by vomiting and watery diarrhea for approximately 3–7 days. Diagnosis is by rapid antigen detection of rotavirus in the stools; electron microscopy of the stools, although less widely available, can also confirm the diagnosis. Prevention and minimization of spread are as for SRSVs, with the isolation and cohort nursing of infected patients. It is vital to ensure that the affected children are adequately rehydrated but no specific therapy is available.

SEPSIS RELATING TO SURGICAL PROCEDURES

Infections are a common complication after surgery on the gastrointestinal tract. They can range from mild superficial wound infections to a fatal septicemia resulting from generalized peritonitis. Many, but not all, of these infections can be prevented by appropriate chemoprophylaxis and/or optimal surgical technique. Patient factors and preoperative diagnosis are also

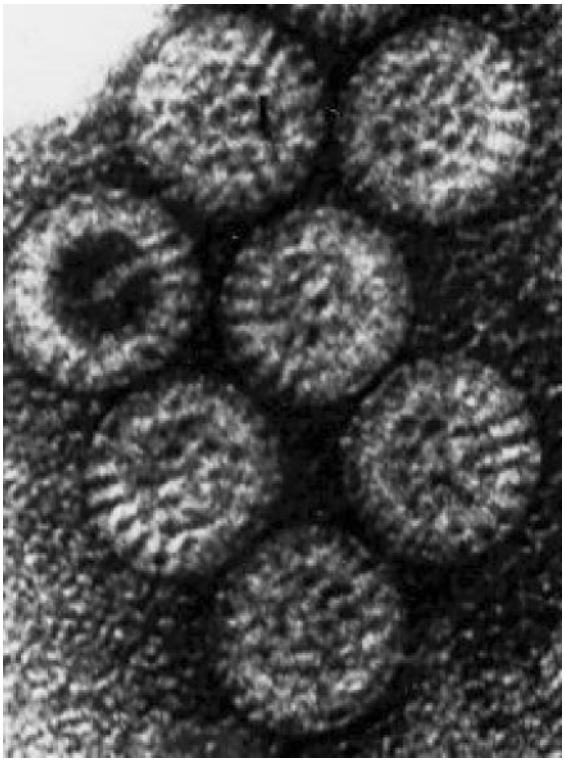


FIGURE 1 An electron micrograph of human rotavirus.

important factors in the likelihood of the patient developing a postoperative infection. Some examples of nosocomial infection post-GI surgery are given below.

Leaking of Anastomoses

Leaking of colonic or rectal anastomoses is more likely to occur than leaking of anastomoses in the small bowel. Sometimes anastomoses are noticed immediately postoperatively but often do not become apparent until several days later when bowel function has returned. A leaking anastomosis may lead to a generalized peritonitis, a local abscess, or fistula formation.

Abdominal Drains

Drains are often required postsurgery but should be removed as soon as they are no longer required because, like any foreign body in a person, they are a potential site of infection. The same applies to drains following biliary tract surgery.

Secondary Fungal Infections

Some patients who have had complicated abdominal surgery and who have presented with severe peritonitis require long courses of antimicrobials to get their bacterial sepsis under control. Unfortunately, this can dramatically change their normal flora, making them more susceptible to both local and systemic fungal infections.

Infection after Biliary Surgery

Operations in the presence of obstructive jaundice are not infrequently followed by wound infections most commonly due to *Escherichia coli*. Preoperative chemoprophylaxis will reduce infection rates, especially if the common bile duct is opened, if acute cholecystitis or obstructive jaundice is present, or if an intestinal anastomosis is performed. Cholangitis and septicemia due to coliforms sometimes occur after surgery for obstructive jaundice. The incidence of these conditions will also be reduced by appropriate preoperative chemoprophylaxis.

Septic Complications of Endoscopic Retrograde Cholangiopancreatography

Ascending cholangitis is a potential complication of endoscopic retrograde cholangiopancreatography (ERCP). Reported mortality rates in different series for post-ERCP cholangitis range between 10 and 16%. Ascending cholangitis results from bacterial infection of an obstructed biliary system, usually from enteric Gram-negative rods, resulting in bacteremia. Incomplete drainage of the biliary system after ERCP occurs in up to 10% of patients who require stenting. It has been suggested that appropriate early antibiotic therapy in this group of patients would probably reduce the frequency of cholangitis post-ERCP by 80%. Therapy should be continued until complete drainage has occurred. An antimicrobial such as ciprofloxacin, which achieves high levels in the biliary tract, should be started empirically but modification of therapy may be required following culture and sensitivity results of biliary fluid and blood, a local abscess, or a fistula.

Sepsis Postpancreatitis

One of the most important risk factors in patients suffering from acute necrotizing pancreatitis is pancreatic infection. The route of infection is probably via the colon and in patients with severe acute pancreatitis the infection rate is thought to be over 40% within the first 3 weeks. Not surprisingly, the bacteria most commonly implicated are those that reside in the gastrointestinal tract, such as *E. coli*, enterococci, and *Pseudomonas* spp. There is increasing evidence that patients with acute necrotizing pancreatitis benefit by starting antimicrobial therapy early, i.e., at initial presentation, as this has been found to decrease the infection rate at 3 weeks. Therapy should be broad spectrum; e.g., a carbapenem has often been recommended.

HEPATITIS

Hepatitis B and hepatitis C are both blood-borne viruses with the potential for nosocomial transmission in surgical or invasive procedures. To prevent this from occurring, all instruments involved in invasive techniques should be adequately sterilized and known carriers of the disease should be placed at the end of a list. All medical staff should be immunized against hepatitis B and in many countries hepatitis B carriers are not allowed to perform invasive procedures in order to prevent doctor-to-patient transmission.

NOSOCOMIAL INFECTIONS RESULTING FROM COLONIZATION OF OTHER AREAS OF THE BODY WITH ORGANISMS ARISING FROM THE GASTROINTESTINAL TRACT

Pneumonia

Pneumonia is the third most common nosocomial infection and is associated with substantial morbidity and mortality. It usually occurs in patients with severe

underlying disease, immunosuppression, or decreased conscious levels, due to either disease or anesthesia/analgesia, patients with underlying cardiac or respiratory disease, and those who have had thoracic or abdominal surgery. Nosocomial pneumonia usually occurs due to aspiration of bacteria colonizing the oropharynx or upper gastrointestinal tract of the patient. Both intubation and mechanical ventilation greatly increase the risk of developing nosocomial pneumonia as they breach the patient's first line of defense (the mucociliary escalator) against invasion of the respiratory tract by microorganisms. The risk of developing nosocomial pneumonia can be minimized by trying to decrease the risk of the patient aspirating, by preventing cross-infection of ventilator tubing, by good hygiene of medical personnel, and by adequate sterilization/disinfection of all devices associated with ventilating patients if they are not "single use." In some instances, there may be a place for reducing oropharyngeal and gastric colonization by potentially pathogenic microorganisms by a process known as "selective gut decontamination."

Sepsis Due to Resistant Organisms in the Bowel Flora

Patients who have received a lot of broad-spectrum antimicrobials for severe ongoing sepsis frequently become colonized with resistant organisms originating from the bowel flora, such as vancomycin-resistant enterococci and multi-resistant Gram negative rods, e.g., multi-resistant *Klebsiella* and *Acinetobacter* spp. Sometimes these colonizations can lead to bacteremias with these organisms, which can be difficult and expensive to treat. Hence, unnecessary antimicrobial treatment should be avoided.

See Also the Following Articles

Antibiotic-Associated Diarrhea • Anti-Diarrheal Drugs • Bacterial Toxins • Colitis, Pseudomembranous • Diarrhea • Hepatitis B • Hepatitis C • Rotavirus

Further Reading

Bergogne-Berezin, E. (2000). Treatment and prevention of antibiotic associated diarrhea. *Int. J. Antimicrob. Agents* **16**, 521–526.

Chadwick, P. R., Beards, G., Brown, D., Caul, E. O., Cheesbrough, J., Clarke, I., Curry, A., O'Brien, S., Quigley, K., Sellwood, J., and Westmorland, D. (2000). Management of hospital outbreaks of gastro-enteritis due to small round structured viruses *J. Hosp. Infect.* **45**, 1–10.

Hurley, B. W., and Nguyen, C. C. (2002). The spectrum of pseudomembranous enterocolitis and antibiotic-associated diarrhea. *Arch. Intern. Med.* **162**, 2177–2184.

Safdar, N., and Maki, D. G. (2002). The commonality of risk factors for nosocomial colonization and infection with antimicrobial-resistant *Staphylococcus aureus*, *Enterococcus*, gram-negative bacilli, *Clostridium difficile* and *Candida*. *Ann. Intern. Med.* **136**, 834–844.

Sganga, G. (2002). New perspectives in antibiotic prophylaxis for intra-abdominal surgery. *J. Hospital Infect.* **50**, S17–S21.

Subhani, J. M., Kibbler, C., and Dooley, J. S. (1999). Review article: Antibiotic prophylaxis for endoscopic retrograde cholangiopancreatography. *Aliment. Pharmacol. Ther.* **13**, 103–106.

NSAID-Induced Injury

BYRON CRYER

Department of Veteran's Affairs Medical Center and University of Texas Southwestern Medical Center, Dallas

cyclooxygenase Enzyme found throughout the body; responsible for the production of prostaglandins and thromboxane; inhibited by nonsteroidal antiinflammatory drugs.

diaphragm Specific type of stricture that develops in the small intestine or colon of chronic users of nonsteroidal antiinflammatory drugs.

leukotriene Fatty acid product of arachidonic acid that contributes to the proinflammatory response.

nonsteroidal antiinflammatory drugs Compounds that are therapeutically effective in reducing inflammation and treating pain and fever; cause injury to the gastrointestinal tract as their principal side effect.

prostaglandins Fatty acids produced by most of the body's cells; within the gastrointestinal tract, protect against injury.

thromboxane Product of cyclooxygenase metabolism in platelets; serum concentrations of thromboxane correlate with platelet activity.

Although nonsteroidal antiinflammatory drugs represent a very effective class of drugs, their use is associated with a broad spectrum of injury, especially in the gastrointestinal tract and kidney. Antiplatelet effects of nonsteroidal antiinflammatory drugs (aspirin particularly) are beneficial for cardiovascular disease prophylaxis, but excessive bleeding is an associated consequence. The upper gastrointestinal side effects constitute the greatest of the untoward effects of nonsteroidal antiinflammatory drugs.

INTRODUCTION

Nonsteroidal antiinflammatory drugs (NSAIDs) are prescribed more frequently than any other group of medicines. The 2002 edition of "The Physician's Desk Reference" lists 26 different NSAIDs, 6 of them salicylate-based compounds and 3 of them classified as cyclooxgenase-2 (COX-2)-specific inhibitors (Table I). A fourth COX-2-specific inhibitor, etoricoxib, is expected to be approved for use soon. All current NSAIDs are orally administered, with the exception of ketorolac, the only currently available parentally administered NSAID. It is anticipated that in the near future another parenterally administered NSAID, parecoxib (a COX-2-specific inhibitor) will be also available.

MECHANISMS OF GI TOXICITY

Irrespective of site of gastrointestinal (GI) damage, the mechanisms through which NSAIDs cause injury are similar throughout the GI tract. The general mechanisms can be grouped into two categories: (1) cyclooxygenase independent and (2) cyclooxygenase dependent. The former category consists largely of topical mucosal toxic processes.

Cyclooxygenase-Independent Toxicity (Topical Damage)

Within a few minutes of ingestion, NSAIDs accumulate at very high concentrations in gastric mucosal cells, causing denudation of surface epithelial cells and increased mucosal permeability to luminal contents. A second topical mechanism of NSAID injury is a reduction of the phospholipid content and surface hydrophobicity of the gastric mucus gel layer. Topical effects of NSAIDs are likely the major mechanism responsible for acute subepithelial hemorrhages and erosions that can be observed soon after NSAID challenge, although such lesions rarely cause clinical problems. Enteric-coated NSAIDs produce considerably less acute topical erosive and hemorrhagic injury compared to plain (non-enteric-coated) formulations during short-term administration (i.e., 1–2 weeks), an observation in support of a local toxic effect of NSAIDs. However, with long-term administration of enteric-coated formulations, gastric ulcers develop at rates that are not different than with non-enteric-coated preparations, presumably as a result the systemic mechanism of injury.

Cyclooxygenase-Dependent Toxicity

The beneficial effect of NSAIDs of decreasing systemic inflammation and their deleterious effects in the GI tract are both, in part, related inhibition of the enzyme cyclooxygenase. Within the GI tract, NSAID-associated reduction in gastroduodenal mucosal prostaglandin concentrations is the major contributor toward NSAID mucosal toxicity. Most NSAIDs (with the exception of cyclooxygnease-2-specific inhibitors)

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I List of NSAIDs Available by Prescription in the United States[a]

Nonsalicylates	Salicylates	COX-2 inhibitors
Diclofenac (Voltaren)[b]	Aspirin[c] (Zorprin, Easprin)	Celecoxib (Celebrex)
Etodolac (Lodine)	Diflunisal (Dolobid)	Rofecoxib (Vioxx)
Fenoprofen (Nalfon)	Salsalate (Disalcid, Salflex)	Valdecoxib (Bextra)
Flurbiprofen (Ansaid)	Choline salicylate (Trilisate)	Etoricoxib[d]
Ibuprofen (Motrin)[c]	Magnesium salicylate (Magan)	Parecoxib[e]
Indomethacin (Indocin)		COX-189[f]
Ketoprofen (Orudis)[c]		
Ketorolac (Toradol)[e]		
Mefenamic acid (Ponstel)		
Meloxicam (Mobic)		
Nabumetone (Relafen)		
Naproxen (Naprosyn, Anaprox)[c]		
Oxaprozin (Daypro)		
Piroxicam (Feldene)		
Sulindac (Clinoril)		
Tolmetin (Tolectin)		

[a] Data from "The Physician's Desk Reference" (2002), 53rd Ed., Montvale, New Jersey.
[b] Combination tablet of diclofenac and a synthetic prostaglandin E1 (misoprostol) is also available.
[c] Also available as over-the-counter preparations in the United States in 2002.
[d] Expected to be available soon.
[e] Parenterally administered.
[f] Orally administered COX-2 inhibitor in clinical development by Novartis (East Hanover, New Jersey).

inhibit GI cyclooxygenase, which reduces gastroduodenal mucosal prostaglandin concentrations, thus removing a major mechanism for protection against mucosal injury. Aspirin, by acetylation of cyclooxygenase, inhibits cyclooxygenase irreversibly, whereas all other NSAIDs inhibit cyclooxygenase in a reversible, concentration-dependent manner. With aspirin, when cyclooxygenase is irreversibly inhibited, the capacity for prostaglandin synthesis does not return to normal for several days, until new enzyme can be synthesized. A recent study has demonstrated that after low daily doses of aspirin, gastric prostaglandins do not fully recover for approximately 5–8 days; in the platelet, prostaglandin synthesis does not return to normal for 14 days. This may explain why aspirin, in comparison to the other NSAIDs, remains one of the most potent inhibitors of prostaglandin and thromboxane synthesis.

In the early 1990s, two structurally related cyclooxygenase isoforms were identified in mammalian cells, cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2). COX-1 is found in most body tissues, including the stomach and platelets. COX-2, by contrast, is the principal COX isoform that participates in inflammation. There is little COX-2 activity present in the normal stomach. This concept has led to development and clinical introduction of COX-2-specific NSAIDs. Recent animal data indicate that for gastric ulceration to occur, both COX-1 and COX-2 must be inhibited. Interestingly, in one model, selective inhibition of COX-1 alone did not cause gastric damage. Thus, the actual explanation for the COX-2 specific inhibitor association with improved GI toxicity may more closely relate to the lack of dual COX isoform inhibition rather than simply the COX-1 sparing effects.

ADVERSE GI EFFECTS OF NSAIDS

NSAID-induced adverse GI effects can have variable presentations, including GI symptoms such as nausea and upper abdominal pain (dyspepsia), asymptomatic mucosal lesions that can be visualized through an endoscope, and, most importantly, ulcers, which are sometimes accompanied by serious GI complications. Serious upper GI complications are most relevant to the clinical morbidity of NSAIDs. Symptomatic GI ulceration (that is, ulcers associated with pain, perforation, bleeding, or obstruction) occurs in approximately 2–4% of patients treated with a NSAID for 1 year.

Low-dose aspirin increases risks of GI bleeding and increases the likelihood of hospitalization for ulcers. Aspirin doses as low as 75 or 300 mg/day have been associated with a two- to fourfold increased risk of GI bleeding. Buffered or enteric-coated aspirin preparations, although probably associated with a reduced incidence

of dyspepsia when compared to plain aspirin, have risks of upper GI bleeding that are similar to those of plain aspirin. There is probably no orally administered dose of aspirin that is both efficacious for cardiovascular disease prophylaxis and is free of GI toxicity.

RISK GROUPS FOR NSAID-INDUCED GI ULCERS

Certain groups of patients taking NSAIDs are at greater risk for development of NSAID ulcer complications (Table II) and should be given greater consideration for strategies to prevent or to reduce ulceration.

THERAPY FOR NSAID-INDUCED ULCERS

Prior to the clinical availability of safer classes of NSAIDs, reduction in NSAID-induced GI toxicity was primarily accomplished by prescribing drugs that, when coadministered with NSAIDs, would protect against mucosal ulceration. Because the majority of patients who chronically take NSAIDs will never develop clinically significant ulceration, the ideal candidates for cotherapy are those considered as high risk for NSAID-induced ulcers (Table II). Various cotherapies have shown to be effective. Misoprostol, a synthetic prostaglandin E1 (PGE1) analogue, reduces NSAID-induced serious GI adverse events. A combination tablet of misoprostol and the NSAID diclofenac has a relatively low ulceration rate.

There are no placebo-controlled studies evaluating whether any proton pump inhibitors (PPIs), including omeprazole, lansoprazole, rabeprazole, pantoprazole, and esomeprazole, can prevent the complications of NSAID-induced ulcers.

TABLE II Definite and Possible Risk Factors for NSAID-Induced Ulcers[a]

Definite	Possible
Prior peptic ulcer disease	*Helicobacter pylori* infection
Prior GI ulcer complication	Cigarette smoking
Advanced age[b]	
Concomitant use of corticosteroids	
Concomitant use of anticoagulants	
Concomitant use of ethanol	
Comorbid diseases (e.g., heart disease, rheumatoid arthritis)	

[a] Risk is also related to dose of NSAID used and to number of NSAIDs used.

[b] Participants in some studies, >65 years old; in others, >75 years old.

When attempting to treat an ulcer that has formed during NSAID use, the first step is always to stop the NSAID. Once the NSAID is stopped, rapid ulcer healing can be achieved by treatment with standard doses of histamine-2 receptor antagonists (H2RAs) or PPIs. With patients for whom NSAIDs cannot be discontinued, use of a PPI will allow ulcer healing, even while NSAID use continues.

NSAIDS WITH IMPROVED GI SAFETY PROFILES

COX-2 Specific Inhibitors

Treatment with rofecoxib or celecoxib is associated with significantly lower incidences of serious upper GI adverse clinical events than treatment with comparator NSAIDs. Two trials of ~8000 arthritis patients each reported a reduction of ~50% in upper GI events with celecoxib and rofecoxib when compared to nonselective NSAIDs. Concurrent use of low-dose aspirin eliminated the GI protective benefit of the COX-2 inhibitor celecoxib.

Older and Safer NSAIDs

In addition to the recently marketed COX-2-specific NSAIDs, several other established NSAIDs (or products in development) have safety evaluations superior to those of conventional NSAIDs. These include etodolac, nabumetone, diclofenac, and nonacetylated salicylates such as salsalate. Salsalate and etodolac have no measurable effects on gastric COX activity.

Nitric Oxide-Releasing NSAIDs

Nitric oxide is a critical mediator of GI mucosal defense, exerting many of the same beneficial actions as prostaglandins within the GI tract. In addition to other properties, nitric oxide increases mucosal blood flow and prevents neutrophil adherence to the vascular endothelium. These observations have led to the development of nitric oxide (NO)-releasing NSAIDs in which the native NSAID has been coupled to a nitric-oxide-releasing moiety. However, NO—NSAIDs are associated with less NSAID-induced gastric and intestinal toxicity than has been associated with the parent compounds. Such drugs are currently in clinical development.

Phospholipid NSAIDs

NSAIDs reduce hydrophobicity of the gastric mucosa by chemically associating with and destabilizing phospholipids (in particular, phosphatidylcholine) within the mucus gel layer. Newer NSAIDs have been

developed in which the native NSAID moiety has been coupled with synthetic phosphatidylcholine (PC). Short-term courses of PC–aspirin are associated with less acute gastric injury when compared with plain aspirin. Should subsequent clinical studies show continued safety and efficacy, PC–NSAIDs may be another future class of safer NSAIDs available to clinicians.

SMALL INTESTINAL AND COLONIC EFFECTS OF NSAIDS

Small Intestine

Ulcers (which may bleed or perforate), strictures (which may obstruct), and an enteropathy have all been described in the small intestines of chronic NSAID users. The pathology of strictures ranges from nonspecific broad-based strictures to intestinal diaphragms. These diaphragms are usually located in the jejunum. Their histology reveals submucosal fibrosis with normal overlying epithelium, except for the central tip of the diaphragm, which contains acute and chronic inflammatory cells. NSAIDs can cause diffuse intestinal inflammation and increased intestinal mucosal permeability, a condition known as "NSAID enteropathy"; this is characterized clinically by occult blood loss, iron deficiency anemia, malabsorption, and a protein-losing enteropathy.

Colon

NSAIDs may also cause ulcers, strictures, and diaphragms, and inflammation in the colon, especially on the right side. Ulceration may be complicated by bleeding or perforation, especially ulcers occurring in the cecum. Cases of eosinophilic, collagenous, pseudomembranous, and nonspecific colitis have been associated with NSAIDs. A disproportionately high number of colitis cases have occurred in patients taking mefenamic and flufenamic acid. Histology usually reveals a nonspecific mild colitis unless one of the other variants is present (eosinophilic, pseudomembranous, or collagenous). Rectal administration of NSAIDs in the form of suppositories has been frequently associated with inflammation, ulcers, and strictures of the anus and rectum.

There are individual case reports and small studies suggesting that NSAIDs are the cause of colonic diverticular perforation and diverticular bleeding. Many cases of diverticular perforation involve a sustained-release preparation of indomethacin, with the perforations attributed to a mechanical effect of the capsule being caught in a diverticulum.

In patients with inflammatory bowel disease (IBD), NSAIDs may exacerbate disease or activate IBD that has been previously quiescent. These occurrences are more frequently observed in ulcerative colitis than in Crohn's disease. One postulated mechanism through which NSAIDs may exacerbate colonic inflammation is by inhibiting cyclooxygenase and shunting arachidonic acid metabolism toward the proinflammatory leukotrienes. Of all patients with IBD, only a subset experience symptomatic flares when placed on a NSAID. These patients, when they relapse, will do so within a few days of receiving a NSAID. It is unknown whether the COX-2-specific inhibitors, compared with the older NSAIDs, are associated with fewer or greater intestinal problems in patients with IBD.

See Also the Following Articles

Arachidonic Acid • Colitis, Collagenous and Lymphocytic • Colitis, Pseudomembranous • Colitis, Ulcerative • Crohn's Disease • Duodenal Ulcer • Gastric Ulcer • Marginal Ulcer • Nitric Oxide • Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) • Rectal Ulcers

Further Reading

Bombardier, C., Laine, L., Reicin, A., *et al.* (2000). Rofecoxib and naproxen in patients with rheumatoid arthritis. *N. Engl. J. Med.* **343**, 1520–1528.

Chan, F. K. L., To, K. F., Wu, J. C. Y., Zakar, T., Leung, W. K., Kwok, T., Hui, Y., Chan, H. L. Y., Chan, F. K. L., Hui, E., Woo, J., and Sung, J. J. Y. (2002). Eradication of *Helicobacter pylori* and risk of peptic ulcers in patients starting long-term treatment with non-steroidal anti-inflammatory drugs: A randomised trial. *Lancet* **359**, 9–13.

Feldman, M., Shewmake, K., and Cryer, B. (2000). Time course inhibition of gastric and platelet COX activity by acetylsalicylic acid in humans. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G1113–G1120.

Silverstein, F. E., Faich, G., Goldstein, J. L., *et al.* (2000). Gastrointestinal toxicity with celecoxib vs. nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: The CLASS study: A randomized controlled trial. *JAMA* **284**, 1247–1255.

Wright, N. J., Gottesdiener, K., Garlick, N. M., Atherton, C. T., Novak, S., Gertz, B., Calder, N. A., Cote, J., Wong, P., Dallob, A., and Hawkey, C. J. (2001). Rofecoxib, a COX-2 inhibitor, does not inhibit gastric mucosal prostaglandin production. *Gastroenterology* **120**, 867–873.

Nuclear Medicine Imaging

MOHAN R. RAMASWAMY AND RANDALL A. HAWKINS
University of California, San Francisco

gamma camera A device capable of detecting and localizing gamma rays, forming an image of the underlying radiopharmaceutical distribution within a patient. Most gamma cameras are capable of producing both planar and tomographic images.

positron emission tomography A technique that utilizes a system especially designed to produce tomographic images of positron-emitting radionuclides (such as ^{18}F and ^{11}C).

radionuclide An atom with an unstable nucleus, which achieves stability by emitting excess energy in the form of gamma rays and subatomic particles (e.g., positrons). The radionuclide most commonly utilized in nuclear medicine is technetium-99m.

radiopharmaceutical A biologically active molecule labeled with a radionuclide.

scintigraphy The process of obtaining images with a gamma camera.

Nuclear imaging techniques rely on a spectrum of radiopharmaceuticals, which are biologically active molecules or cells labeled with radioactivity. By following these radioactive tracers as they are taken up, metabolized, and excreted by various organs and tissues, one is able to obtain a qualitative (and sometimes quantitative) assessment of a wide array of functional disorders of the gastrointestinal tract. Such functional information is often not available from the more conventional anatomic imaging techniques, such as radiography, computed tomography, magnetic resonance imaging, and ultrasound.

GASTROINTESTINAL TRACT SCINTIGRAPHY

Gastrointestinal Bleeding

Using *in vitro* labeling methodology, technetium-99m-pertechnetate ($^{99m}TcO_4^-$) can be introduced into red blood cells, where it binds to the β-subunit of hemoglobin. After withdrawal of a small blood sample, such labeling can be performed within 30 min. This blood is then reinjected into the patient and scintigraphic imaging of the abdomen and pelvis is obtained for a period of 60 to 90 min. Typically, the images are acquired digitally in 1 min frames, which can then be displayed dynamically on a computer screen in a cinematic loop.

Normally, there is uniform distribution of activity within the largest reservoirs of blood: the liver, spleen, and great vessels. Any unbound ^{99m}Tc-pertechnetate is excreted via the kidneys into the bladder. Any other focus of activity developing during the course of the study in the expected location of bowel is usually indicative of an active gastrointestinal bleed (Fig. 1). Due to its irritative effect, such intralumenal blood produces peristalsis and is rapidly conveyed along the gastrointestinal tract. By imaging the transit of activity over time, it is generally possible to determine the site of origin.

To be most useful, labeled erythrocyte studies must be performed at a time when the patient is suspected to be actively bleeding. In this situation, the study is quite sensitive, detecting bleeding rates as low as 0.05 to 0.1 ml/min. When performed properly, scintigraphy will serve to both confirm the presence of active gastrointestinal bleeding and establish the site of origin. This information can be quite useful in guiding subsequent angiographic, endoscopic, or surgical therapy.

Ectopic Gastric Mucosa

When administered in its unbound form, ^{99m}Tc-pertechnetate has a strong affinity for gastric mucosa, both within the stomach and in ectopic sites. This test can be used to localize gastric mucosa within a bleeding Meckel's diverticulum, an enteric duplication, Barrett's esophagus, and a retained gastric antrum after partial gastrectomy. It is reported that at least 2 cm^2 of mucosa must be present for reliable scintigraphic detection. Pretreatment with pharmacologic agents such as cimetidine, pentagastrin, and glucagon may help improve visualization.

Gastric Emptying

A solid meal (e.g., fried eggs) may be easily labeled by cooking with ^{99m}Tc-sulfur colloid. Similarly, a liquid meal (e.g., water or juice) may be labeled by mixing with ^{99m}Tc-sulfur colloid or ^{111}In-DTPA. Images of the upper gastrointestinal tract obtained over a period of 1 to 2 h following oral administration of such a meal

 Copyright 2004, Elsevier (USA). All rights reserved.

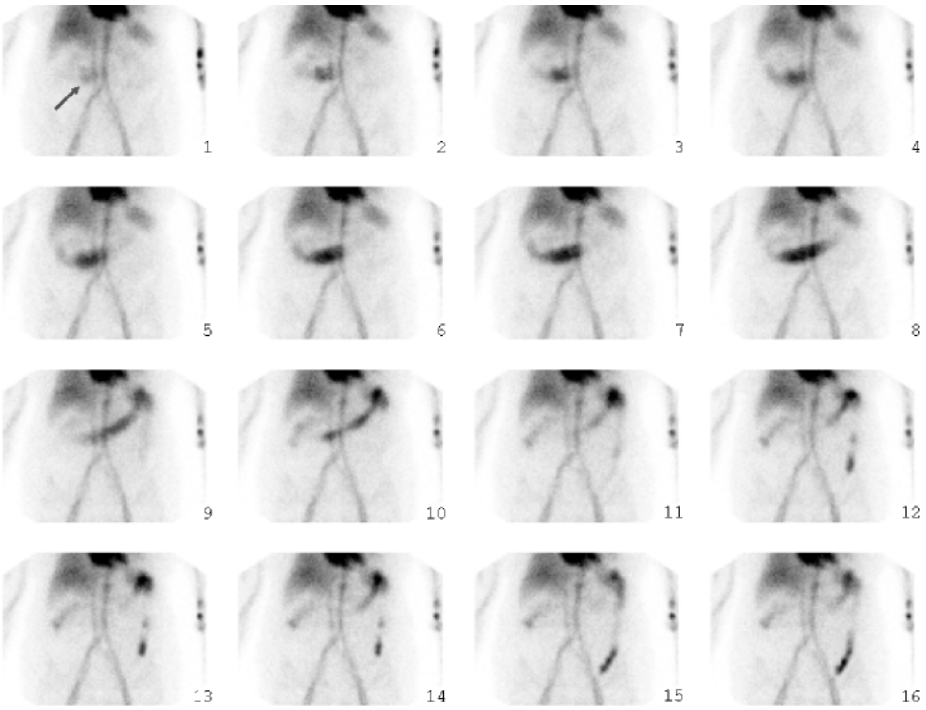


FIGURE 1 Sixteen 1 min frames from a ^{99m}Tc-labeled erythrocyte study performed in an 84-year-old woman presenting with hematochezia. The early images demonstrate a focus of radioactivity developing in the right upper quadrant, roughly in the region of the proximal transverse colon (arrow). Later images demonstrate peristalsis of activity through the distal transverse and descending colon, confirming the site of origin (correlated by subsequent angiography). Note the normal activity in the heart, liver, spleen, and great vessels.

allow for qualitative assessment of solid and/or liquid gastric emptying.

By drawing a region of interest around the stomach, it is possible to quantitate the amount of activity within the stomach. From this value, a quantitative estimate of gastric emptying can be obtained, usually expressed in the form of an emptying half-time (in minutes) or an emptying rate (expressed as the percentage per minute). The rate of gastric emptying may vary greatly with the type of meal (i.e., its quantity, consistency, and nutritional content), the imaging technique, and the method of quantitative analysis. It is therefore imperative for each laboratory to rigorously standardize its imaging protocol and to establish appropriate normal values specific to that protocol.

Gastric emptying studies are most often utilized in assessing the severity of diabetic gastroparesis (Fig. 2), although they can be useful in a wide range of functional gastroparetic disorders. They may also be used to detect rapid gastric emptying, as may be seen in dumping syndromes following upper gastrointestinal tract surgery, and to assess the response to medical or surgical therapies of gastroparesis.

Other Studies

Scintigraphic techniques exist for the detection of gastroesophageal reflux and aspiration and the

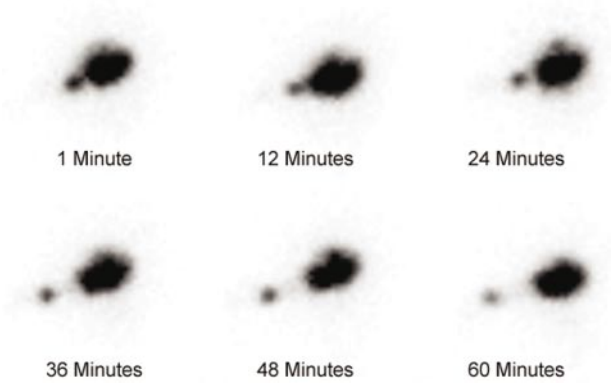


FIGURE 2 Selected images from a ^{99m}Tc-sulfur colloid solid gastric emptying study performed in a 43-year-old woman with a long history of diabetes and early satiety. Very little emptying of the radiolabeled scrambled egg meal is identified within the first 60 min of the study, consistent with the suspected diagnosis of diabetic gastroparesis.

assessment of esophageal, small bowel, and colonic motility. However, these studies are not well standardized and are not widely available.

HEPATOBILIARY SCINTIGRAPHY

Technique

The primary radiopharmaceuticals utilized in hepatobiliary scintigraphy are the ^{99m}Tc-labeled iminodiacetic acid (IDA) derivatives, disofenin and mebrofenin. Following intravenous administration, these agents are taken up by the hepatocytes through the same membrane transport mechanism as bilirubin and, like bilirubin, are then promptly secreted into the bile (without conjugation). This radiolabeled bile then fills the gallbladder through the cystic duct and empties into the duodenum through the common duct.

Immediately following the intravenous administration of one of the ^{99m}Tc-IDA agents, consecutive 1 min scintigraphic images are acquired over the abdomen for a period of 60 to 90 min. Typically, the images are acquired digitally and then displayed dynamically on a computer screen in a cinematic loop. In a normal study, there is prompt extraction of activity from the blood by the hepatocytes in the liver. Secreted activity is identified within the major bile ducts within 10 min, with filling of the gallbladder and transit into the small bowel within 60 min after injection.

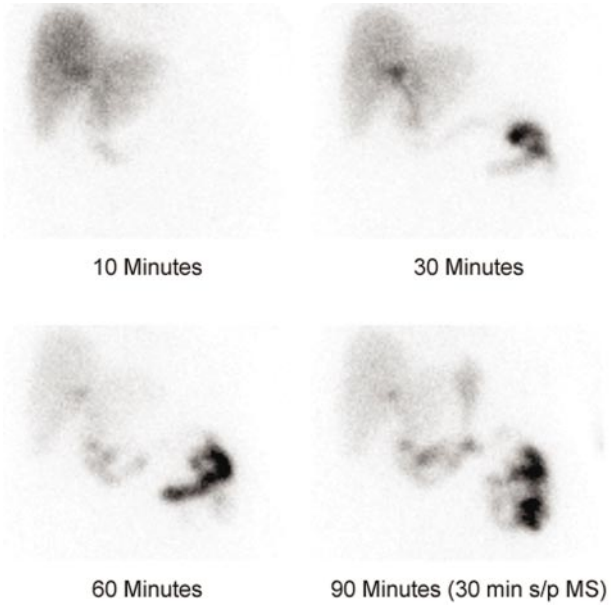


FIGURE 3 Selected images from a ^{99m}Tc-IDA study performed in a 76-year-old woman with right upper quadrant pain and an equivocal ultrasound evaluation. Early images demonstrate uptake of tracer by the liver, with subsequent excretion into the biliary tree and small bowel. Note the lack of gallbladder uptake, even following the intravenous administration of morphine sulfate. Acute gangrenous cholecystitis was found at surgery.

Acute Cholecystitis

Filling of the gallbladder indicates patency of the cystic duct and virtually excludes the possibility of acute cholecystitis. Failure of the gallbladder to fill within 4 h is usually indicative of acute cholecystitis. Morphine sulfate's ability to contract the sphincter of Oddi can be used to shorten the time required to make the diagnosis. Specifically, lack of gallbladder filling within 30 min following the intravenous administration of 0.04 mg/kg of morphine is considered indicative of acute cholecystitis (Fig. 3).

Many studies have found hepatobiliary scintigraphy to have both a sensitivity and a specificity of greater than 90% for the detection of acute cholecystitis. An important cause of false-positive studies is fasting for less than 4 h prior to the study, as a recent meal induces gallbladder contraction and thereby prevents filling. On the other hand, fasting for more than 24 h leads to a sludge-filled gallbladder, which is also unlikely to fill. Other important causes of a false-positive study include poor liver function and severe chronic cholecystitis. False-negative studies are unusual, but may occasionally be seen in the setting of acute acalculous cholecystitis. In this setting, the use of labeled leukocyte imaging may be considered (discussed below).

Chronic Cholecystitis

Fibrotic changes associated with chronic inflammation of the gallbladder can lead to diminished contractile ability, eventually resulting in pain. Contractility of the gallbladder may be assessed as part of a hepatobiliary study by comparing the amount of radioactivity in the gallbladder before and after the induction of gallbladder contraction. This is usually expressed quantitatively, in the form of an ejection fraction:

$$\frac{\left[\begin{array}{l}\text{Radioactive counts in gallbladder before contraction} \\ -\text{Radioactive counts in gallbladder after contraction}\end{array}\right]}{\left[\text{Radioactive counts in gallbladder before contraction.}\right]}$$

Gallbladder contraction is induced physiologically by duodenal secretion of the hormone cholecystokinin. This may be simulated by administration of the active terminal octapeptide of cholecystokinin, sincalide (Kinevac, Bracco Diagnostics, Princeton, NJ), administered intravenously over several minutes at a dose of

0.02 μg/kg. In properly selected patients with recurrent right upper quadrant pain, a gallbladder ejection fraction below 35% provides evidence of chronic cholecystitis or biliary dyskinesia. Furthermore, such patients often demonstrate symptomatic improvement following elective cholecystectomy.

Other Applications

In the jaundiced newborn, hepatobiliary scintigraphy plays an important role in the differentiation of biliary atresia from neonatal hepatitis. Specifically, normal transit of activity from the liver into the small bowel virtually excludes the diagnosis of biliary atresia. Another interesting role is in the differentiation of hepatic masses. Focal nodular hyperplasia tends to demonstrate prompt radiotracer uptake and delayed secretion, whereas many hepatocellular carcinomas tend to demonstrate both delayed uptake and delayed secretion. Most other hepatic tumors fail to demonstrate any radiotracer uptake (although some uptake may be seen in adenomas). Hepatobiliary scintigraphy may also be used to detect enterogastric reflux of bile and sphincter of Oddi dysfunction and may serve as an adjunct to other imaging modalities in the evaluation of biliary obstruction, biliary leaks, and choledochal cysts.

LIVER AND SPLEEN SCINTIGRAPHY

Sulfur Colloid Imaging

^{99m}Tc-sulfur colloid is a particulate radiopharmaceutical, with particle diameters on the order of 100 to 1000 nm. Following intravenous administration, these particles are rapidly phagocytized by the Kupffer cells of the liver (85%) and the macrophages of the spleen (10%) and bone marrow (5%). Although ^{99m}Tc-sulfur colloid imaging of the liver has largely been supplanted by cross-sectional imaging, a small role still exists. A characteristic pattern may be identified in the setting of alcoholic cirrhosis, consisting of a decrease in hepatic uptake with a corresponding increase in splenic and bone marrow uptake (the so-called "colloid shift"). ^{99m}Tc-sulfur colloid can also play a role in the differentiation of hepatic tumors, in that most regenerating nodules, two-thirds of focal nodular hyperplasia, and a small number of adenomas maintain enough Kupffer cell activity to take up the tracer, whereas most other hepatic lesions usually do not.

Hemangioma Imaging

^{99m}Tc-pertechnetate-labeled erythrocytes may be used to differentiate cavernous hemangiomas from other liver masses. This test makes use of the classic blood flow pattern of such hemangiomas, which initially appear as cold defects on images obtained immediately after injection, progressively filling-in to an intensity greater than that of the liver background on subsequent delayed images obtained at 1 to 2 h postinjection. Lesions over 2 to 3 cm in size can be detected with an accuracy of greater than 95%. Somewhat smaller lesions may be detected with high-quality tomographic techniques.

Damaged Erythrocyte Imaging

Although ^{99m}Tc-sulfur colloid may be used for splenic imaging, heat-damaged or chemically damaged ^{99m}Tc-pertechnetate-labeled erythrocytes demonstrate a much stronger affinity for the spleen. This latter imaging method obviously takes advantage of the spleen's role in removing damaged erythrocytes from the blood. Radionuclide imaging of the spleen is most often utilized in detecting recurrent splenic tissue in patients with continuing laboratory abnormalities after splenectomy for thrombocytopenia. It is also useful in differentiating a left upper quadrant mass from an accessory spleen, establishing the diagnosis of splenosis, and evaluating congenital asplenia/polysplenia syndromes.

Hepatic Arterial Infusion Pump Imaging

^{99m}Tc-macroaggregated albumin is a particulate radiopharmaceutical, with particle diameters on the order of 10 to 90 μm (compared to a capillary diameter on the order of 5 μm). When administered through a hepatic arterial chemotherapy infusion pump, these particles will flow through the arterial system until they lodge in the first capillary bed they encounter. Thus, subsequent imaging will demonstrate the arterial system subtended by the pump catheter. This study is often performed after pump placement, to ensure that the entire liver is perfused and that there is no extrahepatic perfusion or significant systemic arteriovenous shunting. Follow-up imaging is often obtained if the patient develops symptoms suggestive of pump occlusion or migration.

INFECTION AND INFLAMMATION SCINTIGRAPHY

Labeled Leukocytes

Using *in vitro* labeling methodology, a mixed population of a patient's white blood cells can be labeled with ^{111}In-oxine or ^{99m}Tc-HMPAO. Following reinjection of the labeled blood sample, the leukocytes are

drawn to areas of active infection and inflammation through the usual chemotactic mechanisms. ^{111}In-labeled leukocytes are typically imaged 18 to 24 h after injection, with normal distribution in the spleen, liver, and bone marrow. Because of better imaging characteristics, ^{99m}Tc-labeled leukocytes provide greater image quality and allow for earlier imaging (at 4 h). Unfortunately, ^{99m}Tc-labeled leukocytes also undergo normal biliary and bowel clearance, somewhat limiting their use in the abdomen.

Labeled leukocytes are not usually the first agent of choice in the imaging of intra-abdominal infection, due to the expense and long imaging delays involved. However, they may be used to detect intra-abdominal abscesses when computed tomography (CT) imaging is unrevealing or equivocal and to verify the presence of cholecystitis when ultrasound and hepatobiliary scintigraphy are equivocal. Other infections that may be investigated with this technique include aortic graft infections, peritonitis, and diverticulitis.

Labeled leukocytes may also play an important role in inflammatory bowel disease, by assessing its extent and severity, by detecting complications (such as abscess formation), and by determining response to therapy. It is possible for leukocytes to be shed into the bowel lumen at sites of active bowel inflammation. These leukocytes may then undergo peristalsis to other parts of the bowel, confounding image interpretation. It is therefore important to perform early imaging in this setting—at 4 h for ^{111}In-labeled leukocytes and at 1 to 2 h for ^{99m}Tc-labeled leukocytes.

Gallium

^{67}Gallium-citrate is an analogue of iron, which tends to accumulate at sites of infection and inflammation through a variety of complex mechanisms. One mechanism appears to involve binding to lactoferrin, an acute-phase reactant released by activated neutrophils. Typically imaged at 48 h, the widespread normal distribution of gallium (which includes the liver, spleen, bowel, bone, and bone marrow) tends to limit its use in the abdomen. However, gallium may have a higher sensitivity for chronic infections (i.e., greater than 1 or 2 weeks old) and may be preferable in this situation.

TUMOR SCINTIGRAPHY

[^{18}F]Fluorodeoxyglucose Positron Emission Tomography Imaging

Like glucose, the radiolabeled glucose analogue [^{18}F]fluorodeoxyglucose (FDG) is taken up in cells through the glucose membrane and undergoes the first step of glycolysis (phosphorylation). However, phosphorylated FDG is unable to pass through the remainder of the glycolytic cycle and remains trapped within the cytoplasm of the cell. Because of the increased glycolysis needed to support rapid mitosis, most malignant tumors appear as "hot spots" on FDG positron emission tomography (PET) imaging.

The current indications for FDG PET imaging supported by Medicare in the gastrointestinal tract are the staging and restaging of esophageal and colorectal carcinoma. In terms of initial staging, PET is limited in its ability to detect the depth of local invasion or the involvement of locoregional lymph nodes. However, it is quite powerful in the detection of distant metastases, with an accuracy of approximately 86% in esophageal carcinoma and 92% in colorectal cancer. This is superior to CT, with accuracies of approximately 62 and 71%, respectively.

FDG PET is also quite useful in detecting suspected residual or recurrent tumor after the completion of therapy. In one recent study, PET identified 79% of recurrences (both local and distant) in patients with an elevated carcinoembryonic antigen (CEA) level after therapy for colorectal carcinoma (Fig. 4).

An exciting new area of investigation in PET is the prediction of tumor response to chemotherapy. By comparing the decline in FDG uptake within a tumor on studies performed before chemotherapy and again following one or two cycles, it may be possible to predict the response to the complete course of chemotherapy. This would allow for midcourse adjustments in the therapeutic regimen for nonresponders.

FDG PET also appears to be quite promising in the imaging of pancreatic adenocarcinoma and gastrointestinal stromal tumors. Results have been mixed in the setting of hepatocellular carcinoma, however, with some studies identifying only 55% of tumors. The development of new PET tracers, such as [^{11}C]choline for potential imaging of primary and metastatic prostate carcinoma, will further increase the power of PET in oncologic imaging.

Other Agents

Radiopharmaceuticals such as ^{67}Ga-citrate, ^{201}Tl-chloride, ^{99m}Tc-sestamibi, and ^{99m}Tc-tetrofosmin tend to accumulate in tumors through a variety of mechanisms. As such, they may be used as nonspecific markers for a variety of tumors. However, the extensive normal physiologic uptake of these tracers in the bowel and/or biliary system tends to limit their use in the abdomen. Nonetheless, they may be of value when PET imaging is unavailable. In-DTPA-Pentreotide (OctreoScan,

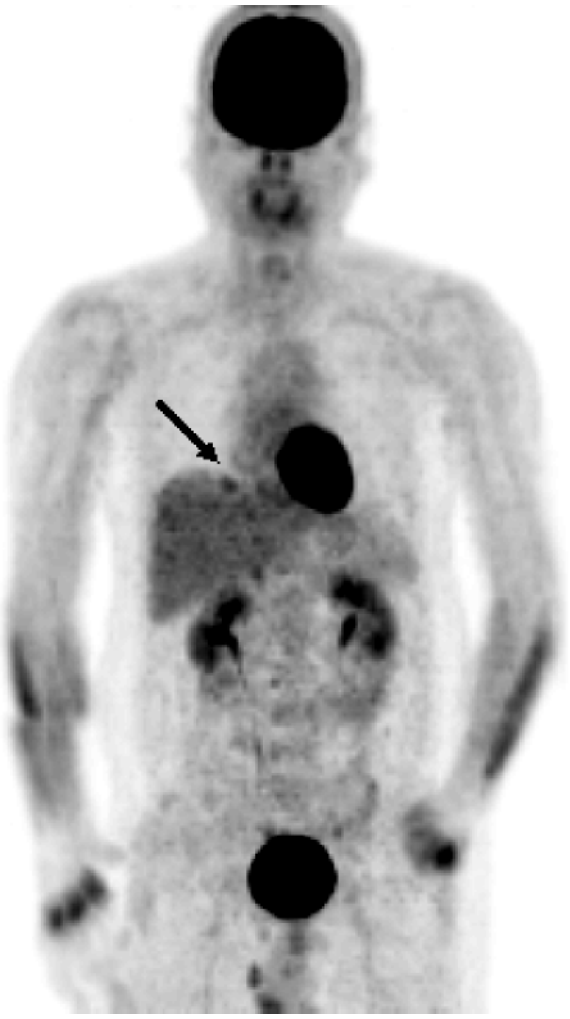


FIGURE 4 Image from a FDG PET study performed in a 51-year-old man with a rising CEA value approximately 1 year following surgery and chemotherapy for colorectal adenocarcinoma. Note the prominent focus of radiotracer uptake at the dome of the liver (arrow), subsequently found at surgery to represent a metastasis.

Mallinckrodt, St. Louis, MO) is a radiolabeled analogue of somatastatin. Typically imaged at 24 h, its normal distribution includes the pituitary, thyroid, liver, spleen, bowel, kidneys, and bladder. This agent has a strong affinity for a variety of somatostatin receptor-bearing tumors, including pancreatic islet cell tumors (with the possible exception of insulinoma) and carcinoid tumors. In this specific setting, OctreoScan has been shown to be as effective as all other imaging modalities combined.

There has been a considerable amount of excitement surrounding the use of radiolabeled antibodies in oncologic imaging. Two agents that are currently commercially available are OncoScint, an ^{111}In-labeled antibody to the antigen TAG-72, and CEA-Scan, a ^{99m}Tc-labeled antibody to the antigen CEA. Both are useful in the imaging of colorectal carcinoma, although FDG PET imaging has largely supplanted their use at the authors' institution.

MISCELLANEOUS SCINTIGRAPHIC STUDIES

Peritoneal Imaging

Direct introduction of a ^{99m}Tc-labeled tracer into the peritoneal cavity allows the assessment of peritoneal flow. This is most useful in detecting the presence of a pleuro-peritoneal leak or in establishing the patency of a peritoneovenous (i.e., LaVeen or Denver) shunt.

Salivary Gland Imaging

In addition to its role as a gastric mucosal agent, ^{99m}Tc-pertechnetate also has an affinity for the salivary glands. By assessing the uptake of this radiopharmaceutical, one can estimate the relative function of the major (parotid and submandibular) salivary glands. Dynamic imaging performed after the administration of a sialagogue (e.g., lemon juice) allows estimation of an ejection fraction for each gland.

CONCLUSION

This has been just a brief overview of the wide range of nuclear techniques available for structural and functional assessment of the gastrointestinal system. The constant development of new radiotracers and imaging techniques will further expand the applications of nuclear imaging in the field of gastroenterology.

See Also the Following Articles

Barium Radiography • Computed Tomography (CT) • Magnetic Resonanace Imaging (MRI) • Radiology, Interventional • Ultrasonography

Further Reading

Alazraki, N. P. (1993). Radionuclide imaging in the evaluation of infections and inflammatory disease. *Radiol. Clin. North Am.* **31**, 783–794.

Drane, W. E. (1991). Nuclear medicine techniques for the liver and biliary system: Update for the 1990s. *Radiol. Clin. North Am.* **29**, 1129–1150.

Hawkins, R. A. (1989). Radionuclide studies of the liver and hepatobiliary system. *Curr. Opin. Radiol.* **1**, 499–507.

Kwekkeboom, D. J., and Krenning, E. P. (2002). Somatostatin receptor imaging. *Semin. Nuclear Med.* **2**, 84–91.

Moadel, R. M., Blaufox, M. D., and Freeman, L. M. (2002). The role of positron emission tomography in gastrointestinal imaging. *Gastroenterol. Clin. North Am.* **31**, 841–861.
Murray, I. P. C., and Ell, P. J. (eds.) (1998). "Nuclear Medicine in Clinical Diagnosis and Treatment," 2nd Ed. Churchill Livingstone, Edinburgh, UK.
O'Hara, S. M. (1996). Pediatric gastrointestinal nuclear imaging. *Radiol. Clin. North Am.* **34**, 845–862.
Pham, K. H., Ramaswamy, M. R., and Hawkins, R. A. (2002). Advances in positron emission tomography imaging for the GI tract. *Gastrointest. Endosc.* **55**, 553–563.
Sandler, M. P., Coleman, R. E., Patton, J. A., Wackers, F. J. T., and Gottschalk, A. (eds.) (2002). "Diagnostic Nuclear Medicine," 4th Ed. Lippincott Williams & Wilkins, Philadelphia, PA.
Suzman, M. S., Talmor, M., Jennis, R., Binkert, B., and Barie, P. S. (1996). Accurate localization and surgical management of active lower gastrointestinal hemorrhage with technetium-labeled erythrocyte scintigraphy. *Ann. Surg.* **224**, 29–36.
Thrall, J. H., and Zeissman, H. A. (eds.) (2001). "Nuclear Medicine: The Requisites," 2nd Ed. Mosby, St. Louis, MO.
Weldon, M. J. (1994). 99mTc-HMPAO planar white cell scanning. *Scand. J. Gastroenterol.* **29**, 36–42.

Nutrient Transport, Regulation of

RONALDO P. FERRARIS
University of Medicine and Dentistry of New Jersey, Newark

borborygmi Rumbling noise produced by movement of gas in the intestine.
electrochemical gradient Driving force underlying the movement of charged compounds (or nonelectrolytes cotransported with charged compounds) across a membrane.
hematochromatosis Disorder of iron metabolism characterized by excessive absorption of ingested iron.
micelles Structures formed by association of bile salts and containing fat digestion products; found in the lumen of the small intestine.
osmotic diarrhea Watery stool induced by abnormal increases in amounts of poorly absorbed, osmotically active solutes in the gut lumen.
transepithelial transport Movement of nutrients across the intestinal mucosal layer, which is made up of epithelial cells joined together by tight junctions.
transphosphorylation Chemical reaction that involves the transfer of a phosphoric group from one compound to another.

The rate of nutrient transport is ultimately regulated by the energy content, the essential nature, or the toxic potential of nutrients. The absorption rate of a nutrient utilized mainly for its energy content increases with dietary concentration, whereas the absorption rate of a nutrient that is essential but otherwise contains little energy will increase only when dietary concentration or body stores decrease. Regulation of nutrients that are both essential and energy yielding or potentially toxic may be complex and may depend on still unidentified factors.

INTRODUCTION

The nutrients that are released from digested foods are absorbed by enterocytes or intestinal absorptive cells. Because many nutrients from food are hydrophilic and cannot diffuse through the lipid bilayer membrane, absorption requires transport systems in the brush border and basolateral membranes of the enterocytes. These transport systems are synthesized by intestinal cells so nutrients can be transferred from the lumen across the brush border into the cytosol, and from the cytosol across the basolateral membrane into the blood.

Why regulate? Regulation of absorption ensures that nutrients are not only completely absorbed but also absorbed in the appropriate regions. Nutrients that are not absorbed eventually reach the large intestine, potentially inducing borborygmi and osmotic diarrhea. Under normal conditions, however, nutrients are completely absorbed when intestinal contents reach the distal small intestine. Even when absorption is complete,

 Copyright 2004, Elsevier (USA). All rights reserved.

Moadel, R. M., Blaufox, M. D., and Freeman, L. M. (2002). The role of positron emission tomography in gastrointestinal imaging. *Gastroenterol. Clin. North Am.* **31**, 841–861.
Murray, I. P. C., and Ell, P. J. (eds.) (1998). "Nuclear Medicine in Clinical Diagnosis and Treatment," 2nd Ed. Churchill Livingstone, Edinburgh, UK.
O'Hara, S. M. (1996). Pediatric gastrointestinal nuclear imaging. *Radiol. Clin. North Am.* **34**, 845–862.
Pham, K. H., Ramaswamy, M. R., and Hawkins, R. A. (2002). Advances in positron emission tomography imaging for the GI tract. *Gastrointest. Endosc.* **55**, 553–563.
Sandler, M. P., Coleman, R. E., Patton, J. A., Wackers, F. J. T., and Gottschalk, A. (eds.) (2002). "Diagnostic Nuclear Medicine," 4th Ed. Lippincott Williams & Wilkins, Philadelphia, PA.
Suzman, M. S., Talmor, M., Jennis, R., Binkert, B., and Barie, P. S. (1996). Accurate localization and surgical management of active lower gastrointestinal hemorrhage with technetium-labeled erythrocyte scintigraphy. *Ann. Surg.* **224**, 29–36.
Thrall, J. H., and Zeissman, H. A. (eds.) (2001). "Nuclear Medicine: The Requisites," 2nd Ed. Mosby, St. Louis, MO.
Weldon, M. J. (1994). 99mTc-HMPAO planar white cell scanning. *Scand. J. Gastroenterol.* **29**, 36–42.

Nutrient Transport, Regulation of

RONALDO P. FERRARIS
University of Medicine and Dentistry of New Jersey, Newark

borborygmi Rumbling noise produced by movement of gas in the intestine.
electrochemical gradient Driving force underlying the movement of charged compounds (or nonelectrolytes cotransported with charged compounds) across a membrane.
hematochromatosis Disorder of iron metabolism characterized by excessive absorption of ingested iron.
micelles Structures formed by association of bile salts and containing fat digestion products; found in the lumen of the small intestine.
osmotic diarrhea Watery stool induced by abnormal increases in amounts of poorly absorbed, osmotically active solutes in the gut lumen.
transepithelial transport Movement of nutrients across the intestinal mucosal layer, which is made up of epithelial cells joined together by tight junctions.
transphosphorylation Chemical reaction that involves the transfer of a phosphoric group from one compound to another.

The rate of nutrient transport is ultimately regulated by the energy content, the essential nature, or the toxic potential of nutrients. The absorption rate of a nutrient utilized mainly for its energy content increases with dietary concentration, whereas the absorption rate of a nutrient that is essential but otherwise contains little energy will increase only when dietary concentration or body stores decrease. Regulation of nutrients that are both essential and energy yielding or potentially toxic may be complex and may depend on still unidentified factors.

INTRODUCTION

The nutrients that are released from digested foods are absorbed by enterocytes or intestinal absorptive cells. Because many nutrients from food are hydrophilic and cannot diffuse through the lipid bilayer membrane, absorption requires transport systems in the brush border and basolateral membranes of the enterocytes. These transport systems are synthesized by intestinal cells so nutrients can be transferred from the lumen across the brush border into the cytosol, and from the cytosol across the basolateral membrane into the blood.

Why regulate? Regulation of absorption ensures that nutrients are not only completely absorbed but also absorbed in the appropriate regions. Nutrients that are not absorbed eventually reach the large intestine, potentially inducing borborygmi and osmotic diarrhea. Under normal conditions, however, nutrients are completely absorbed when intestinal contents reach the distal small intestine. Even when absorption is complete,

 Copyright 2004, Elsevier (USA). All rights reserved.

regulation is important because the rate of absorption and the intestinal region where a nutrient is absorbed determine the time course of appearance of that nutrient in the blood, and therefore of that nutrient's availability to other organs. The pathological consequences of abnormal plasma glucose levels in diabetes are an example of the importance of plasma nutrient concentrations. Obesity in humans may be correlated with chronically elevated rate of uptake of nutrients in the small intestine.

Patterns of Regulation

Although there is an enormous number of types of nutrients and therefore an equally impressive number of transporters, regulatory mechanisms can only increase, decrease, or not change when dietary concentrations or body stores of nutrients change. The patterns of regulation for various nutrients can be predicted, based on four considerations related to the metabolic costs and benefits of nutrient absorption. First, if the metabolic cost of transporter synthesis exceeds the benefits (e.g., in the form of metabolic energy) the transporter provides, the number of transporters should decrease. Conversely, if transporter synthesis will yield energy in terms of quantity of nutrients absorbed, the number of transporters should increase. Hence, for virtually all nonessential nutrients used as calorie source, absorption rate is directly proportional to dietary substrate concentrations (see Fig. 1). Third, for essential nutrients that do not yield calories but are needed at some minimum amounts in the body for essential metabolic reactions, decreases in dietary levels should result in increases in the number of transporters, because transport will yield its greatest benefits when dietary concentrations are low (Fig. 1). Finally, transporter activity should decrease with increasing dietary substrate levels for nutrients that are toxic at high dietary concentrations or body stores. Because the transport step across the brush border is typically the limiting step, and because basolateral transporters also function to obtain nutrients from the blood for enterocytes, these considerations apply mainly to brush border transporters.

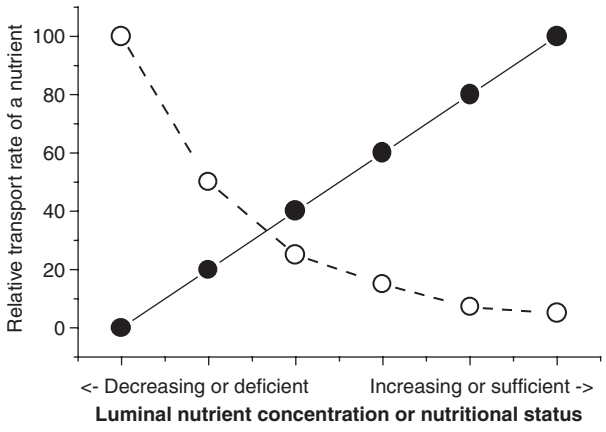


FIGURE 1 Effect of dietary concentration or body stores of a nutrient on intestinal uptake of that nutrient. ●, Nonessential nutrient, calorie source; ○, essential nutrient, not a calorie source.

Regulation can be induced by endogenous signals such as those coming from the blood, or by luminal signals such as those found in the food. These external signals use intracellular signal transduction mechanisms to alter cellular function.

Mechanisms of Regulation

Specific

Changes in absorption rates can be accomplished by specific or nonspecific mechanisms. Specific mechanisms such as changes in affinity for a substrate or in number of transporters are those that increase the absorption rate of a single nutrient or category of nutrients sharing a single transport system. Changes in number of transporters can be accomplished by changes in transporter synthesis and degradation rates or by alterations in rates of translocation of transporters to and from membranes.

Nonspecific

A wide variety of nonspecific mechanisms can alter absorption rate, including number of absorptive cells, ratio of absorptive to nonabsorptive cells, phospholipid composition, fluidity and permeability of the plasma membrane, and permeability of the paracellular pathway. In addition, for nutrients that are cotransported with the sodium or hydrogen ions, changes in electrochemical potential of these ions will also alter transport rate. Different types of mechanisms can be combined to alter nutrient transport rates. The time course of regulation can be brief, in seconds or minutes, or long, measured in hours and days. Rapid rates of regulation are thought to be a response to rapid changes in luminal nutrient concentrations such as those occurring during meals. A longer time course is thought to be caused by alterations in the average luminal concentrations of nutrients initiated by changes in diet, and is the type of regulation emphasized in this article. Diet-induced changes in nutrient absorption typically utilize specific mechanisms to alter transporter activity. There is a caveat regarding the discussion here: a significant number of studies on molecular mechanisms underlying substrate regulation of nutrient transport involve use of

cell cultures or *Xenopus* oocytes expressing the transporter of interest. Unfortunately, regulation of nutrient transporters in these artificial systems may differ from transporter regulation *in vivo* or in animal models. Studies from cell culture systems are therefore excluded from discussion in this article, unless their findings have been replicated *in vivo*. Most of this article is based on knowledge obtained from studies utilizing small animal models.

MACRONUTRIENTS

Carbohydrates

The final products of carbohydrate digestion are glucose, galactose, and fructose. Glucose and galactose enter the intestinal cells across the brush border membrane via the sodium/glucose-dependent cotransporter 1 (SGLT1), whereas fructose is absorbed by the glucose transporter (GLUT5). In mammals with omnivorous diets, the number of SGLT1 and GLUT5 copies typically varies by two- to fourfold, and is correlated with dietary carbohydrate concentrations. In carnivores, the abundance of SGLT1 is low, and little regulation occurs. In most vertebrates, site density of SGLT1 and therefore glucose transport rate increase within 24 hours after increases in levels of dietary carbohydrate.

The proposed mechanism involves an increase in the number of brush border glucose transport systems in young cells located in the lower regions of intestinal villi. Intestinal glucose transport rate gradually increases as these cells migrate up the villus and replace cells that have a lower number of glucose transporters. The change in transport rate is complete when cells with more transporters have replaced all cells with few transporters.

The magnitude of diet-induced changes in site density of GLUT5 is typically greater than that of SGLT1, suggesting that average luminal fructose concentrations undergo wider fluctuations. The proposed mechanism underlying changes in intestinal fructose transport involves changes in the number of GLUT5 copies in all cells lining the villus. Although changes in the time course of glucose transport are dependent on transporter synthesis and cell migration rate, the time course of changes in fructose transport is dependent only on transporter synthesis. As a consequence, the time course of a diet-induced change in fructose transport is typically more rapid, in the range of 4–8 hours.

Changes in number of SGLT1 are brought about by increases in concentrations of many types of sugars, including galactose, fructose, mannose, xylose, glucose, and even some nonmetabolizable analogues of glucose. Because Na^+ is also a substrate of SGLT1, increases in dietary Na^+ concentration have been shown to increase glucose absorption. The signal to increase GLUT5 transporter number is only dietary fructose (or sucrose, which contains fructose) and is therefore quite specific.

Glucose, galactose, and fructose all share the basolateral sugar transporter GLUT2, which mediates the exit of sugars from the cytosol to the blood. Although there is no transepithelial glucose transport in the absence of SGLT1, there is glucose and fructose transport in GLUT2 −/− mice, indicating that an alternative basolateral pathway for sugars exists and that SGLT1 is the rate-limiting step in transepithelial glucose transport. Like GLUT5, site density of GLUT2 transporters also increases within a few hours following increases in dietary carbohydrate, but the mechanisms underlying dietary regulation are not known. Both luminal glucose and fructose can up-regulate GLUT2 mRNA and activity in all enterocytes located in the low to upper regions of the villus, and changes in transporter number have been shown to be dependent mainly on transporter synthesis and subsequent translocation to the basolateral membrane.

Proteins

Amino Acids

Transport of amino acids is accomplished by at least seven families of transporters, each family having several types or isoforms with overlapping specificities for different substrates. Many of these transporters are heterodimeric, such that a heavy-chain subunit is linked by disulfide bridges to a light-chain subunit. There are two known families of heavy subunits and at least five families of light subunits, and any light subunit can pair with any heavy subunit. The result is a large variety of amino acid transporters with overlapping specificities so that many types of amino acids can be transported by a single transporter, and thus a single type of amino acid can be transported by different transporters.

There are two typical ways of categorizing amino acid transporters, including by a transporter's dependence on the Na^+ gradient and by its substrates. Imino and acidic amino acids each have a carrier family with subtypes dedicated to transport these substrates, but basic and neutral amino acids are thought to share several transport systems. The large number of amino acid transporters complicates any study related to their regulation.

Although sugar absorption rates typically vary in parallel with luminal concentration of carbohydrate digestion products, the absorption of amino acids does not always vary monotonically with dietary protein

concentration. There are several explanations for this observation, and they involve the fact that all amino acids can yield energy and that some amino acids are required (essential) nutrients whereas others are toxic at high concentrations. At concentrations above minimum dietary requirements, absorption of essential amino acids increases with dietary protein, unless those amino acids are potentially toxic. Absorption of potentially toxic amino acids either does not increase or increases only modestly with dietary protein levels. Below minimum dietary requirements, absorption of essential amino acids typically increases, suggesting a regulatory mechanism that compensates for deficient dietary levels of a required nutrient. Intestinal absorption rates of nonessential amino acids that are usually not toxic are directly proportional with dietary protein levels. Because the molecular mechanisms of amino acid transporters have just been elucidated, the molecular mechanisms of their regulation are still being studied.

Peptides

Di- and tripeptides are absorbed from the intestinal lumen into the cell by the PEPT transporter family, which relies on the electrochemical gradient of hydrogen to supply the energy required for transport. Dietary free amino acids and dipeptides stimulate the transcription of the PEPT gene and increase PEPT mRNA as well as protein abundance, leading to enhanced peptide absorption rates. Increases in luminal dipeptide concentration have also been shown to increase PEPT mRNA stability. Hence, increases in dietary protein levels typically increase peptide transport across the brush border membrane. Regulation of peptide transport across the basolateral membrane, however, is not known.

Fatty Acids and Bile Acids

After digestion, dietary lipids that are sufficiently hydrophobic are likely to be incorporated into micelles and then passively absorbed by the intestinal cell. Passive transport is dependent on concentration gradients and hence cannot be regulated. However, fatty acids from pancreatic lipase digestion of triglycerides are thought to cross the brush border membrane by carriers that have not yet been extensively characterized at the molecular level. High-fat diets increase fatty acid transporter expression.

Bile acids, though not considered nutrients, are essential for fat digestion and absorption. Synthesis of bile acids is not sufficient to absorb the amount of fat consumed in a typical meal, hence bile acids need to be recycled. Bile acids are amphipathic molecules and require the Na^+-dependent bile acid transporter (ASBT) for transport across the brush border membrane of enterocytes in the distal small intestine. ASBT expression and activity decrease when luminal bile acid concentrations decrease.

MICRONUTRIENTS

Minerals

Iron

Iron is used as a catalytic center in fundamental metabolic reactions, but can be toxic at high concentrations because of its ability to oxidize many cell components. Iron deficiency is primarily caused by lack of dietary iron, whereas iron overload is mainly genetic and in some individuals is attributed to a deficiency in functional hematochromatosis protein. Body iron stores are controlled by regulating intestinal iron absorption via unidentified signals that sense the level of body iron and that communicate body iron status to intestinal crypt cells. Crypt cells are then somehow programmed to absorb dietary iron at rates appropriate for the individual.

The small intestine absorbs both elemental and heme iron. Absorption of elemental iron across the brush border membrane occurs by way of the proton-dependent divalent metal transporter (DMT1). Dietary and body iron deficiency increases mRNA and protein abundance of DMT1 as well as rates of insertion into the apical membrane of enterocytes, resulting in an increase in absorption rate for iron. High dietary levels of iron decrease brush border absorption rate. Expression of the basolateral iron transporter ferroportin 1 (IREG1) either is not affected by dietary iron levels or decreases with dietary iron. In humans with hematochromatosis, regulation of iron absorption is compromised, such that there is increased protein and mRNA expression of DMT1 in spite of excess body iron stores. Both DMT1 and IREG1 mRNA have iron response elements in the 3′ and 5′ untranslated regions that bind iron-response proteins, suggesting that posttranscriptional regulation is important in iron absorption. Hephaestrin, a copper containing iron oxide necessary for basolateral exit of iron, responds to systematic signals of iron status.

The molecular mechanisms of heme iron transport are not well known, but heme iron transport rate is also down-regulated by body iron stores, decreasing in rate when stores are high.

Calcium

Calcium is required for skeleton formation and also plays an important role as intracellular messenger in

virtually all cells. Almost all calcium in the body is stored in the skeleton. Calcium transport in the small intestine is tightly regulated by vitamin D, and is also modulated by dietary calcium levels. Active, transepithelial transport of dietary calcium increases when dietary calcium intake is low and decreases when dietary calcium intake is high.

Calcium enters the cell down an electrochemical gradient via calcium channels. Regulation of these channels is not known. The rate-limiting step of dietary calcium absorption is the intracellular movement of calcium from the apical pole of the enterocyte to the basolateral pole. Intracellular calcium movement is mediated by cytosolic calcium-binding protein (calbindin). When plasma calcium concentration decreases below normal (about 2.5 m*M*), vitamin D increases the synthesis of calbindin, which increases the transfer of dietary calcium from the apical to the basolateral pole, where calcium is extruded against an electrochemical gradient by Ca-ATPase. Vitamin D also increases the rate of entry and the rate of exit of dietary calcium across the brush border and basolateral membranes, respectively.

Phosphate

Like calcium, phosphate is also a major component of bone. Phosphate is absorbed against an electrochemical gradient in the small intestine by a Na^+-dependent phosphate transport system called NaPi, which has many subtypes; NaPi2b is expressed in the small intestine of humans and of many other vertebrates. Phosphate transport rate and the number of NaPi2b transporters increase with chronic consumption of a diet low in phosphate or by increases in plasma vitamin D. Although transporter protein abundance and transport rates increase dramatically, NaPi2b mRNA abundance does not change, indicating that NaPi2b regulation does not involve alterations in transcription rate. Unlike that of the renal phosphate transporter NaPi2a, which adapts within hours following a change in dietary phosphorus concentrations, the time course of adaptation of NaPi2b to changes in dietary phosphorus levels is slow, and typically runs over several days. Exit of phosphate from the enterocyte to the blood occurs by way of a facilitated carrier in the basolateral membrane. Regulation of this transport system is not known.

Vitamins

Most water-soluble vitamins are absorbed by transporters across the brush border and basolateral membranes of intestinal cells, but the molecular mechanisms are not known. Only those vitamins with well-characterized transport systems are mentioned here.

Ascorbic acid is absorbed across the brush border membrane by SVCT1, a high affinity, Na-dependent transporter expressed only in the small intestine and kidney. Chronic consumption of ascorbic acid reduces the rate of intestinal ascorbic acid transport. Low concentrations of dietary ascorbic acid, however, do not increase intestinal transport.

Biotin shares a Na^+-dependent brush border transporter, SMVT, with pantothenic acid. Biotin deficiency leads to a marked increase in intestinal biotin transport, whereas biotin consumption at pharmacological levels decreases biotin transport rate. The regulation of pantothenic acid transport is not known but is likely influenced by dietary or systemic levels of both biotin and pantothenic acid.

Folate is absorbed across the brush border membrane by hIFC-1, located mainly in the enterocytes lining the upper villus regions. Dietary folate deficiency leads to specific increases in intestinal folate transport and in hIFC-1 protein and mRNA abundance. A related thiamine transporter, ThT1, absorbs thiamine across the brush border. In the cell, the thiamine is phosphorylated as it crosses the cytosol. Basolateral exit seems to involve a transphosphorylation process. A thiamine-deficient diet increases transepithelial transport of thiamine.

See Also the Following Articles

Calcium, Magnesium, and Vitamin D Absorption, Metabolism, and Deficiency • Carbohydrate Digestion and Absorption • Dietary Reference Intakes (DRI): Concepts and Implementation • Digestion, Overview • Fat Digestion and Absorption • Iron Absorption • Malabsorption • Protein Digestion and Absorption of Amino Acids and Peptides • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Abumrad, N. A., Sfeir, Z., Connelly, M. A., and Coburn, C. (2000). Lipid transporters: Membrane transport systems for cholesterol and fatty acids. *Curr. Opin. Clin. Nutr. Metab. Care* **3**(4), 255–262.

Bronner, F., and Pansu, D. (1999). Nutritional aspects of calcium absorption. *J. Nutr.* **129**(1), 9–12.

Ferraris, R. P. (2001). Dietary and developmental regulation of intestinal sugar transport. *Biochem. J.* **360**(Pt. 2), 265–276.

Ferraris, R. P. (1994). Regulation of intestinal nutrient transport. *In* "Physiology of the Gastrointestinal Tract," (L. R. Johnson, ed.), Chap. 54, pp. 1821–1844. Raven Press, New York.

Phan, C. T., and Tso, P. (2001). Intestinal lipid absorption and transport. *Front. Biosci.* **6**, D299–D319.

Roy, C. N., and Enns, C. A. (2000). Iron homeostasis: New tales from the crypt. *Blood* **96**(13), 4020–4027.
Said, H. M. (1999). Cellular uptake of biotin: Mechanisms and regulation. *J. Nutr.* **129**(2S Suppl.), 490S–493S.
Shneider, B. L. (2001). Intestinal bile acid transport: Biology, physiology, and pathophysiology. *J. Pediatr. Gastroenterol. Nutr.* **32**(4), 407–417.
Wagner, C. A., Lang, F., and Broer, S. (2001). Function and structure of heterodimeric amino acid transporters. *Am. J. Physiol. Cell Physiol.* **281**(4), C1077–C1093.
Wisen, O., and Johansson, C. (1992). Gastrointestinal function in obesity: Motility, secretion, and absorption following a liquid test meal. *Metabolism* **41**, 390–395.

Nutritional Assessment

KHURSHEED N. JEEJEEBHOY
University of Toronto, Canada

bioimpedance analysis Measurement of body resistance and reactance by the use of a fixed high-frequency alternating current.
bioimpedance spectroscopy Measurement of a range of body resistance and reactance by the use of variable high-frequency alternating current.
body mass index The body weight in kilograms divided by the height expressed in meters squared (17 kg/m^2).
nutrition-associated complications Disease complications that can be influenced by the nutritional status.

Nutritional status is conventionally associated with body mass. This is a concept derived from agricultural science, where the objective was to develop, through nutritional means, large and meaty animals. In humans, the objective of nutritional assessment is to identify individuals who, because of their nutritional status, are at increased risk of disease complications or death. It is hoped that persons likely to benefit (in terms of reduced complications) can be identified by such an assessment.

INTRODUCTION

Nutritional health is maintained by a state of equilibrium in which nutrient intake and requirements are balanced. Malnutrition occurs when net nutrient intakes (nutrient intakes corrected for abnormally large fecal or urinary losses) are less than requirements. The obvious effect of the above processes is a change in body and blood composition and changes in body and blood composition have traditionally been used to assess nutrition. Traditional nutritional science was first developed in the field of agriculture, where the effect of nutrition was entirely judged by animal growth, mass of muscle and fat, and the production of proteins by the liver. It is therefore not surprising that nutritional assessment methods hitherto used were based on changes in body and blood composition.

The prime question when faced with a patient is how can the physician alter the outcome of the disease in a favorable (to the patient) manner. The question is whether measurement of body composition and blood components allows a prediction of outcome and whether these measurements can be used to help devise treatment that is beneficial? To answer these questions, it is necessary to examine the evolution of malnutrition and the ability of nutritional assessment techniques to predict outcome and aid treatment knowing the evolution.

Malnutrition leads to a succession of metabolic abnormalities, physiological changes, reduced organ and tissue function, and loss of body mass. Concurrent stress such as trauma, sepsis, inflammation, and burns accelerates the loss of tissue mass and function. Ultimately, critical loss of body mass and function occurs, resulting in nutrition-associated complications (NAC) and even death.

The assessment of the nutritional status, to be of clinical importance, should be able to predict whether the individual would have increased morbidity and mortality in the absence of nutritional support. In short, can it predict the occurrence of NAC and thus predict outcome? Disease and nutrition interact so that disease in turn may cause secondary malnutrition or malnutrition may adversely influence the underlying

Copyright 2004, Elsevier (USA). All rights reserved.

Nutrition in Aging

NOEL W. SOLOMONS
Center for Studies of Sensory Impairment, Aging and Metabolism, Guatemala City, Guatemala

bioelectrical impedance analysis Estimation of total body water by passing low-amperage electrical alternating current through the body and measuring the electrical properties of resistance and reactance.

***INDY* gene** Recently discovered "I'm not dead yet" gene locus, mutation of which can slow senescence and extend longevity; encodes for a protein that down-regulates cellular energy utilization.

life expectancy Predicted number of years that will pass until only half of any cohort of people will still be alive.

longevity Length of time an individual survives; for humans, 120 years is considered the maximal life span.

sarcopenia Age-dependent loss of skeletal muscle fibers and their replacement by intramuscular fat in the elderly, leading to reduced strength of gait and predisposition to falls.

senescence Age-dependent decline in physiological capacities and anatomic integrity that proceeds throughout adulthood.

The elderly represent a rapidly growing segment of populations around the world. Physiological senescent changes impact on the ability of individuals to feed and nourish themselves, and aging and chronic disease are conducive to both over- and undernutrition. A better understanding of specific nutrient requirements for older persons requires refining the recommendations for selection of foods, food groups, and specific dietary supplements for the aging population.

DEMOGRAPHIC CONSIDERATIONS

Sixty years of age or more is considered to be the criterion for being classified as elderly by the World Health Organization, whereas 65 years is the threshold in North America and the United Kingdom. The number of people over 65 years of age is growing worldwide at an unprecedented rate of 800,000 per month, according to a recent report of the United States Census Bureau and National Institute on Aging. The fastest growing segments of the population are people over 85 years and the centagenarians. Life expectancy at birth, i.e., the expected median survival of a birth cohort, has been advancing steadily in all nations. In developed nations, it now ranges from 75 to 81 years, always with a greater expectancy for women than men. Currently, for instance, 18% of the Italian population is over 65 years of age. Despite much lower life expectancies, the larger total populations in Africa, Asia, and Latin America mean that the majority of people over 60 years of age now live in low-income developing countries.

CHRONOLOGICAL AND BIOLOGICAL AGING

Chronological age refers to the number of years an individual has lived. Survival to 120 years is considered to be the maximum life span for *Homo sapiens*. Biological aging is a process that begins at conception and continues throughout life, first as developmental changes, then, in adulthood, as "senescence." It involves the anatomical changes in tissues and a decline in the capacity for physiological functions of most organ systems.

A series of theories regarding the nature and mediation of the aging process have been advanced, all of which can be rationalized on the basis of an evolutionary advantage to maximize reproductive efficiency at the expense of the mechanisms for genetic and cellular repair. Some theories frame the mechanism as oxidative stress and cumulative mutations associated with the cumulative cellular utilization of metabolic energy. Calorie restriction in animals reduces the rate of senescence and prolongs longevity. A *Drosophila* gene associated with longevity, *INDY*, when mutated to mimic caloric restriction, leads to a longer life span. Thus, longevity may have some genetic basis.

Successful Aging

The classical public health goal for gerontology has been expressed as "adding life to years, rather years to life." This involves the so-called compression of morbidity, in which physical and cognitive capacities are maintained and individuals succumb rapidly in old age after a minimum period of disability and dependency. Such a life evolution has been termed "successful aging," in which chronological age advances more rapidly than biological aging and people maintain normal health and

 Copyright 2004, Elsevier (USA). All rights reserved.

adequate physical and cognitive function. A majority of survivors to the "Third Age," however, exhibit "usual" (normative) aging, in which chronic diseases and some disability intercede. The most burdensome format for individuals, their relatives, and the community is the evolution to "fragile aging," characterized by disease, debility, heavy dependence on caretakers, and the need for institutional care.

Changes in Body Composition with Advancing Age

Basic senescent processes and diseases common to advanced age impact the elemental, molecular, and tissue and organ composition of the body. Conversely, body composition influences resistance and susceptibility to disease and disability. Primary among senescent changes is a progressive decrease in height, beginning around age 30 and continuing thereafter. It is proportionately greater in aging women as compared to men. Settling and desiccation of intervertebral disks combines with compression of vertebral bodies to diminish the length of the spinal column. Curvatures produced by uneven vertebral compression can further reduce stature.

On the aggregate, body weight increases up through the seventh decade and then stabilizes or decreases in later life. The composition of this weight follows a stereotyped modification in the last third of the life span, with an increased percentage of body mass as fat, a decreased percentage of body water, and a lower total cell mass. The deposition of fat added in later life is to some extent visceral, but is largely intramuscular. The term "sarcopenia" (scarcity of muscle) has been assigned to this age-related loss of muscle fibers and marbling of the skeletal muscle, which may be mediated by inflammatory cytokines, e.g., interleukin-6. Sarcopenia contributes to loss of strength in gait and predisposes to falls. Osteopenia (scarcity of bone) is the progressive demineralization of bone and loss of skeletal mass, common in both sexes. A rapid, involutional bone demineralization occurs with menopause, making women more susceptible (but not exclusively so) to a critical reduction in bone mineral density and loss of architecture that leads to pathological fractures of the vertebral bodies, wrists, and hips (osteoporosis).

CAVEATS IN ASSESSING NUTRITIONAL STATUS IN THE ELDERLY

The senescent processes confound the assessment of nutritional status of the elderly to the extent that conventional standards derived from younger adults are employed. As a consequence of the aforementioned body composition changes, anthropometric indicators are confounded. When elderly patients are bedridden or undergo partial amputations, two proxy measurements—arm span and knee height—have been advanced to produce an estimate for height. Moreover, because true (original) height cannot be assessed from the stature of an elderly patient, these same proxy measures can be used to interpolate to what was the likely length in youth. This senescent shortening of stature has implications for the calculation of Quetlet's body mass index (BMI) (weight in kilograms divided by the square of height in meters) in individuals and in groups of elderly individuals. By international BMI standards, underweight begins with a BMI equal to or less than 18.5 kg/m^2. Overweight is a state of BMI from 25 to 29.9 kg/m^2 and obesity ensues at or greater than 30 kg/m^2. The measured BMI for an older person is inevitably higher than what would have been measured for the same weight earlier in life, because of a distortion in the denominator term. This produces an underestimation of true underweight and an overestimation of overweight states in an elderly population. Substituting the arm span or a height term derived from knee height into the BMI formula improves the validity of the diagnostic classification. As a caveat, both arm span and knee height vary in relation to true height across ethnic groups, and this must be compensated for in the clinic or in surveys.

Caliper assessment of skinfolds are confounded both in their measurement and their interpretation in the elderly. Laxity of skin makes the dermis more compressible between the jaws of the calipers, producing a falsely low value. Moreover, because fat deposition shifts from subcutaneous distribution to intramuscular marbling, the relation of skinfolds to total body fat is distorted in measurements made in the elderly, and the formulas for total body fat from skinfolds do not apply.

A global estimation of total body water, from which fat mass and fat-free mass can be derived, applying certain assumptions, can be assessed by bioelectric impedance analysis (BIA). Most investigations have found that nomograms for interpreting BIA data in terms of body composition require an age term.

The Mini-Nutritional Assessment is a four-component diagnostic tool, combining anthropometric indices, dietary intake, examiner's global assessment, dietary evaluation, and patient's subjective assessment points, to be used in general practice to assess the nutritional status of the elderly. Less than a score of 17 out of a 30-point total signifies undernutrition, whereas a score of 17−23 points implies "nutritional risk." When the opportunity for laboratory assessment

comes into play, low albumin and cholesterol concentrations and lymphocyte counts of <1500/μl provide additional evidence of poor geriatric nutritional status.

There is no validity to the oft-cited belief that red cell concentrations are normally reduced with aging. Inflammation and decreased renal mass can reduce erythropoiesis rates, however, independent of any constraints on hematinic nutrients. Hepatic secretion rates of proteins, including albumin, are reduced with increasing age, but also with chronic inflammatory illness. It is not clear if age is an independent variable.

FACTORS INFLUENCING NUTRITIONAL STATUS IN AGING

A generic roster of mechanisms devised by Herbert, which, alone or in combination, can explain all nutrient deficiency states, is useful for understanding undernutrition in the elderly when aging-specific factors are weighed within this framework (Table I). The elderly are also susceptible to overnutrition involving energy and certain micronutrients (Table II).

Undernutrition in the Elderly

Malnutrition (undernutrition) has variously been reported to affect one-fourth of community-dwelling elders, one-half of hospitalized elderly, and up to two-thirds of the residents of long-term chronic-care facilities. Of course, chronic and terminal illness contribute to these statistics, and not all such undernutrition can (or should) be addressed by specific nutritional measures. Taste enhancement with flavorings or monosodium glutamate increases eating satisfaction in the healthy elderly and energy consumption in the sick elderly. Both inappetence and sensory dulling may distort the dietary intake of older persons, i.e., the balance and adequacy of micronutrients are better for dependent institutionalized elderly, fed by caretakers, than for ambulatory persons who select an *ad libitum* diet.

Older persons are much more sensitive to dehydration due to immobility, dysregulated thirst, fear of incontinence, diuretic use, and increased solute loads. For hospitalized and institutionalized elderly, weight loss and dehydration increase risk of pressure sores; nutritional supplementation is generally useful when intake is low, but has not been demonstrated to improve resistance to pressure sores and ulcers.

With respect to micronutrients, the lower energy needs of the elderly oblige the diet to have a higher nutrient density to fulfill the normal intakes of vitamins and minerals. Gastric atrophy (due to *Helicobacter pylori* infection) and increased intraintestinal pH predispose to bacterial overgrowth and decreased absorption of certain micronutrients, such as iron, folic acid, beta-carotene, and vitamin B_{12}. Deficiency of the latter is a threat to the elderly, and its delayed recognition by masking with folic acid may be increasing in risk as public health organization folate fortification policies advance.

Increased requirements for vitamin D collide with reduced food intake and decreased outdoor activities, contributing to vitamin deficiency and greater bone loss. Insufficient intake of zinc and of vitamin K_1 to maximize bone mineralization are also consequences of dietary constraints typical with advancing years.

TABLE I Six Possible Causes of All Nutrient Deficiencies, as Applied to Specific Factors of Chronological Aging[a]

Decreased intake	
	Depression, dementia, economic constraints, immobility, illness, taste and smell deficits, oral disease, edentulousness, decreased thirst sensitivity, altered intestinal motility, medications, alcohol abuse
Decreased absorption	
	Hypochlorhydria, drug interactions, recurrent diarrhea, bacterial overgrowth
Increased wastage	
	Chronic inflammation, glucose intolerance and diuresis, diuretic use, "programmed" tissue remodeling
Decreased utilization	
	Drug interactions, chronic inflammation, alcohol abuse
Increased destruction	
	Drug interactions, chronic inflammation
Increased requirements	
	Possibly for calcium, vitamin D, and vitamin B_6

[a] Modified and adapted from Herbert (1973).

TABLE II Five Possible Causes of All Nutrient Excesses as Applied to Specific Factors or Chronological Aging

Increased intake	
	Depression, self-supplementation, medications
Increased absorption	
	No case examples known
Increased retention	
	Decreased physical activity; lower basal metabolism rates; reduced nephron number, decreased renal blood flow, and diminished glomerular filtration; reduced biliary flow
Decreased mobilization or destruction	
	Decreased postprandial clearance of chylomicrons
Decreased requirements	
	Accumulated lifetime storage of nutrient reserves

Overnutrition in the Elderly

Overweight and obesity are risk factors for cardiovascular disease, hypertension, stroke, type 2 diabetes, osteoarthritis, depression, and certain cancers. Decreased activity and lower basal metabolism contribute to risk in later life. There is ongoing debate, however, as to whether the body mass index associated with lowest all-cause mortality, around 22 kg/m^2, applies to all age groups of adults equally, or whether a higher BMI level is more protective in older age.

Vitamin A toxicity is an issue for the elderly, not by body burden, per se, but because retinyl esters persist an inordinately long time in the circulation. Iron stores increase progressively throughout life in healthy men, and in women after menopause. The upper segment of the population distribution may accumulate sufficient iron to constitute an oxidative burden. Manganese and copper may accumulate in the liver and in other tissue sites if cholestatic conditions are present.

DIETARY REQUIREMENTS, RECOMMENDATIONS, AND FOOD SELECTION PATTERNS FOR THE ELDERLY

In past editions of the Recommended Dietary Allowances (RDAs), all adults above 51 years of age formed a common group; it has only been in the past 10 years that sufficient information to make age-specific recommendations for persons 75 years and older has been developed. Dietary guidelines for the selection of foods and food groups have also proliferated, and differentiation for the specific needs and expectations for the elderly is emerging.

Nutrient Requirements and Recommendations in Later Life

It has long been established, and verified by community survey data, that the energy requirements of the elderly are about two-thirds of what they are in active, young, and middle adulthood. In the United States' and Canada's *Dietary Reference Intakes*, differential increases in recommended levels, higher than those for younger adults, exist for three nutrients: calcium, vitamin D, and vitamin B_6. For two additional nutrients relevant to aging—vitamin B_{12} and vitamin K_1—the general adult recommendations have almost doubled over the previous RDA.

Diet, Function, and Chronic Disease Risk

Alongside the concerns for adequate—but not excessive—intakes of macro- and micronutrients are another series suggestions constituting "dietary guidelines," which are related to selection of and servings of various foods and food groups. A variant Food Pyramid, taking the specific needs of the elderly to consume extra liquid to maintain hydration and more dietary fiber to increase colonic transit, stool frequency, and fecal weight, has been advanced for geriatric preventive nutrition. Most aspects of the conventional Food Pyramid relate to a long-term dietary pattern to avoid the genesis of chronic disease. Doubt surrounds the potential for the *de novo* institution of these healthful eating patterns to prevent or reverse disease processes once an individual has reached advanced age.

Dietary constituents for improved performance of physiological function and disease resistance are being identified. Improved immunoresponsiveness has been documented with chronic intakes of suprarequirement amounts of zinc and vitamin E, levels that could never be achieved from dietary sources alone. Nonnutrient substances such as phytochemicals and carotenoids may also have emerging roles in maintaining health in later life, such as lutein and zeaxanthin for macular degeneration. Reduced incidences of most forms of cataracts are associated with higher dietary intakes and supplements of antioxidant vitamins and some carotenoids. Cognitive function of elderly can be improved by routine use of balanced vitamin and mineral supplements without iron.

See Also the Following Articles

Aging • Dietary Reference Intakes (DRI): Concepts and Implementation • Malnutrition • Nutritional Assessment

Further Reading

Chandra, R. K. (2001). Effect of vitamin and trace-element supplementation on cognitive function in elderly subjects. *Nutrition* **17**, 709–712.

Guigoz, Y., Vellas, B., and Garry, P. J. (1994). Mini-nutritional assessment. A practical tool for grading the nutritional state of elderly patients. *Facts Nutr. Gerontol.* **21** (Suppl. 2), 15–81.

Hartz, S. C., Russell, R. M., and Rosenberg, I. H. (eds.) (1993). "Nutrition in the Elderly: The Boston Nutritional Status Survey." Smith-Gordon, London.

Herbert, V. (1973). The five possible causes of all nutrient deficiency: Illustrated by deficiencies of vitamin B12 and folic acid. *Am. J. Clin. Nutr.* **26**, 77–88.

Rowe, J. W., and Kahn, R. L. (1987). Human aging: Usual and successful. *Science* **237**, 143–149.

Russell, R. M., and Ausman, L. (1994). Nutrition in the elderly. *In* "Modern Nutrition in Health and Disease" (M. E. Shils,

J. A. Olson, and M. Shike, eds.), 8th Ed., pp. 770–780. Lea & Feibiger, Philadelphia.

Russell, R. M., Rasmussen, H., and Lichtenstein, A. H. (1999). Modified food pyramid for people over seventy years of age. *J. Nutr.* **129**, 751–753.

Solomons, N. W. (2000). Demographic and nutritional trend among the elderly in developed and developing regions. *Eur. J. Clin. Nutr.* **54** (Suppl. 3), S1–S13.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine. (2003). Dietary reference intakes for calcium, phosphorus, magnesium, vitamin D, and fluoride (1997), for thiamin, phosphorus, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin, and choline (1998), for vitamin E, vitamin C, selenium, and carotenoids (2000), and for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc (2001). National Academy Press, Washington, D. C.

World Health Organization. (1989). "Health in the Elderly." WHO, Geneva. [Updated information is available from the World Health Report, available online at http://www.who.int/whr.]

Yehh, S.-S., and Schuster, M. W. (1999). Geriatric cachexia: The role of cytokines. *Am. J. Clin. Nutr.* **79**, 183–187.

Roy, C. N., and Enns, C. A. (2000). Iron homeostasis: New tales from the crypt. *Blood* **96**(13), 4020–4027.
Said, H. M. (1999). Cellular uptake of biotin: Mechanisms and regulation. *J. Nutr.* **129**(2S Suppl.), 490S–493S.
Shneider, B. L. (2001). Intestinal bile acid transport: Biology, physiology, and pathophysiology. *J. Pediatr. Gastroenterol. Nutr.* **32**(4), 407–417.
Wagner, C. A., Lang, F., and Broer, S. (2001). Function and structure of heterodimeric amino acid transporters. *Am. J. Physiol. Cell Physiol.* **281**(4), C1077–C1093.
Wisen, O., and Johansson, C. (1992). Gastrointestinal function in obesity: Motility, secretion, and absorption following a liquid test meal. *Metabolism* **41**, 390–395.

Nutritional Assessment

KHURSHEED N. JEEJEEBHOY
University of Toronto, Canada

bioimpedance analysis Measurement of body resistance and reactance by the use of a fixed high-frequency alternating current.
bioimpedance spectroscopy Measurement of a range of body resistance and reactance by the use of variable high-frequency alternating current.
body mass index The body weight in kilograms divided by the height expressed in meters squared (17 kg/m^2).
nutrition-associated complications Disease complications that can be influenced by the nutritional status.

Nutritional status is conventionally associated with body mass. This is a concept derived from agricultural science, where the objective was to develop, through nutritional means, large and meaty animals. In humans, the objective of nutritional assessment is to identify individuals who, because of their nutritional status, are at increased risk of disease complications or death. It is hoped that persons likely to benefit (in terms of reduced complications) can be identified by such an assessment.

INTRODUCTION

Nutritional health is maintained by a state of equilibrium in which nutrient intake and requirements are balanced. Malnutrition occurs when net nutrient intakes (nutrient intakes corrected for abnormally large fecal or urinary losses) are less than requirements. The obvious effect of the above processes is a change in body and blood composition and changes in body and blood composition have traditionally been used to assess nutrition. Traditional nutritional science was first developed in the field of agriculture, where the effect of nutrition was entirely judged by animal growth, mass of muscle and fat, and the production of proteins by the liver. It is therefore not surprising that nutritional assessment methods hitherto used were based on changes in body and blood composition.

The prime question when faced with a patient is how can the physician alter the outcome of the disease in a favorable (to the patient) manner. The question is whether measurement of body composition and blood components allows a prediction of outcome and whether these measurements can be used to help devise treatment that is beneficial? To answer these questions, it is necessary to examine the evolution of malnutrition and the ability of nutritional assessment techniques to predict outcome and aid treatment knowing the evolution.

Malnutrition leads to a succession of metabolic abnormalities, physiological changes, reduced organ and tissue function, and loss of body mass. Concurrent stress such as trauma, sepsis, inflammation, and burns accelerates the loss of tissue mass and function. Ultimately, critical loss of body mass and function occurs, resulting in nutrition-associated complications (NAC) and even death.

The assessment of the nutritional status, to be of clinical importance, should be able to predict whether the individual would have increased morbidity and mortality in the absence of nutritional support. In short, can it predict the occurrence of NAC and thus predict outcome? Disease and nutrition interact so that disease in turn may cause secondary malnutrition or malnutrition may adversely influence the underlying

Copyright 2004, Elsevier (USA). All rights reserved.

disease. Thus, patient outcomes are multifactorial and attempting to formulate the influence of malnutrition based on outcomes of single parameters or simple models fails to consider the many interacting factors. Measurements of body composition and blood components are a snapshot in time and cannot predict the direction of change. Three major factors alter the direction of malnutrition, namely, intake, disease-induced stress, and absorption or losses of nutrients. The progression or regression of these factors ultimately determines whether critical malnutrition will occur in the absence of nutritional support.

Although there are reliable methods to identify patients who will develop NAC, none of the methods or tools can clearly identify patients in whom nutritional support will reduce NAC.

TRADITIONAL NUTRITIONAL ASSESSMENT INDICES

Nutritional status has been traditionally defined by body composition, plasma protein concentrations, immune competence, and multivariate analysis. Assessment of nutritional status based on body composition involves detecting the loss (or gain) of body components relative to previous measurements and relating the values in a given patient to normal standards. The reproducibility and error in the measurements themselves affect the former and the latter is dependent on the normal range of values. A person who starts off at the upper end of the normal range may be classified as "normal" despite considerable changes in the measured value. Therefore, it is possible for a person to be in a negative nutritional state for a long time before anthropometric measurements fall below normal.

Body Weight and Weight Loss

Body weight is a simple measure of total body components and is compared to an "ideal" or desirable weight. This comparison can be made by using formulas or tables. However, a simple approach that gives as much information as these tables is the calculation of the body mass index (BMI) or Quetelet index. Body mass index is calculated as weight in kilograms divided by height in meters squared. A BMI below 14–15 is associated with significant mortality. However, measurements of body weight in patients in hospitals, especially those in intensive care units (ICUs), and of patients with liver disease, cancer, and renal failure are confounded by changes in body water due to underhydration, edema, ascites, and dialysate in the abdomen. Furthermore, the nutritional significance of changes in body weight can again be confounded by changes in hydrational status.

Anthropometry

Triceps and subscapular skinfold thicknesses provide an index of body fat and midarm muscle circumference provides a measure of muscle mass. Although these measurements seem to be useful in population studies, their reliability in individual patients is less clear. The most commonly used standards for triceps skinfold thickness and midarm muscle circumference are not universally applicable. The use of these standards to identify malnutrition in many patients is problematic because of the restricted database and the absence of correction factors for age, hydrational status, and physical activity on anthropometric parameters. Several studies have demonstrated that 20–30% of healthy control subjects would be considered malnourished based on these standards. These measurements do not reflect muscle and fat mass in sick patients, especially those in ICU and those with liver and renal disease, where edema is a major problem in assessing skin folds and arm circumference.

Creatinine–Height Index

The excretion of creatinine in the urine is related to muscle mass. Normalized for height the 24 h creatinine excretion is an index of muscle mass. In theory, it is a good and simple way of assessing the lean body mass. However, it is dependent on complete 24 h urine collections, and urinary losses or oliguria may result in an inappropriate diagnosis of malnutrition. Patients on diuretics such as those with cardiac and liver failure and those with renal disease are especially likely to have low levels of excretion of creatinine.

Serum Albumin

Serum albumin is one of the most extensively studied proteins and there are approximately 20,000 citations referring to this protein in the literature from the past 30 years. Several studies have demonstrated that a low serum albumin concentration correlates with an increased incidence of medical complications. Disease rather than nutrition in adults mainly influences serum albumin.

The concentration of serum albumin represents the net sum of albumin synthesis, degradation, losses from the body, exchange between intra- and extravascular albumin compartments, and the volume in which albumin is distributed. Albumin is highly water soluble and resides in the extravascular space. The total body pool of

albumin in a normal 70 kg man is ~280 g (3.5–5.3 g/kg). Approximately one-third of the total pool constitutes the intravascular compartment and two-thirds constitutes the extravascular compartment. The concentration of albumin in blood is greater than that in lymph or other extracellular fluids and the ratio of intravascular to extravascular albumin concentration varies from tissue to tissue. During steady-state conditions, ~14 g of albumin (200 mg/kg) is produced and degraded daily. Each day, ~5% of the total albumin pool is degraded and replaced by newly synthesized albumin. Protein-calorie malnutrition causes a decrease in the rate of albumin synthesis because adequate nutrient intake is important for polysomal aggregation and maintenance of cellular RNA levels needed for protein synthesis. Within 24 h of fasting, the rate of albumin synthesis decreases markedly. However, a short-term reduction in albumin synthesis has little impact on albumin levels because of a compensatory decrease in albumin degradation and a transfer of extravascular albumin to the intravascular compartment. Prolonged protein-calorie restriction induced experimentally in human volunteers or observed clinically in patients with anorexia nervosa causes marked reductions in body weight but little change in plasma albumin concentration. A protein-deficient diet with adequate calories in elderly persons causes a decrease in lean body mass and muscle function without a change in plasma albumin concentration.

Hospitalized patients may have low levels of plasma albumin for several reasons. Inflammatory disorders cause a decrease in albumin synthesis, an increase in albumin degradation, and an increase in albumin transcapillary losses. Gastrointestinal diseases and some cardiac diseases increase albumin losses through the gut and renal diseases may cause considerable albuminuria. Wounds, burns, and peritonitis cause major losses from the injured surface and in certain circumstances cause an increase in albumin losses through the gut, kidneys, or damaged tissues. Because the exchange between intra- and extravascular albumin is so large, even small changes in the percentage of exchange can cause significant changes in plasma albumin levels.

Prealbumin

Prealbumin is a transport protein for thyroid hormones and exists in the circulation as a retinol-binding–prealbumin complex. The turnover rate of this protein is rapid, with a half-life of 2–3 days. It is synthesized by the liver and is catabolized partly in the kidneys. Protein-energy malnutrition reduces the levels of prealbumin and refeeding restores levels. However, prealbumin levels fall without malnutrition in infections and in response to cytokine and following hormone treatment. Renal failure increases levels and liver failure may cause decreased levels. Although prealbumin is responsive to nutritional changes, it is influenced by several disease-related factors, making it unreliable as an index of nutritional status.

Immune Competence

Immune competence, as measured by delayed cutaneous hypersensitivity (DCH), is affected by severe malnutrition. While it is true that immune competence as measured by DCH is reduced in malnutrition, several diseases and drugs influence this measurement, making it a poor predictor of malnutrition in sick patients. The following factors nonspecifically alter DCH in the absence of malnutrition: (1) infections; (2) uremia, cirrhosis, hepatitis, trauma, burns, and hemorrhage; (3) steroids, immunosuppressants, cimetidine, warfarin, and perhaps aspirin; and (4) general anesthesia and surgery. Hence, in the critically sick patient, many factors can alter DCH and render it valueless in assessing the state of nutrition.

MEASUREMENT OF BODY COMPOSITION

The body consists of compartments or components. There are over 35 well-recognized components and these are organized into five levels of increasing complexity: atomic, molecular, cellular, tissue-system, and whole body. In healthy weight-stable subjects, there are relatively constant relationships among these components, which are correlated with one another.

Isotope Dilution

Total body water, measured by isotope dilution, is usually the largest molecular-level component. Water maintains a relatively stable relationship to fat-free body mass and thus measured water isotope dilution volumes allow prediction of fat-free body mass and fat (i.e., body weight minus fat-free body mass). The relationship between total body water and other body composition components may change with disease and this should be considered when interpreting data from hospitalized or chronically ill patients. The usual approach is to measure a dilution volume using isotopes such as tritium-, deuterium-, or ^{18}O-labeled water and then calculate fat-free mass by assuming that the proportion of fat-free body mass as water is constant at 0.732. However, there are no data about the relationship of fat-free mass and clinical outcome.

Bioimpedance Analysis

Bioimpedance analysis (BIA) is a method of estimating body fluid volumes by measuring the resistance to a high-frequency, low-amplitude alternating electric current (50 kHz at 500 to 800 mA). The amount of resistance measured (*R*) is inversely proportional to the volume of electrolytic fluid in the body and to a lesser extent on the proportions of this volume. A regression equation is then developed based between a reference measurement of fat-free body mass (i.e., isotope dilution) and the measured *R*, height, and other variables.

Recently, the difference in the conductive properties of extra- and intracellular fluid has resulted in the development of bioimpedance spectroscopy (BIS). Direct current will pass through extracellular fluid but will not traverse the lipid membrane of cells. As the frequency of the applied current increases, the cell membrane acts as the dielectric of capacitors and the alternating current will pass partly and easily through the extracellular fluid and with greater resistance through the cells. The total impedance to the current is defined by an equivalent circuit of a resistance and a reactance through a capacitor. As the frequency of the alternating current increases, the reactance at first increases and then decreases until at very high frequencies there is little impedance to the passage of the current through cells. An analysis of the reactance to different frequencies is used to calculate extra- and intracellular water. In healthy adults, it is possible to predict total body water within 2 to 3 liters. Much more variable results are observed in diseased patients, owing in part to the population-specific nature of BIA. BIA measures total body water from which the lean body mass is calculated using the assumption that all tissues have the same degree of hydration. Such an assumption will give erroneous results in patients with edema and ascites. Again, the relationship between body composition, as determined by BIA, and clinical outcome in hospitalized patients has not been established.

Dual-Energy X-Ray Absorptiometry

Dual-energy X-ray absorptiometry (DXA) is a method developed originally for the measurement of bone density and mass. DXA can also be used to measure soft tissue composition. DXA can measure total and regional fat, bone mineral, and bone mineral-free lean components. The method is based on the attenuation characteristics of tissues exposed to X rays at two peak energies. Mathematical algorithms allow calculation of body components using various physical and biological models. The method provides an accurate measurement of bone mineral mass and also of soft tissue mass. Again, there are no data indicating whether DXA can predict outcome in hospital patients.

Whole-Body Counting/Neutron Activation

K, N, P, H, O, C, Na, Cl, and Ca can be measured with a group of techniques referred to as whole-body counting/*in vivo* neutron activation analysis. Shielded whole-body counters can count the gamma-ray emission of naturally occurring ^{40}K. The method is safe and can be used in children and pregnant women. The ^{40}K counts can be used to estimate total body potassium, which in turn can be used to calculate body cell mass and fat-free body mass. Prompt gamma neutron activation analysis can be used to measure total body N and H. Nitrogen can be used to calculate total body protein. Delayed gamma neutron activation measures total body Ca, Na, Cl, and P. These elements can be used to calculate bone mineral mass and extracellular fluid volume. Finally, inelastic neutron-scattering methods measure total body O and C. Carbon is useful in models designed to quantify total body fat. Whole-body-counting/neutron activation methods are important because they provide a means of estimating all major chemical components *in vivo*. These methods are considered the standard for evaluating the body composition components of nutritional interest, including body cell mass, fat, fat-free body mass, skeletal muscle mass, and various fluid volumes. Refeeding the malnourished subject by mouth or by TPN results in a rapid increase in TBK but not TBN. In animal studies, it has been shown that this increase in TBK is the result of improved cell membrane voltage and an increase in the intracellular ionic potassium. The findings are consistent with improved cell energetics demonstrated by nuclear magnetic resonance spectroscopy and the improved muscle function shown concurrently. Loss of TBK is a good predictor of poor outcome in a variety of conditions associated with malnutrition.

Computerized Axial Tomography and Magnetic Resonance Imaging

These methods measure components at the tissue-system level of body composition, including skeletal muscle, adipose tissue, visceral organs, and brain. A large number of studies in phantoms, cadavers, and *in vivo* validate these methods. There are no studies of imaging methods that show that they can predict clinical outcome in hospital patients.

TABLE I Nutritional Indices

Index	Features	Classes
Prognostic Nutritional Index (PNI)	PNI = 158 − 1.6 × albumin (g/dl) − 0.78 × TSF (mm) − 0.2 transferrin (mg/dl) − 5.8 × DCH	<40 predicts sepsis
Nutritional Risk Index	NRI = 1.519 × albumin (g/dl) + 41.7 × present weight/usual weight	Normal nutrition >100 Severe malnutrition <83.5
Maastricht Index	MI = 20.68 − 0.24 × albumin (g/liter) − 19.21 × prealbumin (g/liter) − 1.86 × lymphocyte count (10^6/liter) − 0.04 × ideal body weight	>0 Malnourished

Note. Reprinted from Schneider, S. M., and Hebuterne, X. (2000). Use of nutritional scores to predict clinical outcome in chronic disease. *Nutr. Rev.* **58**, 31–38, with permission.

Body Composition and Outcomes

Although the above methods of body composition can accurately assess different components, they are difficult to apply in the clinical setting except in special units. The only methods of nutritional assessment available for wide application in populations are BIA and BIS. There are few data to show that these methods can predict outcome in hospital patients. In patients on renal dialysis, it has been shown that a reduction of the reactance from 70 to 43 Ω increased morbidity by 9% and a reduction of reactance to 31 Ω increased morbidity by 14%. In cancer patients, a lean body mass below the normal range as indicated by BIA was associated with increased morbidity.

NUTRITIONAL INDICES

There are a number of nutritional indices based on measurements of serum proteins, immune competence, and anthropometry. These are the Prognostic Nutritional Index (PNI), a simplified version of the PNI called the Nutritional Risk Index, and the Maastricht Index, all of which predict postoperative complications (Table I).

SUBJECTIVE GLOBAL ASSESSMENT

A clinical method for evaluating nutritional status, termed subjective global assessment (SGA), encompasses historical, symptomatic, and clinical parameters as given below. This approach defines malnourished patients as those who are at increased risk for medical complications and who will presumably benefit from nutritional therapy. The basis of this assessment is to determine whether nutrient assimilation has been restricted because of decreased food intake, maldigestion, or malabsorption, whether reduction in intake has adversely influenced organ function and body composition, and whether the patient's disease has altered nutrient requirements. The specific features of the history and physical examination used in the SGA are listed in Table II.

History

The nutritional history should evaluate the following questions:

1. Has there been a recent change in body weight and was the change intentional or unintentional? The percentage of body weight lost in the previous 6 months is characterized as mild (<5%), moderate (5–10%), or severe (>10%). The pattern of loss is also important and it is possible for a patient to have significant weight loss but still be considered well nourished if body weight (without edema or ascites) recently increased. For example, a patient who has had a 10% body weight loss but regained 3% of that weight over the past month would be considered well nourished.

2. Is dietary intake adequate? The patient should be questioned about his or her habitual diet and any change in diet pattern. Have the number, size, and contents of meals changed? Are nutrient supplements being taken? A diary documenting food intake may be useful when the history is inconclusive.

3. What is the reason(s) for the change in dietary intake? Has appetite changed? Is there a disturbance in taste, smell, or the ability to chew or swallow food? Has there been a change in mental status or increased depression? Has there been a change in the ability to prepare meals? Are there gastrointestinal symptoms, such as early satiety, postprandial pain, nausea, or vomiting? Is the patient taking medications that affect food intake?

4. The presence of persistent gastrointestinal symptoms, such as anorexia, nausea, vomiting,

TABLE II Subjective Global Assessment

A. History
  1. Weight and height
    Current height ____ cm, weight ____ kg
    Overall weight loss in past 6 months: ____ kg, ____%
    Change in weight in past 2 weeks (use + or −): ____ kg, ____%
  2. Dietary intake change (relative to usual intake)
    No change
    Change: Duration = ____ days
      Type: Suboptimal solid diet
        Hypocaloric liquids
        Starvation
        Supplement (circle): Nil, vitamin, minerals
  3. Gastrointestinal symptoms that persisted for >2 weeks
    None
    Nausea
    Vomiting
    Diarrhea
    Pain (circle): At rest, on eating
  4. Functional capacity
    No dysfunction
    Dysfunction: Duration ____ days
      Type: Working suboptimally
        Ambulatory but not working
        Bedridden
  5. Disease and its relation to nutritional requirements
    Primary diagnosis: ____
    Metabolic demand (stress)
      No stress
      Moderate stress
      High stress (burns, sepsis, severe trauma)
B. Physical status (for each trait specify: 0 = normal, 1 = mild deficit, 2 = established deficit)
    Loss of subcutaneous fat
    Muscle wasting
    Edema
    Ascites
    Mucosal lesions
    Cutaneous and hair changes
SGA Grade ____

diarrhea, and abdominal pain, which have occurred almost daily for at least 2 weeks is recorded. Is there evidence of malabsorption?

5. The patient's functional capacity is defined as bedridden, suboptimally active, or full capacity. Are there symptoms of specific nutrient deficiencies including macrominerals, micronutrients, and water?

6. The metabolic demands of the patient due to the underlying disease state should be determined. Examples of high-stress illnesses are burns, major trauma, and severe inflammation, such as acute colitis. Moderate-stress diseases might be a mild infection or limited malignant tumor.

Physical Examination

The physical examination corroborates and adds to the findings obtained by history:

Anthropometric Assessment

Current body weight should be compared with previously recorded weights, if available. Weight for height should be compared with standard normal values. A search for evidence demonstrating depletion of body fat and muscle mass should be made. Clearly defined bony, muscular, and venous outlines and loose skin folds can judge a general loss of adipose tissue. A fold of skin, pinched between the forefinger and thumb, can detect the adequacy of subcutaneous fat. The presence of hollow cheeks, buttocks, and perianal area suggests body fat loss. An examination of the temporalis, deltoids, and quadriceps muscles should be made to search for muscle wasting.

Assessment of Muscle Function

Strength testing of individual muscle groups should be made to evaluate generalized and localized muscle weakness. In addition, a general evaluation of respiratory and cardiac muscle function should be made.

Fluid Status

An evaluation for dehydration (hypotension, tachycardia, postural changes, mucosal xerosis, dry skin, and swollen tongue) and excess body fluid (edema, ascites) should be made.

Evaluation for Specific Nutrient Deficiencies

Rapidly proliferating tissues, such as oral mucosa, hair, skin, and bone marrow, are often more sensitive to nutrient deficiencies than are tissues that turn over more slowly.

In contrast to the traditional approach, the SGA identifies the initial nutritional state and the interplay of the factors influencing the progression or regression of nutritional abnormalities. Therefore, SGA is a dynamic process that is not limited to a single snapshot at the moment of measurement but provides a picture of current nutritional status and insight into the patient's future status. The clinical assessment of nutritional status involves a focused history and physical examination. The information is then used as a composite observation to judge whether the intake and quality of food eaten, the status of the gastrointestinal tract, metabolic requirements, and existing habitus are critical or would result in critical malnutrition in the future if nutritional support is not given.

It is a mistake to create many subcategories of nutritional status. The findings from the patient history and physical examination are used to categorize patients as follows: being well nourished and not having risk factors for progressive malnutrition (category A); currently having no significant wasting or minimal weight loss but having changes in intake, absorption, stress, and function, as judged globally, that will induce, and cause the patient to progress toward, increased weight loss and functional deficit (category B); or significant weight loss, functional deficit and risk factors for progression (category C). The rank is assigned on the basis of subjective weighing. Equivocal information is given less weight than definitive data. Fluid shifts related to onset or treatment of edema or ascites must be considered when interpreting changes in body weight. In general, a patient who has experienced weight loss and muscle wasting but is currently eating well and is gaining weight is classified as well nourished. A patient who has experienced moderate weight loss, continued compromised food intake, continued weight loss, and progressive functional impairment and who has a "moderate-stress" illness is classified as moderately malnourished. A patient who has experienced severe weight loss and who continues to have poor nutrient intake, progressive functional impairment, and muscle wasting is classified as severely malnourished independent of disease stress. Several studies have shown that the use of SGA in evaluating hospitalized patients gives reproducible results and there was more than 80% agreement when two blinded observers assessed the same patient.

Such a structured approach allows an assessment of whether there is a role for nutritional support or whether the weight loss and fatigue are entirely due to disease. For example, for a patient with thyrotoxicosis, increased appetite, and no gastrointestinal symptoms restricting intake but only signs of hypermetabolism, the role of nutritional support is minimal. In contrast, a person with marked weight loss due to obstruction caused by upper gastrointestinal cancer with compromised swallowing, marked functional deficit, and poor intake will benefit from nutritional support. On the basis of a technical review of the literature, the American Gastroenterological Society stated that "preoperative parenteral nutrition has a more profound effect in reducing the rate of major postoperative complications in patients undergoing major surgery for cancer of the esophagus or stomach." On the basis of the same review, they concluded that "[p]reoperative parenteral nutrition should also be considered in those patients who are severely malnourished (as defined in the technical review). A retrospective subgroup analysis of one large RCT suggests that it does have a greater therapeutic effect in decreasing postoperative complications in this group."

Illustrative Cases

Case 1: A 60-year-old woman was admitted to the hospital for elective resection of a colon carcinoma. She had lost 10% of her initial weight over 8 months before admission. However, she recently gained weight after therapy with nutritional supplements was initiated. She continued to work and was active. On physical examination, there was no loss of muscle or fat. She is SGA A.

Case 2: A 40-year-old man with an acute exacerbation of Crohn's disease had lost 10% of his body weight within the previous 2 weeks and was ingesting mostly liquids to avoid gastrointestinal discomfort. He was ambulatory but was not going to work. On physical examination, he had slight loss of subcutaneous tissue manifested by a reduced buccal fat pad and loose skin folds over the arms. He is SGA B.

Case 3: A 67-year-old man with esophageal cancer had minimal food intake for almost 3 months. He lost 15% of his body weight during the previous 4 months and is continuing to lose weight. He was able to move around the house but had marked muscle weakness and fatigue and did not walk outdoors. On physical examination, he lacked subcutaneous tissue and had hollow temples, deltoid wasting, and mild pitting edema. He is SGA C.

SGA has been shown to be a better predictor of postoperative infectious complications than serum albumin, serum transferrin, delayed cutaneous, hypersensitivity, anthropometry, creatinine–height index, and the prognostic nutritional index. How does SGA perform in predicting complications in conditions other than preoperative patients? SGA has been shown to predict complications in general surgical patients, cancer patients, patients on dialysis, and liver transplant patients. Two recent large-scale studies by Perman *et al*. and Pirlich *et al*. have shown that SGA independently identifies the increased mortality and morbidity from malnutrition even when the data are statistically adjusted for other factors influencing survival and complications.

GERIATRIC NUTRITIONAL ASSESSMENT

SGA is a reliable tool for the assessment of malnutrition in the elderly. In addition, two other tools have been developed for the assessment of the nutritional status in the elderly. They are the Mini Nutritional Assessment (MNA) and the Malnutrition Risk Scales (SCALES). The MNA consists of 18 items. Weight, height, and weight

loss are recorded. Questions are asked regarding lifestyle, medications, and degree of mobility. Then, a diet history of number of meals, food and fluid intake, and ability to self-feed is recorded. Finally, a self-assessment of health and nutritional status is noted. The answers to these questions are converted into a score. A score ≥ 24 is considered well nourished, a score of 17–24 indicates a risk of malnutrition, and a score <17 indicates that the patient is malnourished. A score below 17 has been associated with increased mortality and a longer hospital stay. SCALES is composed of sadness, cholesterol, albumin <40 g/liter, loss of weight, eating problems, and shopping problems. Each component that is abnormal is given a point. If the sum of the points is ≥ 3, then there is a high risk of malnutrition.

FUNCTIONAL TESTS OF MALNUTRITION

It was shown that muscle function and mitochondrial function are very sensitive to nutritional manipulations and are influenced before changes in body composition. The measurement of muscle function has been shown to predict complications in a number of studies. The association between outcome and muscle function was shown by Klidjian using hand-grip strength and by Windsor and Hill using stimulated and respiratory muscle function. Elderly subjects with a hand-grip strength of <5 kg have a very high mortality. Functional tests (the simplest in practice is the test of hand grip strength) have been shown to predict NACs. They are also a snapshot in time and are best used to follow the effects of nutritional support. Since they improve before changes in body composition, they could be used to determine whether the individual is progressively responding to nutrition before there are changes in body composition. They could be used to examine corroborate the functional changes observed during evaluation of SGA.

See Also the Following Articles

Diet and Environment, Role in Colon Cancer • Dietary Reference Intakes (DRI): Concepts and Implementation • Digestion, Overview • Malnutrition • Nutrient Transport, Regulation of • Nutrition in Aging • Pancreas, Nutritional Effects on the • Parenteral Nutrition • Protein-Calorie Deficiency—"Kwashiorkor"

Further Reading

American Gastroenterological Association (2001). American Gastroenterological Association Medical Position Statement: Parenteral nutrition. *Gastroenterology* **121**, 966–969.

Briet, F., Twomey, C., and Jeejeebhoy, K. N. (1999). Refeeding increases mitochondrial complex activity in malnourished patients: Evidence for a sensitive nutritional marker. *Gastroenterology* **116**, A541.

Detsky, A. S., McLaughlin, J. R., Baker, J. P., Johnston, N., Whittaker, S., Mendelson, R. A., and Jeejeebhoy, K. N. (1987). What is subjective global assessment of nutritional status? *JPEN* **11**, 8–13.

Enia, G., Sicuso, C., Alati, G., and Zoccali, C. (1993). Subjective global assessment of nutrition in dialysis patients. *Nephrol. Dial. Transplant.* **8**, 1094–1098.

Fenton, S. S. A., Johnston, N., and Delmore, T. (1987). Nutritional assessment of continuous ambulatory peritoneal dialysis patients. *Trans. Am. Soc. Artif. Organs* **23**, 650–653.

Guigoz, Y., Vellas, B., and Garry, P. J. (1996). Assessing the nutritional status of the elderly: The mini nutritional assessment as part of the geriatric assessment. *Nutr. Rev.* **54**, S59–S65.

Hasse, J., Strong, S., Gorman, M. A., and Liepa, G. (1993). Subjective global assessment: Alternative nutrition-assessment technique for liver-transplant candidates. *Nutrition* **9**, 339–343.

Hirsch, S., de Obaldia, N., Petermann, M., Rojo, P., Barrientos, C., Iturriaga, H., and Bunout, D. (1991). Subjective global assessment of nutritional status: Further validation. *Nutrition* **7**, 35–37.

Koretz, R. L., Lipman, T. O., and Klein, S. (2001). AGA Technical Review: Parenteral nutrition. *Gastroenterology* **121**, 970–1001.

Perman, M., Crivelli, A., and Khoury, M. (2002). Nutritional prognosis in hospitalized patients. *Am. J. Clin. Nutr.* **75**, 426S.

Pikul, J., Sharpe, M. D., Lowndes, R., and Ghent, C. N. (1994). Degree of preoperative malnutrition is predictive of postoperative morbidity and mortality in liver transplant recipients. *Transplantation* **57**, 469–472.

Pirlich, M., and Lochs, H. (2001). Nutrition in the elderly. *Best Pract. Res. Clin. Gastroenterol.* **15**, 869–884.

Pirlich, M., Schuetz, T., Gasttell, S., and Lochs, H. (2002). Malnutrition affects long-term prognosis in hospitalized patients. *Gastroenterology* **122**(Suppl.), A636.

The Veterans Affairs Total Parenteral Nutrition Cooperative Study Group (1991). Perioperative total parenteral nutrition in surgical patients. *N. Engl. J. Med.* **325**, 525–532.

Windsor, J. A., and Hill, G. L. (1988). Weight loss with physiologic impairment: A basic indicator of surgical risk. *Ann. Surg.* **207**, 290–296.

Obesity, Treatment of

GEORGE A. BRAY
Pennington Biomedical Research Center, Baton Rouge, Louisiana

body mass index A measure of normal weight and overweight [weight (in kilograms)]/[height (in centimeters)]^2.
gastric bypass Construction of a small (10–30 ml) proximal gastric pouch by stapling across the stomach or complete resection from the rest of the stomach, connected to a segment of jejunum brought up at a Roux-en-Y limb.
obesity Body mass index is >30 kg/m^2.
overweight Body mass index in the range of 25–29.9 kg/m^2.

Obesity produces complications that involve nearly all major organ systems. The major gastrointestinal medical complications associated with obesity include gallstones, nonalcoholic fatty liver disease, and possibly gastroesophageal reflux disease.

INTRODUCTION

As the epidemic of obesity has grown over the past 20 years, the need to understand its causes and apply this knowledge to the treatment of obesity has become ever more urgent. As medical science's understanding of obesity increases, novel ideas about techniques, strategies, molecules, and procedures that can be used to curtail the devastating effects of this epidemic will emerge. This article will focus on five different approaches to the treatment of obesity. These are behavioral therapy, diet, exercise, medications, and surgery. As there is only space for a brief discussion of each, the reader is referred to other sources for more details (see Further Reading). Since obesity is an imbalance between the energy that is ingested and the energy that is expended, any treatment must have an effect on one or both sides of the energy balance equation. This concept is easily applied to diet, which aims to reduce food intake, and exercise, which aims to raise energy expenditure. The subtleties of this concept, when applied to medications, are more complex.

BEHAVIOR THERAPY

Over the past 30 years, there has been a steady improvement in behavioral techniques that were originally introduced only in 1967 by Stuart in a classic paper. The basic elements of any behavioral program focus on three things: identifying the antecedents to eating and changing them; characterizing an individual's eating style and improving it; and providing rewards for improved eating behavior. With programs that last up to 24 weeks, the retention is usually more than 75% and weight losses can be up to 10% of initial body weight.

Retrospective analysis has shown that several strategies are particularly useful. These include self-monitoring of one's eating habits, eating a lower fat diet, being more physically active, and developing a support system to help reward successful behavior. Providing structured meal plans has proven to be an additional useful strategy. These can be meal plans, written menus, or portion-controlled foods. The use of the internet has also proven to be a potentially valuable approach to delivering behavioral therapy. The major problem with behavioral strategies is that patients tend to regain weight after leaving the program. Recent focus has concentrated on ways of continuing the impact of the program over longer periods of time.

Overweight adolescents are a good target for preventive strategies using behavior therapy, because good 10-year data show that intervention for this group can reduce the degree of overweight into adult life. Data on the efficacy of behavior programs carried out in controlled settings show that weight losses average nearly 10% in trials lasting more than 16 weeks. The problem with behavioral therapy is that regaining weight once the behavior treatment ends is frequent. At least one long-term study showed that behavioral therapy could provide long-term weight loss, providing it was continued.

DIET

Calories

The essential element of any diet is that it reduces the total caloric intake. Many different diets have been recommended over the past 150 years since Mr. Banting published his famous pamphlet in 1863. One of the most

 Copyright 2004, Elsevier (USA). All rights reserved.

popular themes has been the low-carbohydrate diet. The fact that new diets continue to appear suggests that none of them have a magic formula because if one did, none of the others would have succeeded in the marketplace.

Fat

There are a number of elements of an effective diet that anyone can incorporate with potential benefit. The observation from behavioral therapy that a low-fat diet is what successful patients adopt suggests that this would be a useful strategy. Several reviews have shown that low-fat diets are associated with weight loss. The initial weight is one variable predicting the magnitude of the response (overweight people lose more than do normal weight people). A second variable is the amount of fat removed from the diet—the greater the reduction, the more weight is lost. Palatability of the diet may be another factor. A recent study showed that replacing fat with olestra, a nondigestible fat substitute, reduced body weight by 6% and body fat by nearly 20% over 9 months, whereas a standard reduced-fat diet with the same available fat was relatively ineffective. However, a very-low-fat diet can be difficult for many people to adhere to. For this reason, the best advice is to reduce saturated and *trans*-fatty acids and to lower total fat to as near 25% as possible for the individual.

Carbohydrate and Fiber

Carbohydrate and both digestible and indigestible fiber intake can affect food intake. Although recent studies show that substituting starch for sugar does not produce greater weight loss, the type of digestibility of that carbohydrate may play a role. The glycemic load, which is the glycemic index times the carbohydrate intake, may be important. In the Nurses Health Study, the glycemic load was related to the risk of developing heart disease and diabetes. Diets with a low glycemic index, i.e., with a lower rise in glucose, produce more satiety than do diets with a high glycemic index.

Breast Feeding

Several recent studies suggest that the length of breastfeeding affects childhood obesity. In a large German study of more than 11,000 children, von Kries *et al*. showed that the duration of breastfeeding as the sole source of nutrition was inversely related to the incidence of obesity, defined as a weight above the 95th percentile, when children entered the first grade. In this study, the incidence was 4.8% in children with no breastfeeding, falling in a graded fashion to 0.8% in children who were fed solely from the breast for 12 months or more. A second large report also showed that breastfeeding reduced the incidence of overweight adolescents. The third report with fewer subjects and more ethnic heterogeneity failed to show this effect. However, the possibility that lengthening the duration of breastfeeding could reduce the future risk of obesity is another reason to recommend breastfeeding for at least 6 to 12 months.

Dietary Calcium

Nearly 20 years ago, McCarron *et al*. reported that there was a negative relationship between body mass index and dietary calcium intake in the data collected by the National Center for Health Statistics. More recently, in 1999, Zemel *et al*. found that there was a strong inverse relationship between calcium intake and the risk of being in the highest quartile of body mass index. These studies have prompted a reevaluation of clinical trials that have given calcium supplements orally. In the prospective trials, subjects receiving calcium had a greater weight loss than did subjects receiving placebos. Increasing calcium from 0 to nearly 2000 g/day was associated with a reduction in body mass index (BMI) of approximately 5 BMI units. These data might suggest that low calcium intake plays a role in the current epidemic of obesity.

EXERCISE

Exercise is an obvious way to increase energy expenditure. This is desirable for all Americans and would be beneficial for reducing cardiovascular risks. However, for many people who are overweight, exercise can be a challenge, because they are already expending more energy doing everyday activities. Thus, for many such patients, a simple walking program can be the best approach. That this is valuable is shown by the Diabetes Prevention Program, where a lifestyle program of diet and exercise produced a 58% reduction in the conversion of patients with impaired glucose tolerance to diabetes mellitus.

With substantial amounts of supervised exercise, significant weight loss can be obtained. However, for many people, exercise adds little extra weight loss to a dietary program aimed at weight loss, probably because people do not maintain the amount of exercise that is prescribed. Some reports suggest that exercise may conserve body protein during dieting, but others do not.

The most beneficial part of an exercise program comes when trying to maintain a lower body weight. In a survey of people who were successful at maintaining

weight, exercise was maintained at a level significantly above that in people without a weight problem.

MEDICATIONS

Medication should be seriously considered for clinically overweight individuals, defined as individuals with a body mass index above 30 kg/m^2, or those with diabetes, impaired glucose tolerance, hypertension, or heart disease combined with a BMI above 27 kg/m^2. Two strategies can be used to treat individuals who are clinically overweight. If they have co-morbidities, individual drugs can be used to treat each co-morbidity; i.e., patients can be treated for their diabetes, hypertension, dyslipidemia, and sleep apnea. Alternatively, or in addition, patients with a BMI > 30 kg/m^2 could be treated with anti-obesity drugs. Current drugs include appetite suppressants that act on the central nervous system and orlistat, which blocks pancreatic lipase. The availability of these agents differs from country to country and any physician planning to use them should be familiar with the local regulations. Most of the drugs on the market were reviewed and approved more than 20 years ago and are approved for short-term use only. The basis for the short-term use is twofold. First, almost all the studies of these agents are short-term. Second, the regulatory agencies are concerned about the potential for abuse and thus have restricted their prescription to "only a few weeks," which is usually interpreted as up to 12 weeks. The withdrawal of fenfluramine and dexfenfluramine from the market in 1997 following the development of valvular heart disease in patients treated with these drugs further compounds the concern of health authorities about the safety of appetite suppressants. Because of the regulatory limitations and the lack of longer-term data on safety and efficacy, the use of the drugs approved for short-term treatment must be carefully justified. These drugs may be useful in initiating treatment and in helping a patient who is relapsing, but only for a few weeks.

Sibutramine (Meridia; Reductil) is approved in many countries for long-term use. The evidence shows that weight loss of 10% or more can be produced with this drug. The side effect profile includes dry mouth, asthenia, insomnia, and constipation. It also produces a small increase in heart rate of 2–5 beats per minute and a small rise in blood pressure of 2–4 mm Hg. Clinical data show no evidence of valvulopathy. Blood pressure should be followed carefully and the drug may be inappropriate in patients with stroke, congestive heart failure, or recent myocardial infarction. It should not be used with other serotonergic drugs or drugs that inhibit monoamine oxidase.

Orlistat (Xenical), a drug that blocks intestinal lipase, has been approved for long-term use in many countries. In clinical trials lasting up to 4 years, orlistat was associated with a mean weight loss of up to 10% at the end of 1 year in patients who were prescribed a 30% fat diet. As might be expected, because the drug blocks pancreatic lipase in the intestine, fecal fat loss is increased. Major side effects related to the release of undigested triglyceride occurred in some patients, usually within the first month, and were reported with reduced frequency over time, indicating that patients learned to use the drug effectively in relation to dietary intake of fat. The effective use of this medication requires that physicians and their staff members provide good dietary counseling to patients.

The combination of ephedrine and caffeine is a third compound for which a randomized clinical trial has been published by Astrup *et al*. Over 6 months, the patients treated with ephedrine and caffeine lost more weight than did the placebo-treated group or the two groups treated individually with ephedrine or caffeine. This research has been used for marketing of herbal preparations that contain ephedra alkaloids with or without caffeine. Two clinical trials by Boozer *et al*. have shown significantly greater weight loss with an herbal ephedra/caffeine preparation than with placebo.

The epidemic of obesity, the discovery of genes that produce obesity, and a slim armamentarium for treatment of obesity have spurred many pharmaceutical companies to search for new agents. These agents can be divided into two categories: compounds that are in clinical trials with suggestive data and new molecules just entering the clinical arena. In the former category are bupropion, leptin, and topiramate. Bupropion is approved by the Food and Drug Administration (FDA) as an antidepressant. Two recent studies show that it produces weight loss. Leptin is the peptide produced primarily in adipose tissue. When deficient, it produces massive obesity. Leptin-deficient patients respond to leptin with weight loss, suggesting that leptin might be clinically useful if the proper way of delivering it can be found. The third molecule, topiramate, is an anticonvulsant approved by the FDA for this purpose. Several studies now suggest that topiramate produces weight loss in patients receiving it for their epilepsy. Clinical trials with this compound are currently under way.

A number of other molecular targets could serve as the basis for clinically useful drugs. These molecules can be divided into those that have a peripheral mechanism of action and those that act on the brain. Thermogenic β-adrenergic receptors are one target that have been investigated for more than 15 years, but as yet

no clinically useful molecule has been identified. Cholecystokinin reduces food intake in animals and humans. To date, no useful molecules that work through this receptor have been identified. Glucagon-like peptide 1 (GLP-1) is processed from the enteroglucagon molecule by gastrointestinal (GI) cells. Infusion of GLP-1 into lean and obese human beings reduces food intake and molecule agonists to this receptor system are under development. Ghrelin, produced primarily in the stomach, has been recently identified and shown to increase food intake when given peripherally or centrally. Antagonists to this receptor system offer potential for future molecules with which to treat obesity. Finally, the pentapeptide enterostatin, cleaved from pro-colipase in the intestinal lumen, reduces food intake and primarily fat intake. This molecule or similar molecules might have interesting potential for modulation of clinical fat intake.

Neuropeptide Y is one of the most potent stimulators of food intake when it is injected into the brain. Several abstracts suggest that this may be a useful target, with molecules aimed at either the Y-1 or the Y-2 receptor. Loss of the melanocortin-4 receptor produces massive obesity in mice, suggesting that molecules aimed at this receptor could be useful. The endogenous signal for this system is probably α-melanocyte-stimulating hormone (α-MSH), which is produced from proopiomelanocortin. Another endogenous molecule, agouti-related peptide (AGRP), inhibits the effect of α-MSH. Modulation of the receptors for AGRP or α-MSH would be two potential targets. The peptides mentioned above are under the control of circulating leptin acting on receptors in the arcuate nucleus of the hypothalamus. Cocaine–amphetamine-regulated transcript (CART) is a fourth molecule in the brain that modulates, and reduces, food intake. A receptor agonist that mimicked CART might also be an attractive target. At least two other hypothalamic peptides need to be considered in this discussion. The first is melanocortin-concentrating hormone. Animals overexpressing this peptide are heavier than controls and disabling the production of this peptide produces leanness. Other hypothalamic peptides are corticotropin-releasing hormone (CRH) and urocortin, two variants that differentially bind to CRH R1 and CRH R2. CRH reduces food intake and modulating its activity would be another potential mechanism to reduce food intake.

SURGERY

Surgical intervention was initiated more than 40 years ago with operations that shortened the absorptive surface available to GI contents by various bypass operations. At present, the principal operations are the gastric bypass, the vertical banded gastroplasty, and the gastric band, which allows a constrictive band around the stomach to be expanded by injecting saline into a subcutaneous reservoir. With the introduction of laparoscopic surgery and advancement in the skill of surgeons using this surgery during the 1990s, the safety of these procedures improved. Although originally recommended for people with a body mass index above 40 kg/m^2, the studies showed marked benefits to patients having this surgery, which suggests that the BMI should be reduced to 35 kg/m^2 or even lower if there are significant risks associated with the obesity.

BENEFITS OF INTENTIONAL WEIGHT LOSS

Because the weight loss after gastric surgery is more sustained than after other methods of weight reduction, much of the data on the benefits of weight loss are derived from those patients. Although the data are not entirely clear, studies indicate that intentional weight loss $\geq$20 pounds has been associated with an approximately 25% reduction in obese and overweight patients who have an obesity-related illness, especially type 2 diabetes mellitus. Other benefits of modest sustained weight loss include improved cardiovascular function, better control of type 2 diabetes mellitus, dyslipidemia, and hypertension. Improved respiratory function with less obstructive sleep apnea and improved reproductive and urinary tract function in women require considerably more weight loss.

Although the relationship between obesity and factors that dispose to gastroesophageal reflux disease is not clear, significant weight loss after gastric surgery caused resolution of symptoms, even before significant weight loss occurred. Gallstones may increase in incidence in obesity and rapid weight loss increases bile cholesterol supersaturation, increasing the risk of new gallstones. Nonalcoholic fatty liver disease (NAFLD) is associated with the metabolic syndrome (increased insulin, hypertension, and hyperlipidemia), perhaps as an earlier stage before obesity or diabetes mellitus develops. A gradual weight loss may improve hepatic chemistries and liver size, but rapid weight loss (e.g., after gastric surgery) may exacerbate NAFLD.

See Also the Following Articles

Appetite • Diabetes Mellitus • Dietary Reference Intakes (DRI): Concepts and Implementation • Gastric Surgery • Minimally Invasive Surgery • Prader–Willi Syndrome • Satiety

Further Reading

Bray, G. A., Blackburn, G. L., Ferguson, J. M., Greenway, F. L., Jain, A. K., Mendel, C. M., Mendels, J., Ryan, D. H., Schwartz, S. L., Scheinbaum, M. L., and Seaton, T. B. (1999). Sibutramine produces dose-related weight loss. *Obesity Res.* **7**, 189–198.

Bray, G. A., and Greenway, F. L. (1999). Current and potential drugs for treatment of obesity. *Endocr. Rev.* **20**, 805–875.

Bray, G. A., and Popkin, B. M. (1998). Dietary fat intake does affect obesity! *Am. J. Clin. Nutr.* **68**, 1157–1173.

Bray, G. A., and Tartaglia, L. A. (2000). Medicinal strategies in the treatment of obesity. *Nature* **404**, 672–677.

Davies, K. M., Heaney, R. P., Recker, R. R., Lappe, J. M., Barger-Lux, M. J., Rafferty, K., and Hinders, S. (2000). Calcium intake and body weight. *J. Clin. Endocrinol. Metab.* **85**, 4635–4638.

Freedman, M. R., King, J., and Kennedy, E. (2001). Popular diets: A scientific review. *Obesity Res.* **9**, 1S–40S.

Klein, S., Wadden, T., and Sugerman, H. J. (2002). AGA technical review on obesity. *Gastroenterology* **123**, 882–932.

Ravussin, E., and Bouchard, C. (2000). Human genomics and obesity: Finding appropriate drug targets. *Eur. J. Pharmacol.* **410**, 131–145.

Roberts, S. B. (2000). High-glycemic index foods, hunger, and obesity: Is there a connection? *Nutr. Rev.* **58**, 163–169.

Sjostrom, C. D., Lissner, L., Wedel, H., and Sjostrom, L. (1999). Reduction in incidence of diabetes, hypertension and lipid disturbances after intentional weight loss induced by bariatric surgery: The SOS Intervention Study. *Obesity Res.* **7**, 477–484.

Sjostrom, L., Rissanen, A., Andersen, T., Boldrin, M., Golay, A., Koppeschaar, H. P., and Krempf, M. (1998). Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. *Lancet* **352**, 167–172.

Tuomilehto, J., Lindstrom, J., Eriksson, J. G., Valle, T. T., Hamalainen, H., Ilanne-Parikka, P., Keinanen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M., Salminen, V., and Uusitupa, M. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **344**, 1343–1350.

von Kries, R., Koletzko, B., Sauerwald, T., von Mutius, E., Barnert, D., Grunert, V., and von Voss, H. (1999). Breast feeding and obesity: Cross sectional study. *Br. Med. J.* **319**, 147–150.

Wing, R. (2002). Behavioral approaches to the treatment of obesity. *In* "Handbook of Obesity" (G. A. Bray and C. Bouchard, eds.), 2nd Ed. Dekker, New York, in press.

Occult Gastrointestinal Bleeding

DON C. ROCKEY
Duke University Medical Center

capsule wireless endoscopy Use of a small camera containing a capsule that is ingested to evaluate the gastrointestinal tract. Typically used to investigate bleeding sources in the small intestine.

hematemesis Vomiting of blood; indicates an upper gastrointestinal site of bleeding.

hematochezia Bright red blood per rectum; may or may not be mixed with stool.

Meckel's diverticulum A 2- to 6-cm outpocketing of the ileum along the antimesenteric border, usually located 50 to 180 cm proximal to the ileocecal valve.

melena Tarry, foul-smelling black stools.

Gastrointestinal bleeding is one of the most common problems in medicine. Bleeding from the upper and lower gastrointestinal tracts is recognized as a major source of disease-related morbidity and, on occasion, mortality. Occult gastrointestinal bleeding is even more common than upper and lower gastrointestinal bleeding. Patients may present with bleeding that is truly hidden (occult); this form of bleeding is extremely common and can even be considered to be pervasive. Patients may instead have clinically obvious bleeding, but the blood is from a source in the gastrointestinal tract that is difficult to identify; this type of occult bleeding has been termed "obscure gastrointestinal bleeding." The clinical spectrum of occult gastrointestinal bleeding is therefore expansive and may encompass any of a number of clinical scenarios. Because patients may present with massive obvious bleeding or may be entirely unaware that they are bleeding, gastrointestinal bleeding afflicts inpatients as well as outpatients.

INTRODUCTION

Protocols for evaluation and management of patients with gastrointestinal bleeding have undergone many recent advances. Nonetheless, the fundamental clinical

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Bray, G. A., Blackburn, G. L., Ferguson, J. M., Greenway, F. L., Jain, A. K., Mendel, C. M., Mendels, J., Ryan, D. H., Schwartz, S. L., Scheinbaum, M. L., and Seaton, T. B. (1999). Sibutramine produces dose-related weight loss. *Obesity Res.* **7**, 189–198.

Bray, G. A., and Greenway, F. L. (1999). Current and potential drugs for treatment of obesity. *Endocr. Rev.* **20**, 805–875.

Bray, G. A., and Popkin, B. M. (1998). Dietary fat intake does affect obesity! *Am. J. Clin. Nutr.* **68**, 1157–1173.

Bray, G. A., and Tartaglia, L. A. (2000). Medicinal strategies in the treatment of obesity. *Nature* **404**, 672–677.

Davies, K. M., Heaney, R. P., Recker, R. R., Lappe, J. M., Barger-Lux, M. J., Rafferty, K., and Hinders, S. (2000). Calcium intake and body weight. *J. Clin. Endocrinol. Metab.* **85**, 4635–4638.

Freedman, M. R., King, J., and Kennedy, E. (2001). Popular diets: A scientific review. *Obesity Res.* **9**, 1S–40S.

Klein, S., Wadden, T., and Sugerman, H. J. (2002). AGA technical review on obesity. *Gastroenterology* **123**, 882–932.

Ravussin, E., and Bouchard, C. (2000). Human genomics and obesity: Finding appropriate drug targets. *Eur. J. Pharmacol.* **410**, 131–145.

Roberts, S. B. (2000). High-glycemic index foods, hunger, and obesity: Is there a connection? *Nutr. Rev.* **58**, 163–169.

Sjostrom, C. D., Lissner, L., Wedel, H., and Sjostrom, L. (1999). Reduction in incidence of diabetes, hypertension and lipid disturbances after intentional weight loss induced by bariatric surgery: The SOS Intervention Study. *Obesity Res.* **7**, 477–484.

Sjostrom, L., Rissanen, A., Andersen, T., Boldrin, M., Golay, A., Koppeschaar, H. P., and Krempf, M. (1998). Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. *Lancet* **352**, 167–172.

Tuomilehto, J., Lindstrom, J., Eriksson, J. G., Valle, T. T., Hamalainen, H., Ilanne-Parikka, P., Keinanen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M., Salminen, V., and Uusitupa, M. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **344**, 1343–1350.

von Kries, R., Koletzko, B., Sauerwald, T., von Mutius, E., Barnert, D., Grunert, V., and von Voss, H. (1999). Breast feeding and obesity: Cross sectional study. *Br. Med. J.* **319**, 147–150.

Wing, R. (2002). Behavioral approaches to the treatment of obesity. *In* "Handbook of Obesity" (G. A. Bray and C. Bouchard, eds.), 2nd Ed. Dekker, New York, in press.

Occult Gastrointestinal Bleeding

DON C. ROCKEY
Duke University Medical Center

capsule wireless endoscopy Use of a small camera containing a capsule that is ingested to evaluate the gastrointestinal tract. Typically used to investigate bleeding sources in the small intestine.

hematemesis Vomiting of blood; indicates an upper gastrointestinal site of bleeding.

hematochezia Bright red blood per rectum; may or may not be mixed with stool.

Meckel's diverticulum A 2- to 6-cm outpocketing of the ileum along the antimesenteric border, usually located 50 to 180 cm proximal to the ileocecal valve.

melena Tarry, foul-smelling black stools.

Gastrointestinal bleeding is one of the most common problems in medicine. Bleeding from the upper and lower gastrointestinal tracts is recognized as a major source of disease-related morbidity and, on occasion, mortality. Occult gastrointestinal bleeding is even more common than upper and lower gastrointestinal bleeding. Patients may present with bleeding that is truly hidden (occult); this form of bleeding is extremely common and can even be considered to be pervasive. Patients may instead have clinically obvious bleeding, but the blood is from a source in the gastrointestinal tract that is difficult to identify; this type of occult bleeding has been termed "obscure gastrointestinal bleeding." The clinical spectrum of occult gastrointestinal bleeding is therefore expansive and may encompass any of a number of clinical scenarios. Because patients may present with massive obvious bleeding or may be entirely unaware that they are bleeding, gastrointestinal bleeding afflicts inpatients as well as outpatients.

INTRODUCTION

Protocols for evaluation and management of patients with gastrointestinal bleeding have undergone many recent advances. Nonetheless, the fundamental clinical

Copyright 2004, Elsevier (USA). All rights reserved.

issues surrounding acute gastrointestinal bleeding remain unchanged. First, attention must be directed at the patient's hemodynamic status. Subsequently, the source of bleeding should be determined, any active bleeding should be stopped, and recurrent bleeding should be prevented.

CLINICAL PRESENTATION

Types of Manifest Bleeding

Clinical signs of gastrointestinal bleeding reflect the site, etiology, and rapidity of bleeding. Clinically obvious bleeding from the gastrointestinal (GI) tract is usually manifest as hematemesis, melena, melenemesis, or hematochezia (coffee-ground emesis). Hematemesis is defined as the vomiting of blood and indicates an upper gastrointestinal site of bleeding—essentially a site proximal to the ligament of Treitz. Such blood may be fresh (bright red) or old (with the appearance of coffee grounds). Melena describes black stools that are tarry and foul smelling. The black, tarry character of melena is classic and must not be confused with the greenish character of ingested iron or the black, non-foul-smelling stool typical of that caused by ingestion of bismuth (i.e., as in bismuth salicylate compounds such as Peptobismol); it is important to review the patient history. Hematochezia refers to bright red blood per rectum that may or may not be mixed with stool. Occult bleeding (bleeding that is not noted by the patient) is usually diagnosed by tests that detect small amounts of blood in the stool (i.e., guaiac tests) or because of iron deficiency anemia.

Directed History

A directed history helps localize the source of bleeding and is critical in making a preliminary diagnosis (Table I). Hematemesis and melena are the most common symptoms and signs of gastrointestinal bleeding. Melena is caused by at least 50 ml of blood in the upper gastrointestinal tract (although larger volumes of blood can be clinically silent). Although vomiting of bright red blood usually signifies aggressive bleeding, vomiting of small amounts of blood can be alarming to patients and care providers alike; therefore, a careful history about volumes of vomited blood is important (as is assessment of vital signs). Patients exhibiting coffee-ground emesis typically have had a recent or remote bleed. Hematochezia can result from bleeding anywhere in the gastrointestinal tract. When upper gastrointestinal tract bleeding results in hematochezia, bleeding is brisk and often hemodynamically significant. Chronic, occult blood loss may lead to end-organ symptoms such as lightheadedness and orthostasis, dyspnea, and angina (or symptoms of myocardial infarction).

TABLE I Historical Features Important in Assessing the Etiology of Gastrointestinal Bleeding

History	Diseases suggested
Age	Malignancy, vascular ectasia
Prior bleeding	Vascular ectasia, ulcer disease, diverticula
Previous surgery	Ulcer disease
Liver disease	Varices, portal hypertensive gastropathy
Renal failure	Vascular Ectasia
NSAIDs/ASA[a]	Ulcer disease
Abdominal pain	Ischemia
Change in bowel habits	Colon Malignancy, celiac sprue
Weight loss/anorexia	Malignancy, celiac sprue
Dysphagia, odynophagia	Esophageal malignancy, achalasia

[a] NSAIDs, nonsteroidal antiinflammatory drugs; ASA, acetylsalicylic acid.

An important historical feature in gastrointestinal bleeding is age. Elderly patients are at risk for bleeding from a number of disorders that are less common in younger patients. Bleeding in very young patients is most often due to Meckel's diverticula (which does not occur in patients older than 25 to 30 years). A history of liver disease raises the possibility of bleeding associated with portal hypertension. A history of prior bleeding raises the likelihood of ulcer disease or diverticular bleeding. A history of aspirin or nonsteroidal antiinflammatory drug usage raises the possibility of bleeding from gastric or gastroduodenal ulceration. A history of reflux points to the esophagus as a potential site of bleeding. Other important historical features include the presence of abdominal pain (mesenteric ischemia), retching (Mallory–Weiss tear), change in bowel habits (malignancy), anorexia, and weight loss (the latter two can indicate malignancy).

The Physical Examination

Like the patient history, simple physical examination provides important information. Cutaneous spider angiomata and Dupytren's contractures should suggest the presence of liver disease; other signs of liver disease (splenomegaly, ascites, caput) may suggest the possibility of portal hypertension. Acanthosis nigricans raises the possibility of underlying cancer (especially gastric cancer), cutaneous spider angiomata in the absence of liver disease raises the possibility of hemorrhagic telangiectasia (Osler–Weber–Rendu syndrome), pigmented

lip lesions may be seen with Peutz–Jeghers syndrome, cutaneous tumors suggest neurofibromatosis, and purpura are consistent with vascular disease (Henloch–Schönlein syndrome or polyarteritis nodosa). Abdominal pain (peptic ulcer, pancreatitis, ischemia), abdominal masses, lymphadenopathy (malignancy), and splenomegaly (cirrhosis, splenic vein thrombosis) are important to detect.

Initial Patient Assessment

The first step in all forms of gastrointestinal bleeding is to assess the urgency of bleeding. Therefore, hemodynamics are the initial focal point and remain the gold standard for assessment of the overall clinical well being of the patient. Features of the history and physical examination provide critical clues not only about the cause of bleeding, but also about the severity of bleeding. "Bedside" tools are critical in initial assessment of the patient. Vital signs are essential in helping to gauge the severity of bleeding. Additionally, features of the history and physical exam provide critical information. For example, patients with brown stool are unlikely to have aggressive bleeding. Patients passing numerous stools containing melena—even in the absence of a positive nasogastric lavage—are likely to have aggressive ongoing bleeding. Patients with infrequent stools are unlikely to be actively bleeding. Patients with a history of coffee-ground emesis or even hematemesis, normal appearing stools, and normal vital signs have had a trivial bleed.

Laboratory Evaluation

The laboratory evaluation most often focuses on the hematocrit. However, it is important to recognize that the hematocrit, particularly when determined soon after the onset of bleeding, may not accurately reflect blood loss. Equilibration with extravascular fluid and subsequent hemodilution require several hours. In contrast, patients who bleed small amounts of blood over long periods of time develop iron deficiency, and despite the presence of a low hematocrit may be entirely hemodynamically stable. The mean corpuscular volume (MCV) is often an important clue in these patients; additionally, the ferritin level is used to establish a diagnosis of iron deficiency. It is important to emphasize that ferritin levels of even 50 μg/ml are consistent with iron deficiency anemia.

The blood urea nitrogen (BUN) may be mildly elevated in patients with upper GI bleeding. The elevation is typically out of proportion to elevation in serum creatinine, due to breakdown of blood proteins to urea by intestinal bacteria as well as from a mild reduction in glomerular filtration rate.

Clinical Localization of Bleeding

It is useful to determine the source of bleeding so as to direct investigation; this can often be accomplished with a careful history and physical examination. Hematemesis indicates an upper gastrointestinal source of bleeding. Melena indicates that blood has been in the gastrointestinal tract for extended periods of time. Melena is usually the result of upper gastrointestinal bleeding, but can be a result of distal small bowel and even right-sided colonic bleeding. The latter occurrence, which is relatively uncommon, requires the volume of bleeding to be too small to cause hematochezia but sufficient to supply enough hemoglobin for degradation. Hematochezia typically is a result of colonic bleeding, although an upper gastrointestinal lesion may bleed briskly enough to cause hematochezia. Somewhere around 10% of all patients with rapid bleeding from an upper source will pass bright red blood per rectum. Other clues to an upper gastrointestinal source of bleeding include hyperactive bowel sounds and an elevation in BUN out of proportion to creatinine. The use of nasogastric lavage can further help localize bleeding.

OCCULT BLEEDING

Obscure Bleeding

A standard diagnostic evaluation with upper and lower endoscopy, tagged red blood cell (RBC) scintigraphy, and visceral angiography is able to identify the source of bleeding in the majority of patients. However, in approximately 5% of patients, the source will remain obscure after a diagnostic evaluation. Even more problematically, some of these patients experience recurrent gastrointestinal bleeding that can not be easily ascribed to a definite site. These patients represent a considerable management challenge.

Differential Diagnosis

The differential diagnosis in patients with obscure gastrointestinal bleeding encompasses all of the lesions that can bleed in the upper and lower gastrointestinal tracts. The most common causes of obscure gastrointestinal bleeding are highlighted in Table II. The most common lesion is vascular ectasia. These lesions are often found in the small intestine, but can be identified in any location.

TABLE II Causes of Obscure Gastrointestinal Bleeding[a]

Vascular ectasias[b]
Small bowel neoplastic lesions
Hemosuccus pancreaticus
Hemobilia
Aortoenteric fistula
Dieulafoy's ulcer (stomach more frequently than other sites)
Meckel's diverticulum
Extraesophageal varices (gastric, small bowel, colonic)
Diverticula (especially small intestinal)

[a] From Rockey, D.C. (2002). Used with permission from W. B. Saunders.

[b] Small intestinal lesions are particularly important.

Evaluation

The approach to evaluation should proceed along a standard algorithm (Fig. 1). Initially, easily "overlooked" lesions should be considered. Such lesions include Dieulafoy's lesions, gastric and duodenal varices, diverticula, aortoenteric fistula, hemobilia, hemosuccus pancreaticus, and, in young patients, Meckel's diverticulum. Thus, repeat endoscopy directed at the most likely site of bleeding is usually warranted. Reexamination of the upper gastrointestinal tract within reach of a standard gastroscope will identify lesions in a substantial proportion of patients. Importantly, however, familiarity with rare and/or subtle bleeding lesions is required. If a lesion cannot be identified, further evaluation depends on the briskness of bleeding. In those with active bleeding, technetium-99 radionuclide scanning or angiography should be performed. Technetium scanning, although sensitive (reportedly, as little as 0.1 ml/minute can be detected), is useful only to confirm the site of bleeding, and available literature assessing its impact on management has been disappointing. Mesenteric angiography is less sensitive (requiring a bleeding rate >0.5 ml/minute) than technetium scanning but reportedly more often identifies the site of bleeding, perhaps due to selection bias in published studies. In selected situations, other diagnostic tests (computed tomography, Meckel's scan) may be helpful. In those with subacute bleeding in whom repeat endoscopies are negative, the focus of investigation should be broadened to include the small intestine. The lesions most commonly identified as bleeding sites in the small bowel

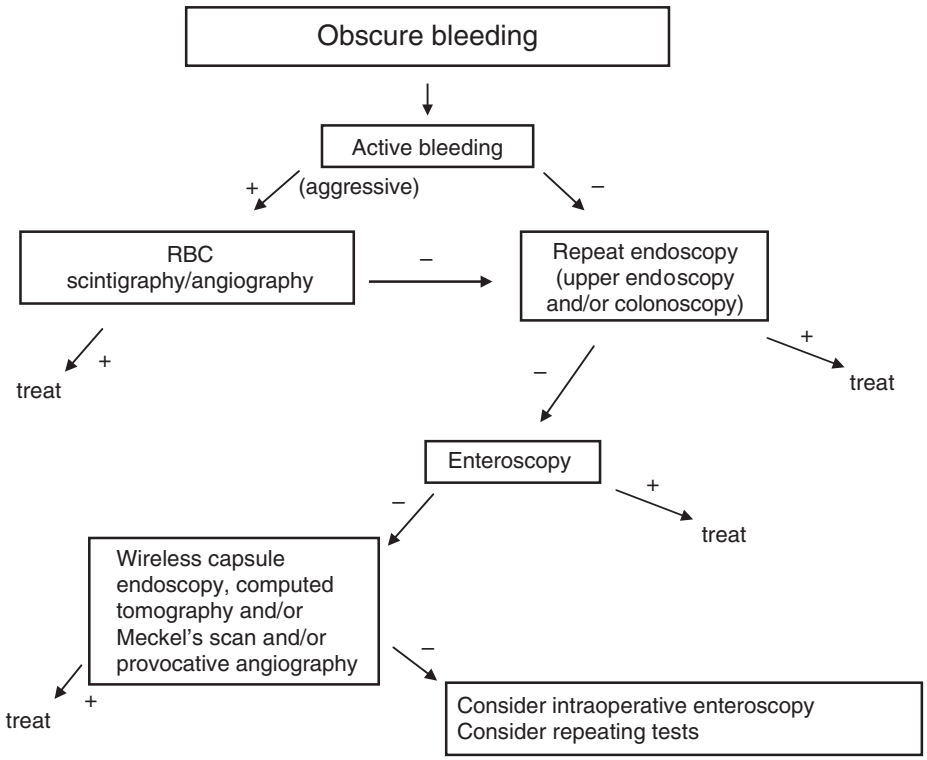


FIGURE 1 Scheme for evaluation of obscure gastrointestinal bleeding. RBC, Red blood cell. Adapted with permission from Rockey, D. C. (2002). Gastrointestinal bleeding. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman *et al.*, eds.), 7th Ed. W. B. Saunders, Philadelphia.

include tumors and vascular ectasias, which vary in frequency depending on age. In patients less than 25 years of age, Meckel's diverticula are the most common source of small bowel bleeding; in those between 30 and 50 years old, tumors are the most common abnormalities, and vascular ectasias predominate in the elderly.

Once the upper and lower gastrointestinal tracts have been excluded as a source of bleeding, the focus of evaluation should shift to the small bowel. This can be accomplished with standard small bowel follow-through, enteroclysis, push enteroscopy, sonde enteroscopy, intraoperative enteroscopy, or, most recently, wireless capsule endoscopy. We prefer endoscopic evaluation over radiologic evaluation because of the ability to identify flat lesions such as vascular ectasia, which may be missed by barium studies. Push enteroscopy should probably be the first test; several recent reports suggest that push enteroscopy is able to identify a substantial number of gastrointestinal lesions in patients with obscure gastrointestinal bleeding. Sonde enteroscopy is limited by the inability to biopsy lesions or make therapeutic interventions. Wireless capsule endoscopy is rapidly gaining favor. Initial results with this test are encouraging and demonstrate that it is capable of identifying lesions that could be considered a source of bleeding in up to 30–40% of patients.

An alternative approach to the diagnosis of recurrent, obscure bleeding is to reactivate or augment bleeding with the use of vasodilators, anticoagulants, and/or thrombolytics in association with tagged RBC scintigraphy or visceral angiography. This approach is attractive in principle, but may be dangerous in practice and needs to be studied further.

Treatment and Outcome

Specific treatment for the obscure bleeder depends on the abnormality identified. Enteroscopy, which often demonstrates putative bleeding lesions, has not always led to improved outcomes. Enteroscopic cauterization of vascular ectasias has been shown to lead to an improvement in hemoglobin and a reduction in blood transfusion requirements, and can be effective in selected patients. Additionally, only about 50% of patients treated at the time of intraoperative enteroscopy will stop bleeding, indicating that this intervention is not always ideal. Finally, wireless capsule endoscopy is an attractive diagnostic approach in patients with obscure bleeding. However, it and other diagnostic interventions in obscure bleeding have yet to be shown to improve meaningful outcomes. It is important to emphasize that care of these patients requires an experienced and dedicated team.

Small bowel vascular ectasias are the single most common source of bleeding in patients with obscure gastrointestinal bleeding. In general, specific endoscopic and surgical therapy is most successful in those with large focal vascular ectasias. However, because vascular ectasias are often diffuse, endoscopic and surgical interventions are often limited. Hormonal therapy with estrogen/progesterone compounds has been advanced as a medical alternative; whereas some clinicians have reported positive experiences with such pharmacological therapy, controlled trials have failed to show an advantage.

Fecal Occult Blood

Occult gastrointestinal bleeding is not clinically apparent but rather is manifest by biochemical evidence of blood in the stool or by laboratory evidence of iron deficiency. In these situations, blood loss is often entirely unknown to the patient. When iron deficiency occurs in men or postmenopausal women, it is usually due to occult gastrointestinal blood loss. Occult bleeding can occur from any location in the gastrointestinal tract.

Fecal occult blood is the most common form of occult gastrointestinal bleeding; when fecal occult blood tests have been applied to large populations, up to 16% of those tested are positive. Although fecal occult blood tests were designed to detect colonic bleeding, they detect blood from other lesions in the gastrointestinal tract. The likelihood that these tests will detect blood resulting from gastrointestinal bleeding varies and is dependent on the mechanistic features of the test as well as the anatomic level of bleeding (which influences whether hemoglobin is degraded; Fig. 2). A further critical issue, unrelated to characteristics of fecal occult blood tests per se, is the biology of bleeding gastrointestinal tract lesions, many of which bleed in an irregular fashion. Thus, multiple factors contribute to the ability of fecal occult blood tests to detect bleeding lesions.

Types of Fecal Occult Blood Tests

The prototypical fecal occult blood tests are guaiac based and take advantage of the fact that hemoglobin possesses pseudoperoxidase activity. Guaiac tests are more sensitive for detecting bleeding from the lower than from the upper gastrointestinal tract because hemoglobin is degraded in the gastrointestinal tract (Fig. 2). The likelihood that occult blood will be detected by a positive guaiac test is generally proportional to the quantity of fecal heme, which in turn is related to the size and location of the bleeding lesion.

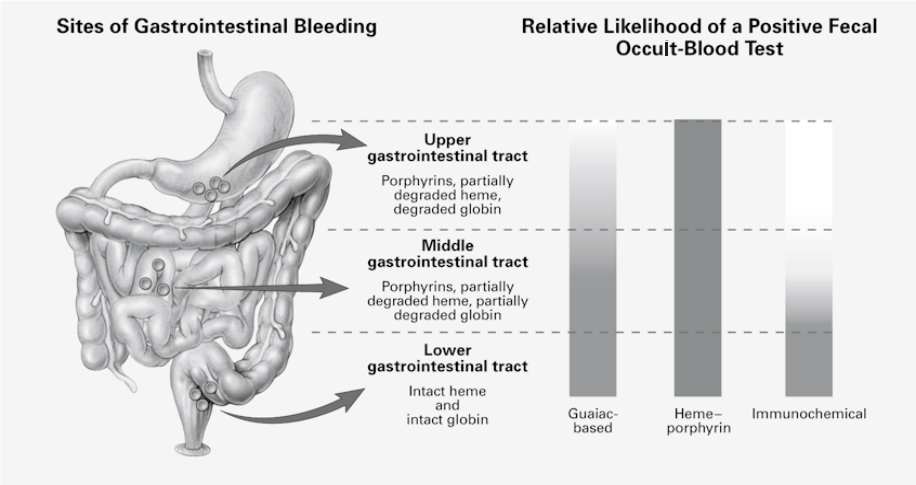


FIGURE 2 Intraluminal hemoglobin metabolism and fecal occult blood tests. In the upper gastrointestinal tract, hemoglobin is degraded by gastric pepsin and/or pancreatic proteases in the proximal small intestine, resulting in heme and globin. Some (generally $<15\%$) intraluminal heme is reabsorbed in the small intestine. A portion of heme that is not absorbed is converted to porphyrins and iron via poorly understood mechanisms; this portion of heme has been termed the "intestinal converted fraction" of heme. This fraction is not detected by guaiac tests but is detected by the heme–porphyrin assay (HemoQuant), which measures both heme and porphyrins, and is therefore a highly accurate indicator of bleeding, regardless of level. Globin in the upper gastrointestinal tract is digested by pepsin and pancreatic and intestinal proteases and is thus not detected by immunochemical fecal occult blood tests. Used with permission from Rockey (1999). Occult gastrointestinal bleeding. *N. Engl. J. Med.* **341**, 38–46.

Thus, guaiac-based tests are best at detecting large, more distal lesions. Fecal hemoglobin levels must exceed 10 mg/g (approximately 10 ml daily blood loss) before most guaiac tests are positive at least 50% of the time. It is important to emphasize that a number of other factors, many of which are dietary (Table III), influence guaiac test results. Another factor affecting the reactivity of guaiac-based tests is fecal rehydration, which markedly raises sensitivity but significantly reduces specificity. It is commonly believed that oral iron causes false positive guaiac tests. Care must be taken not to confuse the dark-green or black appearance of iron in stool with the typical blue color of a positive guaiac reaction; it has been confirmed that orally administered iron, even in large amounts, does not cause false positive guaiac reactions. Bismuth-containing antacids and antidiarrheals cause the stool to become black and may confound reading of guaiac-based tests, or may even cause confusion with melena.

Immunochemical tests (using antibodies directed against human globin) readily detect colonic blood and do not detect small quantities of blood from upper gastrointestinal sources (Fig. 2). Thus, they are more specific than guaiac-based tests. However, they are

TABLE III Factors Affecting Fecal Occult Blood Tests[a]

Variable	Guaiac	Heme–porphyrin	Immunochemical
False positives			
Animal hemoglobin	++++	++++	0
Dietary peroxidases	+++	0	0
False negatives			
Hemoglobin degradation	++	0	++
Storage	++	++++	++
Vitamin C	++	0	0

[a] Adapted with permission from Rockey (1999). Occult gastrointestinal bleeding. *N. Engl. J. Med.* **341**, 38–46. Relative comparisons are shown on a scale of 0 to ++++, with 0 being negative and ++++ highly positive.

limited by technical problems such as loss of hemoglobin antigenicity at room temperature and the requirement for laboratory processing. Finally, the heme–porphyrin test (HemoQuant, Mayo Medical Laboratories, Rochester, MN) measures porphyrin spectrofluorometrically and therefore allows precise determination of total stool hemoglobin. Substances that may interfere with or cause false positive guaiac tests (i.e., vegetable peroxidases) do not affect this test. However, this test detects myoglobin, an important source of nonhuman heme found in red meats. This test is the most sensitive method to detect occult gastrointestinal bleeding.

Differential Diagnosis and Approach to Evaluation

Fecal occult blood tests are widely used to screen the colon for cancer, but also help to investigate symptoms. Virtually any gastrointestinal lesion can bleed and cause a positive fecal occult blood test result (Table IV). Once a patient is determined to have fecal occult blood, investigation is initially focused on the colon. However, the choice of colonic imaging modality (colonoscopic or radiographic) is controversial, both in terms of diagnostic accuracy and in terms of cost. Flexible sigmoidoscopy is necessary for patients undergoing air-contrast barium enema to fully evaluate the rectosigmoid colon. Air-contrast barium enema can accurately detect colonic malignancy and large adenomas; however, under certain circumstances, it may also be inaccurate. Colonoscopy is generally considered to be more accurate but may also miss important lesions. Computed tomography and computerized rendering of the colon (virtual colonoscopy or colonography) may eventually play a role in colon evaluation, but are currently not widely available. At present, it is generally believed that the best choice for evaluation of the colon in patients with fecal occult blood is colonoscopy.

Patients with fecal occult blood but a normal colon often harbor upper gastrointestinal pathology. A number of studies have addressed this issue and have emphasized that upper endoscopy will detect abnormalities in around one-third of patients. This finding is somewhat surprising because the guaiac-based tests used in these studies have a relatively low sensitivity for detecting upper gastrointestinal blood. Nonetheless, guaiac-based tests are clearly capable of detecting even small amounts of upper gastrointestinal tract blood, and many of the types of lesions identified in the upper gastrointestinal tract in these studies bleed enough to produce positive guaiac-based tests. The cost-effectiveness of upper gastrointestinal tract investigation in patients with fecal occult blood and normal colons is unknown.

Whether patients with occult blood found in the stool after digital rectal examination should be evaluated is controversial; anorectal trauma and/or dietary factors may lead to false positive tests. However, in both symptomatic and asymptomatic patients with fecal occult blood detected by digital rectal examination, the number of new lesions identified by gastrointestinal

TABLE IV Differential Diagnosis of Occult Gastrointestinal Bleeding[a]

Mass lesions	**Vascular**
Carcinoma (any site)[a]	Vascular ectasia (any site)[b]
Large (>1.5 cm) adenoma (any site)	Portal hypertensive gastropathy/colopathy
Inflammation	Watermelon stomach
Erosive esophagitis[b]	Hemangioma
Ulcer (any site)[b]	Dieulafoy's ulcer[d]
Cameron lesions[c]	**Infectious**
Erosive gastritis	Hookworm
Celiac sprue	Whipworm
Ulcerative colitis	Stronglyoidiasis
Crohn's disease	Ascariasis
Colitis (nonspecific)	Tuberculous enterocolitis
Idiopathic cecal ulcer	Amebiasis
Miscellaneous	**Surreptitious**
Long-distance running	Hemoptysis
Fictitious	Oropharyngeal (including epistaxis)
	Pancreaticobiliary

[a] Adapted with permission from Rockey (1999). Occult gastrointestinal bleeding. *N. Engl. J. Med.* **341**, 38–46. Potential lesions leading to fecal occult blood or iron deficiency anemia are shown. Lesions that may lead to recurrent obscure bleeding are not listed (see Table I).

[b] Most common abnormalities.

[c] Linear erosions within a hiatus hernia.

[d] Large superficial artery underlying mucosal defect.

evaluation is substantial and therefore evaluation is indicated; if symptoms are present, further evaluation should be directed accordingly. Whether testing stool obtained by digital rectal examination is a viable cancer screening option is an important question and would be influenced by variables such as improved compliance versus inadequate dietary preparation.

Occult gastrointestinal bleeding generally should not be attributed to anticoagulant or aspirin therapy. Indeed, fecal blood levels in patients therapeutically anticoagulated have been reported to be normal and low-dose aspirin alone does not result in significantly elevated fecal blood levels. Further, neither warfarin nor low-dose aspirin alone appears to cause positive guaiac-based fecal occult blood tests. Thus, a positive fecal occult blood test should not be attributed to the effect of anticoagulation or aspirin alone, but rather should lead to standard evaluation.

Treatment and Outcome

Treatment of patients with fecal occult blood varies and depends on the abnormalities identified. Likewise, outcomes are directly related to specific findings. Non-steroidal antiinflammatory drugs cause mucosal injury and should be discontinued if possible. Patients with vascular ectasias are problematic, and if they bleed chronically should be managed carefully (see earlier discussion of obscure gastrointestinal bleeding). The prognosis of patients with positive fecal occult blood tests but no identifiable gastrointestinal pathology is generally favorable, but long-term outcome data on this subject are lacking.

Iron Deficiency Anemia

Iron deficiency anemia is the most common form of anemia in the world. In the United States, 5–11% of adult females and 1–4% of adult males are iron deficient whereas approximately 5 and 2% of adult women and men, respectively, have iron deficiency anemia. Although iron deficiency anemia in women of reproductive age is typically assigned to menstrual- and pregnancy-associated iron losses, in groups other than premenopausal women, iron deficiency anemia is assumed to be due to chronic occult gastrointestinal bleeding, resulting depletion of the iron pool and ultimately anemia (Fig. 3).

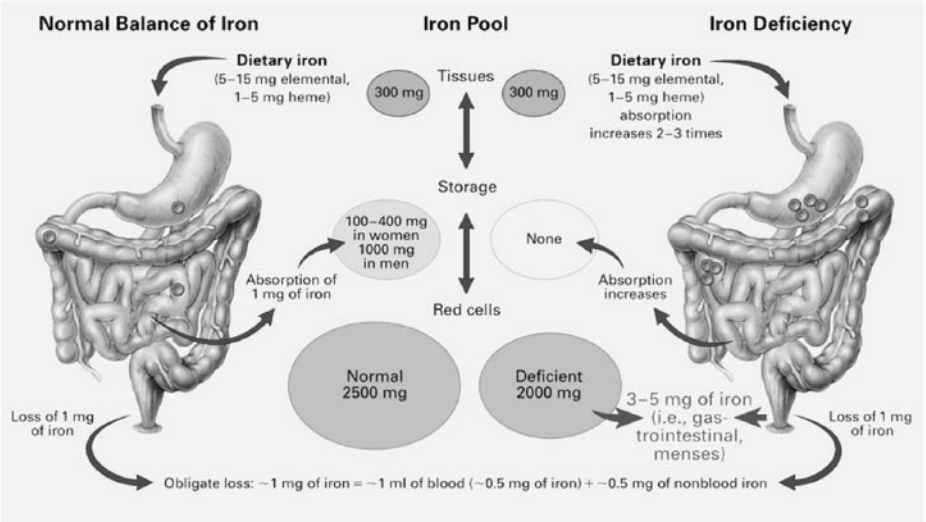


FIGURE 3 Gastrointestinal blood loss and iron balance. Normal obligate daily iron loss is from (1) blood loss (presumably from gastrointestinal mucosal microerosions or microulcerations) and (2) iron in sloughed gut epithelial cells. Total daily iron loss is thus approximately 1 mg. The usual Western diet contains mostly elemental iron, of which about 10% is absorbed. Heme-iron, derived primarily from myoglobin in meats, is preferentially absorbed and accounts for 60–80% of the iron absorbed per day. Under normal circumstances, iron homeostasis is tightly regulated and daily iron loss is precisely balanced by iron absorption. Iron deficiency results only when the dynamic, but limited, absorptive capacity of the small intestine is exceeded by iron loss. The time required to develop iron deficiency depends on the size of initial iron stores, the rate of bleeding, and intestinal iron absorption. Iron deficiency generally occurs only with increased loss of over 5 ml of blood daily. Importantly, anemia is a late manifestation of the iron-depleted state. With permission from Rockey (1999). Occult gastrointestinal bleeding. *N. Engl. J. Med.* **341**, 38–46.

Differential Diagnosis and Approach to Evaluation

The diagnosis of iron deficiency and iron deficiency anemia is most often confirmed by a low serum ferritin level or a bone marrow examination. Effort should be made to verify the diagnosis in equivocal cases, because a diagnosis of iron deficiency anemia leads to extensive and often costly evaluation. Iron deficiency without anemia is often also associated with significant gastrointestinal tract abnormalities.

As with all occult gastrointestinal bleeding, virtually any gastrointestinal tract lesion can bleed in an occult fashion and lead to iron deficiency anemia (Table IV). Although the colon is traditionally viewed as the source of bleeding in most patients with iron deficiency anemia, a number of studies have now documented prominent abnormalities in the upper gastrointestinal tract in patients without colonic lesions. The most common abnormalities have been erosive disease of the esophagus, stomach, and duodenum (i.e., severe esophagitis, presumably reflux mediated, gastric, or duodenal ulcer). Colon cancers and large adenomas are the most commonly identified lesions in the colon. Importantly, only about 5% of patients have had simultaneous pathology identified in both upper and lower gastrointestinal sites.

A large body of literature has helped change the way patients with iron deficiency anemia are currently managed. The emphasis is on the importance of evaluation of the upper gastrointestinal tract. It is important to emphasize that although mass lesions and large ulcerative upper gastrointestinal lesions can lead to significant blood loss (up to 20 ml/day), it is unlikely that trivial lesions such as mild inflammation and small adenomas bleed enough to lead to iron deficiency. Thus, care must be taken when attributing iron deficiency anemia to lesions not expected to cause significant bleeding.

Gastrointestinal symptoms in patients with iron deficiency anemia typically help focus gastrointestinal tract evaluation. Directed gastrointestinal tract evaluation is practical and desirable as a cost- and risk-containing strategy. Although many patients with iron deficiency anemia are entirely asymptomatic, those with classic symptoms (change in stool caliber, epigastric pain, or heartburn) should undergo site-directed investigation. The initial investigation should be directed toward the location of specific symptoms, if present (Fig. 4). Because dual lesions are rare, identification of an abnormality clearly consistent with bleeding (mass lesion, large ulceration, severe inflammation) makes further evaluation unnecessary. In the absence of symptoms, particularly in elderly patients, evaluation should

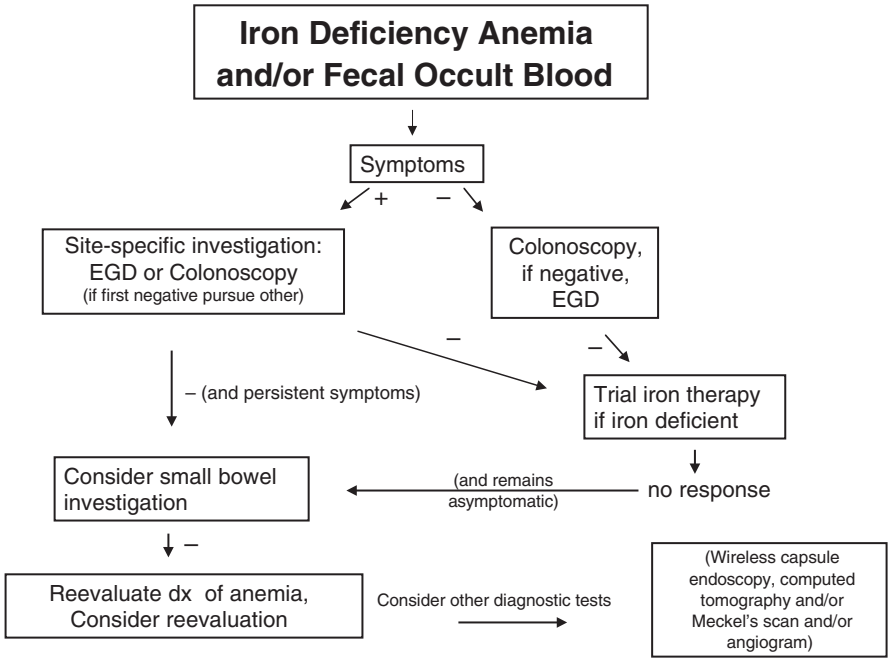


FIGURE 4 Scheme for evaluation of fecal occult blood and iron deficiency anemia. EGD, Esophagogastroduodenoscopy. From Rockey, D. C. (2002). Gastrointestinal bleeding. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman *et al.*, eds.), 7th Ed. W. B. Saunders, Philadelphia). Used with permission.

begin with the colon, but if this is negative, evaluation of the upper gastrointestinal tract is indicated.

The degree of gastrointestinal evaluation in premenopausal women with iron deficiency anemia is controversial. On one hand, few iron deficiency anemia premenopausal women undergo evaluation. However, in studies focusing on those referred to gastroenterologists, it has been reported that up to 20% of patients will have significant gastrointestinal tract abnormalities, including important lesions such as colon cancer, celiac sprue, and inflammatory bowel disease. Clearly, a management algorithm is required for these patients. A practical approach is as follows: for those with specific gastrointestinal symptoms, weight loss, fecal occult blood, or perhaps severe anemia, evaluation is indicated. For asymptomatic women or those with abnormal menses, whether to pursue gastrointestinal tract evaluation should be individualized.

The small bowel is an important potential site of bleeding in patients with negative colon and upper gastrointestinal tract examinations. For example, celiac sprue, a classic small bowel disorder, can not only lead to malabsorption of iron, but may cause occult bleeding and should be considered. This is especially true in patients of Northern European descent. Radiographic examination of the small bowel (enteroclysis or small bowel follow-through) is of limited value and is generally not recommended. On the other hand, push enteroscopy has a greater sensitivity for mucosal abnormalities and detects abnormalities in approximately 25% of patients. However, it remains unknown whether push enteroscopy leads to improved outcomes. Likewise, the use of wireless capsule endoscopy in patients with iron deficiency anemia remains unsettled.

Nearly one-third of patients with iron deficiency anemia have no identifiable gastrointestinal tract abnormality by routine endoscopic examination. In these patients, explanations for iron deficiency anemia include non-gastrointestinal blood loss, misdiagnosis of the type of anemia, missed lesions, or nutritional deficiency. Additionally, these patients may have gastric achlorhydria and atrophy, suggesting that lack of acid in this subgroup could contribute to iron malabsorption.

Treatment and Outcome

All patients with iron deficiency anemia should be started on iron therapy. Oral ferrous sulfate is the best option because it is inexpensive and effective (the recommended dose is 300 mg three times daily). Intolerance to ferrous sulfate is common; in these patients, ferrous gluconate and fumarate are acceptable alternatives. Parenteral iron therapy is reserved only for patients with severe malabsorption or intolerance to all iron supplements. The prognosis for patients with iron deficiency anemia and readily treatable lesions (i.e., duodenal ulcer, esophagitis, large adenoma) is excellent. Likewise, the prognosis for patients who do not have lesions identified during gastrointestinal evaluation is favorable; long-term followup of patients in this group suggests that significant lesions do not surface at a later date. In patients who do not respond to iron therapy, the diagnosis of iron deficiency anemia should be reevaluated and repeat gastrointestinal evaluation should be contemplated. A careful reexamination of the gastrointestinal tract for easily missed lesions (in the esophagus, for Cameron lesions within hiatus hernia; in the stomach, for atrophic gastritis; in the small bowel, celiac sprue) is indicated when evaluating persistent unexplained iron deficiency anemia.

CONCLUSIONS

Occult gastrointestinal bleeding is most often manifest as occult blood in the stool, typically by the use of guaiac-based fecal occult blood tests. Occult gastrointestinal bleeding may also be manifest as iron deficiency anemia, which often results from chronic occult gastrointestinal bleeding. These forms of occult gastrointestinal bleeding are common. In contrast to fecal occult blood and iron deficiency anemia, obscure gastrointestinal bleeding is uncommon but represents a considerable diagnostic and therapeutic challenge. Evaluation of patients with fecal occult blood and iron deficiency anemia often parallels similar algorithms and in asymptomatic patients should usually begin with investigation of the colon. Colonoscopy is the preferred method, but flexible sigmoidoscopy plus air-contrast barium enema may be acceptable. If evaluation of the colon does not reveal a bleeding site, evaluation of the upper gastrointestinal tract should be considered, and in patients with iron deficiency anemia, this is mandatory. In patients with iron deficiency anemia, the role of small bowel investigation is controversial and is probably best reserved for those with persistent gastrointestinal symptoms or those who fail to respond to appropriate therapy. Celiac sprue should be considered as a potential cause of iron deficiency anemia in certain epidemiological groups. The treatment and prognosis of patients with occult blood in the stool and/or iron deficiency anemia depends on the gastrointestinal tract abnormalities identified. Those without identifiable bleeding sites generally respond to conservative management and have a favorable prognosis. For patients with obscure gastrointestinal bleeding, investigation is typically focused on the small bowel, typically with enteroscopy,

but perhaps also with wireless capsule endoscopy. In these patients, vascular ectasias are a major consideration. The prognosis for these patients is less clear, and outcome studies are required. This group requires a committed and experienced team approach to diagnosis and therapy.

Acknowledgment

This work was supported by the Burroughs Wellcome Fund (BWF Clinical Scientist Award).

See Also the Following Articles

Colonoscopy • Endoscopy, Complications of • Hemobilia • Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Meckel's Diverticulum • Sigmoidoscopy • Trauma, Overview • Upper Gastrointestinal Bleeding • Variceal Bleeding

Further Reading

Ahlquist, D. A., McGill, D. B., Schwartz, S., Taylor, W. F., and Owen, R. A. (1985). Fecal blood levels in health and disease. A study using HemoQuant. *N. Engl. J. Med.* **312**, 1422–1428.

Ahlquist, D. A., McGill, D. B., Fleming, J. L., *et al.* (1989). Patterns of occult bleeding in asymptomatic colorectal cancer. *Cancer* **63**, 1826–1830.

Appleyard, M., Fireman, Z., Glukhovsky, A., *et al.* (2000). A randomized trial comparing wireless capsule endoscopy with push enteroscopy for the detection of small-bowel lesions. *Gastroenterology* **119**, 1431–1438.

Bloomfeld, R. S., Smith, T. P., Schneider, A. M., and Rockey, D. C. (2000). Provocative angiography in patients with gastrointestinal hemorrhage of obscure origin. *Am. J. Gastroenterol.* **95**, 2807–2812.

Chak, A., Koehler, M. K., Sundaram, S. N., Cooper, G. S., Canto, M. I., and Sivak, M. V., Jr. (1998). Diagnostic and therapeutic impact of push enteroscopy: Analysis of factors associated with positive findings. *Gastrointest. Endosc.* **47**, 18–22.

Chalasani, N., Clark, W. S., and Wilcox, C. M. (1997). Blood urea nitrogen to creatinine concentration in gastrointestinal bleeding: A reappraisal [see comments]. *Am. J. Gastroenterol.* **92**, 1796–1799.

Foutch, P. G., Rex, D. K., and Lieberman, D. A. (1995). Prevalence and natural history of colonic angiodysplasia among healthy asymptomatic people. *Am. J. Gastroenterol.* **90**, 564–567.

Hara, A. K., Johnson, C. D., Reed, J. E., *et al.* (1997). Detection of colorectal polyps with CT colography: Initial assessment of sensitivity and specificity. *Radiology* **205**, 59–65.

Iddan, G., Meron, G., Glukhovsky, A., and Swain, P. (2000). Wireless capsule endoscopy. *Nature* **405**, 417.

Lewis, B. S. (1994). Small intestinal bleeding. *Gastroenterol. Clin. North Am.* **23**, 67–91.

Rockey, D. C. (1999). Occult gastrointestinal bleeding. *N. Engl. J. Med.* **341**, 38–46.

Rockey, D. C. (2002). Gastrointestinal bleeding. In "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman *et al.*, eds.), 7th Ed. W. B. Saunders, Philadelphia, PA.

Rockey, D. C., and Cello, J. P. (1993). Evaluation of the gastrointestinal tract in patients with iron deficiency anemia. *N. Engl. J. Med.* **329**, 1691–1695.

Rockey, D. C., Koch, J., Cello, J. P., Sanders, L. L., and McQuaid, K. (1998). Relative frequency of upper gastrointestinal and colonic lesions in patients with positive fecal occult-blood tests. *N. Engl. J. Med.* **339**, 153–159.

van Cutsem, E., Rutgeerts, P., and Vantrappen, G. (1990). Treatment of bleeding gastrointestinal vascular malformations with oestrogen-progesterone. *Lancet* **335**, 953–955.

Waye, J. D. (1997). Enteroscopy. *Gastrointest. Endosc.* **46**, 247–256.

Winawer, S. J., Stewart, E. T., Zauber, A. G., *et al.* (2000). A comparison of colonoscopy and double-contrast barium enema for surveillance after polypectomy. National Polyp Study Work Group. *N. Engl. J. Med.* **342**, 1766–1772.

Zuckerman, G. R., Prakash, C., Askin, M. P., and Lewis, B. S. (2000). AGA technical review on the evaluation and management of occult and obscure gastrointestinal bleeding. *Gastroenterology* **118**, 201–221.

Over-the-Counter Drugs

ESAM Z. DAJANI*,†, NOURA E. DAJANI*,‡, AND THOMAS G. SHAHWAN*,‡

*International Drug Development Consultants Corporation, Long Grove, Illinois,
†Stritch School of Medicine, Loyola University Chicago, and ‡University of Illinois School of Medicine, Peoria

acid-neutralizing capacity Amount of acid (mEq) neutralized by a unit dose (teaspoonful or tablet) of an antacid.

antacids Soluble and/or insoluble substances capable of buffering or neutralizing gastric (stomach) acid.

antiflatulents Drugs effective for the relief of painful bloating or sensations of pressure and fullness, commonly referred to as gas in the digestive tract.

constipation Commonly defined as related to decreased bowel movement frequency, difficulty in initiating passage of feces, passage of firm or small-volume feces, or a feeling of incomplete evacuation.

diarrhea Symptom of several gastrointestinal disorders affecting water and electrolyte transport; characterized by alteration of the frequency and consistency of stools.

dyspepsia Mild pain present in the upper abdomen.

heartburn Specific symptom of gastroesophageal reflux disease, typically manifested by a burning sensation in the upper abdomen, behind the chest, and as high as the throat.

histamine-2 (H2) receptor antagonists Drugs that block action of histamine at a specific histamine (type 2) receptor located in the stomach, which results in effective inhibition of gastric acid secretion.

Increasing numbers of drugs are obtainable without a prescription, i.e., over the counter (OTC), for self-treatment of mild gastrointestinal symptoms. The consumer must be able to recognize the symptoms of the condition that the product can treat easily, without medical assistance. Unlike prescription-based drugs, the ingestion of which requires physicians' supervision and management for treating the underlying disease process, OTC drugs are primarily used for self-treatment of well-recognized symptoms by consumers. It is increasingly important to review the therapeutics of gastrointestinal drugs specifically available for self-use and to remain vigilant about the safe use of drugs sold on an OTC basis in the United States.

INTRODUCTION

The availability of OTC drugs lessens the economic burden to our health care system and clearly provides much needed convenience to consumers. However, despite these advantages, there are clear risks related to the use of OTC drugs. For example, aspirin and acetaminophen are still associated with serious gastrointestinal toxicity despite their availability as OTC drugs for many years. In addition, chronic use of antacids and OTC gastric antisecretory drugs can delay the diagnosis of gastric cancer. Therefore, it is most essential that consumers are made aware of the indications and limitations of OTC drugs through improved education by the pharmaceutical manufacturers and easily understood labels, affixed to the containers, which should explain the desired uses, treatment duration, dosage, and adverse effects of these drugs. Clear drug labeling, as mandated by the Food and Drug Administration (FDA), and perhaps pharmacist counseling are most essential for the safe use of OTC drugs.

AGENTS BUFFERING OR INHIBITING GASTRIC ACIDITY

Antacids

Antacids are drugs capable of buffering stomach acid. Antacids raise the pH of the stomach contents toward neutrality. An antacid that raises the pH from 1.5 to 3.5 produces a 100-fold reduction in the concentration of gastric acid. Furthermore, the reduction of acidity is accompanied by inhibition of pepsin activity, another important component of the digestive juice. According to their approved labeling, the only symptoms that can be safely diagnosed and self-treated with OTC antacids are those caused by excess stomach acid. These symptoms have been described as burning sensations in the upper abdomen, behind the chest, and high as the throat. The official FDA-approved claims for antacids are for the relief of "heartburn, sour stomach, acid indigestion, and upset stomach." Based on the many years that OTC antacids have been used for heartburn, it is apparent that a consumer who has heartburn can determine when he or she has heartburn, and in most cases can predict what foods, situations, or life stresses can cause heartburn.

Great disparity exists in the acid-neutralizing capacity (ANC) of various antacids (Table I). The more

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Acid-Neutralizing Capacity of Selected Antacids[a]

Product	Acid-neutralizing capacity in mEq acid/dose	Standard dose to neutralize 152 mEq acid/dose
Alka-Seltzer (tablet)	10.6 tablets	15 tablets
Amphogel (liquid)	6.5 tsp	25 tsp
Amphogel (tablet)	9 tablets	17 tablets
Maalox (liquid)	13.5 tsp	11 tsp
Maalox Therapeutic Concentrate (liquid)	28.5 tsp	5 tsp
Mylanta (liquid)	12.5 tsp	12 tsp
Mylanta II (liquid)	25.5 tsp	6 tsp
Mylanta II (tablet)	23 tablets	7 tablets
Titralac (liquid)	19 tsp	8 tsp
Tums (tablet)	10 tablets	16 tablets

[a] The amount of acid (mEq) being neutralized by a unit dose (teaspoonful or tablet) of an antacid; tsp, teaspoonful.

Note: Data adapted from Zimmerman, 1983.

potent antacids (e.g., Mylanta II) require a smaller dose, compared to a weaker product (e.g., Amphogel), and this is advantageous for treating ulcers. However, ulcer treatment is clearly outside the scope of OTC product labeling and should be treated only by physicians, not by self-treatment. The dose of antacid that will relieve pain in a patient with gastric ulcer must provide an ANC of 152 mEq. A weaker product, such as Amphogel, would require 25 teaspoonfuls to neutralize gastric acid, whereas a stronger product, such as Mylanta II, requires only 6 teaspoonfuls (Table I).

It should be noted that liquid antacid preparations generally provide faster buffering action than do tablet preparations. However, from a therapeutic perspective, there is a dissociation between the duration of buffering capacity of antacids, which is relatively short (30 minutes), and the duration of the pain relief required for heartburn or peptic ulcer (approximately 4 hours). Thus, on the basis of pain relief, antacids administered at a low frequency of about three times daily would not be therapeutically sufficient for healing gastric or duodenal ulcers. However, this low frequency of administration does provides pain relief in mild to moderate heartburn.

The neutralization of gastric contents by antacids promotes antral gastrin release, which promotes gastric emptying and hence contributes to the short duration of their buffering action. Due to the gastrin-induced gastric emptying, the more potent antacids have durations of buffering action essentially similar to those of weaker antacids. Therefore, the increased potency of an antacid should not be a criterion for a patient's product selection for the self-treatment of heartburn.

Several principal ingredients are present in antacids: bicarbonate (sodium, potassium, and calcium), aluminum, magnesium, phosphate, and silicates. Bismuth subsalicylate (Pepto-Bismol), which is marketed for the treatment of heartburn, has minimal antacid action and therefore is not pharmacologically classified as an antacid. All antacid products contain at least one ingredient; most contain at least two ingredients. Sodium bicarbonate and calcium carbonate are more potent antacids, compared to magnesium compounds. Magnesium compounds are more potent than aluminum compounds. Sodium and potassium bicarbonate are soluble antacids that are readily absorbed into the blood, and thus are particularly risky for patients with impaired kidney function. Calcium compounds (calcium carbonate and calcium phosphate) are potent and fast-acting antacids that are readily absorbed. However, these drugs may form calcium kidney stones. Paradoxically, calcium antacids promote antral gastrin release to a greater degree than do noncalcium antacids, which in turn stimulate gastric acid production. The aluminum compounds induce constipation and the magnesium compounds induce diarrhea, thus these two ingredients are combined together, although the net effect of the combined product is laxation. This is especially noted when the combined drugs are taken at increased dosages. The silicate antacids (e.g., magnesium trisilicate) interfere with the absorption of some drugs and this needs to be considered by patients and health care providers. For specific details about all available antacid products and their pharmacological characteristics, the readers are encouraged to consult the *Physicians' Desk Reference for Nonprescription Drugs*.

TABLE II Pharmacological Characteristics of Selected OTC Histamine-2 Receptor Antagonists

Drug	Trade name	Doses	Half-life (hours)	Drug interactions
Cimetidine	Tagamet HB 200, others	200 mg up to twice daily	2.0	Yes[a]
Ranitidine	Zantac 75, others	75 mg up to twice daily	2.1	No
Famotidine	Pepcid AC	10 mg up to twice daily	2.6	No

[a] The OTC dosage of cimetidine has a small potential for pharmacokinetic interaction with theophylline. Prescription dosage of cimetidine has shown pharmacokinetic interactions with phenytoin, theophylline, and warfarin.

Histamine-2 Receptor Antagonists

Histamine-2 (H2) receptor antagonists were previously marketed as prescription-based products for the healing and prevention of gastric ulcers and duodenal ulcers, and for the management of gastroesophageal reflux disease (GERD) due to their well-known gastric antisecretory actions. However, patients with GERD represent a spectrum ranging from mild to severe disease states and many patients with heartburn symptoms have been accustomed to the use of OTC antacids and H2 receptor antagonists. Therefore, three of the four H2 receptor antagonists that were previously marketed as prescription-based products were switched to OTC use at reduced doses, which are one-half of their prescription doses (Table II). The currently available OTC drugs, cimetidine, ranitidine, and famotidine, have essentially two similar medical claims in their labeling: (a) "for the relief of heartburn associated with acid indigestion and sour stomach" and (b) "for the prevention of heartburn associated with acid indigestion and sour stomach brought on by certain foods or beverages." The H2 antagonist nizatidine, which is currently available as a prescription-based product, has not yet been switched to an OTC status.

The frequency of the administration of all three H2 antagonists is one to two doses per day and the duration of self-use should not exceed 2 continuous weeks, except under the advice and supervision of a physician. All three drugs have similar biological half-lives of about 2 hours (Table II). However, as is the case with prescription-based products, ranitidine and famotidine, but not cimetidine, are marketed at doses that exceed their median effective gastric antisecretory doses (ED_{50}). Thus, when administered at the recommended OTC dosage of twice daily, it is expected that ranitidine and famotidine would provide a much longer duration of gastric antisecretory action than would cimetidine. However, despite the differences in the magnitude of their doses, all three drugs appear to have similar degree of efficacy for the treatment and prevention of heartburn.

Unlike antacids, which provide immediate relief of heartburn, there is some delay (15–30 minutes) in the onset of pain relief provided by H2 antagonists, because they are systemically acting drugs and require absorption into the bloodstream. However, one H2 antagonist combines famotidine with the antacids calcium carbonate and magnesium hydroxide (Pepcid Complete) to provide both immediate and prolonged gastric antisecretory action.

The tolerability of the three drugs appears to be similar. However, given the fact that cimetidine is known to inhibit various P450-metabolizing isoenzymes, which could affect metabolism of other drugs and thereby increase their blood concentration, patients should consult with their physicians if they are taking theophylline, warfarin, and phenytoin with cimetidine.

ANTIFLATULENT (ANTIGAS) AGENTS

An antiflatulent is a drug that expels gas from the stomach and intestines. Simethicone (contained in products such as Mylicon, Phazyme, GAS-X, Gas Aid, and others) is the only drug approved as antiflatulent by the FDA. The drug reduces the surface tension of small gas bubbles and is often combined with antacids. The approved OTC indication for the product is "for the relief of painful bloating, and the sensation of pressure and fullness, commonly referred to as gas in the digestive tract." Such gas is frequently caused by excessive swallowing of air or by eating certain foods. The antiflatulent property is useful in conditions in which the retention of gas may be problematic, such as air swallowing, functional dyspepsia, postoperative gaseous distension, peptic ulcers, spastic or irritable colon, or diverticulosis. Tablet or liquid formulations of the drug are used in dosages of 80 to 125 mg, six times daily after meals and at bedtime, or as directed by a physician.

ANTIDIARRHEAL AGENTS

Diarrhea is well recognized by ordinary individuals to be a disorder of "too rapid evacuation of too fluid stools." Diarrhea is a symptom of several gastrointestinal (GI) diseases affecting water and electrolyte transport. Several types of drugs exist for the treatment of diarrhea.

In some situations, diarrhea is an adaptive and defensive process that expedites the removal of toxins, bacteria, fungi, viruses, and protozoa from the GI tract. Under these conditions, it is probably best not to interfere with their removal using an antidiarrheal drug, because this might prolong disease. Moreover, antidiarrheal drugs do not affect the underlying GI disease process, but rather provide control of symptoms.

Antipropulsive Drugs

Loperamide hydrochloride (e.g., Imodium A-D) was previously available only as a prescription-based drug; after about 10 years, it was introduced for OTC use. Loperamide is the only available OTC antidiarrheal drug; it works by inhibiting intestinal propulsive activity and by affecting water and electrolyte transport by the bowel, probably secondary to the inhibition of propulsive motility. Loperamide is an opiate receptor agonist similar to morphine and specifically binds to receptors in the brain and in the gut (myenteric plexus). Loperamide stimulates intestinal circular muscle contractile activity to induce segmentation or mixing, but, unlike morphine, it has reduced ability to cross the blood–brain barrier.

Loperamide is indicated for the symptomatic relief of acute nonspecific diarrhea and should not be used if diarrhea is accompanied by high fever (greater than 101°F), or if blood is present in the stool, suggesting an infectious etiology. The dosage for adults is 4 mg (two caplets or four teaspoonfuls) after the first loose bowel movement. If needed, 2 mg can be used after each additional loose bowel movement. An adult should not exceed a total of 8 mg in a 24-hour period unless directed by a physician. A lower dosage is recommended for children. Warnings appear on the package label and instruct the user not to take the drug for more than 2 days unless directed by a physician. Following systemic absorption, loperamide has a long half-life of 18 hours, and the recommended dosage must therefore be adhered to in order to avoid potential toxicity.

Overdose of loperamide may result in constipation, central nervous system depression, and nausea. A slurry of activated charcoal administered promptly after ingestion may reduce the amount of the drug that is absorbed. In the event of overdosage, patients should be monitored for signs of central nervous system (CNS) depression for at least 24 hours. In addition, children may be more sensitive to the CNS effect of loperamide. If CNS depression is observed, the opiate antagonist naloxone should be used. If responsive to nalaxone, vital signs must be monitored carefully for the recurrence of symptoms of drug overdose for at least 24 hours after the last dose of naloxone.

Bismuth Drugs

Bismuth subsalicylate (Pepto-Bismol) is the only colloidal bismuth salt drug approved for OTC use in multiple GI conditions, including "diarrhea, heartburn, indigestion without constipation, nausea, and upset stomach." The mechanism of action of bismuth subsalicylate in these gastrointestinal disorders is not completely understood. Although colloidal bismuth compounds have no significant acid-neutralizing capacity, they inhibit the action of pepsin, increase the secretion of mucus, and interact with protein in necrotic ulcer craters, presumably forming a barrier to the diffusion of acid. Bismuth colloids have an antibacterial action, which is relevant for the treatment of infectious diarrhea and peptic ulcers (e.g., *Helicobacter pylori*). The salicylate component of bismuth subsalicylate may exert intestinal antiinflammatory and antisecretory actions. Thus, the combined antibacterial and intestinal antisecretory action of bismuth subsalicylate is relevant for the OTC treatment of mild to moderate diarrhea. Pepto-Bismol controls diarrhea within 24 hours, and also relieves associated abdominal cramps.

Pepto-Bismol is available in two liquid suspension forms (262 and 525 mg per 15 ml) and in a tablet dosage form (102 mg per tablet). The recommended dosage for adults is 262 to 1050 mg, to be repeated every hour, if needed, to a maximum of four doses in 24-hour period. Lower dosages are provided for children of various age groups. Pepto-Bismol should not be used in patients who have a known allergy to aspirin. Caution is also advised when administering to patients taking medications for anticoagulation, diabetes, and gout. A warning statement advises caution in the administration of the drug to children, including teenagers, during or after recovery from chicken pox or flu. In addition, the medication may cause a temporary and harmless darkening of the tongue and/or stool.

Adsorbent Drugs

Adsorbents have been so named because of their potential to adsorb intestinal luminal toxins and bacteria associated with some types of infectious diarrhea, and theoretically they should enhance fecal elimination of the toxins or bacteria. Although this adsorption principle has been demonstrated in *in vitro* studies, there have been no meaningful confirmatory *in vivo* studies. Furthermore, adsorbents are of little or no value for the treatment or prevention of acute infectious diarrhea. Clearly, additional clinical studies are needed to

support the utility of these drugs for the treatment of diarrhea. Several types of adsorbent drugs are commercially available for OTC use and include kaolin (e.g., Kaopectate, Donagel), attapulgite (e.g., Diasorb, Rheaban), and pectin (e.g., Kaopectate, Donagel). The official claim for these products is "for the relief of diarrhea and cramps."

Oral Rehydration Therapy

Although healthy adults with episodes of mild to moderate diarrhea do not need extensive oral hydration therapy, they are advised to eat easily digested food and to drink noncarbonated beverages such as fruit juices. Fruit juices contain easily digested sugars and water. It is well established that glucose promotes the intestinal absorption of water, sodium, and other electrolytes, thus overcoming the losses associated with diarrhea. In patients with severe diarrhea, however, and especially young children, the use of oral hydration solutions containing glucose, electrolytes, and amino acids in proportions similar to the World Health Organization formula is advised. Such solutions are available OTC for use in infants and children when recommended by physicians (e.g., Pedialyte).

LAXATIVES

Laxatives are drugs or fibers that relieve the symptoms of constipation. Constipation is understood by lay individuals to be related to decreased fecal frequency, difficulty in initiating fecal passage, the passage of firm or small-volume feces, or a feeling of incomplete evacuation. Although there are several types of laxatives, most laxatives increase intestinal water content via distinct pharmacological mechanisms (summarized in Table III). All laxatives have a general warning statement that they should not be used if abdominal pain, nausea, or vomiting is present.

TABLE III Pharmacological Classifications of Major Drugs Used for the Treatment of Constipation

Pharmacological class	Drug examples
Bulk-forming agents	Hydrophillic colloids, fibers
Osmotic agents	Nonabsorbable magnesium salts
Emollient agents	Mineral oil
Nonspecific stimulants or irritants	Docusate salts, anthraquinone (senna, cascara, aloe), diphenylmethane (bisacodyl), castor oil

Bulk-Producing Laxatives

Bulk-forming hydrophilic colloids such as psyllium husk derived from the plantago seed (e.g., Metamucil), methylcellulose (e.g., Citrucel), and calcium polycarbophil (Fibercon) are most popular with consumers. Psyllium is one of the oldest bulk laxatives, with more than 70 years of clinical application. Psyllium is available in various oral dosage forms, including powder and cookies. The bulking effect of the psyllium husk is due both to the water-holding capacity (gel formation) of the undigested fibers and to an increased bacterial mass following extensive colonic bacterial fermentation (about 85%) of the undigested fibers. Colonic fermentation of undigested psyllium fibers and other natural dietary fibers leads to the formation of short-chain fatty acids, which are osmotically active and add water to the bulk of the stools. This fermentation results in the formation of intestinal gases, which some individuals may find objectionable. However, with dose adjustments, most individuals are able to tolerate chronic treatment with psyllium and other natural dietary fibers. In contrast, the synthetic bulk laxatives methylcellulose and calcium polycarbophil act to increase colonic water content primarily by gel formation, because they are poorly fermentable (15%).

All bulk laxatives produce bowel movements after 12 to 72 hours, and are well tolerated with good safety records. Furthermore, the approved labeling indications for Metamucil (3.4 g psyllium/dose) and Citrucel (2 g/dose) are rather broad, and include the treatment of occasional constipation and, when recommended by a physician, chronic constipation and constipation associated with irritable bowel syndrome, diverticulosis, hemorrhoids, convalescence, senility, and pregnancy. In contrast, the approved labeling for calcium polycarbophil (625 mg/dose) is more limited to "the relief of constipation and to help restore and maintain regularity."

Osmotically Active Agents (Saline Laxatives)

This group contains the drugs magnesium hydroxide (e.g., Phillips Milk of Magnesia, others), magnesium sulfate, magnesium citrate (e.g., Citrate of Magnesia), and sodium biphosphate (e.g., Fleet Phospho-Soda, Fleet Enema). Their cathartic action results from osmotically mediated water retention in the bowel. The increased intestinal volume stimulates intestinal peristalsis and promotes defecation. The magnesium-containing cathartics stimulate release of the intestinal hormone cholecystokinin, which promotes intestinal secretion and increased intestinal motility. The usual adult dose of magnesium hydroxide is 800 to 1600 mg at bedtime, which produces a bowel movement

within 6 hours. Phosphate salts are better absorbed than are magnesium salts and therefore require higher dosage to effect catharsis. Saline cathartics should not be used for more than 1 week unless directed by a physician. Both magnesium- and phosphate-containing preparations are reasonably well tolerated, but they need to be used with caution in patients with renal insufficiency or cardiac disease.

Emollient Laxatives

Mineral oil is the only drug available as an indigestible emollient. When used for 2 to 3 days, mineral oil (e.g., Fleet Mineral Oil Enema, 118 ml/dose; or as an oral liquid) can soften very dry stools. However, its poor taste and possible oil leakage from the anus precludes its regular use. Furthermore, the chronic use of mineral oil could result in foreign body reactions and interference with the absorption of fat-soluble substances, such as some vitamins.

Nonspecific Stimulants (Irritant Laxatives)

Stimulant laxatives induce bowel movement by virtue of their ability to alter intestinal mucosal permeability and stimulate intestinal water and electrolyte secretion in the bowel. This pharmacological class of drugs constitutes four distinct subgroups:

1. Surfactants: Anionic surfactants containing the docusate salts dioctyl sodium sulfosuccinate or dioctyl calcium sulfosuccinate (e.g., Surfak Liqui-Gels, Colace, Doxinate) are mild stimulant laxatives. These surfactants were initially believed to act by lowering the surface tension of stool to allow mixing of aqueous and fatty substances in the stool, thereby permitting easier defecation. However, more recent evidence indicates that these drugs act in a manner similar to that of other stimulant drugs in this class, by stimulating intestinal water and electrolyte secretion. However, despite their widespread use, these surfactants have marginal, if any, efficacy in most cases of constipation.

2. Diphenylmethane derivatives: Bisacodyl (e.g., Dulcolax, 10-mg tablets or 5-mg rectal suppositories) is the only currently available drug in this class, and when taken orally once daily it produces a bowel movement in about 6 hours. A suppository form, however, can produce a bowel movement within 30 to 60 minutes. Bisacodyl is a prodrug and requires *in vivo* hydrolysis for its activation. To avoid activation of the drug in the stomach and consequential gastric irritation, enteric-coated tablets should be swallowed whole without chewing. The laxative effects may be accompanied with cramps and excessive fluid loss. The drug should not be used for more than 10 days, unless advised by a physician.

3. Anthraquinone laxatives: These are plant-derived drugs that have been used as laxatives since ancient times. This group includes senna (e.g., Senokot, Nature Remedy, Ex-Lax), cascara sagrada, and aloe. All of these drugs act by inducing a low-grade inflammatory state in the small and large bowel; this inflammation induces the secretion of water and electrolytes and this is accompanied by giant migrating colonic contractions. The drugs are poorly absorbed in the small intestine and require metabolic activation by colonic bacteria to form monoanthrone derivatives. The usual dose of sennosoides, derived from senna, is 8.6 to 15 mg, which produces a bowel movement in 6 to 12 hours. The chronic use of these drugs may results in laxative dependence and a "cathartic colon" manifested by dilatation of the colon, a relative absence of myenteric plexus neurons, atrophy of the muscularis propria, and the inability to evacuate without their use. This condition is observed typically in women following chronic use of these drugs. Given this colonic pathology, anthraquinone laxatives cannot be recommended for chronic use. In fact, the official labeling for this pharmacological class mandates that these products should not be used for more than 1 week, unless directed by a physician.

4. Castor oil: Castor oil (Neoloid, Purge) is derived from the castor bean plant and contains two noxious ingredients, an extremely toxic protein, ricin, and an oil composed of the triglycerides of ricinoleic acid. In the upper intestine, the triglycerides are hydrolyzed by the pancreatic lipase enzyme to liberate ricinoleic acid, which acts in the small intestine to stimulate fluid and electrolyte secretion and thereby reduce intestinal transit time. Catharsis is produced within 1 to 3 hours. However, due to its unpleasant taste and the potential for toxicity toward the intestinal epithelium and myenteric neurons, similar to the senna, castor oil is now seldom used.

ANTIEMETIC DRUGS

Dimenhydrinate (Dramamine, 50 mg/dose) and meclizine hydrochloride (Bonine, 25 mg/dose) are centrally acting histamine-1 (H1) receptor antagonists that are effective for the "prevention and treatment of nausea, vomiting, or dizziness associated with motion sickness." Although their mechanism of action is not well understood, it is believed that these drugs act by blocking muscarinic receptors in the brain rather than by the blockade of central histamine receptors. These drugs

possess central nervous depressant activity and consumers should be warned not to take these drugs when driving motor vehicles or operating machinery, or when taking sedatives, tranquilizers, or alcohol.

Emetrol, a phosphorylated carbohydrate solution, is an oral preparation formulated for the "relief of nausea associated with upset stomach." Each 5-ml of solution contains 1.87 g of glucose, 1.87 g of fructose, and 21.5 mg of phosphoric acid. The mechanism of the beneficial action is not fully understood, although the official labeling states that this product "has a local action on the hyperactive GI tract." The product should not be diluted. The adult dosage is 15 to 30 ml to be taken every 15 minutes until distress subsides. However, not more than five doses should be taken per hour without consulting a physician.

Of interest is the observation that bismuth subsalicylate (Pepto-Bismol) is also indicated for the treatment of "upset stomach and nausea."

GASTROINTESTINAL DIETARY SUPPLEMENTS

Dietary supplements, which are available on an OTC basis, are frequently used for GI disorders. Unlike drugs, for which there are requirements addressing the proof of safety, efficacy, and good manufacturing practices (GMPs) to assure quality and standardization, dietary supplements do not have these legal requirements. Dietary supplements are therefore marketed on the basis of the 1994 Dietary Supplement Health and Educational Act (DSHEA) passed by the United States Congress. The DSHEA requires that the label state that the product "must not be intended to diagnose, treat, cure, or prevent any disease." Furthermore, dietary supplements do not need premarket approval or review by the FDA, as is the case for drugs, and hence their use by consumers is not based on solid scientific evidence. However, a premarket safety notification by the manufacturers is required for new ingredients. The sponsor can only claim the role of the nutrient or the dietary ingredient "intended to affect structure and function in humans."

Several digestive products and probiotics are available OTC; however, these products have not been evaluated for their clinical efficacy in controlled, double-blind studies. Digestive products are substances that promote the process of digestion in conditions characterized by lack of one or more of the specific substances that digest food. Two digestive (lactase enzyme and α-galactosidase) and one probiotic (*Lactobacillus reuteri*) products are commercially available.

Lactase Enzymes

Lactaid and other similar preparations represent a family of products that contain a lactase enzyme derived from *Aspergillus oryzae*. Lactase aids in the digestion of lactose present in dairy products and converts it to the simple sugars, glucose and galactose. Many adults, particularly those from specific ethnic groups, such as African-Americans or Asian-Americans, have low levels of intestinal lactase. If lactose is not fully digested, it can be fermented by colonic bacteria to induce gas, bloating, cramps, and diarrhea. In such patients, ingestion of lactase may be desirable. Caplets or chewable tablets of Lactaid must be taken with the first bite of dairy food and the dosage needs to be adjusted depending on the desired response. The labeling instructs the user to seek medical attention should he/she experience any unusual symptoms, or symptoms seemingly unrelated to the condition for which the product was taken. As a dietary supplement, lactase enzyme is also added to milk to yield finished products for use by individuals who have lactose intolerance.

α-Galactosidase

The product Beano contains the enzyme α-galactosidase, which is derived from *Aspergillus niger* and aids in the digestion of the sugars raffinose, stachyose, and/or verbascose. These sugars are present in almost all legumes (e.g., beans, peas, chickpeas, lentils, oats) and all or most of the cruciferous vegetables (e.g., cabbage, broccoli). Adverse reactions listed on the label of Beano include cramping and diarrhea as well as allergic-type reactions.

Lactobacillus reuteri

The product Probiotica is a digestive product containing the bacteria *Lactobacillus reuteri*. The product label indicates that this dietary supplement provides "friendly bacteria" to the digestive system. However, there is no specific information on the label to indicate how this product provides improved digestive health. In fact, the mechanisms of action of probiotic bacteria in a variety of intestinal disorders, while supported by anecdotal clinical reports, remain the subject of investigation.

SUMMARY AND CONCLUSIONS

The availability of drugs on an over-the-counter basis provides patients with improved access to effective therapies. However, optimal therapy with OTC drugs

requires that consumers correctly diagnose the underlying condition and safely use a desirable drug.

The OTC gastrointestinal drugs comprise six broad therapeutic categories consisting of antacids and gastric antisecretory agents, antiflatulents, antidiarrheals, laxatives, antiemetics, and dietary supplements. Antacids and histamine (H2) receptor antagonists are effective for the resolution of mild symptoms of occasional heartburn. However, H2 antagonists have a longer duration of action than do antacids when used for the self-treatment of heartburn. The product bismuth subsalicylate, which is not an antacid, is also indicated for the treatment of heartburn. Simethicone is the only approved antiflatulent drug used for the relief of painful bloating, commonly referred to as gas, in the digestive tract. The antidiarrheal drugs loperamide and bismuth subsalicylate are effective for the resolution of diarrheal symptoms. However, adsorbent antidiarrheal drugs are of no value for the treatment and prevention of acute infectious diarrhea. Laxatives of various pharmacological classes are effective for the treatment of constipation. The bulk laxatives have the most physiologic action on the colon, whereas stimulant laxatives have potential for inducing intestinal toxicity and their use should be restricted to short time periods. The histamine (H1) receptor antagonists dimenhydrinate and meclizine are effective antiemetics for the prevention and treatment of nausea and vomiting associated with motion sickness. Since 1994, diverse dietary supplements have become available and consist of either digestive enzymes (e.g., lactase, α-galactosidase) or probiotics.

Clearly, as drug patents expire for prescription-based drugs, many such drugs will likely be switched to an OTC status, thus benefiting consumers and reducing the burden of health care costs. However, it is essential that continuous vigilance is maintained to verify that consumers are safely using OTC drugs and dietary supplements.

See Also the Following Articles

Antacids • Anti-Diarrheal Drugs • H2-Receptor Antagonists • Laxatives • Pharmacology, Overview

Further Reading

Anonymous. (2002). "Physicians' Desk Reference for Nonprescription Drugs and Dietary Supplements," 23rd Ed. Thomson Medical Economics, Montvale, New Jersey.

Brass, E. P. (2001). Changing of the status of drugs from prescription to over-the-counter availability. *N. Engl. J. Med.* **345**, 810–816.

Dajani, E. Z. (2001). Gastroesophageal reflux disease: Pathophysiology and pharmacology overview. *J. Assoc. Acad. Minor. Phys.* **11**, 7–11.

Dajani, E. Z., Bianchi, R. G., East, P. F., Bloss, J. L., Adelstein, G. W., and Yen, C. H. (1977). The pharmacology of SC-27166: A novel antidiarrheal agent. *J. Pharmacol. Exp. Therap.* **203**, 512–526.

Engel, L. W., and Straus, S. E. (2002). Development of therapeutics: Opportunities within complementary and alternative medicine. *Nat. Rev. Drug Discov.* **1**, 229–236.

Hoogerwerf, W. A., and Pasricha, P. J. (2001). Agents for control of gastric acidity and treatment of peptic ulcers and gastroesophageal reflux disease. *In* "Goodman and Gilman's The Pharmacological Basis of Therapeutics" (J. G. Hardman, L. E. Limbird, and A. G. Gilman, eds.), 10th Ed., pp. 1005–1020. McGraw-Hill, New York.

Jafri, S., and Pasricha, P. J. (2001). Agents for diarrhea, constipation, and inflammatory bowel disease; agents used for biliary and pancreatic disease. *In* "Goodman and Gilman's The Pharmacological Basis of Therapeutics" (J. G. Hardman, L. E. Limbird, and A. G. Gilman, eds.), 10th Ed., pp. 1037–1058. McGraw-Hill, New York.

Krishnan, H. S., and Schaefer, M. (2000). Evaluation of the impact of pharmacist's advice giving on the outcomes of self-medication in patients suffering from dyspepsia. *Pharm. World Sci.* **22**, 102–108.

Mackerer, C. R., Clay, G. A., and Dajani, E. Z. (1976). Loperamide binding to opiate receptor sites of brain and myenteric plexus. *J. Pharmacol. Exp. Therap.* **199**, 131–140.

Shaw, M. J., Fendrick, A. M., Kane, R. L., Adlis, S. A., and Talley, N. J. (2001). Self-reported effectiveness and physicians consultation rate in users of over-the-counter histamine-2 receptor antagonists. *Am. J. Gastroenterol.* **96**, 673–676.

Sheen, C. L., and Colin-Jones, D. G. (2001). Over-the-counter drugs and the gastrointestinal tract. *Aliment. Pharmacol. Ther.* **15**, 1263–1270.

Sirgo, M. A., and Schachtel, B. P. (1997). Development of gastrointestinal drugs for over-the-counter indications. *In* "Gastrointestinal Pharmacology and Therapeutics" (G. Friedman, E. D. Jacobson, and R. W. McCallum, eds.), pp. 723–741. Lippincott-Raven Publ., Philadelphia.

Zimmerman, D. R. (1983). "The Essential Guide to Nonprescription Drugs." Harper & Row, New York.

Pancreas, Anatomy

NINA W. FLAY AND FRED S. GORELICK
Yale University and Veterans Affairs Medical Center

ampulla of Vater A common channel receiving the contents of the main pancreatic duct and the common bile duct.
common bile duct Vessel that receives the drainage of the cystic duct of the gallbladder and common hepatic duct of the liver; it drains into the ampulla of Vater.
duct of Santorini A small accessory pancreatic duct located cephalad to the main pancreatic duct.
duct of Wirsung The main pancreatic duct.
pancreatic islet–acinar portal system An arterial supply network that connects the islet cells with the exocrine cells.
sphincter of Oddi Smooth muscle surrounding the major duodenal papilla, where the ampulla of Vater releases its contents into the descending portion of the duodenum.
uncinate process That portion of the pancreas that lies between the descending aorta posteriorly and the superior mesenteric artery anteriorly.

The pancreas is a mixed endocrine and exocrine gland that crosses the midline at the transpyloric plane (L1), extending between vertebral level T10 on the left and L2 on the right. In humans, the pancreas weighs approximately 85 ± 15 g in the adult female, 90 ± 16 g in the adult male, and approximately 5 g in the newborn. Located just posterior to the stomach, the pancreas is mainly a retroperitoneal organ and not readily palpated. Imaging techniques can demonstrate pancreatic anatomy and pathology. Computed tomography (CT) and ultrasound detect inconsistencies in the pancreatic texture or masses (as observed with inflammation or tumors) and define its relationships to neighboring structures. CT, magnetic resonance imaging, and intravenous dyes that can be visualized by X-ray are used to define blood flow and vascular abnormalities. Endoscopic retrograde cholangiopancreatography is a combined endoscopic and X-ray technique in which dye is injected into the pancreatic and/or bile ducts.

ANATOMICAL RELATIONSHIPS

In cross section, the pancreas forms an anterior convex curve, with the central portion of the pancreas located on the midline ridge formed by the upper lumbar vertebrae (Fig. 1). Abdominal trauma can lead to fracture of the pancreatic duct at two sites: where the pancreas crosses the lumbar vertebrae (the most common site) and in the pancreatic tail, where it is attached to the splenic hilum by the splenorenal ligament.

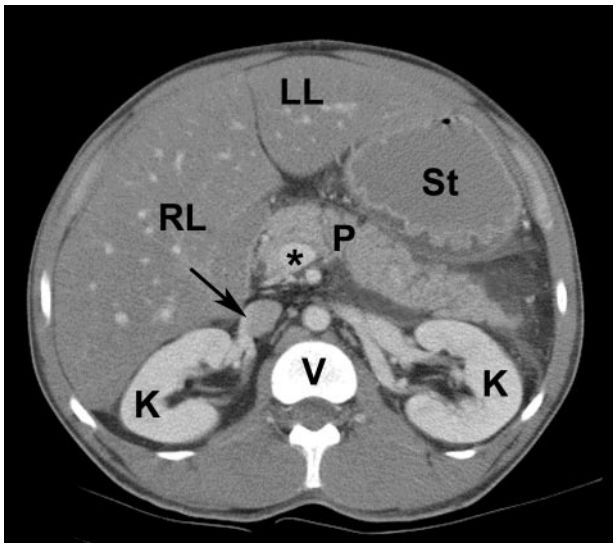


FIGURE 1 CT scan. This axial view reveals relationships of the pancreas (P) to neighboring major structures: liver (right lobe, RL, and left lobe, LL), stomach (St), kidneys (K), vertebral body of lumbar vertebra (V), portal confluence (*), and inferior vena cava (arrow). The body of the pancreas crosses the midline and the tail of the pancreas lies in close proximity to the spleen (not seen) on the left. The distal part of the body and the tail of the pancreas are anterior to the left kidney. The body of the pancreas is anterior to the inferior vena cava and the vertebral body. Note that the pancreas is posterior to the stomach and left lobe of the liver. The portal confluence, visible in the head of the pancreas, is in close approximation to the right lobe of the liver.

The pancreas is related to the duodenum (Fig. 2). The head and uncinate process of the pancreas rest against the duodenal bulb and descending duodenum. Rarely, ulcers may penetrate the duodenal wall and cause pancreatitis. Pancreatic tumors, inflammation, or fibrosis may cause obstruction of the duodenum. The anterior surfaces of the pancreas and duodenum are covered with peritoneum, with the exception of

 Copyright 2004, Elsevier (USA). All rights reserved.

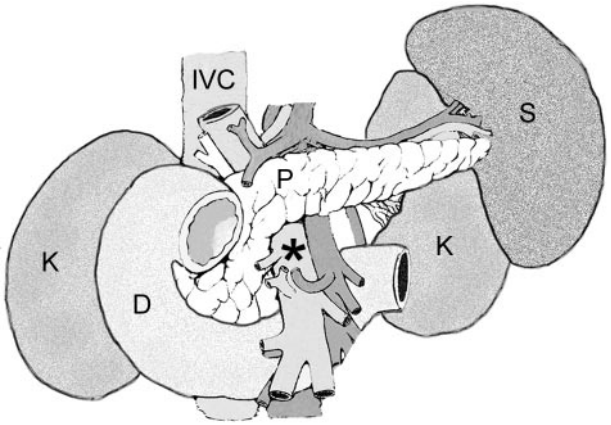


FIGURE 2 *In situ* view of the pancreas. The body of the pancreas (P) crosses the midline of the abdomen anterior to the superior mesenteric vein (*) and artery. The head of the pancreas is in the bulb of the duodenum (D) and its tail is on the left, in the splenorenal ligament that extends between the spleen (S) and the kidney (K). The inferior vena cava (IVC) is the most posterior structure.

the area at the midline where the transverse mesoderm originates.

Structures anterior to the pancreas include the stomach, omental bursa, and transverse colon. The omental bursa (lesser peritoneal sac) lies between the peritoneal coverings of the stomach (anterior) and the pancreas (posterior). The transverse colon crosses anterior to the descending duodenum and the head of the pancreas. The transverse mesocolon is formed by the peritoneal reflections off the posterior abdominal wall and anterior surfaces of the duodenum and pancreas. Pancreatic inflammation may extend into the colon and result in obstruction or bleeding. Gas in the overlying colon or small intestine may obscure visualization of the pancreas by ultrasonography.

Many structures lie posterior to the pancreas (Fig. 2). The tail and distal part of the body of the pancreas are anterior and to the right of the left kidney, with the tail encased in the splenorenal ligament. The splenic vein passes through, or is adjacent to, the pancreas. Chronic pancreatic inflammation can lead to splenic vein thrombosis, causing engorgement of the splenic vessels that connect to the stomach through the left gastric vessels. Enlargement of veins connected to the left gastric vein can result in gastric varices, which are prone to spontaneous rupture and life-threatening bleeding. The superior and inferior mesenteric vessels pass adjacent to the pancreas. If they are involved in pancreatic cancer, resection of the tumor is precluded. The portal vein originates at the junction of the splenic vein and the superior mesenteric veins. In some cases, it originates at a junction composed of three veins: the splenic vein, the superior mesenteric vein, and the inferior mesenteric vein. In either case, the portal vein drains directly into the liver. The common bile duct passes through the head of the pancreas and usually joins the main pancreatic duct to form a common channel that empties into the duodenum. Tumors, inflammation, or fibrosis within the head of the pancreas can obstruct the intrapancreatic portion of the common bile duct. This can cause jaundice and secondary biliary cirrhosis. Since the pancreas does not have a capsule, and there is no peritoneum between the dorsal part of the pancreas and structures posterior to it, tumor cells can spread to all of the structures located posterior to the pancreas.

REGIONS OF THE PANCREAS

The pancreas has four parts: the head, neck, body, and tail. The head of the pancreas is located in the cap of the duodenum. The uncinate process is an extension of the pancreatic head that is located between the superior mesenteric artery and the abdominal part of the descending aorta. Constriction of these vessels compresses the uncinate process. The neck lies anterior to the origin of the portal vein. The body crosses the midline and lies anterior to the aorta, the splenic vein, the left suprarenal gland, the left renal vessels, the left kidney, and the left crus of the diaphragm. The tail is the only intraperitoneal part of the pancreas and lies within the splenorenal ligament.

PANCREATIC DEVELOPMENT AND THE DUCT SYSTEM

The structure of the adult pancreatic duct system is directly related to the dual embryonic origin of the pancreas. The dorsal bud is larger than the ventral bud and gives rise to the major portion of the pancreas. It supplies all of the tail and the body and some of the head and the uncinate process. Initially, the ventral bud is a paired structure, with the left portion atrophying and the right portion continuing to grow. After rotation of the duodenum and the pancreatic buds, fusion of the dorsal and ventral ducts occurs at approximately 6 weeks of human gestation. The main pancreatic duct, called the duct of Wirsung, is surrounded by its own smooth muscle sphincter and arises from the ventral bud. It generally becomes the main conduit in the pancreas. It supplies parts of the head and the uncinate process. Approximately 80% of the time, the main pancreatic duct fuses with the common bile duct, which also

has a smooth muscle sphincter, forming a common channel called the ampulla of Vater. The ampulla of Vater releases its contents into the descending portion of the duodenum at the major duodenal papilla, which is surrounded and innervated by the smooth muscle of the sphincter of Oddi (Fig. 3). Most pancreatic duct contents are released into the duodenum at the major duodenal papilla. In addition to the main pancreatic duct, there is a small accessory duct, the duct of Santorini, which is located cephalad. In the majority of cases, it connects to the main pancreatic duct, but in a small percentage of cases, it has a separate opening into the minor duodenal papilla. The small minor duodenal papilla is cephalad to the major duodenal papilla (Fig. 3).

The pancreatic duct diameter becomes smaller from the head to the tail. Increases in the diameter of the main pancreatic duct are observed with some pathologic processes, such as obstruction by tumors or in chronic pancreatitis. However, the diameter of the pancreas may also increase with age.

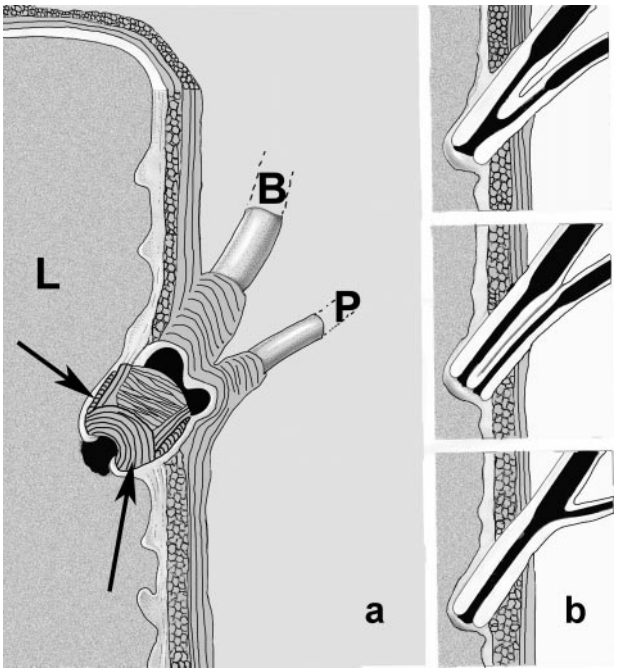


FIGURE 3 Pancreatic duct system. (a) The common bile duct (B) and the main pancreatic duct (P) join to form a common duct (ampulla of Vater). The ampulla of Vater empties its contents into the lumen (L) of the descending portion of the duodenum at the major duodenal papilla, which is surrounded by the sphincter of Oddi (arrows). Note that the main pancreatic duct and the common bile duct are surrounded by their own sphincters. (b) Three variations of the common bile duct and main pancreatic duct system. (Top) Short common channel of the two ducts. (Middle) Two separate ducts. (Bottom) Long common channel of the two ducts. Adapted from Gerard Pucher (1991). *In* "Pancreatitis (Morgenroth and Kozuschek, eds.), Walter de Gruyter.

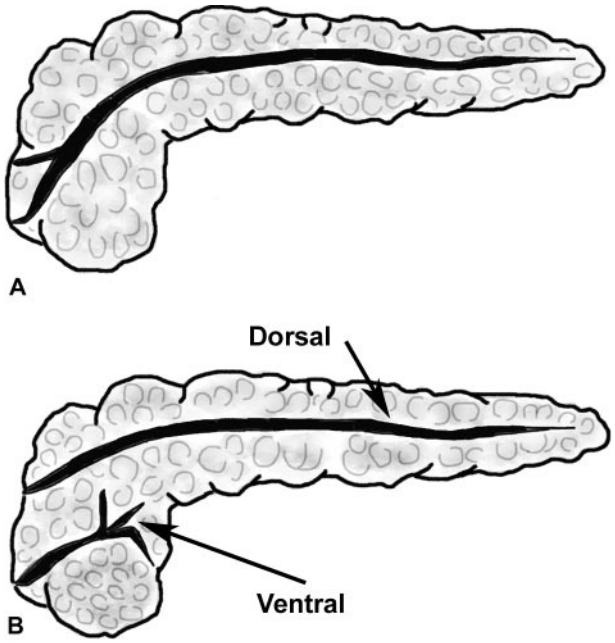


FIGURE 4 Comparison between normal pancreatic duct structure (A) and pancreas divisum (B). Note that in pancreas divisum, there are separate dorsal and ventral pancreatic ducts. There are several variations on this duct pattern.

Variations in the scheme of the pancreatic ducts are frequent. Most often, there is a small accessory duct. However, in approximately 10% of individuals, the dorsal and ventral pancreatic ducts do not fuse. This is called pancreas divisum. Most pancreatic secretions drain though a small minor papillae in pancreas divisum. In a small minority of patients, this may cause relative obstruction of the pancreatic duct that occasionally leads to pancreatitis (Fig. 4). Annular pancreas, a rare condition, may lead to obstruction in the duodenum. It is caused by the lack of atrophy of one of the portions of the ventral bud and the resulting emergence of a bifid ventral pancreatic bud that can surround and constrict the descending duodenum.

VASCULATURE AND LYMPHATICS

Arterial supply to the pancreas is provided by the splenic artery, the pancreatic branches of the gastroduodenal and superior mesenteric arteries, the superior posterior pancreaticoduodenal artery (a branch of the gastroduodenal artery), and the inferior pancreaticoduodenal artery (a branch of the superior mesenteric artery). The splenic artery, a branch of the celiac trunk, lies along the upper border of the pancreas and forms

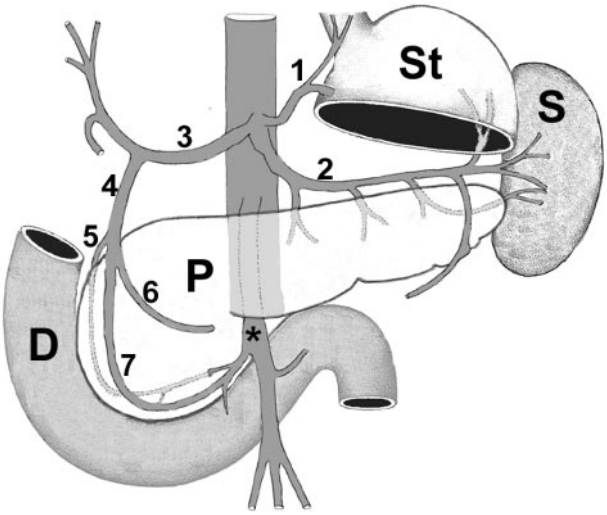


FIGURE 5 Arterial supply to the pancreas. The major arterial supply to the pancreas is from the splenic artery (2) on the superior border of the pancreas and the posterior (5) and anterior (7) superior pancreaticoduodenal arteries, which are branches of the gastroduodenal artery (4). Other vessels in the region include the left gastric (1), common hepatic (3), right gastroepiploic (6), and superior mesenteric (asterisk) arteries.

arcades with the pancreatic branches of the gastroduodenal and superior mesenteric arteries (Fig. 5). The arcades supply the body and tail of the pancreas. The intimate association of the pancreas with critical vascular structures often leads to direct involvement of pancreatic cancer with these blood vessels. This is often the reason that pancreatic cancers cannot be removed surgically. A pancreatic portal system connects the endocrine and exocrine pancreas, allowing arteries supplying the islets to flow directly to the acini located in the immediate vicinity of the islets. Venous drainage of the pancreas occurs by way of the splenic vein, which lies posterior to the pancreas. Venous blood exiting the pancreas flows directly into the liver through the portal vein.

Lymphatic drainage from the pancreas is carried to the pancreaticosplenic, pancreaticoduodenal, subpyloric, and hepatic lymph nodes. Lymph from these nodes flows through the celiac lymph nodes into the intestinal lymphatic trunk and the thoracic duct, ultimately draining into the junction of the left jugular and the left subclavian veins.

NERVOUS INNERVATION

The autonomic nervous system (ANS) plays a major role in the sensory and motor innervations of the pancreas. Nerve fibers are unevenly distributed throughout the pancreas, with the ANS supply being richer in the head of the pancreas than in the tail. The right celiac, hepatic, and superior mesenteric nerve plexuses innervate the head and neck of the pancreas. The celiac plexus and splanchnic neurological networks supply the pancreatic body and tail.

General sensation from the pancreas is carried by visceral afferent fibers of the vagus nerve. Pancreatic visceral pain carried by sympathetic fibers is referred to dermatomes T5–T10, which mark the upper abdomen in the area of the stomach. Severe abdominal pain is a characteristic of both acute and chronic pancreatitis. With chronic disease, the pain is associated with changes in neuronal architecture and changes in neurotransmitter content. Disruption of the perineural sheath may allow toxic substances to come into direct contact with nerve fibers. In patients suffering from chronic pancreatitis, inflammatory cells (lymphocytes, granulocytes, and macrophages) are found around nerves and ganglia supplying the pancreas and the size of nerves is greatly increased. In pancreatic cancer, tumor cells disrupt the perineural sheath and invade the underlying nerve fibers. Severe back pain may be due to the activation of nerve fibers located in the posterior abdominal wall. Changes in neurotransmitter content include the release of increased amounts of two pain transmitter substances: calcitonin gene-related peptide (CGRP) and substance P.

Recent research suggests that some pancreatic pain may be the result of proteolytic activation of the protease-activated receptor PAR-2, a member of the G-protein-coupled receptor family. PAR-2 expression has been detected on a subset of peripheral peptidergic neurons and is involved in the neurogenic component of inflammation. In pancreatitis, PAR-2 may be activated by pathologically generated trypsin from acinar cells or tryptase from mast cells. Pancreatic pain can also be generated by edema, pancreatic duct distension, or ischemia.

Pancreatic secretion is controlled by the parasympathetic and sympathetic nerve fibers as well as peptidergic nerve fibers. Nerve fibers secrete neurotransmitters (acetylcholine from parasympathetic fibers, norepinephrine from sympathetic fibers, and various peptides from peptidergic fibers) along the length of their axons. The neurotransmitters diffuse to target cells, bind to cell surface receptors, and work through signal transduction pathways to stimulate or inhibit pancreatic secretion. Since sympathetic terminals are predominantly associated with blood vessels, decreased blood flow is also related to decreased secretion. Corticotropin-releasing factor (CRF) and CGRP also exert effects on pancreas secretion indirectly, through sympathetic pathways. CRF and CGRP cause the decreases in pancreatic secretion associated with stress. Vagal parasympathetic fibers are also major regulators of

interdigestive secretion, supplying pancreatic acini and islets. They stimulate secretion by releasing acetylcholine, which interacts with the M3 receptor on the acinar cell. A major component of pancreatic secretion stimulated by either cholecystokinin (CCK) or secretin is probably indirect. In response to a meal, CCK is released from I cells and secretin is released from S cells, in the duodenum. These ligands bind to neural pathways and release neurotransmitters that stimulate receptors on acinar and duct cells. They also may have direct effects on exocrine cells. Acetylcholine plays a major role in this pathway. Another major pathway regulating pancreatic secretion is mediated by serotonin, released when food is in the duodenum and interacts with 5-hydroxtryptamine receptors.

Peptidergic nerves in the pancreas exert effects on exocrine tissue but are vulnerable to proteases, so they act in a paracrine manner by releasing peptides that stimulate secretion by neighboring acinar cells. Somatostatin, enkephalin, and pancreastatin are examples of peptides that act in a paracrine manner.

In summary, the pancreas is a mixed endocrine and exocrine gland that crosses the midline at the transpyloric plane (L1) of the abdomen. It consists of four parts: the head, neck, body, and tail. Except for the tail, which is intraperitoneal, the gland is retroperitoneal. This fact, combined with the absence of a capsule, explains the various pathological conditions that arise due to the spread of pancreatic tumors to other retroperitoneal organs. The duct system, of dual embryonic origin, releases its contents into the descending portion of the duodenum. Anastomosis of several major arteries supplies the pancreas. Venous blood drains into the portal vein. Innervation to the pancreas is supplied by the vagus nerve and the ANS. Imaging techniques used to examine the pancreas include computed tomography, ultrasound, endoscopic retrograde cholangiopancreatography, and X-rays.

See Also the Following Articles

Automonic Innervation • Circulation, Overview• Endocrine Pancreas • Exocrine Pancreas • Gastrointestinal Tract Anatomy, Overview • Pancreatic Enzyme Secretion (Physiology)

Further Reading

Adler, G., Beglinger, C., Braun, U., Reinshagen, M., Koop, I., Schafmayer, A., Rovati, L., and Arnold, R. (1991). Interaction of the cholinergic system and cholecystokinin in the regulation of endogenous and exogenous stimulation of pancreatic secretion in humans. *Gastroenterology* **100**, 537–543.

Buchler, M., Weihe, E., Friess, H., Malfertheiner, P., Bockman, E., Muller, S., Nohr, D., and Beger, H. G. (1992). Changes in peptidergic innervation in chronic pancreatitis. *Pancreas* **7**, 183–192.

Friess, H., Shrikhande, S., Shrikhande, M., Martignoni, M., Kulli, C., Zimmermann, A., Kappeler, A., Ramesh, H., and Buchler, M. (2002). Neural alterations in surgical stage chronic pancreatitis are independent of the underlying aetiology. *Gut* **50**, 682–686.

Gorelick, F. S., and Jamieson, J. D. (1994). The pancreatic acinar cell structure–function relationships. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1353–1376. Raven Press, New York.

Hoogerwerf, W. A., Zou, L., Shenoy, M., Sun, D., Micci, M. A., Lee-Hellmich, H., Xiao, S. Y., Winston, J. H., and Pasricha, P. J. (2001). The proteinase-activated receptor 2 is involved in nociception. *J. Neurosci.* **21**, 9036–9042.

Li, Y., Wu, X. Y., Zhu, J. X., and Owyang, C. (2001). Intestinal serotonin acts as paracrine substance to mediate pancreatic secretion stimulated by luminal factors. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G916–G923.

Pancreas, Development

J. K. MAROSKY, A. BHATIA, T. SPILDE, B. PREUETT, K. PRASADAN, AND G. K. GITTES
Children's Mercy Hospital, Kansas City, Missouri

homeobox/homeodomain Specific sequences of transcription factor nucleotides and amino acids that confer a DNA binding patterning/regulation capacity.
lateral inhibition Process by which a single, centrally located cell self-determines its differentiation and, in doing so, inhibits a similar fate selection by adjacent lateral cells.
lineage selection Process by which developmental fate choices are made during organogenesis between different cell lineage types.
Pdx-1 Patterning transcription factor required for early development of the embryonic pancreas.
protodifferentiated Early state of differentiation wherein cells exhibit low-level expression of lineage-specific genes, but have not acquired the higher expression levels characteristic of fully differentiated cells.

The pancreas develops through the evagination of early endoderm. Cells then remain in the epithelium to become the ductal–acinar network or migrate out of the epithelium to give rise to the endocrine cells. These processes are controlled through a complex array of intracellular and extracellular molecular influences. Defects in these pathways can lead to various pathologic states.

INTRODUCTION

The pancreas develops from the caudal foregut, where endodermal cells give rise to a ventral and dorsal pancreatic bud. These two buds then fuse and mesenchymal factors induce the early pancreatic epithelium to undergo a complex pattern of cellular differentiation and lineage selection to yield both epithelial exocrine (acinar and ductal) and nonepithelial endocrine cells. Studies of pancreatic organogenesis have identified several morphogenetic signals, patterning transcription factors, and pathogenetic mechanisms of pancreatic abnormalities. The following overview is a description of the present knowledge of pancreatic organogenesis and morphogenesis.

EMBRYONIC ANATOMY

At approximately 5 weeks of gestation in the human, the pancreatic dorsal and ventral buds evaginate from the endodermal lining of the caudal foregut. The ventral bud moves with the axial rotation of the gut tube during week 6, as the C-loop of the duodenum takes its final position and comes to lie behind and below the dorsal bud. By gestational week 7, fusion of the two pancreatic buds occurs (see Fig. 1).

The most clinically relevant gross anatomy of the exocrine pancreas is the duct system. The main

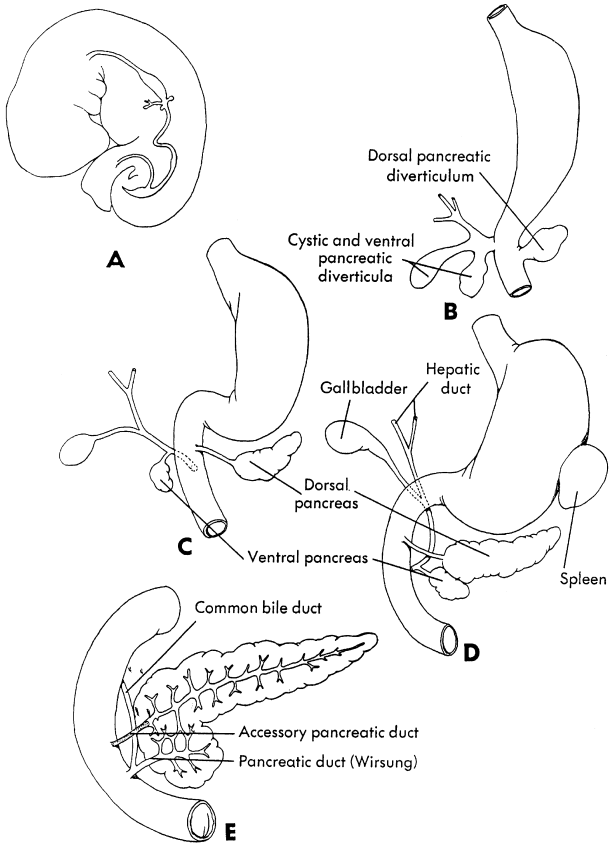


FIGURE 1 Development of the derivatives of the human caudal foregut. (A) Orientation of the gut within a human embryo at about 30 days of gestation. (B–D) The stomach, duodenum, and pancreatic and hepatic diverticula at approximately 30, 33, and 36 days. (E) The definitive relationships of the pancreatic and common bile ducts. Reproduced with permission from Allan, F.G. (1969). "Essentials of Human Embryology." New York: Oxford University Press.

 Copyright 2004, Elsevier (USA). All rights reserved.

pancreatic duct (duct of Wirsung) is formed by fusion of the distal dorsal duct with the entire ventral duct. This main pancreatic duct drains into the duodenum via the ampulla of Vater. The proximal dorsal duct usually persists as an accessory duct, communicates with the main duct, and opens into the duodenum at the minor papilla. When present, this accessory duct is called the duct of Santorini. Approximately 10% of humans with otherwise normally developed pancreata do not have duct fusion, and the entire dorsal duct drains via the minor papilla. The ventral bud develops into the inferior head and uncinate process of the mature pancreas; the dorsal bud provides the remainder. The pancreas is highly vascular and has extensive lymphatic drainage. The connective tissue, septae, and lymphatics of the adult gland are derived from the splanchnic mesoderm.

Histologically, the pancreas is composed of two distinct tissue types, exocrine and endocrine. The exocrine pancreas consists of lobules of acinar cells at the tips of branched ducts. The acinar cells develop zymogen granules containing proenzymes for over 20 digestive enzymes. Zymogens release these proenzymes (nucleases, proteases, amylase, and lipase), into the gastrointestinal tract, where they are activated and participate in digestion. The function of the healthy exocrine pancreas is under complex regulation by hormones, including cholecystokinin (CCK), neurohormones, and secretin.

The cells of the endocrine pancreas constitute only 1–2% of the adult gland, but early in differentiation these represent a major component of the developing pancreas. Clusters of endocrine cells, the islets of Langerhans, form from cells that bud off of the exocrine ducts. Four types of endocrine cells secrete their peptide hormones into the bloodstream from the islets. The majority of pancreatic endocrine cells are beta cells, which produce insulin and amylin (an insulin antagonist). The other three cell types are glucagon-secreting alpha cells, somatostatin-producing delta cells, and the pancreatic polypeptide (PP)-secreting cells. These islets do not arise from a single progenitor cell, and the cellular content of islets is variable. For example, islets with a higher concentration of pancreatic polypeptide-producing cells are found in the head of the pancreas, which derives from the embryologic ventral bud.

CELLULAR DIFFERENTIATION AND MORPHOGENESIS

The normal structural development of the embryonic pancreas during early gestation has been well described. Specifically, Wessells and Cohen in 1967 and Pictet and Rutter in 1972 delineated the morphogenesis of the pancreas in the rat. The first morphologic evidence of the embryonic pancreas, an evagination of the foregut dorsal and ventral endoderm, forms the dorsal and the ventral pancreatic buds, respectively. Dorsal bud evagination occurs first and requires previous contact with overlying notochord, followed by dorsal aorta. The process of evagination and subsequent further growth and differentiation of the pancreatic buds seems to require the presence of the overlying splanchnic mesoderm. This early evagination occurs at 9–10 days of gestation in the mouse.

The developing pancreatic bud starts as a simple sheet of epithelium, which then quickly becomes highly folded. Progressive growth and branching lead to the exocrine (ducts and acini) network. This ductal–acinar structure of the pancreas is evident by 14.5 days in the mouse. Detected throughout the early development of the pancreas, endocrine cells appear to break away and form nonepithelial clusters of endocrine cells that will become vascularized and form the islets of Langerhans.

Pancreas-specific cytodifferentiation has been described by Pictet and Rutter as beginning with a "protodifferentiated" epithelial cell. Exocrine and endocrine cell lineages both originate from these morphologically undifferentiated cells. Expression of lineage-specific mRNA, such as insulin and amylase RNA in these cells, suggests an early commitment to a lineage. Insulin and glucagon genes are expressed prior to dorsal pancreatic bud evagination at embryonic day 9.5 in the mouse. Amylase and other acinar enzyme genes are first expressed around days 11–12 in the mouse.

MOLECULAR INFLUENCES

Although the morphologic and functional development of the pancreas has been well described, the details of the molecular influences regulating pancreatic cell proliferation and differentiation remain elusive. As mentioned previously, contact of early endoderm with the notochord is necessary to induce evagination and subsequent pancreatic differentiation. This interaction is mediated through notochord production of fibroblast growth factor-2 (FGF-2) and activin B, which inhibit the production of sonic hedgehog (SHH) by the endodermal cells that form the pancreas. This SHH suppression is specific to the prepancreatic area. This backdrop of extracellular signaling, accompanied by expression of patterning transcription factors (Pdx-1 and Hlbx9), leads to the early development of the embryonic pancreas. Differentiation pathways determined by homeodomain proteins such as Pdx-1 and Hlbx9 are further modulated by the expression of secondary patterning genes. For example, neurogenin 3, a helix–loop–helix

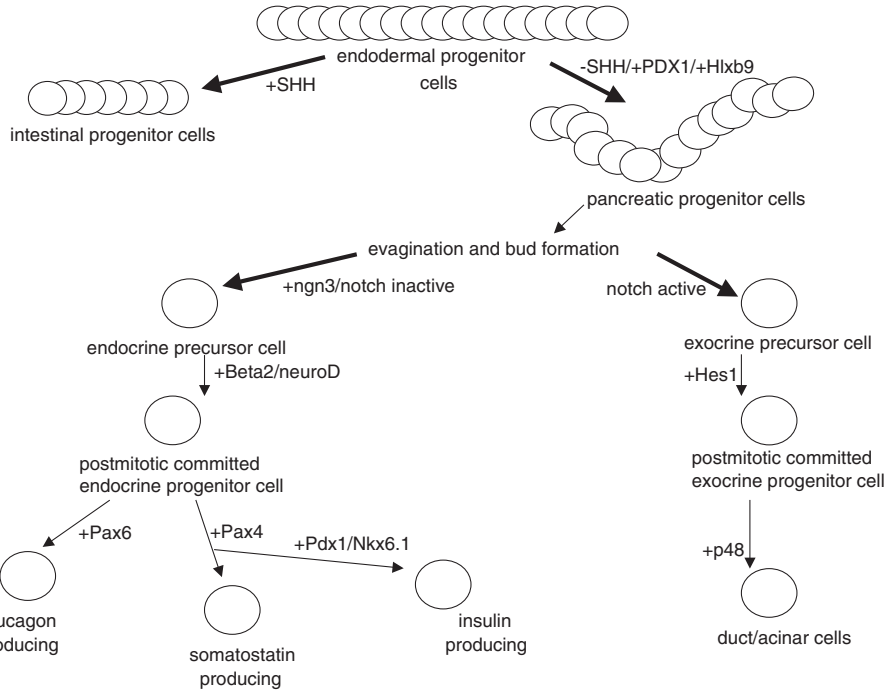


FIGURE 2 Molecular influences and cell lineage map for differentiation of the pancreas. SHH, Sonic hedgehog; ngn3, neurogenin 3.

(HLH) protein, has been found to be necessary for the commitment of Pdx-1-positive cells to become pancreatic endocrine cells. Notch signaling, through mechanisms similar to pathways of neurogenesis called "lateral inhibition," opposes neurogenin 3 signaling and leads to exocrine differentiation. Further levels of lineage selection occur after these events (see Fig. 2).

In endocrine differentiation, Pax4 and Pax6 confer specific differentiation of the glucagon cell (alpha cell) and the insulin cell (beta cell), respectively. In exocrine differentiation, p48 is an HLH protein that is necessary for exocrine differentiation and activates acinar genes. Much of this secondary differentiation beyond the initial pancreatic commitment is mediated by factors in the surrounding pancreatic mesenchyme. FGFs and laminins are examples of such critical mesenchyme-derived factors.

DEVELOPMENTAL ABNORMALITIES

A rare embryologic abnormality of the pancreas is complete agenesis; this has been associated with mutation of the Pdx-1 gene and is frequently fatal in the newborn. A heterozygous Pdx-1 mutation leads to one form of mature-onset diabetes of the young (MODY). Partial pancreatic agenesis describes an otherwise normal pancreas with a portion absent, typically the dorsal pancreas. Hypoplasia of the pancreas, or "lipomatous pseudohypertrophy of the pancreas," is the congenital absence of secondary exocrine structure development. The pancreas has normal islets, but the ductal structures have been replaced by fat.

Pancreas divisum is noncommunication of the ducts in the dorsal and ventral pancreas. Failure of fusion between the main duct (ventral) and the accessory duct (dorsal) results in most of the pancreas draining through the minor papilla via the persistent duct of Santorini. Pancreas divisum is the most common anomaly of the pancreas (~5–10%), and may be more prominent in patients with pancreatitis (16–25%).

Failure of proper pancreatic rotation is thought to lead to a complete ring of pancreatic tissue around the second portion of the duodenum (annular pancreas). Annular pancreas is frequently associated with other anomalies, such as trisomy 21 and duodenal atresia.

Persistent hyperinsulinemic hypoglycemia of infancy (PHHI), also termed nesidioblastosis, is a condition wherein the insulin cells have faulty glucose sensing and thus overproliferate and overproduce insulin. PHHI is often due to mutations in the sulfonylurea receptor 1 (SUR1) () or Kir6.2, and is sometimes associated with the Beckwith–Wiedemann syndrome and multiple endocrine neoplasia type 1(MEN 1). Duct–endocrine proliferation, with new

islet formation from the pancreatic duct epithelium, characterizes nesidioblastosis, but these findings can be seen in the normal neonate as well. Frequently a difficult clinical and pathologic diagnosis, nesidioblastosis requires prompt medical and surgical treatment to avoid brain damage from severe hypoglycemia.

See Also the Following Articles

Development, Overview • Endocrine Pancreas • Exocrine Pancreas • Pancreas, Anatomy

Further Reading

Allan, F. G. (1969). "Essentials of Human Embryology." Oxford University Press, New York.

Edlund, H. (2001). Developmental biology of the pancreas. *Diabetes* **50**(Suppl. 1), S5–S9.

Gittes, G. K., and Rutter, W. J. (1992). Onset of cell-specific gene expression in the developing mouse pancreas. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1128–1132.

Hill, D. J., and Duvillié, B. (2000). Pancreatic development and adult diabetes. *Pediatr. Res.* **48**, 269–274.

Maldonado, T. S., Crisera, C. A., Kadison, A. S., Longaker, M. T., and Gittes, G. K. (2000). Role of mesenchymal–epithelial interactions in pancreas development. *In* "Molecular Basis of Pancreatic Endocrine Development and Function" (M. Hussain, C. Miller, and J. Habener, eds.), pp. 170–188. Kluwer Academic Publ., Boston.

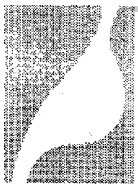
Peters, J., Jürgensen, A., and Klöeppel, G. (2000). Ontogeny, differentiation and growth of the endocrine pancreas. *Virchows Arch.* **436**, 527–538.

Pictet, R., and Rutter, W. J. (1972). Development of the embryonic endocrine pancreas. *In* "Handbook of Physiology" (D. F. Steiner and N. Frenkel, eds.), Sect. 7, Vol. 1. American Physiological Society. Williams & Wilkins, Philadelphia, PA.

Polak, M., Bouchareb-Banaei, L., and Scharfmann, C. P. (2000). Early pattern of differentiation in the human pancreas. *Diabetes* **49**, 225–232.

Schwitzgebel, V. M., Scheel, D. W., Connors, J. R., Kalamaras, L. J. E., Anderson, D. J., Sussel, L., Johnson, J. D., and German, M. S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542.

Slack, J. M. W., (1995). Developmental biology of the pancreas. *Development* **121**, 1569–1580.



Pancreas, Nutritional Effects on

C. S. PITCHUMONI

St. Peter's University Hospital, New Brunswick, New Jersey

kwashiorkor Protein deficiency state with adequate caloric intake.

marasmus Deficiency state resulting from deprivation of both proteins and calories.

nutritional pancreatitis Form of nonalcoholic chronic pancreatitis prevalent in India and other tropical countries.

Inadequate nutrition, especially of protein, may lead to pancreatic atrophy or in some cases chronic pancreatitis or diabetes. Both the exocrine and endocrine components of the pancreas can be injured. Factors causing injury to the pancreas directly or indirectly interfere with both components to a varying degree, although pancreatic exocrine and endocrine components have considerable functional reserve and malabsorption is not seen until 90% of the pancreas is lost.

MALNUTRITION AND EXOCRINE PANCREATIC INJURY

Protein Deficiency

Among the specialized organs of the body, the pancreas (along with liver and small intestine) has the highest rate of protein synthesis. The acinar cells synthesize and secrete between 6 and 20 g of digestive enzymes in 24 hours. Consequently, the pancreas is extremely vulnerable to short- or long-term protein deficiency. Protein energy malnutrition (PEM) is the most important public health problem in developing countries. Malnutrition in affluent nations is often occult, resulting from chronic alcoholism, drug abuse, immunodeficiency states, and problems associated with old age.

Copyright 2004, Elsevier (USA). All rights reserved.

islet formation from the pancreatic duct epithelium, characterizes nesidioblastosis, but these findings can be seen in the normal neonate as well. Frequently a difficult clinical and pathologic diagnosis, nesidioblastosis requires prompt medical and surgical treatment to avoid brain damage from severe hypoglycemia.

See Also the Following Articles

Development, Overview • Endocrine Pancreas • Exocrine Pancreas • Pancreas, Anatomy

Further Reading

Allan, F. G. (1969). "Essentials of Human Embryology." Oxford University Press, New York.

Edlund, H. (2001). Developmental biology of the pancreas. *Diabetes* **50**(Suppl. 1), S5–S9.

Gittes, G. K., and Rutter, W. J. (1992). Onset of cell-specific gene expression in the developing mouse pancreas. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1128–1132.

Hill, D. J., and Duvillié, B. (2000). Pancreatic development and adult diabetes. *Pediatr. Res.* **48**, 269–274.

Maldonado, T. S., Crisera, C. A., Kadison, A. S., Longaker, M. T., and Gittes, G. K. (2000). Role of mesenchymal–epithelial interactions in pancreas development. *In* "Molecular Basis of Pancreatic Endocrine Development and Function" (M. Hussain, C. Miller, and J. Habener, eds.), pp. 170–188. Kluwer Academic Publ., Boston.

Peters, J., Jürgensen, A., and Klöeppel, G. (2000). Ontogeny, differentiation and growth of the endocrine pancreas. *Virchows Arch.* **436**, 527–538.

Pictet, R., and Rutter, W. J. (1972). Development of the embryonic endocrine pancreas. *In* "Handbook of Physiology" (D. F. Steiner and N. Frenkel, eds.), Sect. 7, Vol. 1. American Physiological Society. Williams & Wilkins, Philadelphia, PA.

Polak, M., Bouchareb-Banaei, L., and Scharfmann, C. P. (2000). Early pattern of differentiation in the human pancreas. *Diabetes* **49**, 225–232.

Schwitzgebel, V. M., Scheel, D. W., Connors, J. R., Kalamaras, L. J. E., Anderson, D. J., Sussel, L., Johnson, J. D., and German, M. S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542.

Slack, J. M. W., (1995). Developmental biology of the pancreas. *Development* **121**, 1569–1580.

Pancreas, Nutritional Effects on

C. S. PITCHUMONI

St. Peter's University Hospital, New Brunswick, New Jersey

kwashiorkor Protein deficiency state with adequate caloric intake.

marasmus Deficiency state resulting from deprivation of both proteins and calories.

nutritional pancreatitis Form of nonalcoholic chronic pancreatitis prevalent in India and other tropical countries.

Inadequate nutrition, especially of protein, may lead to pancreatic atrophy or in some cases chronic pancreatitis or diabetes. Both the exocrine and endocrine components of the pancreas can be injured. Factors causing injury to the pancreas directly or indirectly interfere with both components to a varying degree, although pancreatic exocrine and endocrine components have considerable functional reserve and malabsorption is not seen until 90% of the pancreas is lost.

MALNUTRITION AND EXOCRINE PANCREATIC INJURY

Protein Deficiency

Among the specialized organs of the body, the pancreas (along with liver and small intestine) has the highest rate of protein synthesis. The acinar cells synthesize and secrete between 6 and 20 g of digestive enzymes in 24 hours. Consequently, the pancreas is extremely vulnerable to short- or long-term protein deficiency. Protein energy malnutrition (PEM) is the most important public health problem in developing countries. Malnutrition in affluent nations is often occult, resulting from chronic alcoholism, drug abuse, immunodeficiency states, and problems associated with old age.

Copyright 2004, Elsevier (USA). All rights reserved.

Severe nutritional deficiency causes initially reversible and finally irreversible changes. In an experimental study on Bonnet monkeys, animals on protein-deficient, normal carbohydrate diets showed atrophy of pancreatic tissue with replacement with adipose tissue. Animals on low-protein, high-carbohydrate diets showed severe changes. The additional carbohydrate seemed to harm the pancreas more than normal carbohydrate consumption. The low-protein, high-carbohydrate diet mimics the usual diet of the population groups in most developing nations. Other experimental studies have shown recovery of pancreatic function when protein deficiency is corrected, depending on the severity and duration of malnutrition. Clinical observations support experimental studies. Children dying of kwashiorkor have small fibrosed pancreases along with atrophic intestinal mucosa. In children suffering from marasmus and kwashiorkor, pancreatic enzyme output is decreased with no change in HCO_3^- output. The pancreatic ductules, which produce HCO_3^-, are usually well preserved in kwashiorkor.

Micronutrient Deficiencies

Clinical malnutrition is seldom a pure protein deficiency state. Multiple deficiencies of trace elements and vitamins occur. Unopposed free radicals (FRs) are potential mediators of injury to many organs, including the pancreas. The roles played by antioxidant enzymes, which include a mineral in their structure (metalloenzymes), and antioxidant vitamins (such as vitamins A, E, C, and β-carotene) are increasingly clear. Zinc, an essential micronutrient, is a component of enzymes such as DNA polymerase, RNA polymerase, and reverse transcriptase, which are involved in protein synthesis. Experimental studies have shown that zinc deficiency promotes acinar cell degeneration. Clinical zinc deficiency occurs in chronic alcoholism, cirrhosis of the liver, sickle cell disease, and other conditions. Similarly, selenium is an important trace element because it is a component of the enzyme glutathione peroxidase. A selenium-deficient diet in chicks causes pancreatic atrophy. Clinical selenium deficiency occurs in chronic alcoholics and cigarette smokers.

There is no pancreatic disease that can be solely attributed to malnutrition. However, a form of nonalcoholic chronic pancreatitis, prevalent in India and other tropical countries, known as tropical or nutritional pancreatitis, is reported to occur in children and young adults of low-income groups. Although the etiology for tropical pancreatitis is not yet elucidated, experimental data coupled with epidemiological observations have indicated that malnutrition might play a major role in its pathogenesis.

Malnutrition and Endocrine Pancreatic Injury

The role of malnutrition as a cause for endocrine pancreatic injury has only recently been recognized. Although diabetes is often recognized to be a complication of being overweight, abnormal glucose tolerance is a feature of malnutrition. A reduction in insulin secretory capacity is noted in malnourished animals in experimental studies. The functional damage to B cells in malnutrition initially starts as high levels of insulin (hyperinsulinemia) in early stages of subclinical malnutrition and progresses to low levels of insulin (hypoinsulinemia) with the onset of frank malnutrition. In kwashiorkor, there is islet cell hypertrophy in early stages, followed by atrophy. Malnutrition-related diabetes mellitus is recognized to be a subtype of diabetes. It is not clear whether protein deficiency is the sole cause of this diabetes.

See Also the Following Articles

Dietary Reference Intakes (DRI): Concepts and Implementation • Malnutrition • Pancreatitis, Chronic • Protein-Calorie Deficiency—"Kwashiorkor"

Further Reading

Pitchumoni, C. S., and Scheele, G. (1993). Interdependence of nutrition and exocrine pancreatic function. *In* "The Pancreas Biology: Pathobiology and Disease" (Go, *et al.*, eds.). Raven Press, New York.

Rao, H. R. (1988). Diabetes in the undernourished: Coincidence or consequence. *Endocr. Rev.* **9**, 67–87.

Sandhyamani, S., Vijayakumari, A., and Nair, B. (1999). Bonnet monkey model for pancreatic changes in induced malnutrition. *Pancreas* **18**, 84–95.

Pancreatic Anomalies

THEODORE R. LIN AND MICHAEL W. MULHOLLAND
University of Michigan Medical Center, Ann Arbor

endoscopic retrograde cholangiopancreatography A procedure whereby pancreatic and biliary ducts are visualized by endoscopic injection of contrast medium.
stenosis An area of narrowing.

The pancreas develops during the fourth week of gestation, arising from the endodermal lining of the duodenum. From this anlage, two pancreatic pouches, dorsal and ventral, develop, which ultimately give rise to the body/tail and head/uncinate processes of the pancreas, respectively. By the sixth week of gestation, the ventral pouch assumes a position adjacent to the dorsal pouch through the process of clockwise migration as the primitive duodenum rotates to assume its characteristic C-shaped configuration. Fusion of the dorsal and ventral primordia and their ductal systems is achieved by the eighth week. At the time of birth, the pancreatic parenchyma is unified and the accessory and main pancreatic ducts are fused. Malformations of pancreatic embryology include annular pancreas, pancreas divisum, and heterotopic pancreas.

ANNULAR PANCREAS

Annular pancreas occurs when the ventral pancreatic pouch fails to properly rotate clockwise posteriorly around the duodenum (see Fig. 1). Thus, the ventral pouch lies anterior to the duodenum in this anomaly. The duodenum can become partially or completely obstructed by the encircling pancreatic ring (Fig. 2). The band of pancreatic tissue commonly lies proximal to the major duodenal papilla. Microscopically, pancreatic tissue often is found to invade the duodenal wall into the muscularis layer. Other congenital defects associated with annular pancreas include duodenal atresia, Down's syndrome (trisomy 21), intracardiac defects, intestinal malrotation, and tracheoesophageal fistula. Annular pancreas is the most common anomaly obstructing the duodenum in infants, who may present with feeding intolerance and vomiting. On examination, visible peristalsis and distension may be appreciated due to the obstruction. Less commonly, the existence of annular pancreas may not become manifest until adulthood. Patients may complain of bloating, pain, or vomiting. Upper gastrointestinal obstruction, peptic ulceration, and pancreatitis may result from annular pancreas. Diagnosis is facilitated by plain abdominal radiograph, which classically reveals the "double-bubble" sign characteristic of gastric and duodenal dilation secondary to duodenal obstruction. Often, however, upper gastrointestinal contrast study is necessary to document obstruction. Endoscopic retrograde cholangiopancreatogram (ERCP) may also be useful in demonstrating ductal anomalies found with this condition.

Treatment is operative when obstructive symptoms develop. Bypass of the obstruction by duodenojejunostomy is preferred over resection of the annular pancreatic tissue due to the high predisposition to develop pancreatic and duodenal fistulas.

PANCREAS DIVISUM

Failure of the ventral and dorsal pancreatic ductal systems to fuse results in pancreas divisum. In this disorder, the majority of the pancreas is drained through the accessory pancreatic duct (duct of Santorini) into the minor duodenal papilla (Fig. 3). Pancreatic secretions from the uncinate process and portions of the pancreatic head, derived from the ventral anlage, drain via the major duodenal papilla (via the duct of Wirsung), separate from the remainder of the pancreas. Pancreas divisum is relatively common, demonstrated in 5 to 10% of subjects by autopsy or by ERCP.

Manifestations of pancreas divisum occur uncommonly in childhood. Controversy exists as to whether pancreas divisum has a causative role in recurrent idiopathic acute pancreatitis, chronic pancreatitis, and chronic abdominal pain. Some investigators believe that stenosis of the minor duodenal papilla must coexist with pancreas divisum for pancreatitis to develop. Diagnosis of pancreas divisum is made by ERCP; on cannulation of the accessory papilla, the duct of Santorini is found to span the length of the pancreas without communicating with the duct of Wirsung.

Given the uncertainty of a causal relationship between pancreas divisum and pancreatitis, no specific intervention is advised for mild or single episodes of pancreatitis. Severe or recurrent bouts of pancreatitis may warrant therapeutic invention. Both endoscopic and surgical approaches have been

 Copyright 2004, Elsevier (USA). All rights reserved.

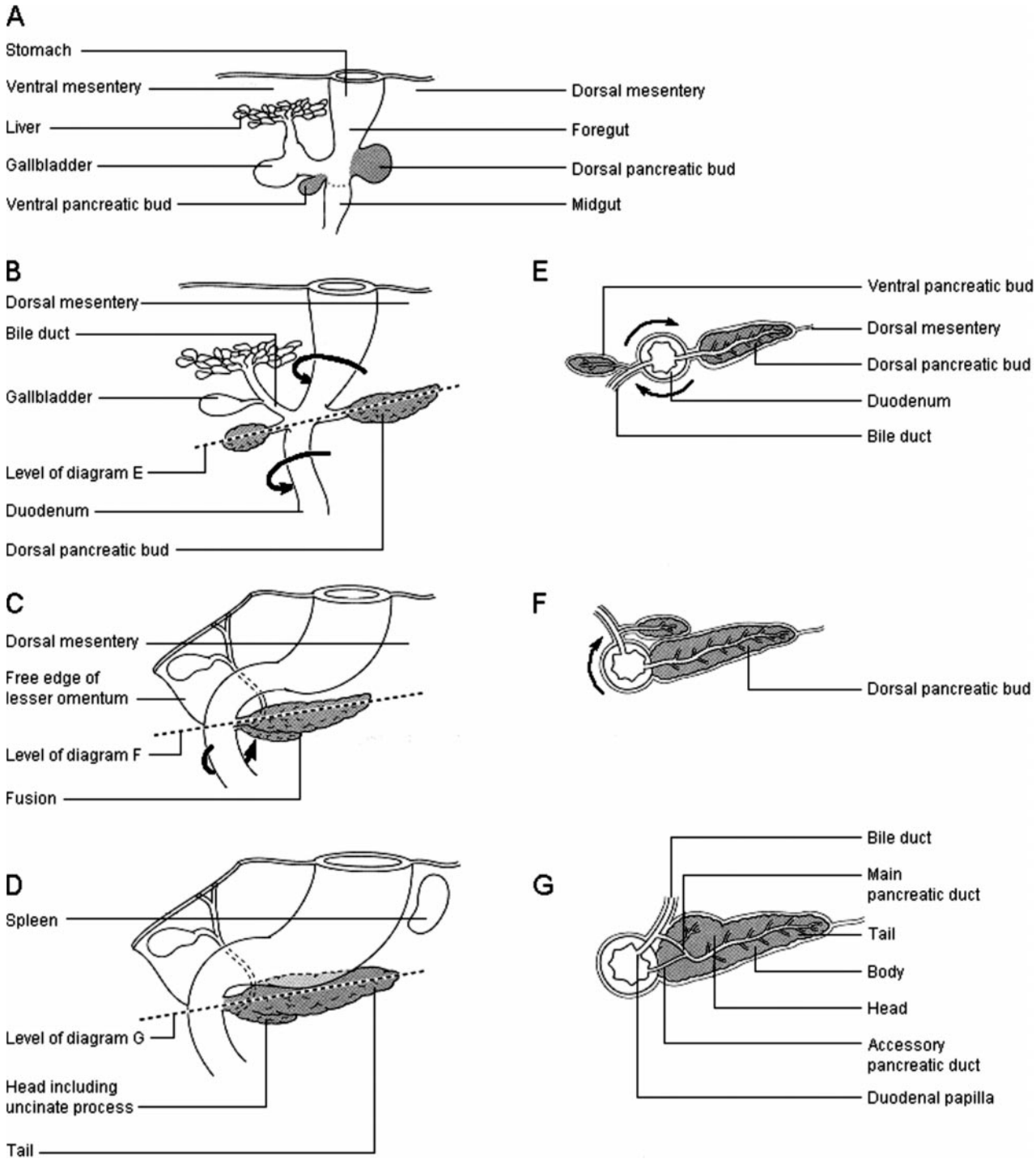


FIGURE 1 Development of the fetal pancreas from the fifth to the eighth week. Rotation (arrows) of the duodenum brings the ventral bud into apposition with the dorsal bud, with subsequent fusion. Reprinted from Anderson, D. and Brunicardi, F. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., with permission. Copyright Lippincott-Raven Publishers, 1997.

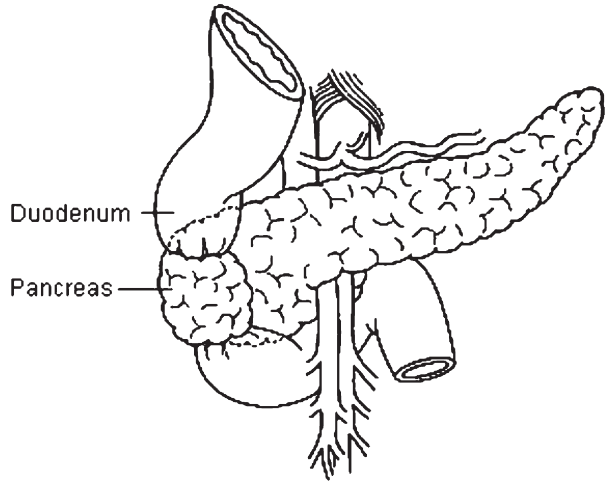


FIGURE 2 Annular pancreas with consequent duodenal obstruction secondary to the encircling pancreatic ring. Reprinted from Oldham, K. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., with permission. Copyright Lippincott-Raven Publishers, 1997.

described. Stenosis of the minor duodenal papilla or stricture of the accessory pancreatic duct may be approached endoscopically with dilation, stent placement, or sphincterotomy. Surgical approaches include transduodenal sphincteroplasty of the minor papilla, pancreatic head resection, and pancreaticojejunostomy.

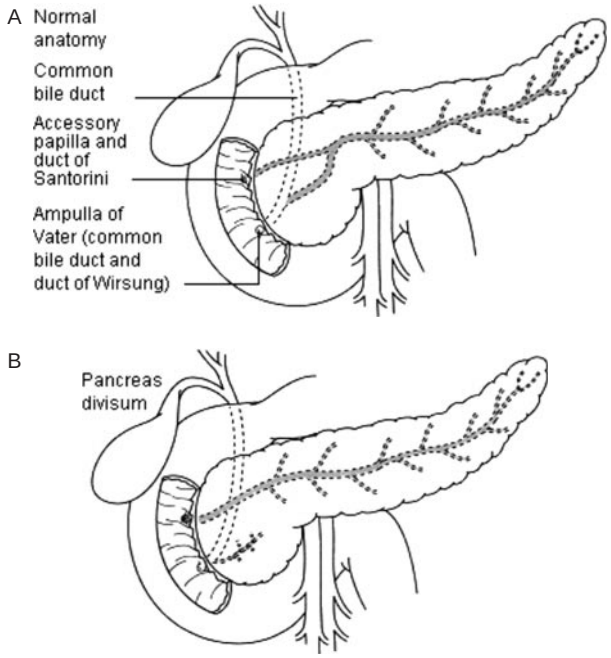


FIGURE 3 (A) Normal ductal anatomy. (B) Pancreas divisum. No communication exists between the duct of Wirsung and the duct of Santorini. Most of the pancreas is drained through the duct of Santorini. Reprinted from Coran, A. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., with permission. Copyright Lippincott-Raven Publishers, 1997.

HETEROTOPIC PANCREAS

When pancreatic tissue develops outside of the main pancreatic body, these aberrant rests of tissue are referred to as heterotopic pancreas. Heterotopic pancreatic tissue may be found anywhere along the gastrointestinal tract, most commonly in the stomach, duodenum, small bowel, and Meckel's diverticulum. Rarely, ectopic pancreatic tissue may be found in the gallbladder, omentum, umbilicus, colon, or lungs. The etiology of heterotopic pancreas is unclear. Aberrations in stem cell differentiation or migration of the pancreatic pouches during development are thought to be involved in this anomaly.

Heterotopic pancreas is usually an asymptomatic condition found incidentally at autopsy or surgery. Bowel obstruction may result when rests serve as lead points for intussusception or as space-occupying lesions. Other complications include hemorrhage and ulceration. Upper endoscopy is helpful in diagnosis. Submucosal nodules with central depression characteristic of heterotopic pancreas may be discerned. Biopsy is required to distinguish these lesions from polyps, leiomyomas, and lymphoma. Treatment is indicated for patients with complications from heterotopic pancreas. Excision with histologic examination is required to exclude malignancy.

See Also the Following Articles

Duodenal Obstruction • Duodenal Ulcer • Exocrine Pancreas • Pancreatitis, Acute • Pancreatitis, Chronic • Pancreatitis, Pediatric

Further Reading

Grendell, J. H., and Ermak, T. H. (1998). Anatomy, histology, embryology, and developmental anomalies of the pancreas. *In* "Sleisenger and Fordtrans's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, B. Scharschmidt, M. Sleisenger, and R. Zoras, eds.), 6th Ed., pp. 761–770. W. B. Saunders Co., Philadelphia, PA.

Liddle, R. A. (1991). Congenital and hereditary diseases of the pancreas. *In* "Textbook of Gastroenterology" (T. Yamada, D. Alpers, C. Owyang, D. Powell, and F. Silverstein, eds.), 1st Ed., Vol. 2., pp. 1937–1942. J. B. Lippincott Co., Philadelphia, PA.

Oldham, K. (1997). Pediatric abdomen. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., pp. 2038–2040. Lippincott-Raven, Philadelphia, PA.

Yeo, C. J., and Cameron, J. L. (2001). Exocrine pancreas. *In* "Sabiston Textbook of Surgery" (C. Townsend, R. Beauchamp, B. Evers, and K. Mattox, eds.), 16th Ed., pp. 1112–1114. W. B. Saunders Co., Philadelphia, PA.

Pancreatic Bicarbonate Secretion

R. MAYNARD CASE*, HIROSHI ISHIGURO†, AND MARTIN C. STEWARD*
*University of Manchester, United Kingdom and †Nagoya University, Japan

cholecystokinin A peptide hormone produced by endocrine cells of the upper small intestine.
vasoactive intestinal peptide A polypeptide from the small intestine with cardiovascular and gastrointestinal effects.

Following appropriate stimulation, the pancreas secretes a bicarbonate (HCO_3^-)-rich fluid that is derived largely from duct cells. Understanding how the pancreatic ductal epithelium is able to secrete HCO_3^- at a concentration five- to sixfold greater than that in plasma is of intrinsic interest. Moreover, it is important from a clinical viewpoint: defects in ductal secretion underlie the pancreatic pathology that characterizes cystic fibrosis and perhaps certain forms of pancreatitis, whereas 90% of pancreatic cancers are of ductal origin.

INTRODUCTION

Pancreatic juice is the product of two distinct secretory processes: protein (enzyme) secretion and electrolyte secretion. Enzymes are secreted by exocytosis. Electrolyte secretion is achieved by the vectorial transport of ions across the secretory epithelium accompanied by water in isotonic proportions. The most significant of these ions is bicarbonate (HCO_3^-). Each day, these two secretory processes result in the human pancreas delivering 6–20 g of digestive enzymes to the duodenum in approximately 2.5 liters of HCO_3^--rich fluid. Although the role of pancreatic enzymes is defined, that of pancreatic HCO_3^- secretion is less precise. Clearly, the fluid acts as a vehicle for transporting enzymes to the duodenum where the HCO_3^- neutralizes gastric acid. Pancreatic HCO_3^- may also aid disaggregation of secreted enzymes following their exocytosis.

PATTERNS OF BICARBONATE SECRETION

The regulation of pancreatic electrolyte secretion and the volume and composition of the secreted fluid differ considerably from species to species (Table I). It is important to recognize these differences for two reasons: (1) observations made in one experimental species, especially those involving models of disease, cannot be assumed to be relevant to humans; (2) a regulatory or secretory mechanism dominant in one experimental species may be present to a small, perhaps unrecognized, extent in humans where it could account for otherwise inexplicable symptoms and/or be explored therapeutically. The major differences and their relevance are as follows:

TABLE I Species-Dependent Patterns of HCO_3^- Secretion

Species	Stimulus	Volume	Maximum $[HCO_3^-]$ (mM)
Dog, cat, and human	Spontaneous	0 (+)	—
	+ Secretin	+++++	145
	+ CCK	+	60
	+ Vagus	+	?
Rat	Spontaneous	+	25
	+ Secretin	++	70
	+ CCK	+++	30
	+ Vagus	++	?
Guinea pig	Spontaneous	+	95
	+ Secretin	+++++	150
	+ CCK	+++	140
	+ Vagus	+++	120

Note. This table gives an idea of the response to stimuli given alone: potentiation often occurs when stimuli are given together. Most data were obtained from studies on anesthetized animals: quantitative differences may occur in conscious animals, especially in the rat, in which secretion is increased fivefold in conscious animals. CCK, cholecystokinin.

1. In all species, secretin evokes the secretion of HCO_3^--rich pancreatic juice. However, the amount of fluid that is secreted varies; an especially small amount is secreted in the rat (approximately fivefold less than that in cat, per gram of tissue). During maximal stimulation, HCO_3^- concentration reaches

 Copyright 2004, Elsevier (USA). All rights reserved.

130 mmol/liter or more in all species except the rat, in which 70 mmol/liter is approximately the maximum value observed.
2. There is a reciprocal relationship between juice HCO_3^- and Cl^- concentrations: as flow rate increases, so does HCO_3^- concentration, with a corresponding reduction in Cl^- concentration. This reciprocal relationship results from a flow rate-dependent loss of HCO_3^- from the primary secretion in exchange for Cl^- as the fluid passes down the ductal tree.
3. The effect of cholecystokinin (CCK) on fluid secretion is very variable. In rat, it evokes a relatively large volume of Cl^--rich fluid that is secreted by acinar cells. In guinea pig, it evokes a HCO_3^--rich fluid. Studies on duct segments isolated from guinea pig pancreas indicate that the ducts are a source of this secretion, but do not exclude an acinar component. In other species, the effect of CCK lies somewhere between these two extremes. Where CCK is a weak stimulant of HCO_3^- secretion, it usually potentiates the action of secretin.
4. The influence of vagal stimulation is also complex. In some species, muscarinic cholinoceptor activation evokes HCO_3^- secretion. In some species (notably pig and guinea pig), vagal stimulation evokes a copious HCO_3^--rich secretion because in these species the vagus nerves contain many VIPergic neurons and the vasoactive intestinal peptide released from these neurons acts in a manner similar to secretin.

In addition to these classical stimulatory mechanisms, many other candidate stimulatory and inhibitory control mechanisms undoubtedly influence pancreatic HCO_3^- secretion to a greater or lesser extent.

ORIGIN OF BICARBONATE SECRETION

Studies on isolated pancreatic duct segments confirm the generally held view, first proposed 50 years ago, that pancreatic duct cells are the principal site of HCO_3^- secretion, without necessarily excluding a possible contribution from acinar cells. Given that duct cells constitute approximately 5% of gland mass, if all pancreatic fluid secretion was derived from duct cells, they would secrete their own volume of fluid in approximately 2 min. This assumes that all duct cells contribute equally to secretion. However, there is some evidence that the terminal duct cells (centroacinar cells) and small ducts contribute disproportionately to secretion, in which case the fluid secretory rate in that region is even faster.

MECHANISM OF DUCTAL BICARBONATE SECRETION

The "textbook" model of HCO_3^- secretion by pancreatic ducts has the following components: (1) generation of intracellular HCO_3^- by the hydration of CO_2 under the influence of carbonic anhydrase; (2) extrusion from the cell of the residual protons by means of a Na^+/H^+ exchanger in the basolateral membrane; (3) secretion of HCO_3^- across the apical (lumenal) membrane in exchange for Cl^- on an anion exchanger working in parallel with an anion channel that allows entry of Cl^- into the lumen. This lumenal anion channel is usually regarded as being the cystic fibrosis transmembrane conductance regulator protein (CFTR).

This model arose from early studies on perfused whole glands and, more recently (and more importantly), from experiments on isolated duct segments from rat pancreas. Although the model can successfully generate the HCO_3^- concentrations observed in the rat (i.e., approximately 70 mmol/liter), on thermodynamic grounds it seems most unlikely to be able to generate the higher concentrations (up to 150 mmol/liter) seen in other species. However, recent observations have suggested a number of modifications to the model, which help to explain how the ducts of these species achieve such high secretory HCO_3^- concentrations.

At the basolateral membrane, the situation is relatively clear. Although the Na^+/H^+ exchanger may be involved to a small extent in HCO_3^- accumulation across the basolateral membrane, experimental studies on duct segments isolated from guinea pig pancreas show that it is largely achieved by a $Na^+-HCO_3^-$ co-transporter (NBC) located in this membrane. Furthermore, the electrogenicity of the NBC contributes to the driving force for HCO_3^- secretion across the lumenal membrane. Immunohistochemical studies confirm that NBC is expressed in the basolateral membrane of human duct cells. However, since NBC is also present in rat ducts, its presence alone cannot account for the much higher HCO_3^- concentrations secreted by the guinea pig pancreas.

At the lumenal membrane, the situation is less clear. Although spontaneous secretion in guinea pig ducts involves anion exchange across the lumenal membrane (Fig. 1A), secretin-evoked secretion can occur in the nominal absence of lumenal Cl^- or when the activity of the anion exchanger is blocked. Furthermore, raising the lumenal HCO_3^- concentration (as occurs during secretin stimulation) actually inhibits lumenal anion exchanger activity, thereby helping to prevent the reabsorption of secreted HCO_3^-. There must, therefore, be an

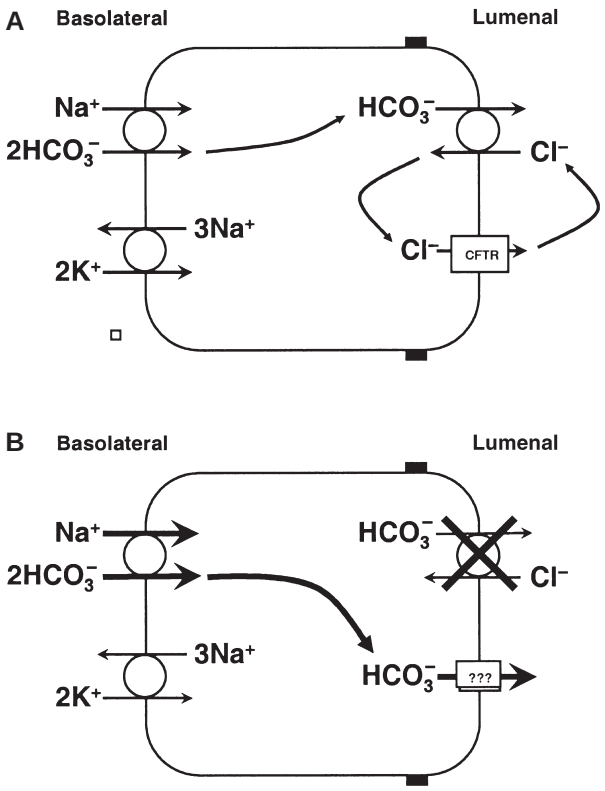


FIGURE 1 HCO_3^- secretion in guinea pig pancreatic duct. (A) During spontaneous secretion, HCO_3^- enters the lumen by anion exchange with Cl^-, which recycles across the lumenal membrane via the CFTR Cl^- channel. (B) Following stimulation with secretin, the anion exchanger becomes inhibited by the high lumenal HCO_3^- concentration and HCO_3^- secretion occurs mainly by diffusion through a lumenal anion channel, probably CFTR. In both situations, HCO_3^- uptake across the basolateral membrane is achieved by Na^+–HCO_3^- co-transport and, to a lesser extent, through H^+ extrusion by Na^+/H^+ exchange (not shown). Both processes derive their energy from the inward Na^+ gradient maintained by the basolateral Na^+,K^+-ATPase.

alternative pathway for HCO_3^- secretion across the lumenal membrane under these conditions.

In contrast to its effect in rat ducts, secretin does not greatly depolarize cells in guinea pig ducts; instead the membrane potential remains at approximately −60 mV. Therefore, even at a lumenal HCO_3^- concentration of 125 mmol/liter and a cytoplasmic HCO_3^- concentration calculated to be 20 mmol/liter, the electrochemical gradient for HCO_3^- would favor diffusion from the cell to the lumen via an anion channel. Furthermore, under these conditions intracellular Cl^- dips to very low values (approximately 7 mmol/liter). Consequently, a HCO_3^--rich secretion is favored by the lack of driving force for Cl^- provided that the anion conductance at the lumenal membrane also conducts HCO_3^- (Fig. 1B). Whether CFTR or another anion conductance provides such a pathway for HCO_3^- remains to be clarified.

In other words, given a large enough HCO_3^- conductance at the lumenal membrane and a low intracellular Cl^- concentration, there may be no need to invoke any other membrane transport protein in the lumenal membrane to explain the high HCO_3^- concentration observed in pancreatic juice from the guinea pig. If so, what is true for the guinea pig may presumably be true of other species, including humans, that show a similar pattern of HCO_3^- secretion.

Finally, why does the rat not achieve such high HCO_3^- concentrations? The answer may lie in the basolateral rather than the lumenal membrane. In rat (and mouse) ducts, fluid secretion can be inhibited by bumetanide. It thus seems likely that a Na^+–K^+–$2Cl^-$ co-transporter is present on the basolateral membrane of these species, which acts to increase intracellular Cl^- concentration. As a result, there are driving forces for both Cl^- and HCO_3^- secretion via the lumenal anion conductance. This causes the production of a secretion containing comparable concentrations of the two anions. The absence or inactivity of this co-transporter in the guinea pig (and other species?) ensures that the secretion is HCO_3^--rich.

See Also the Following Articles

Cholecystokinin (CCK) • Pancreatic Enzyme Secretion (Physiology) • Secretin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Case, R. M. (1998). Pancreatic exocrine secretion: Mechanisms and control. *In* "The Pancreas" (H. G. Beger, *et al.*, eds.), pp. 63–100. Blackwell, Oxford.

Case, R. M., and Argent, B. E. (1993). Pancreatic duct secretion: Control and mechanisms of transport. *In* "The Pancreas: Biology, Pathobiology, and Disease" (G. L. W. Go, E. P. Di Magno, J. D. Gardner, E. Lebenthal, H. A. Reber, and G. A. Scheele, eds.), pp. 301–550. Raven Press, New York.

Ishiguro, H., Naruse, S., San Román, J. I., Case, R. M., and Steward, M. C. (2001). Pancreatic ductal bicarbonate secretion: Past, present and future. *J. Pancreas* **2**, 192–197. Available at http://www.joplink.net.

Ishiguro, H., Steward, M. C., Sohma, Y., Kubota, T., Kitagawa, M., Kondo, T., Case, R. M., Hayakawa, T., and Naruse, S. (2002). Membrane potential and bicarbonate secretion in isolated interlobular ducts from guinea-pig pancreas. *J. Gen. Physiol.*, **120**, 617–628.

Sohma, Y., Gray, M. A., Imai, Y., and Argent, B. E. (2000). HCO_3^- transport in a mathematical model of the pancreatic duct. *J. Membr. Biol.* **176**, 77–100.

Pancreatic Cancer

CHARLES E. BINKLEY AND DIANE M. SIMEONE
University of Michigan Medical Center, Ann Arbor

CA 19–9 Widely used serum marker for pancreatic cancer; has limited specificity because it is also expressed in other malignancies as well as in pancreatitis, hepatitis, and biliary obstruction.

chemical splanchnicectomy Chemical blockage of the celiac nerve plexus, performed either intraoperatively or percutaneously, in order to palliate pain associated with pancreatic adenocarcinoma.

double duct sign Dilation of the distal pancreatic and common bile ducts, with proximal stricturing of both structures evident on endoscopic retrograde cholangiopancreatography and highly suggestive of adenocarcinoma affecting the head of the pancreas.

endoscopic ultrasonography Imaging technique whereby an ultrasound probe attached to an endoscope is used to assess structures through the gastric and duodenal wall.

gemcitabine Chemotherapeutic agent and potent radiosensitizer that has recently been added to the multimodality treatment regimen for pancreatic adenocarcinoma.

hereditary pancreatitis Autosomal dominant trait, leading to recurrent acute pancreatitis and a 40-fold increased risk of pancreatic adenocarcinoma.

K-ras Oncogene; the mutated form is found in over 90% of pancreatic cancers.

laparoscopic staging Minimally invasive means of evaluating the resectability of pancreatic adenocarcinoma by visualization through a tube inserted into the abdomen.

neoadjuvant therapy Strategy of preoperative administration of chemoradiation aimed at increasing the number of patients able to complete multimodality therapy.

palliation Treatment of symptoms such as abdominal pain to optimize quality of life, limit morbidity, and increase survival.

pancreatic intraepithelial neoplasia Areas of focal ductal proliferation adjacent to infiltrating pancreatic cancers that may be precursor lesions to pancreatic adenocarcinoma.

pancreaticoduodenectomy Surgical excision of the head and uncinate process of the pancreas by en bloc resection of the distal stomach and duodenum to the ligament of Treitz, common bile duct, and head of the pancreas. Also known as a Whipple procedure.

Pancreatic adenocarcinoma is one of the deadliest known human malignancies, with an overall 5-year survival rate of less than 5%. Among gastrointestinal malignancies, pancreatic adenocarcinoma is second to colorectal cancer in terms of incidence in the United States, with 30,300 estimated new cases in 2002. Pancreatic cancer deaths in the United States in 2002 were estimated to approach 29,700, making pancreatic cancer the fifth most common cause of cancer-related mortality. Delayed diagnosis, relative chemotherapy and radiation resistance, and an intrinsic biologic aggressiveness all contribute to the abysmal prognosis associated with pancreatic adenocarcinoma. The risk of pancreatic adenocarcinoma is twice as great in men compared to women, and African Americans as well as Japanese Americans have a higher incidence compared to other ethnic groups, suggesting an as-yet undetermined specific genetic or environmental association.

EPIDEMIOLOGY

Many risk factors for pancreatic adenocarcinoma have been identified, with cigarette smoking having the strongest overall association and thought to account for one-quarter of all patients diagnosed. This may also in part explain the greater number of men diagnosed with pancreatic adenocarcinoma. The mechanism believed to be responsible for the association between cigarette smoking and pancreatic cancer involves the *N*-nitroso compounds present in cigarette smoke. Exposure to these agents leads to pancreatic ductal hyperplasia, a possible precursor to adenocarcinoma.

Other factors associated with increased risk of pancreatic adenocarcinoma include saturated fat intake, exposure to nonchlorinated solvents, and the pesticide dichlorodiphenyl trichloroethane (DDT), although the overall contribution of these is likely small. The risk of pancreatic adenocarcinoma increases with age, with most patients diagnosed between 60 and 80 years old. Studies examining the risk between alcohol consumption and pancreatic adenocarcinoma are equivocal, except in the case of heavy alcohol use leading to chronic pancreatitis. Chronic pancreatitis clearly increases the risk of pancreatic adenocarcinoma, although the direct or indirect role of alcohol is not yet defined.

 Copyright 2004, Elsevier (USA). All rights reserved.

MOLECULAR GENETICS

Genetic analysis has shown two distinct patterns of inherited risk for developing pancreatic adenocarcinoma: those inheriting mutations specifically in the *PRSS1* gene, which leads to hereditary pancreatitis, and those inheriting mutations in general cancer susceptibility genes. Hereditary pancreatitis is an autosomal dominant trait, with complete penetrance. Individuals inheriting a mutated *PRSS1* gene, which codes for cationic trypsinogen, develop recurrent acute pancreatitis, frequently beginning in childhood. These individuals have a 40-fold increased risk of developing pancreatic adenocarcinoma when compared to the general population. The mechanism seems to involve recurrent bouts of acute pancreatitis, with increased pancreatic ductal cell turnover, leading to increased rates of DNA synthesis and accompanying random mutations. It is believed that the accumulation of random genetic mutations in key oncogenes and tumor suppressor genes finally leads to the development of pancreatic adenocarcinoma.

Increased risk for developing pancreatic adenocarcinoma also occurs when individuals inherit germ-line mutations in cancer susceptibility genes. A germ-line mutation in the *BRCA2* gene is present in approximately 5–10% of pancreatic adenocarcinomas and predisposes an individual to pancreatic as well as breast and ovarian cancer. The p16 tumor suppressor gene is mutated in over 95% of the cases of pancreatic adenocarcinoma, usually by a somatic mutation, yet germ-line mutations have been described and are associated with an increased risk of both pancreatic malignancy and melanoma.

Inherited mutations in the *LKB1/STK11* gene leads to Peutz–Jeghers syndrome, which is associated with the development of benign gastrointestinal polyps as well increased risk of gastrointestinal cancers, including pancreatic adenocarcinoma. Other familial cancer syndromes involving increased risk of pancreatic adenocarcinoma include hereditary nonpolyposis colorectal cancer (Lynch II), Gardner's syndrome, and familial and atypical multiple mole melanoma (FAMMM). In all of these familial cancer syndromes, one of the two genes inherited at birth is mutated, with cancer arising after the loss of the remaining normal gene.

In contrast to inherited or germ-line gene mutations, several genes have been found to be commonly spontaneously mutated in pancreatic adenocarcinoma. Mutations in the oncogene *K-ras* confer constitutive activation of the gene and are found in over 90% of pancreatic cancers. Mutations in *K-ras* have also been detected in hyperplastic foci within the duct of patients with chronic pancreatitis, and may predict which of these patients is at risk to develop pancreatic adenocarcinoma. As might be expected, mutations in tumor suppressor genes, both spontaneously arising and inherited, are common in pancreatic tumors, with over 95% of cases harboring mutations in p16, 50–75% with p53 mutations, and approximately 55% with mutations in Smad4/DPC4 (deleted in pancreatic cancer, locus 4). Other less frequent mutations occur in tumor suppressors *MKK4* (4%), transforming growth factor-β (TGF-β) receptors I or II (<5%), and *RB1* (<5%).

PATHOLOGY

The pancreas has both exocrine and endocrine cell types, and as such, neoplasms can arise from multiple cellular structures. Approximately 90% of tumors of the exocrine pancreas have a ductal phenotype, whereas less than 1% have an acinar cell phenotype, with the balance being of uncertain histogenesis. Overall, 75% of pancreatic neoplasms are ductal adenocarcinoma, 65% of which arise from the head, neck, or uncinate process, 15% from the body or tail, and 20% are diffuse. Hallmarks of pancreatic adenocarcinoma include an intense desmoplastic reaction and perineural invasion (Fig. 1).

Areas of focal ductal proliferation have been noted both adjacent to infiltrating pancreatic cancers and in the setting of chronic pancreatitis. Molecular analysis of these lesions, termed pancreatic intraepithelial neoplasia (PanIN), has shown progressive accumulation of genetic changes with increasing severity of histologic atypia. The genetic alterations in PanIN lesions are similar to those seen in pancreatic adenocarcinomas,

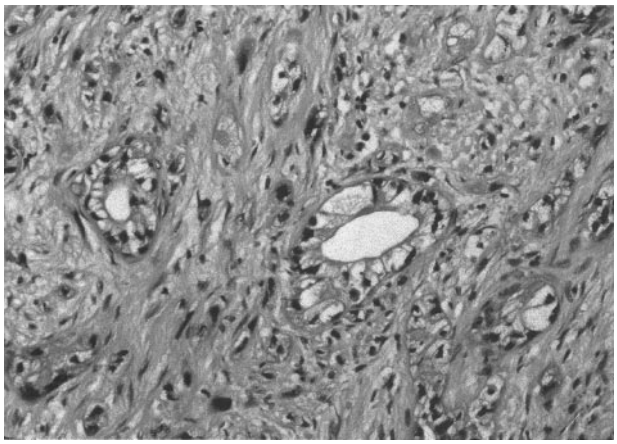


FIGURE 1 Histologic section of an adenocarcinoma of the pancreas, revealing the characteristic appearance of tumor cells embedded in dense regions of fibrosis.

supporting the hypothesis that they are precursor lesions. There appears to be a step-wise accumulation of specific genetic alterations in the continuum from normal tissue to infiltrating carcinoma, with mutations in K-ras and Her-2/neu occurring in low-grade lesions, mutations in p16 present in intermediate-grade lesions, and p53, DPC4, and BRCA2 mutations present in high-grade lesions. Although the progressive accumulation of genetic alterations has been identified, some controversy remains as to whether these lesions originate from fully differentiated duct cells or metaplastic conversion of either islet or acinar to ductal cells.

CLINICAL PRESENTATION

The early symptoms associated with pancreatic adenocarcinoma are nonspecific and therefore patients often delay seeking medical attention until the disease is advanced. A common presenting symptom, present in over 90% of patients with pancreatic cancer, is cachexia, which usually precedes a diagnosis of pancreatic adenocarcinoma by many months. A major contributor to eventual mortality, cachexia is due to weight loss from local obstructive factors causing nausea, vomiting, and anorexia as well as to elaboration of tumor factors such as tumor necrosis factor α (TNFα) and cytokines such as interferon α (IFNα).

Abdominal pain is reported at presentation in 75–90% of patients with pancreatic cancer. The pain is believed to be caused by compression of, or invasion into, perineural and splanchnic neuronal structures as well as contiguous organs and the retroperitoneum. Because of the proximity of the common bile duct and duodenum to the pancreas, tumors located in the head of the organ may grow and compress these structures. Bile duct compression leads to obstructive jaundice in 70–85% of patients with pancreatic head lesions. Compression of the duodenum leads to delayed gastric emptying and early satiety, contributing to nausea and vomiting, which is initially present in 35–45% of patients. Up to 5% of patients present with advanced tumors causing complete duodenal obstruction.

The onset of diabetes mellitus is also associated with pancreatic adenocarcinoma, with 10–15% of patients developing glucose intolerance 6–12 months prior to cancer diagnosis. The onset of diabetes appears to be due to tumor elaboration of a yet undefined factor that stimulates islet cells to secrete the prodiabetic polypeptide amylin. Amylin levels are higher in patients with pancreatic adenocarcinoma compared to patients with diabetes and other gastrointestinal malignancies, and serum amylin levels and insulin resistance abate after tumor resection.

DIAGNOSIS AND STAGING

Although the constellation of symptoms of abdominal pain, weight loss, nausea, vomiting, and obstructive jaundice is highly suggestive of pancreatic malignancy, other processes that can mimic these symptoms must also be considered. Benign biliary strictures and common bile duct stones as well as carcinoma of the bile ducts or gallbladder can cause these symptoms, as can ampullary and duodenal tumors. The possible diagnosis of pancreatic cancer is often proposed during the evaluation of obstructive jaundice by transabdominal ultrasonographic imaging. Because of multiple variables, including skill of the examiner, patient habitus, and overlying loops of gas-filled bowel limiting complete imaging, the sensitivity of transabdominal ultrasound for diagnosing pancreatic cancer has ranged from 44 to 94% in different studies. Even with optimal results, transabdominal ultrasound is not able to stage patients or determine resectablity, thus it must always be accompanied by another imaging modality.

Once a diagnosis of pancreatic cancer is suspected, the modality of choice for confirmation is a helical computed tomography (CT) scan of the abdomen with dual-phase scan acquisition. Dual-phase scans, acquired during both the arterial and the portal phases, allow for pancreatic parenchymal and arterial enhancement during the former phase, and hepatic parenchymal as well as peripancreatic venous enhancement during the later phase (Fig. 2). Enhancement of the pancreatic parenchyma allows for detection of small, hypodense

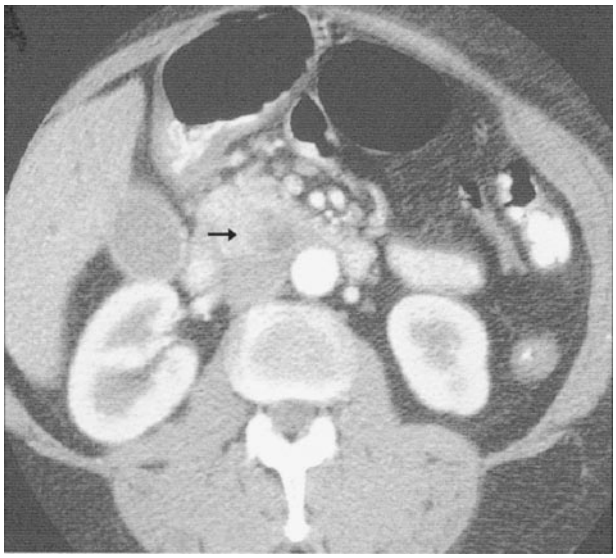


FIGURE 2 Dual-phase computer tomography scan of a patient with a mass in the head of the pancreas (arrow) representing an adenocarcinoma.

carcinomas as well as arterial involvement by the tumor, which directly affects resectability and possible cure. Tumor involvement of venous vessels and liver metastasis are delineated during the portal phase. The sensitivity of helical CT scanning to detect lesions greater than 2 cm in diameter is approximately 89%, decreasing to approximately 71% for lesions less than 2 cm in diameter, with an overall diagnostic accuracy of 97% for pancreatic adenocarcinoma.

Besides providing a diagnosis, CT scanning is useful to determine resectability of the tumor. The criteria for resectability include absence of extrapancreatic disease, the absence of tumor extension to the superior mesenteric artery and celiac axis, and a patent superior mesenteric–portal vein confluence. The accuracy of CT for predicting unresectability approaches 100%, but the accuracy for predicting resectability is only approximately 30% due to the limited ability to detect small metastases to the surface of the liver, peritoneum, lymph nodes, and peripancreatic soft tissue.

Other modalities may be used to better define resectability prior to laparotomy. However, magnetic resonance imaging (MRI) and magnetic resonance cholangiopancreatography (MRCP) offer no diagnostic advantage over CT, having an accuracy of 90–100% in several small series. The only role for magnetic resonance imaging in evaluating for potential pancreatic malignancy is in patients allergic to the iodinated contrast material used for CT imaging, because the contrast agent used, gadolinium, is associated with a lesser incidence of severe allergic reactions. Positron emission tomography (PET) scanning with fluodeoxyglucose F 18 offers promise for delineating inflammatory from neoplastic pancreatic processes, although experience is limited and data are unavailable for assessing overall accuracy.

Endoscopic ultrasonography (EUS) is highly sensitive for diagnosing pancreatic cancer, with an overall sensitivity of approximately 93% when combining data from various studies. EUS is particularly useful for detecting small (< 2 cm) pancreatic tumors that may not be well visualized by CT, as well as in assessment of local vascular involvement. Combining EUS with fine needle aspiration (FNA) of the tumor increases the specificity of EUS alone for diagnosing pancreatic cancer.

ERCP allows for visualization of the biliary and pancreatic ductal structures and may be useful for diagnosing the cause of biliary obstruction when no mass is evident by CT scanning, such as in ampullary tumors or cholangiocarcinoma, and for differentiating focal pancreatitis from neoplasm. Dilation of the distal pancreatic and common bile ducts with proximal stricturing of both structures, the "double duct sign," is highly suggestive of adenocarcinoma affecting the head of the pancreas (Fig. 3). Data summarized from 16 studies suggest that ERCP has a sensitivity of 92% and a specificity of 96% for diagnosing pancreatic cancer. Bile aspiration for cytologic evaluation and tissue sampling by biopsy and ductal brushing can also be performed at the time of ERCP. Biliary decompression can also be achieved at ERCP if a tumor has been deemed unresectable or if a patient with symptomatic biliary obstruction will have a delay prior to resection. Multiple studies have failed to show a benefit from routine preoperative biliary decompression, and some studies have shown increased operative morbidity related to infectious complications.

In addition to radiographic and endoscopic assessment, laparoscopy has been proposed as an accurate means of evaluation for resectability prior to laparotomy. Evaluation begins with a thorough inspection of the abdomen, including the liver and peritoneal surfaces. Any suspicious nodules should be biopsied for histological assessment. Metastatic disease not identified by spiral CT can be detected in as many as 30% of patients undergoing laparoscopic staging of pancreatic cancer. The addition of laparoscopic ultrasonography to evaluate for intrahepatic and lymph node metastasis as well as unrecognized vascular invasion may further alter the management of additional patients.

Because of the limitations of detecting very early pancreatic cancers, development of a biological marker that could detect these early lesions may translate into

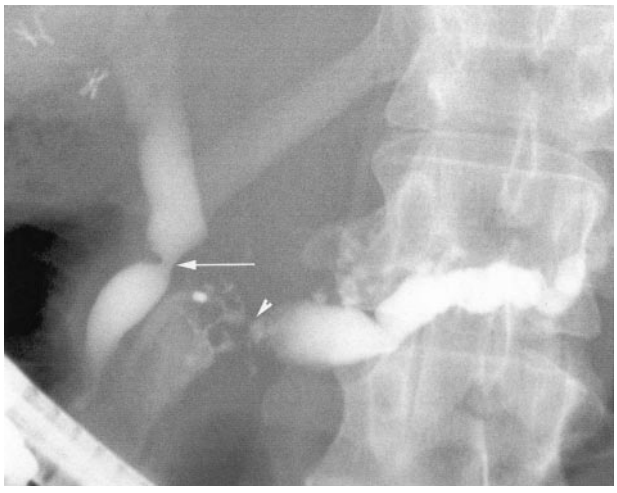


FIGURE 3 The "double duct sign." Endoscopic retrograde cholangiopancreatography of a patient with adenocarcinoma of the head of the pancreas reveals that both the common bile duct (arrow) and the pancreatic duct (arrowhead) are compressed by the tumor.

decreased mortality. Although many tumor markers have been investigated, currently none exists with acceptable specificity either for confirmation of pancreatic adenocarcinoma in equivocal cases or for routine screening.

CA 19-9 is presently the most widely used serum marker for pancreatic cancer. CA 19-9 is a sialylated Lewis^a antigen associated with circulating mucins and is expressed in normal pancreatic, biliary, and gastric epithelial cells. Although it is most frequently elevated in pancreatic adenocarcinoma, it may also be expressed in biliary, gastric, and colonic malignancies as well as in acute and chronic pancreatitis, hepatitis, and biliary obstruction. Marked elevations are found in acute cholangitis and hepatic cirrhosis. The reported sensitivities and specificities of CA 19-9 for diagnosing pancreatic adenocarcinoma are related to the serum cutoff level selected. A cutoff of 15 U/ml produces a sensitivity of 92% and a specificity of 60%, whereas a cutoff of 1000 U/ml yields a sensitivity of 40% and a specificity of 99%. When using the usual cutoff of 37 U/ml, combined studies have shown a sensitivity of 81–85% and a specificity of 81–90%.

Carcineoembryonic antigen (CEA) is commonly used as a serum marker for colon cancer. It has been investigated as a marker for pancreatic adenocarcinoma, but its sensitivity and specificity from combined studies is 58 and 75%, respectively. A higher diagnostic accuracy of 83% is observed when measuring CEA levels in pancreatic juice rather than in serum. The oncogene *K-ras* is mutated in more than 90% of pancreatic adenocarcinomas and can be detected in the pancreatic juice from 55–77% of these patients. Several reports retrospectively examining pancreatic juice specimens from patients subsequently diagnosed with pancreatic adenocarcinoma have suggested that screening for *K-ras* mutations may be effective for the early detection of pancreatic cancer, but the utility of this is limited by the finding that *K-ras* mutations are also present in some cases of chronic pancreatitis. Whether *K-ras* mutations in the setting of chronic pancreatitis predict those who are at increased risk for developing pancreatic adenocarcinoma is currently unknown.

In some cases, even after multiple diagnostic modalities have been applied, it is still impossible to differentiate pancreatic adenocarcinoma from some other inflammatory process, often chronic pancreatitis, prior to laparotomy. In these instances, particularly when a pancreatic mass is identified in the setting of antecedent chronic pancreatitis, a tissue diagnosis is not necessary prior to resection. Although this approach risks overtreating a benign process, the greater concern is undertreating a potentially curable malignancy.

SURGICAL TREATMENT

Only approximately 20% of pancreatic tumors are resectable at the time of diagnosis due to local vascular invasion in 40% of patients and metastatic disease present in 40%. Currently, the only potential cure for pancreatic adenocarcinoma is surgical resection. Because lesions originating in the head and uncinate process of the pancreas obstruct the bile duct and cause jaundice, they tend to present at an earlier stage and have a higher resectability rate than do cancers in the tail and body. The uncommon lesion in the pancreatic body or tail that is resectable may be treated by distal or subtotal pancreatectomy and splenectomy, whereas pancreatic head and uncinate lesions are treated by pancreaticoduodenectomy.

Although survival rates following pancreaticoduodenectomy are approximately 25% at 5 years with negative margins of resection, given the low postoperative mortality of 1–4% and the chance for cure, most patients with resectable disease and no prohibitive comorbidities undergo this operation. The procedure involves division of either the distal stomach, in the classic Whipple operation, or the proximal duodenum 2–3 cm distal to the pylorus in the pylorus-preserving modification, with en bloc resection of the distal common bile duct and involved portion of the pancreas along with the duodenum to the ligament of Treitz. Reconstruction consists of pancreatic, biliary, and either gastro- or duodenoenteric anastomoses, usually in that anatomical order. Typically, the proximal jejunum is used to reestablish pancreatic–enteric continuity, although the stomach is preferred by some surgeons, with roughly equivalent results.

Several important controversies surround the technical aspects of this operation. In Whipple's original description, the stomach was transected proximal to the antrum, whereas in the pylorus-preserving modification, the duodenum is transected 2–3 cm distal to the pylorus. The more extensive gastric resection was performed not only on oncologic grounds, but also to reduce the acid burden and subsequent incidence of marginal ulceration. Pylorus preservation maintains normal gastrointestinal physiology, specifically in terms of acid production, gastric reservoir and emptying functions, and hormone secretion. The rates of early postoperative delayed gastric emptying are similar for the two procedures, and pyloric preservation shortens operative time and is associated with no adverse early or late sequelae.

The extent of peripancreatic dissection is also the source of some controversy. The extended or radical pancreaticoduodenectomy entails en bloc wide

retroperitoneal lymphadenectomy and often resection of the superior mesenteric vein–portal vein confluence along with the tumor. Arguments that favor this approach claim improve resectability and cure rates, yet sufficient data are currently lacking to make a final determination. Although operative mortality rates associated with pancreaticoduodenectomy are typically 1–4%, up to half of the patients undergoing this operation will experience a complication. Approximately 30% of patients undergoing a pancreaticoduodenectomy will experience delayed gastric emptying, though to be related to decreased motilin levels, removal of the duodenal pacemaker and disruption of gastroduodenal neural connections. Erythromycin, a motilin agonist, has been found to improve gastric emptying of both solids and liquids when administered intravenously during the postoperative period.

Pancreatic fistula resulting from failure of healing at the pancreatic–enteric anastomosis, with subsequent intraperitoneal leakage of pancreatic secretions, can usually be managed conservatively if there is no evidence of abdominal sepsis. Isolated fluid collections should be drained, percutaneously if possible, and the patient should have nutrition provided parenterally and remain without oral intake. Intraoperatively, a drain is usually placed in the vicinity of the pancreatic–enteric anastomosis and should be maintained as the fistula is allowed to heal.

Intraabdominal abscesses can result from leakage at the pancreatic–enteric, gastroenteric, or hepaticoenteric anastomoses. Patients with evidence of systemic infection should be evaluated for an intraabdominal abscess and the collection should be drained, preferably percutaneously, with the initiation of appropriate antibiotics. In addition, hepaticoenteric anastomotic leaks may require a transhepatic catheter to allow for external biliary drainage.

PALLIATION OF UNRESECTABLE PANCREATIC ADENOCARCINOMA

At the time of diagnosis, 80% of patients are not candidates for potentially curative resection, narrowing their options to medical palliation for the primary disease and treatment of specific complications of pancreatic cancer. The majority of patients diagnosed with pancreatic adenocarcinoma will experience one or more of its complications, including biliary obstruction, gastric outlet obstruction, and severe abdominal pain. As well, with improvements in determining preoperative resectability of pancreatic cancer, fewer patients undergo exploratory laparotomy during which palliative procedures are performed. Palliation of the complications of pancreatic cancer may be achieved nonsurgically in most cases.

Up to 70% of patients with pancreatic cancer will develop obstructive jaundice and accompanying pruritis with an increased risk of cholangitis. In patients deemed unresectable intraoperatively, a biliary–enteric bypass may be performed for decompression. The preferred procedures involve hepatico- or choledocojejunostomy; cholecystoenterostomy is associated with a rate of recurrent jaundice of 20%. In patients deemed unresectable who do not undergo an operation, endoscopic or transhepatic radiographic placement of a biliary stent may be accomplished. Radiographically placed transhepatic catheters with exclusive external biliary drainage result in large fluid and electrolyte losses and are less desirable for palliation of obstructive jaundice in patients with pancreatic cancer. This procedure is reserved for patients who fail internal endoscopic drainage. Biliary decompression prior to planned resection should be limited to cases with severe symptoms of obstructive jaundice in which surgery is delayed.

Gastric outlet obstruction from duodenal compression will affect up to 25% of patients with pancreatic cancer and may require surgical gastric bypass for palliation. Controversy exists as to whether patients deemed unresectable at exploratory laparotomy should undergo a prophylactic gastroenteric bypass. Proponents cite the higher mortality rate, approaching 25%, for patients requiring a second operation for palliation of gastric outlet obstruction following exploratory laparotomy, as well as no increase in mortality when gastric bypass is performed at the initial operation. Increased morbidity associated with gastric bypass, most notably delayed gastric emptying, as well as increased hospital stay and overestimation of the need for gastric bypass at the time of exploration is cited by its opponents. A confounding factor in the discussion of pancreatic cancer-related gastric outlet obstruction is whether common symptoms of nausea and vomiting represent gastroparesis or true mechanical obstruction. Clearly, any patient with radiographically confirmed gastric outlet obstruction associated with pancreatic adenocarcinoma who is fit for surgery should undergo gastroenterostomy, using either an open or laparoscopic technique. Limited data are available regarding the use of endoscopically deployed duodenal stents in this setting.

Severe, debilitating abdominal pain and back pain are a frequent complication of pancreatic

adenocarcinoma, often requiring significant analgesia for adequate palliation. Chemical splanchnicectomy with 50% ethanol, performed either at the time of exploration or subsequently through the percutaneous route, has been shown to palliate pain associated with pancreatic adenocarcinoma, although its duration is limited. When performed intraoperatively, this procedure has not been accompanied by increases in morbidity, mortality, return to oral intake, or length of hospital stay.

Thoracoscopic splanchnectomy, using video-assisted thoracoscopy, transects the pain fibers in the posterior mediastinum as they course cephalad through the sympathetic chain. The short-term efficacy is clear, with 99% of patients reporting excellent analgesia and 50% reporting a sustained effect at 4 months. This procedure does involve a general anesthetic, a hospital stay, and an associated risk of complications that occur more often and are more severe compared to chemical splanchnicectomy. The role of palliative radiotherapy for pain control has been used, but has several limitations, including extended length of time for pain control to be achieved and need for repeated hospital visits.

For the 80% of patients deemed unresectable, palliative chemoradiation has been shown to increase survival and possibly to reduce severity of pain compared to untreated patients. In 1981, the Gastrointestinal Tumor Study Group (GITSG) published results of a trial in which patients surgically staged to confirm unresectability and no evidence of peritoneal or liver metastases were randomized to receive either external beam radiation alone or external beam radiation with 5-fluorouracil (5-FU). Patients treated with chemoradiation fared better in terms of median survival (49 weeks) than did those receiving radiation alone (22 weeks). Other studies have confirmed the benefit of multimodality therapy over either radiotherapy or chemotherapy alone. In a subsequent GITSG trial, patients received either combined streptozotocin, mitomycin C, and 5-FU (SMF) alone or radiotherapy with 5-FU followed by SMF. The median survival in patients receiving chemoradiation was 42 weeks, compared with 32 weeks in those assigned to the chemotherapy alone group, and 1-year survival was 41% in the chemoradiation group and 19% in the SMF alone group.

A recent advance in the chemotherapeutic regimen to treat pancreatic adenocarcinoma is the use of gemcitabine, a potent radiosensitizer. Gemcitabine has been shown to improve survival when compared to 5-FU, with reported 1- year survival rates of 18 and 2%, respectively. As well, gemcitabine appears to confer clinical benefit in terms of decreased pain intensity and analgesic consumption with improvement in overall functional status.

NEOADJUVANT THERAPY

The observation that 25–30% of patients do not receive adjuvant therapy following pancreatic resection due to prolonged recovery, perioperative complications, or patient refusal has led to the investigation of preoperative chemoradiation. Besides increasing the number of patients able to complete multimodality therapy, neoadjuvant therapy has been studied for its ability to convert locally unresectable pancreatic cancer to resectable disease. Other advantages of this modality include more effective radiation therapy to well-oxygenated cells not devascularized by surgery, reducing tumor dissemination during surgical manipulation, and targeting retroperitoneal margins of excision that may not be adequately treated surgically. As well, preoperative chemoradiation would allow an interval window for restaging prior to surgical resection so that previously occult metastatic disease could be detected, thus saving patients an unnecessary laparotomy. The utility of this approach is currently under intense investigation.

ADJUVANT THERAPY FOR RESECTABLE PANCREATIC ADENOCARCINOMA

The basis for adjuvant therapy following surgical resection with negative margins comes from a GITSG study that showed a median survival nearly twice as long for patients treated postoperatively with radiation and concurrent 5-FU compared to those randomized to observation alone (20 vs. 11 months, respectively). Single-institution studies investigating the benefit of postoperative chemoradiation compared to observation have also shown a survival benefit. Thus, the current data support the use of postoperative adjuvant chemoradiation in patients who undergo surgical resection.

FUTURE STRATEGIES

Although many recent advances have been made in the treatment of pancreatic adenocarcinoma, including lower postoperative mortality rates, neoadjuvant therapy, and addition of gemcitabine to the multimodality armamentarium, none is expected to dramatically alter the high mortality associated with this disease. The single advance that offers the most promise is elucidation of the molecular events involved in pancreatic oncogenesis, metastasis, and resistance to radiation and chemotherapy.

Understanding pancreatic adenocarcinoma on a molecular level allows for the development of targeted therapies that may be employed either alone or in combination with traditional chemoradiation. The finding that the *K-ras* oncogene is mutated and activated in over 90% of pancreatic adenocarcinomas, and that this is an early event in the accumulation of genetic damage, has prompted investigation into strategies that would block its activation. SCH 66336, a farnesyl transferase inhibitor that renders *K-ras* functionally inactive, as well as *K-ras* antisense therapy, are currently being investigated.

The p53 tumor suppressor is often mutated and functionally inactive in pancreatic adenocarcinoma, causing inhibition of a major apoptotic pathway induced by most chemoradiation therapies. Strategies to correct functional losses are more challenging than are those that inhibit gene expression or function, but one promising method is to use genetically modified adenoviruses that replicate only in p53-deficient cells. An additional challenge for pancreatic cancer is delivering the virus to the tumor target, although intratumoral injection using endoscopic ultrasound is an area of investigation. Other molecular therapeutic targets are angiogenic factors such as vascular endothelial growth factor (VEGF), which is up-regulated in pancreatic adenocarcinoma, growth stimulatory factor receptors such as the epidermal growth factor receptor and human growth factor receptor 2 (HER2), as well as matrix metalloproteinases that degrade the extracellular matrix, allowing cancer cells to metastasize. Strategies to antagonize these factors and their function have been developed and are currently being tested in clinical trials. Although the prognosis of pancreatic adenocarcinoma remains abysmal, the outlook for novel molecular strategies that would significantly reduce the mortality and suffering associated with this disease is very promising.

See Also the Following Articles

Cancer, Overview • Computed Tomography (CT) • Pancreatic Ductal Adenocarcinoma • Pancreatic Tumors, Other • Pancreatitis, Hereditary • Smoking, Implications of

Further Reading

Barkin, J., and Goldstein, J. (1999). Diagnostic approach to pancreatic cancer. *Gastroenterol. Clin. North Am.* **28**, 709–722.

Evans, D. (2001). Preoperative chemoradiation for resectable and locally advanced adenocarcinoma of the pancreas. *J. Gastrointest. Surg.* **5**, 2–5.

Freeny, P. (1999). Pancreatic imaging: New modalities. *Gastroenterol. Clin. North Am.* **28**, 723–746.

Jemal, A., Thomas, A., Murray, T., and Thun, M. (2002). Cancer statistics, 2002. *CA Cancer J. Clin.* **52**, 23–47.

Kern, S. (2000). Molecular genetic alterations in ductal pancreatic adenocarcinomas. *Med. Clin. North Am.* **84**, 691–695.

Kern, S., Hruban, R., Hollingsworth, M., Brand, R., Adrian, T., Jaffee, E., and Tempero, M. (2001). A white paper: The product of a pancreas cancer think tank. *Cancer Res.* **61**, 4923–4932.

McGinn, C., and Zalupski, M. (2001). Combined-modality therapy in pancreatic cancer: Current status and future directions. *Cancer J.* **7**, 338–348.

Sener, S., Fremgen, A., Menck, H., and Winchester, D. (1999). Pancreatic cancer: A report of treatment and survival trends for 100,313 patients diagnosed from 1985–1995, using the National Cancer Database. *J. Am. Coll. Surg.* **189**, 1–7.

Sohn, T., Lillemoe, K., Cameron, J., Huang, J., Pitt, H., and Yeo, C. (1999). Surgical palliation of unresectable periampullary adenocarcinoma in the 1990s. *J. Am. Coll. Surg.* **188**, 658–666.

Wolff, R. (2001). Novel therapies for pancreatic cancer. *Cancer J.* **7**, 349–358.

Yeo, C. (1999). The Whipple procedure in the 1990s. *Adv. Surg.* **32**, 271–303.

Pancreatic Digestive Enzymes

JOHN A. WILLIAMS
University of Michigan, Ann Arbor

translational control The regulation of protein synthesis at the level where the mRNA sequence is translated into the amino acid sequence on the ribosome.
zymogen granules Pancreatic acinar cell secretory granules.
zymogens Inactive precursors of certain pancreatic digestive enzymes that are activated by proteolytic cleavage.

The pancreatic digestive enzymes are a mixture of approximately 20 enzymes synthesized by pancreatic acinar cells and secreted into the pancreatic ducts. These enzymes are responsible for almost all of the activity during the lumenal phase of digestion in the intestine.

INTRODUCTION

Pancreatic digestive enzymes can be divided into four groups according to the major class of macronutrients upon which they act (Table I). There is only one major form of amylase, DNase, and RNase, possibly due to the relatively limited number of distinct chemical bonds to cleave. There is a much more diverse array of proteases, due to the increased complexity of peptide bonds as a result of the larger number of amino acids. Multiple closely related forms, termed isoenzymes, are present for many proteases, such as trypsinogens 1, 2, and 3. The biological significance of these multiple gene products is not clear. All of the pancreatic proteases, as well as phospholipase A2 and colipase, are synthesized as inactive precursors known as zymogens and designated by either the prefix "pro" or the suffix "ogen." These have little or no enzymatic activity in the pancreatic juice but are activated in the intestinal lumen by a cascade in which the intestinal enzyme enterokinase cleaves and activates trypsinogen and the released trypsin rapidly activates the other enzymes. This article addresses the general features of pancreatic digestive enzymes.

TABLE I Human Pancreatic Digestive Enzymes

Enzyme	Molecular weight (daltons)
Proteases	
Trypsinogen 1	25,000
Trypsinogen 2	25,000
Trypsinogen 3	23,400
Chymotrypsinogen	24,000
Proelastase 1	33,000
Proelastase 2	26,600
Kallikreinogen	35,000
Procarboxypeptidase A1	44,500
Procarboxypeptidase A2	47,000
Procarboxypeptidase B1	47,300
Procarboxypeptidase B2	47,300
Glycosidase	
Amylase	57,000
Lipases	
Triglyceride lipase	48,000
Colipase	10,000
Carboxyl ester hydrolase	100,000
Phospholipase A2	14,000
Nucleases	
DNase I	30,000
RNase	15,000

SYNTHESIS AND PACKAGING OF DIGESTIVE ENZYMES

All of the digestive enzymes contain a sequence in the nascent peptide that induces translation on ribosomes attached to the endoplasmic reticulum (ER). The nascent peptide is inserted through a protein channel into the lumen of the ER. There the proteins are processed and folded with the help of intralumenal chaperone proteins. The proenzymes move to the Golgi by vesicular transport and in the *trans*-Golgi network are sorted into newly forming secretory granules, termed condensing vacuoles, which are distinguishable in electron microscopic images by their lower density. With the possible help of granule membrane proteins, such as GP-2, the contents become more densely packed and lose some of their water. Mature zymogen granules then move from the Golgi to the apical region of the cell in a microtubule-dependent manner, where they await the signal for exocytosis, the process by which their contents are released into the acinar lumen. Although the concept was somewhat controversial in the past, all the different digestive enzymes synthesized in each cell are believed to be packaged in the same granules. When granules differ in enzyme content, it is believed to be due

 Copyright 2004, Elsevier (USA). All rights reserved.

to different levels of mRNA being expressed in different cells. For example, the acinar cells surrounding the islets of Langerhans have larger amounts of certain enzymes such as amylase, possibly due to locally high concentrations of insulin. The stimulation of secretion from such different regions of the gland can result in the phenomenon of "nonparallel secretion," in which the composition of secreted digestive enzymes changes in response to a physiological stimulus.

REGULATION OF DIGESTIVE ENZYME SYNTHESIS BY DIET AND HORMONES

The synthesis of digestive enzymes is regulated at several levels. Dietary adaptation occurs when the pancreas modifies its composition of different digestive enzymes in response to changes in the level of dietary protein, fat, and carbohydrate. In general, enzyme levels increase in response to an abundance of their specific dietary substrate. Thus, a high-starch diet induces the presence of more amylase and a high-protein diet induces the presence of more proteases. This effect takes 2–10 days to occur and is accompanied by changes in mRNA levels for the individual digestive enzymes. Some of these effects are well established as being hormonally mediated. Thus, insulin, which increases in response to a high-carbohydrate diet, increases amylase mRNA levels through enhanced gene transcription. Cholecystokinin (CCK) has similarly been related to a high-protein diet and secretin to a high-fat diet. These changes have been studied primarily in rodents and their relevance to the human pancreas is unknown.

The second type of regulation is the increased rate of protein synthesis due to enhanced mRNA translation that occurs following a meal. In this case, because mRNA levels are relatively unchanged, synthesis of all the digestive enzymes increases in parallel and the main function is to increase the amount of new enzyme in the cell to be ready for the next meal. This effect is mediated by the pancreatic secretagogues, CCK and acetylcholine (the latter is released from the vagal nerve), insulin, and dietary amino acids, particularly branched chain amino acids such as leucine. These all activate a biochemical pathway that involves protein kinase B (Akt) and mTOR (mammalian target of rapamycin) and leads to the activation, by phosphorylation or dephosphorylation, of specific translation factors and ribosomal proteins that are involved in the initiation and elongation stages of translation.

See Also the Following Articles

Amylase • Digestion, Overview • Pancreatic Enzyme Secretion (Physiology) • Pancreatic Triglyceride Lipase • Trypsin

Further Reading

Adelson, J. W., and Miller, P. E. (1989). Heterogeneity of the exocrine pancreas. *Am. J. Physiol.* **256**, G817–G825.

Lowe, M. E. (1993). The structure and function of pancreatic enzymes. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1531–1542. Raven Press, New York.

Scheele, G. A. (1993). Extracellular and intracellular messengers in diet-induced regulation of pancreatic gene expression. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1543–1554. Raven Press, New York.

Williams, J. A. (2001). Intracellular signaling mechanisms activated by cholecystokinin regulating synthesis and secretion of digestive enzymes in pancreatic acinar cells. *Annu. Rev. Physiol.* **63**, 77–97.

Pancreatic Disease, Pediatric

MARK E. LOWE
Washington University, St. Louis and St. Louis Children's Hospital

acute pancreatitis An acute condition due to inflammatory disease of the pancreas that typically presents with abdominal pain and elevated pancreatic enzymes in the blood.

chronic pancreatitis A persistent inflammatory disease of the pancreas characterized by irreversible morphological change that often causes pain or loss of exocrine function or both.

exocrine pancreas The portion of the pancreas that synthesizes and secretes the components of pancreatic juice. The juice contains digestive enzymes, water, and bicarbonate. The pancreatic acinar and ductal cells constitute the cellular components of the exocrine pancreas.

pancreatic insufficiency Decreased secretion of digestive enzymes or insulin to the extent that malabsorption or diabetes mellitus appears. Malabsorption typically manifests with steatorrhea, the presence of more than 7% of dietary fat in the stool.

Disorders of the exocrine pancreas represent a small segment of pediatric disease, but are common enough that all physicians caring for children should have a working knowledge of the pancreatic diseases that affect children. Cystic fibrosis is the most common ailment of the exocrine pancreas in children, but there are many other causes of pancreatic dysfunction and inflammation in pediatrics that present difficult diagnostic challenges. This article focuses on the disorders of the exocrine pancreas that affect children except for the pancreatic dysfunction associated with cystic fibrosis.

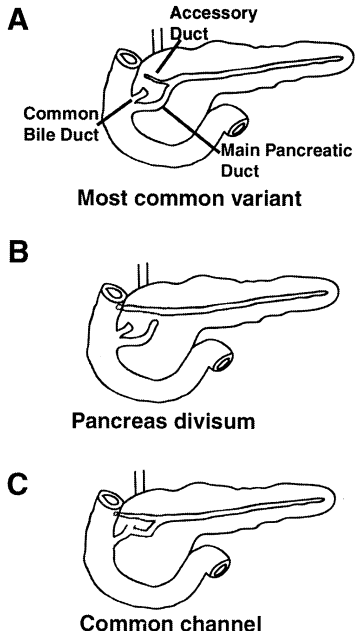


FIGURE 1 Three variations in ductal anatomy. (A) In 40–50% of people, the main pancreatic duct enters the duodenum with the common bile duct and the accessory duct regresses. (B) Five to 10% of people have pancreas divisum. The main and accessory pancreatic ducts do not connect and each duct drains into the duodenum through its own papilla. (C) The main pancreatic duct enters the common bile duct approximately 5 to 15 mm before the ampulla of Vater in approximately 5 to 10% of individuals.

CONGENITAL ANOMALIES

Pancreas Divisum

Incomplete fusion of the dorsal and ventral pancreas may result in the substitution of the dorsal pancreatic duct, which empties through the relatively small accessory papilla, for the ventral duct as the main conduit for pancreatic fluids (Fig. 1). This anatomical variant, pancreas divisum, is the most common congenital anomaly of the pancreas and the most controversial. Some believe that an anatomic or functional stenosis at the accessory papilla produces increased ductal pressure, which, in turn, causes bouts of acute and recurrent pancreatitis. Others point out that there is no causal relationship between pancreas divisum and pancreatitis and that there is disagreement among clinical studies about the incidence of pancreatitis in patients with pancreas divisum. Some studies have found an increased incidence, whereas other studies have failed to confirm those findings.

The diagnosis of pancreas divisum is best made by endoscopic retrograde cholangiopancreatography (ERCP). The role of magnetic resonance cholangiopancreatography (MRCP) has not been established in pediatric patients. Once pancreas divisum is diagnosed,

 Copyright 2004, Elsevier (USA). All rights reserved.

the decision to treat, and how to treat, is not straightforward. If the patient has had recurrent episodes of pancreatitis, most gastroenterologists opt for ERCP therapies, such as sphincterotomy with or without stenting, particularly if objective evidence of papillary stenosis exists. Rarely, a surgical drainage procedure is offered.

Common Channel Syndrome

Abnormal junctions of the common bile duct and the main pancreatic duct represent another frequently encountered ductal anomaly. In these cases, the common bile duct joins the main pancreatic duct 5 to 15 mm before the ampulla of Vater to create a common channel, which has been associated with pancreatitis, choledochal cysts, and, in adults, bile duct or gallbladder cancer. Because the juncture occurs before the sphincter of Oddi, bile can reflux into the main pancreatic duct and cause pancreatitis. Similarly, pancreatic juice can reflux into the common bile duct, a situation that may cause some choledochal cysts. Animal models support both relationships. The presentation of these patients relates to the symptoms of pancreatitis or of choledochal cysts. The latter presents with right upper quadrant pain, jaundice, and, sometimes, a palpable mass below the liver. The diagnosis can be made by ERCP, MRCP, or intraoperative cholangiogram. Treatment for choledochal cysts is excision of the cysts with a Roux-en-Y hepatojejunostomy. Therapy for an isolated common channel has traditionally been surgical, but, recently, ERCP treatments have been advocated although long-term studies of outcome after ERCP therapy are lacking.

Annular Pancreas

Annular pancreas is an uncommon developmental anomaly that is generally associated with other congenital malformations and has a particularly high incidence in patients with trisomy 21. The exact pathogenesis of annular pancreas remains unclear although most theories invoke a failure in one of the steps leading to fusion of the ventral and dorsal pancreatic buds. Symptoms of partial or complete upper gastrointestinal obstruction present at any age, but newborns represent the largest group. Radiographic studies, such as upper gastrointestinal series, ultrasound, and ERCP, can suggest the diagnosis of annular pancreas, but definitive diagnosis occurs at surgery. During surgery, the obstruction is bypassed by a duodenoduodenostomy or duodenojejunostomy.

Pancreatic Agenesis and Aplasia

Disorders of pancreatic development are rare and range from complete to partial agenesis of the pancreas. Only two cases of pancreatic agenesis and five cases of partial agenesis were described in an autopsy series of 2000 fetuses. These disorders almost certainly arise from mutations in the genetic or signaling pathways responsible for the early development of the fetal pancreas. In one patient with pancreatic agenesis, a homozygous mutation in the PDX1 gene was identified. PDX1 is a transcription factor that is required for pancreatic development in rodents and almost certainly contributed to the lack of pancreatic development in this patient.

Complete absence of the pancreas is incompatible with life and is usually diagnosed at autopsy. Newborns with this disorder have significant intrauterine growth retardation and insulin-dependent hyperglycemia. In contrast, most cases of partial pancreatic agenesis remain asymptomatic and may be diagnosed at surgery or autopsy. Only those with less than approximately 5% of normal tissue will manifest evidence of exocrine dysfunction with poor growth and fat malabsorption or evidence of endocrine dysfunction with hyperglycemia and low glucagon levels and response. With treatment of the endocrine insufficiency with pancreatic enzymes and fat-soluble diets and insulin if required, these patients may do well.

HEREDITARY DISORDERS

Shwachman–Diamond Syndrome

Shwachman–Diamond syndrome (SDS) is the second most commonly recognized cause of pancreatic insufficiency in children. Estimates of incidence for this autosomal recessive disease range from 1 in 10,000 to 1 in 50,000 live births. Although the constellation of pancreatic insufficiency and hematological abnormalities in the face of normal sweat electrolytes typifies this disorder, many other organ systems may be affected (Table I). Among these, the prominent features include short stature, skeletal abnormalities, and learning or behavioral abnormalities.

The pancreas varies in size from normal to small and has a prominent fatty infiltration, which replaces acinar tissue and can be appreciated in computed tomography of the abdomen. The ducts appear normal and pancreatic secretions have relatively normal volume with normal levels of bicarbonate and chloride. The primary defect is in the acinar cells and may involve a defective transcription factor or growth factor, which would also account for the wide range of organ involvement in SDS.

TABLE I Features of Shwachman–Diamond Syndrome

Pancreatic exocrine insufficiency
Short stature
Hematological changes
Neutropenia
Thrombocytopenia
Anemia
Myelolymphoproliferative disorders
Skeletal changes
Metaphyseal chondrodysplasia
Short or flared ribs
Clinodactyly
Delayed bone age
Recurrent infections
Liver changes
Elevated serum aminotransferases
Periportal fibrosis
Miscellaneous
Psychomotor retardation
Icthyosis
Dental abnormalities

Linkage studies suggest that the defective gene or genes locate in the peri-centrometric region of chromosome 7. At present, a specific gene has not been identified.

Most patients present in infancy with symptoms of pancreatic insufficiency. Interestingly, pancreatic exocrine function improves with age and as many as 50% of patients have normal fecal fat excretion by 4 years of age. Even so, these patients still have decreased secretion of pancreatic enzymes in pancreatic stimulation tests. This is also true of the patients who do not have steatorrhea. Analysis of pancreatic secretions in this group still reveals deficient secretion of pancreatic enzymes but at levels that permit efficient digestion of dietary fats. Thus, normal fecal fat excretion does not exclude the diagnosis of SDS.

Likewise, the bone marrow dysfunction, generally neutropenia, anemia, or thrombocytopenia, is not always found initially. Both the neutropenia and thrombocytopenia can be intermittent and multiple blood counts should be performed to aid diagnosis. The anemia is less common than the other two abnormalities and is often mild. Bone marrow biopsies in patients with SDS vary considerably and may be normal or show decreased cellularity, fat infiltration, and myeloid maturation arrest.

Therapy is symptomatic and includes optimal pancreatic enzyme replacement, fat-soluble vitamin supplementation, and aggressive treatment of febrile episodes, particularly during periods of neutropenia. Even with optimal treatment, SDS patients may not reach normal adult height. Growth hormone therapy does not appear to alter this outcome. Complications of bone marrow failure produce the most morbidity and mortality in SDS. Infections may be life-threatening in patients with neutropenia, and myelodysplasia, aplastic anemia, and acute myeloid leukemia may be significantly increased in SDS patients.

Johansson Blizzard Syndrome

This syndrome also causes pancreatic lipomatosis, but it is considerably less common than SDS, with fewer than three dozen cases described. Like SDS, the exocrine pancreatic insufficiency in this syndrome most likely results from a defect in acinar cell development, whereas duct function remains intact. In addition to exocrine pancreatic insufficiency, the syndrome includes imperforate anus, hypoplastic alae nasi, hypothyroidism, ectodermal scalp defects, deafness, and mental retardation. Other endocrine abnormalities and genitourinary defects have also been described. Unlike SDS, this syndrome does not include hematological or skeletal abnormalities.

Pearson's Marrow–Pancreas Syndrome

Described in 1979, this syndrome consists of refractory sideroblastic anemia with vacuolization of erythroid and myeloid precursors. The described patients all required repeated blood transfusions and all had failure to thrive. With increased age, many organs, such as the liver, kidney, intestine, skin, and nervous system, show abnormalities. The patients who survived infancy underwent pancreatic stimulation testing, which showed deficient secretion of digestive enzymes, decreased volume, and low bicarbonate concentrations, indicating defects in both acinar and duct cell function in distinction from SDS. At autopsy, fibrosis and acinar cell atrophy were found. Fatty infiltration was not present. Mutations, often deletions, in mitochondrial DNA have been consistently found in patients with this syndrome. The connection between mutations in mitochondrial DNA and pancreatic dysfunction has not been established, but may relate to impaired energy production in the pancreas, which must meet the high energy requirements required to produce large amounts of proteins.

Jeune Syndrome

Pancreatic insufficiency, abnormalities of the thorax, short-limbed dwarfism, cystic dysplasia of the kidneys, and hepatic portal lesions characterize this rare syndrome. The skeletal abnormalities in the thorax may

produce respiratory distress in neonates. Fibrosis and cystic lesions are found in the pancreas.

Isolated Enzyme Deficiencies

A small number of papers report patients with deficiencies of individual pancreatic digestive enzymes and of enterokinase, an intestinal brush-border enzyme that is required for activation of pancreatic zymogens. The patients with decreased levels of triglyceride lipase or of colipase, an essential cofactor for triglyceride lipase, all had evidence of fat malabsorption that was improved by pancreatic enzyme supplementation. These patients presented with oily bowel movements, but otherwise had no other symptoms and were well grown. In contrast, the two reported patients with decreased trypsin activity presented with severe growth failure, hypoproteinemia, and edema. Although the authors speculated that the patients had trypsin deficiency, the report lacked any supporting evidence for a defect in trypsin. A later report of patients with identical clinical symptoms and similarly decreased protease levels raised additional questions about the existence of isolated trypsin deficiency. In these patients, enterokinase was deficient in small bowel biopsies, endogenous enterokinase corrected the protease deficiencies, and the symptoms resolved with pancreatic enzyme supplementation. No gene defects have been described in patients with any of these purported deficiencies and until that time the existence of these isolated enzyme deficiencies is uncertain.

Hereditary Pancreatitis

First described in 1952, hereditary pancreatitis is a rare disorder causing recurrent episodes of acute pancreatitis and, in approximately 75% of patients, chronic pancreatitis. Multiple pedigrees have been well described and it is clear that hereditary pancreatitis has an autosomal dominant inheritance pattern with approximately 80% penetrance and a variable clinical course even among family members.

Recently, mutations in the cationic trypsinogen gene were linked to the development of hereditary pancreatitis in most kindreds. Two mutations, R122H and N29I, are the most prevalent and have been described in families from North America, Europe, and Japan. A few patients have other substitutions at these same amino acid positions. Additional mutations in the same gene, A16V, D22G, and K23R, have been described in a small number of patients, but these are weakly associated with pancreatitis.

The mechanism behind the increased incidence of pancreatitis in patients with mutations in the cationic trypsinogen gene remains speculative. After the description of the R122H mutation, it was speculated that the mutation interfered with the normal protective mechanisms of the acinar cells against premature activation of trypsinogen (Fig. 2). Intracellular activation of trypsinogen would permit the activation of other zymogens,

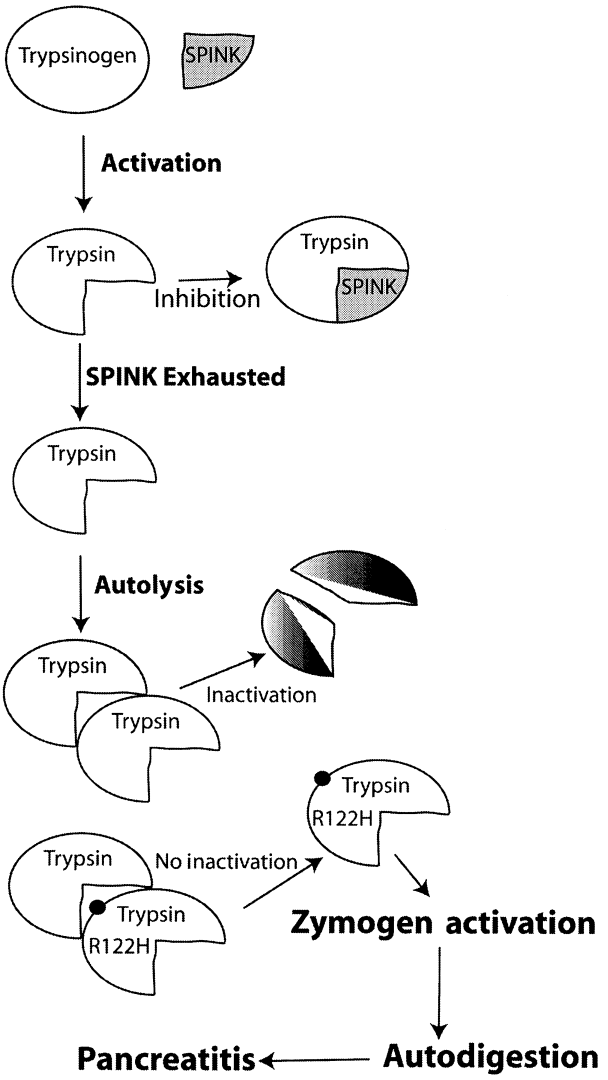


FIGURE 2 Mechanisms that protect the pancreas from premature trypsinogen activation. Acinar cells synthesize and package trypsinogen and SPINK1 or pancreatic secretory trypsin inhibitor at a 5 to 1 ratio. When trypsinogen is activated to trypsin, the first line of defense is inhibition by SPINK1. If there is enough trypsinogen activation to overwhelm the capacity of SPINK1, then trypsin autolysis provides a second line of defense. Autolysis begins with cleavage of the bond after arginine 122. In hereditary pancreatitis, the R122H mutation slows autolysis and allows trypsin to activate other zymogens. Once the zymogen activation cascade begins, autodigestion of the pancreatitis can occur and produce pancreatitis.

resulting in damage to the acinar cells. To protect against premature trypsinogen activation, the acinar cell contains an inhibitor of trypsin. This first line of defense is limited and can be overwhelmed by excessive trypsin activation. In that case, the acinar cell has a second line of defense, trypsin autolysis. Degradation begins with hydrolysis at Arg-122. In hereditary pancreatitis, the histidine substitution at position 122 prevents autolysis. *In vitro* studies on human cationic trypsinogen confirm the increased autolytic resistance of the R117H mutant and also demonstrate that the mutation interferes with multiple protective mechanisms. For instance, the mutation increases trypsinogen autoactivation. Similar studies on the R29I human cationic trypsinogen reveal that the mutation results in faster autoactivation and increased trypsin stability. Thus, both mutants can increase the trypsin levels in the acinar cells and cause increased activation of other zymogens.

Episodes of pancreatitis begin before the age of 20 in approximately 80% of patients. Pain is generally the presenting symptom and may be accompanied by nausea and vomiting. Although some authors suggest that fasting, fatty meals, alcohol, and stress may precipitate pancreatitis, the triggers for individual episodes remain ambiguous, despite the considerable progress made in understanding the pathophysiology of hereditary pancreatitis in the past 5 years. The attacks occur with variable frequency and may become less frequent and less severe with age, but this progression is not universal.

With time, the pancreas atrophies and calcifications develop in approximately 50% of patients. Pancreatic insufficiency develops in a sizable proportion of patients, perhaps as many as 75%. Diabetes mellitus is also relatively common. In some, glucose intolerance may be more troublesome during episodes of pancreatitis. As in other causes of pancreatitis, pseudocysts and hemorrhage can complicate the clinical course. Importantly, the risk of pancreatic cancer is increased above that for the general population and the cumulative risk may be 40% by age 70. The incidences of these complications have not been clearly established, in part because, until recently, identification of patients was limited to clinical diagnoses.

Treatment of acute episodes is supportive, as discussed later. Little can be done to prevent episodes. If triggers for acute pancreatitis are identified, they should be avoided. In most patients, the triggers are elusive. Even so, patients should be advised to avoid alcohol and medications known to cause pancreatitis. Antioxidant cocktails have been suggested as a way to prevent attacks, but no study has demonstrated clear benefit. Regular monitoring for complications is an important facet of patient management. Patients who develop pancreatic insufficiency should receive pancreatic enzyme and fat-soluble vitamin supplementation. Some may require insulin. Chronic pain develops in many patients and presents difficult management decisions. Narcotic analgesia may be required for long periods of time. Treating structural complications may relieve pain in some patients. Pseudocysts can be observed over time for spontaneous resorption or they may require drainage (Fig. 3). Currently, endoscopic internal drainage and external drainage by interventional radiologists offer viable options to surgery. The chronicity, size, location, and complexity of the pseudocyst all contribute to decisions about the best treatment. Generally, pancreatic duct strictures and pancreatic stones can be treated by balloon dilation and stent placement during ERCP. In some patients, surgical drainage procedures, such as a modified Puestow (longitudinal pancreaticojejunostomy), are performed in an attempt to relieve chronic pain. The success of these procedures is difficult to assess because available reports group hereditary pancreatitis with other forms of chronic pancreatitis.

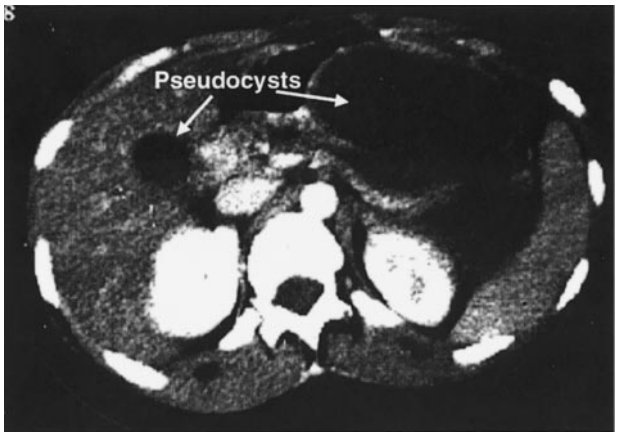


FIGURE 3 Pancreatic pseudocysts. Fluid collections are common complications of acute and chronic pancreatitis. Larger collections may persist and develop a pseudomembrane. These collections can become infected, obstruct other organs, and be a source of hemorrhage.

ACUTE PANCREATITIS

Pancreatitis in children is not common, although the true incidence has never been reliably determined. Earlier published series reported 2 to 9 cases per year in each institution. A more recent series from a large children's hospital shows a steady increase in incidence from 5 cases per year to 113 cases per year 5 years later. Whether this change represents an actual increase in incidence or in diagnosis is uncertain.

TABLE II Etiology of Pancreatitis in Childhood

Trauma
Multisystem disease
Shock
Sepsis
Rheumatological disease
Inflammatory bowel disease
Infection
Pancreaticobiliary abnormalities
Gallstones
Pancreas divisum
Ductal strictures
Common channel
Choledochal cyst
Drugs
Azathioprine
Valproic acid
L-Asparaginase
Sulfonamides
Thiazides
Metabolic or genetic
Hyperlipoproteinemia I, IV, or V
Cystic fibrosis
Hereditary pancreatitis
Hypercalcemia
Protein-calorie malnutrition
Miscellaneous
Juvenile tropical pancreatitis
Perforated duodenal ulcer
Spinal surgery
Idiopathic fibrosing pancreatitis
Enteric duplication cysts
Idiopathic

There are multiple causes of acute pancreatitis in children. Trauma has always been mentioned as a leading cause of pancreatitis and accounts for approximately 20% of acute pancreatitis in childhood. The largest proportion of cases, approximately 50%, occur in patients with associated systemic illness. Approximately 15–20% of patients have idiopathic pancreatitis, 10% have pancreaticobiliary disease such as gallstones or duct anomalies, and approximately 5% of pancreatitis cases are associated with medications. Other causes contribute a small percentage of cases (Table II). Even though hereditary pancreatitis affects a small fraction of patients, these patients account for a disproportionate number of episodes and hospital admissions.

As in adults, the predominant symptom of acute pancreatitis in children is pain. Although the pain can be severe, it may be mild or even absent. Radiation to the back may occur. Nausea and vomiting may be present and sometimes the diagnosis is first suspected because of feeding intolerance when feeding is introduced in patients with systemic illness. Transient fever or jaundice can also be present. Jaundice should raise concerns about biliary tract involvement. Rarely, patients present with an abdominal mass or ascites.

The diagnosis of acute pancreatitis can be difficult because no readily available test confirms the diagnosis. Most patients are screened with serum amylase and lipase levels. There are no studies of the diagnostic performance of either amylase or lipase levels in children. Although there have been multiple attempts to determine the sensitivity and specificity of elevations in both enzymes in adults, the studies all suffer from the absence of a method to separately and absolutely document pancreatitis. It is clear that both enzymes can be normal when there is radiographic and clinical evidence of pancreatitis. Also, both enzymes can be elevated by other conditions unrelated to pancreatitis (Table III). The level of elevation is also not diagnostic although the higher the level is above the upper reference limit, the more likely there is to be pancreatic inflammation. Levels just above the upper reference limit may still be secondary to pancreatitis especially in patients who present several days after the onset of symptoms. Other pancreatic products, such as phospholipase A2, trypsin, trypsinogen activation peptide, and elastase, are also elevated in pancreatitis, but none has found widespread use in the clinical laboratory. Serum transaminases are elevated in some patients and the combination of elevated amylase or lipase and elevated serum transaminases may be more predictive of pancreatitis than elevated amylase or lipase alone.

Radiographic studies help in the diagnosis of acute pancreatitis and in defining complications of acute pancreatitis. Both ultrasound and computed tomography (CT scan) have been used to document pancreatitis. Ultrasound findings include enlargement of the pancreas, altered echogenicity of the pancreas, dilated main pancreatic duct, gallstones, pancreatic calcifications, choledochal cysts, and fluid collections, either peripancreatic or cystic. CT scan will show similar findings,

TABLE III Causes of Elevated Amylase or Lipase

Pancreatic disease	Nonpancreatic causes
Acute pancreatitis	Salivary adenitis
Chronic pancreatitis	Salpingitis
Pseudocyst	End-stage renal disease
Pancreatic ascites	Burn injury
Pancreatic cancer	Acute cholecystitis
	Upper GI endoscopy
	Diabetic ketoacidosis
	Macroamylasemia
	Macrolipasemia

TABLE IV Complications of Acute Pancreatitis

Local	Systemic
Edema	Shock
Inflammation	Pulmonary edema
Fat necrosis	Pleural effusions
Phlegmon	Acute renal failure
Hemorrhage	Coagulopathy
Abscess	Hypocalcemia
Fluid collections	Hyperglycemia
Extension to nearby organs	Distant fat necrosis

except that abnormal attenuation, rather than altered echogenicity, is seen. If the pancreas does not perfuse with intravenous contrast during the CT scan, pancreatic necrosis is likely. MRCP may be helpful in defining abnormalities of the ductal system, but its utility in pediatric patients has not been carefully studied. ERCP should be reserved for patients with unexplained recurrent episodes of pancreatitis, a prolonged episode of pancreatitis, or gallstone pancreatitis.

Medical management of acute pancreatitis is supportive and consists of providing adequate fluids and analgesia and of monitoring for metabolic complications (Table IV). Nutritional therapy should be started early in the hospitalization. Until recently, parenteral nutrition was considered the only option, but several studies show that adult patients with acute pancreatitis tolerate jejunal feeding with fewer complications than those given parenteral nutrition.

Fortunately, pancreatitis is generally not severe in children. Hemorrhagic pancreatitis and infected pancreatic necrosis can occur, but the incidence has not been adequately established. Hypovolemic shock and severe underlying systemic illness are risk factors for severe pancreatitis in children. Typically, patients with severe pancreatitis develop ominous clinical findings such as peritoneal signs, respiratory distress, or circulatory collapse. The utility of either the Ranson or APACHE II scoring system for the severity of pancreatitis has not been tested in pediatrics. The prophylactic use of broad-spectrum antibiotics in severe pancreatitis remains an unsettled issue because current evidence supporting this approach has many uncertainties and leaves unanswered a number of questions about details of therapy. Still, many clinicians would commence a course of broad-spectrum antibiotics in critically ill patients. Surgical debridement is rarely required in children, but should be considered in patients with necrosis who are deteriorating or failing to progress despite maximal supportive care.

Other complications of pancreatitis may require specific treatment. Pancreatic pseudocysts occur frequently in children and are managed as discussed earlier. Abscesses can often be treated with external drainage and intravenous antibiotics and surgical drainage is rarely necessary. Surgery may be necessary for traumatic rupture of the duct although endoscopic stenting across the disrupted duct is another option.

CHRONIC PANCREATITIS

The distinction between acute and chronic pancreatitis may present diagnostic problems, particularly in the early stages of chronic pancreatitis. To make a diagnosis of chronic pancreatitis, irreversible morphological changes in the pancreas must be present. These changes are most often determined by radiographic studies or by ERCP (Fig. 4). The presence of permanent exocrine insufficiency and diabetes mellitus in association with the morphological changes greatly aids the diagnosis. Patients with recurrent pancreatitis that is secondary to ductal obstruction may show radiographic changes similar to those seen in chronic pancreatitis and may even have evidence of decreased function, but these can improve on correction of the obstruction.

In pediatrics, cystic fibrosis is by far the commonest cause of chronic pancreatitis. Other causes, hereditary pancreatitis and SDS, have been discussed. The remaining causes occur quite infrequently. Many cases remain idiopathic despite advances in the diagnosis of genetic and structural defects. In a few affected patients, biopsy of the pancreas reveals marked fibrosis of the parenchyma with relative sparing of islets. This entity has been called juvenile idiopathic fibrosing pancreatitis and may lead to pancreatic or bile duct obstruction by fibrotic tissue. Other identifiable causes of chronic pancreatitis include hyperparathyroidism (in which the hypercalcemia is thought to cause repeated episodes of pancreatitis) and hyperlipidemias, in particular types I, IV, and V. Tropical pancreatitis is found only in tropical areas including south India and Africa and its etiology remains unknown. Autoimmune pancreatitis causes chronic pancreatitis associated with increased serum gamma globulin, autoantibodies, enlargement of the

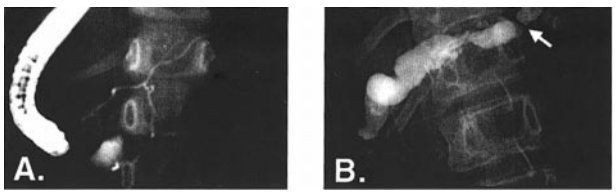


FIGURE 4 ERCP in chronic pancreatitis. (A) Normal ductal anatomy is visualized by contrast injection. (B) The dilated and tortuous main pancreatic duct in a patient with chronic pancreatitis. A stricture is present near the tail (arrow).

pancreas, which may be diffuse or localized, and a narrow main pancreatic duct. Although rare, the diagnosis is important because these patients respond to steroid therapy. To date, there have been no reports of autoimmune pancreatitis in children. The diagnosis and treatment of childhood chronic pancreatitis are the same as discussed for hereditary pancreatitis.

See Also the Following Articles

Cystic Fibrosis • Exocrine Pancreatic Insufficiency • Pancreatitis, Acute • Pancreatitis, Chronic • Pancreatitis, Hereditary • Pancreatitis, Pediatric

Further Reading

Durie, P. R., Forstner, G. G., Gaskin, K. J., Kopelman, H. R., Pitchumoni, C. S., and Robertson, M. A. (1996). The pancreas. *In* "Pediatric Gastrointestinal Disease" (P. R. Durie, J. R. Hamilton, W. A. Walker, J. A. Walker-Smith, and J. B. Watkins, eds.), 2nd Ed., Vol. 2, pp. 1247–1513. Mosby, New York.

Lopez, M. J. (2002). The changing incidence of acute pancreatitis in children: A single-institution perspective. *J. Pediatr.* **140**, 622–624.

Storman, M. O., and Durie, P. R. (2002). Pathophysiologic basis of exocrine pancreatic dysfunction in childhood. *J. Pediatr. Gastroenterol. Nutr.* **35**, 8–21.

Werlin, S. L. (1998). Disorders of the pancreas in children. *Curr. Opin. Pediatr.* **10**, 507–511.

Pancreatic Ductal Adenocarcinoma

MICHAEL BOUVET, JEANNIE SHEN, AND A. R. MOOSSA
University of California, San Diego

adenocarcinoma Malignant neoplasm of epithelial origin that is predominantly glandular or ductal.

ascites the accumulation of free fluid in the peritoneal cavity.

Blumer's shelf Palpable rectovaginal or rectovesical nodularity that may represent metastatic disease from an intra-abdominal or retroperitoneal malignancy.

carcinomatosis The presence of widespread peritoneal implants from tumor.

cholangitis Infection of the biliary tract.

Courvoisier's sign A nontender, palpable gallbladder, often associated with a malignant obstruction of the common bile duct.

double duct sign A radiographic sign seen on computed tomography scan or endoscopic retrograde cholangiopancreatography that results from the simultaneous obstruction of the common bile duct and pancreatic duct by a mass from the head of the pancreas.

endoscopic retrograde cholangiopancreatography A method of cholangiography that involves cannulating the sphincter of Oddi under direct vision through a fiberoptic endoscope.

endoscopic ultrasound A method of ultrasound performed via an endoscope that can be useful in evaluating the presence and character of a gastric, duodenal, biliary, or pancreatic mass.

gastric outlet obstruction A clinical condition that consists of epigastric abdominal pain, early satiety, and postprandial emesis that can be associated with, but is not specific to, pancreatic malignancy.

gastric varices Dilated gastric veins that occur as a result of portal hypertension; may also occur as a result of splenic vein thrombosis.

jaundice Yellowish discoloration of the skin.

oncogene The altered form of a proto-oncogene that promotes uncontrolled cell proliferation.

proto-oncogene Cellular genes that are involved in the regulation of proliferation, development, and differentiation.

Sister Mary Joseph's nodule Periumbilical mass that may contain lymph node metastases from a peritoneal source of malignancy.

steatorrhea Excess fecal fat, which may be a manifestation of pancreatic exocrine insufficiency and fat malabsorption.

Trousseau's sign Recurrent or migratory superficial thrombophlebitis that may be an early manifestation of abdominal cancer or other systemic illnesses.

Copyright 2004, Elsevier (USA). All rights reserved.

pancreas, which may be diffuse or localized, and a narrow main pancreatic duct. Although rare, the diagnosis is important because these patients respond to steroid therapy. To date, there have been no reports of autoimmune pancreatitis in children. The diagnosis and treatment of childhood chronic pancreatitis are the same as discussed for hereditary pancreatitis.

See Also the Following Articles

Cystic Fibrosis • Exocrine Pancreatic Insufficiency • Pancreatitis, Acute • Pancreatitis, Chronic • Pancreatitis, Hereditary • Pancreatitis, Pediatric

Further Reading

Durie, P. R., Forstner, G. G., Gaskin, K. J., Kopelman, H. R., Pitchumoni, C. S., and Robertson, M. A. (1996). The pancreas. *In* "Pediatric Gastrointestinal Disease" (P. R. Durie, J. R. Hamilton, W. A. Walker, J. A. Walker-Smith, and J. B. Watkins, eds.), 2nd Ed., Vol. 2, pp. 1247–1513. Mosby, New York.

Lopez, M. J. (2002). The changing incidence of acute pancreatitis in children: A single-institution perspective. *J. Pediatr.* **140**, 622–624.

Storman, M. O., and Durie, P. R. (2002). Pathophysiologic basis of exocrine pancreatic dysfunction in childhood. *J. Pediatr. Gastroenterol. Nutr.* **35**, 8–21.

Werlin, S. L. (1998). Disorders of the pancreas in children. *Curr. Opin. Pediatr.* **10**, 507–511.

Pancreatic Ductal Adenocarcinoma

MICHAEL BOUVET, JEANNIE SHEN, AND A. R. MOOSSA
University of California, San Diego

adenocarcinoma Malignant neoplasm of epithelial origin that is predominantly glandular or ductal.

ascites the accumulation of free fluid in the peritoneal cavity.

Blumer's shelf Palpable rectovaginal or rectovesical nodularity that may represent metastatic disease from an intra-abdominal or retroperitoneal malignancy.

carcinomatosis The presence of widespread peritoneal implants from tumor.

cholangitis Infection of the biliary tract.

Courvoisier's sign A nontender, palpable gallbladder, often associated with a malignant obstruction of the common bile duct.

double duct sign A radiographic sign seen on computed tomography scan or endoscopic retrograde cholangiopancreatography that results from the simultaneous obstruction of the common bile duct and pancreatic duct by a mass from the head of the pancreas.

endoscopic retrograde cholangiopancreatography A method of cholangiography that involves cannulating the sphincter of Oddi under direct vision through a fiberoptic endoscope.

endoscopic ultrasound A method of ultrasound performed via an endoscope that can be useful in evaluating the presence and character of a gastric, duodenal, biliary, or pancreatic mass.

gastric outlet obstruction A clinical condition that consists of epigastric abdominal pain, early satiety, and postprandial emesis that can be associated with, but is not specific to, pancreatic malignancy.

gastric varices Dilated gastric veins that occur as a result of portal hypertension; may also occur as a result of splenic vein thrombosis.

jaundice Yellowish discoloration of the skin.

oncogene The altered form of a proto-oncogene that promotes uncontrolled cell proliferation.

proto-oncogene Cellular genes that are involved in the regulation of proliferation, development, and differentiation.

Sister Mary Joseph's nodule Periumbilical mass that may contain lymph node metastases from a peritoneal source of malignancy.

steatorrhea Excess fecal fat, which may be a manifestation of pancreatic exocrine insufficiency and fat malabsorption.

Trousseau's sign Recurrent or migratory superficial thrombophlebitis that may be an early manifestation of abdominal cancer or other systemic illnesses.

Copyright 2004, Elsevier (USA). All rights reserved.

tumor suppressor gene Altered form of anti-oncogene whose loss-of-function or deactivation has a permissive role in the development of a neoplasm.

Virchow's node Lymph node located at the terminus of the thoracic duct in the left supraclavicular region that may contain lymph node metastases from distant primary sites via the retroperitoneal and postmediastinal lymph channels.

Pancreatic ductal adenocarcinoma is a malignancy of the exocrine pancreas that is the fifth leading cause of cancer-related deaths. Although the exact cause of this disease is unknown, factors thought to contribute its etiology include cigarette smoking and other environmental risks. The molecular mechanisms involved in the pathogenesis of these tumors include mutations in the K-ras oncogene, p53 and DPC4 tumor suppressor genes, and the p16 cell cycle regulatory protein. The clinical symptoms of pancreatic cancer include jaundice, abdominal pain, and weight loss. There is often a delay in diagnosis that contributes to the high rate of mortality resulting from the disease. Newer diagnostic and staging modalities involve the use of tumor markers, helical computed tomography scanning, magnetic resonance imaging, endoscopic ultrasound, and laparoscopy. If the tumor is localized to the pancreas, surgical resection offers the only chance of long-term cure. Adjuvant chemoradiation strategies can improve survival and decrease local recurrence. In advanced disease, palliation of obstructive jaundice, gastric outlet obstruction, and pain is important.

INTRODUCTION

In the United States, pancreatic cancer (ductal adenocarcinoma) accounts for approximately 30,000 deaths per year. The majority of patients present in the late stages of the disease with locally advanced or metastatic tumors. Only 10–20% of patients are candidates for resection and hence have any potential for cure. The signs and symptoms of pancreatic cancer vary from vague nonspecific abdominal complaints to severe jaundice and often the diagnosis can be difficult, especially in the early stages. Until the past decade or so, the traditional approach was surgical exploration for tissue diagnosis, staging, and assessment of resectability. More recently, sophisticated tests, including thin-cut helical computed tomography (CT) scan, magnetic resonance imaging (MRI), endoscopic ultrasound (EUS), biochemical tumor markers, fine-needle aspiration (FNA) cytology, and laparoscopy, have added a new dimension to the diagnosis and staging of pancreatic carcinoma. The treatment of pancreatic cancer is surgical resection, when possible. Adjuvant chemoradiation has been shown by several, but not all, studies to improve survival. The overall prognosis for these patients depends on pathologic variables of the tumor. Future strategies for earlier diagnosis and treatment are focused on the evolving knowledge of the genetic mutations that characterize pancreas cancer.

EPIDEMIOLOGY AND ETIOLOGY

Cancer of the pancreas is distinctly more common in older people with most patients between the ages of 65 and 80 at diagnosis. There is a slight male predilection for the disease with the male to female ratio being 1.3 to 1.0. The strongest association is between pancreatic cancer and cigarette smoking. Current estimates suggest that 30% of pancreatic cancer cases are due to cigarette smoking. A high-protein and high-fat diet, characteristic of the Western population, has been proposed as a possible factor. Exposure to industrial carcinogens, especially betanaphthylamine and benzidine, has been documented in pancreatic cancer patients. A higher that normal incidence rate of the neoplasm has also been reported in chemists, workers in metal industries, and coal and gas plant employees.

Diabetes mellitus has been proposed as a risk factor for pancreatic cancer. However, a large cohort study examining this issue showed that the diabetes was typically of new onset, indicating that it was brought on by the tumor. Another known risk factor is longstanding chronic pancreatitis. Such patients have a 1.8 and 4.0% chance of developing pancreatic cancer 10 and 20 years, respectively, after their initial diagnosis of chronic pancreatitis. Finally, pancreatic cancer is associated with certain inherited diseases, such as the atypical mole melanoma syndrome and hereditary pancreatitis.

PATHOLOGY AND MOLECULAR BIOLOGY

Most (95%) malignant neoplasms of pancreatic origin arise from the exocrine portion of the gland, with ductal adenocarcinomas being the most common subtype. The predominant histologic feature of these tumors is a dense collagenous stroma with atrophic acini, remarkably preserved islet cell clusters, and a slight to moderate increase in the number of ducts, both normal-appearing and cancerous (Fig. 1). The diagnosis of ductal adenocarcinoma rests on the identification of mitoses, nuclear and cellular pleomorphism, discontinuity of ductal epithelium, and evidence of perineural, vascular, or lymphatic invasion. Much more infrequent are tumors arising from the islets of Langerhans (endocrine cells

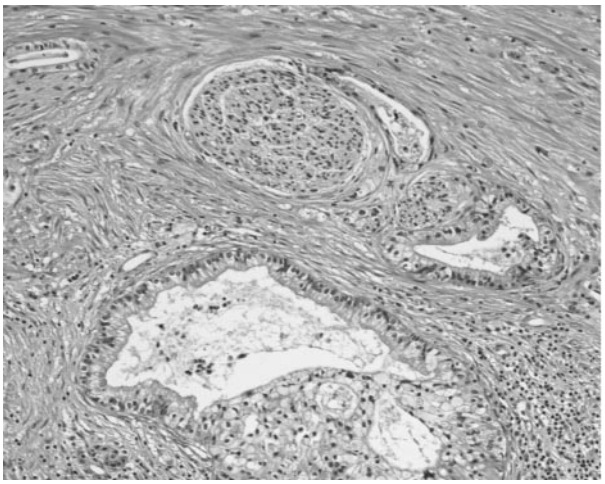


FIGURE 1 Hematoxylin and eosin section of pancreatic ductal adenocarcinoma. Note extensive fibrosis and perineural invasion.

of the pancreas). Primary nonepithelial tumors of the pancreas (e.g., lymphomas or sarcomas) are extremely rare.

Sporadic cancers of the pancreas are frequently associated with the activation of an oncogene, K-ras, and the inactivation of multiple tumor suppressor genes, including p53, *DPC4* (deleted in pancreatic carcinoma, locus 4), p16, and Rb. Up to 90% of human pancreatic cancers have ras gene mutations, most of which are K-ras mutations, and K-ras mutations seem to be an early event in pancreatic carcinogenesis. A number of studies have indicated that p53 mutations are relatively common in adenocarcinoma of the pancreas, occurring in 70% of patients. In addition, p53 mutation may be an independent prognostic factor in patients with pancreatic ductal adenocarcinoma. Mutations in the p16 gene are found in approximately 60% of pancreatic adenocarcinomas and may be associated with short patient survival. *DPC4* is inactivated in approximately half of all pancreatic cancers and, in contrast to p53 and p16, appears to be relatively specific to pancreatic cancer.

CLINICAL PRESENTATION

The clinical features of pancreatic cancer vary from vague symptoms of abdominal discomfort to anorexia, weight loss, and obstructive jaundice (Table I). The initial symptoms and signs depend on the site and extent of the pancreatic cancer. Not surprisingly, tumors of the pancreatic body and tail can grow to a very large size before obvious symptoms alert the patient or physician to the diagnosis. On the other hand, small tumors, when located in the pancreatic head, can cause obstructive jaundice or "pancreatitis," alerting the clinician to suspect pancreatic cancer.

Progressive jaundice occurs in over 75% of patients with carcinoma of the head of the pancreas and the incidence of jaundice decreases as the location of the lesion progresses to the left toward the tail of the pancreas. Occasionally, a tumor may invade and compress the third or fourth part of the duodenum without actually obstructing the common bile duct. This is often seen in cancer originating in the lower part of the uncinate process that tends to extend inferiorly into the root of the superior mesenteric vessels. Pain is extremely frequent and the classic description of painless jaundice is rarely encountered. Weight loss and anorexia are also common symptoms even in early stages. Nausea, epigastric bloating, change in bowel habits, and vomiting are occasionally present. Hematemesis and melena may sometimes occur in late cases as a result of direct invasion into the duodenal or gastric mucosa by tumor or portal vein–splenic vein compression by the tumor, leading to gastric varices. Chills and fever due to ascending cholangitis rarely occur even in longstanding biliary obstruction unless the biliary tree is contaminated by catheterization at endoscopic retrograde cholangiopancreatogram (ERCP).

Physical examination may not be pathognomonic but may aid in the diagnosis. Approximately 75% of patients with carcinoma of the head of the pancreas are jaundiced. The palpable gallbladder (Courvoisier's sign) may be present in 25% of patients with a malignant obstruction of the common bile duct. Patients with advanced disease may present with ascites, palpable supraclavicular lymphadenopathy (Virchow's node), a periumbilical mass (Sister Mary Joseph's nodule), or a palpable rectovaginal or rectovesical nodularity (Blumer's shelf). Diabetes of recent onset in elderly

TABLE I Signs and Symptoms of Pancreatic Cancer

Jaundice
Pain
Weight loss
Anorexia
Nausea
Bloating
Emesis
Courvoisier's gallbladder
Diabetes of recent onset
Advanced disease
Ascites
Virchow's node
Sister Mary Joseph's nodule
Blumer's shelf
Trousseaus's sign (migratory thrombophlebitis)

patients with vague gastrointestinal symptoms should alert the physician to a possible underlying pancreatic carcinoma. Migratory thrombophlebitis (Trousseau's sign) can be present in any patient with advanced cancer, is not specifically indicative of pancreatic carcinoma, and, by itself, does not merit diagnostic laparotomy or laparoscopy.

DIAGNOSIS AND STAGING

It is clear that no clinical feature by itself is sufficiently accurate to make a definite diagnosis of, or to exclude, a pancreatic cancer. The following list should serve as a guideline to choose those patients who may warrant further investigation of the pancreas. The clinical suspicion should be increased in patients who are over 40 years of age, who are heavy cigarette smokers, and who present with any of the following symptoms: (1) obstructive jaundice; (2) unexplained recent weight loss greater than 10% of body weight; (3) unexplained upper abdominal or lumbar back pain; (4) unexplained dyspepsia; (5) sudden onset of diabetes mellitus without any predisposing factors, such as a family history of diabetes mellitus or obesity; (6) one or more attacks of "idiopathic" pancreatitis; and (7) unexplained steatorrhea. It should be remembered that the earlier the cancer, the more difficult it is to achieve a positive diagnosis, even at operation.

CA 19–9 and Other Tumor Markers

CA 19–9 was first described in 1979 as a tumor-associated antigen and later identified as a Lewis blood group-related mucin. Although initially CA 19–9 was thought to be useful in the management of patients with colorectal carcinoma, its role in pancreatic cancer became more evident. Over the past 10 years, several hundred papers have subsequently documented the utility of CA 19–9 in the diagnosis, prognosis, and monitoring of pancreatic cancer. However, as is the case with most tumor markers, the sensitivity of CA 19–9 is not perfect. Therefore, the proper role for CA 19–9 is as a diagnostic adjunct or simply as another piece of critical diagnostic information to be integrated into the diagnostic decision-making process.

When CA 19–9 reference values of >90 or >200 U/ml are used, the diagnostic accuracy for pancreatic cancer is 85 or 95%, respectively. However, there are several factors that influence CA 19–9 interpretation, most importantly the degree of jaundice. The false increase is likely because of hepatic insufficiency to degrade and excrete CA 19–9 metabolically. Additionally, CA 19–9 levels can be elevated in other tumor types, such as lung and ovarian cancers. Even in severe cases of pancreatitis, high CA 19–9 levels are noted. Interferon significantly elevates CA 19–9. Finally, patients with the Lewis blood phenotype (-a, -b) and pancreatic cancer may not have elevated levels of CA 19–9. The combined use of CT and CA 19–9 in nonicteric patients provides a positive predictive value of 99 to 100% in the diagnosis of pancreatic cancer when a reference value of 100 U/ml is used. Likewise, better results are obtained if CA 19–9 is used in conjunction with ultrasound or ERCP. A host of other tumor markers, such as CA 125, CEA, CA 494, CA 242, CA 50 SPAN-1, DU-PAN 2, and CA12−5, have been described for detection of pancreatic adenocarcinoma. Although some of these markers appear promising, most of them are not widely available and have not been tested to same degree as CA 19–9.

Ultrasonography

Real-time ultrasonography is an excellent modality for the initial evaluation of upper abdominal pain or obstructive jaundice. It does not entail exposure to ionizing radiation and is less costly and time consuming than CT scan. However, 15 to 20% of ultrasound examinations are technically suboptimal, as a result of bowel gas interference, obesity, or previous operations. Furthermore, visualization of the body, tail, and uncinate process is often marginal. Bile duct dilation defined as greater than 7 mm is commonly seen with carcinoma of the pancreatic head. However, bile duct dilation is not specific for carcinoma and may be indicative of advancing age, prior cholecystectomy, common duct stones, ampullary stenosis, and pancreatitis. Pancreatic ductal dilation, defined as ductal diameter exceeding 2 to 3 mm, is one of the most frequent secondary signs, seen in 20 to 60% of cases of pancreatic cancer. Most pancreatic cancers are hypoechoic relative to the normal parenchyma, whereas an increase in echogenicity may be found in focal pancreatitis. Ultrasonography is also useful in the detection of ascites and liver metastases. The yield of ultrasonography depends mostly on the skill of the ultrasonographer and thus may vary from institution to institution.

Computed Tomography Scanning

Thin-section helical CT scanning through the pancreas with an intravenous bolus injection of contrast remains the test of choice to evaluate the extent of disease in pancreatic cancer and to assess tumor resectability (Fig. 2). Typical CT findings of pancreatic cancer are those of a mass that deforms the size and contour of the gland. Most tumors have a central area of decreased attenuation (hypodense compared to

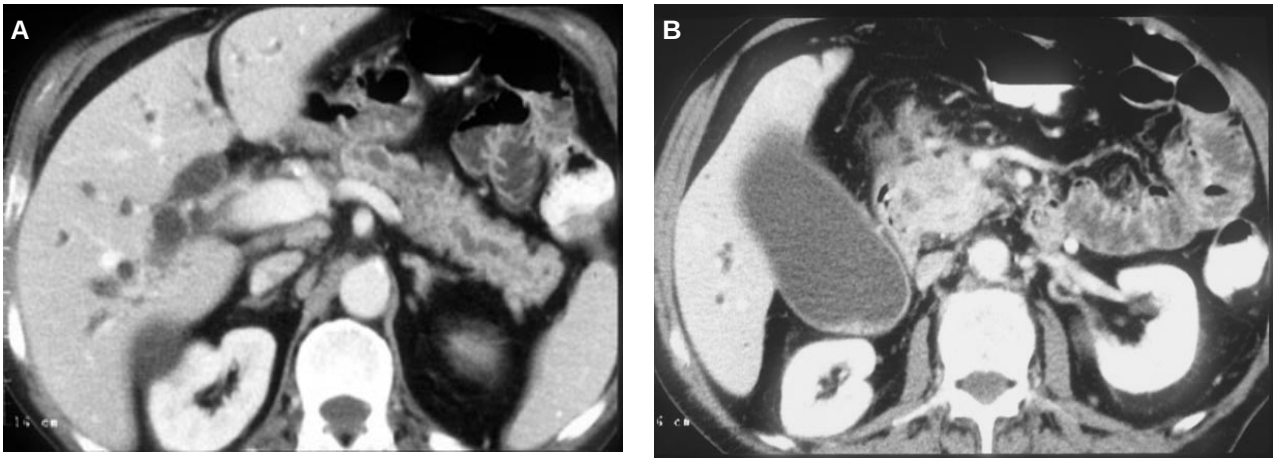


FIGURE 2 (A) CT scan showing both common bile duct and pancreatic duct dilation (double duct sign) in a patient with adenocarcinoma of the pancreatic head. (B) Note the massively enlarged, easily palpable gallbladder (Courvoisier's gallbladder).

normal pancreas). Dilation of the pancreatic duct proximal to the tumor is present in 70–80% of cases. Other common findings include pancreatic atrophy proximal to the tumor and a double duct sign if the tumor is located in the pancreatic head (Fig. 2).

The radiologic criteria for potentially resectable disease are defined as (1) the absence of extrapancreatic disease; (2) the absence of direct tumor extension to the superior mesenteric artery and celiac axis as defined by the presence of a fat plane between the low-density tumor and these arterial structures; and (3) a patent superior mesenteric–portal vein confluence (Fig. 3). In addition, CT is useful in determining the distant spread of pancreatic cancer. Liver metastases from pancreatic ductal carcinoma are hypovascular and are hypoattenuating to the liver on portal-phase scans (Fig. 3). Other common sites of metastases include the mesentery, the omentum, and the lungs. The presence of ascites usually indicates carcinomatosis and unresectability. False-negative CT scanning can be caused by small (less than 1 cm) metastasis to the liver and small peritoneal seedings.

Magnetic Resonance Imaging

Although CT scan is typically the first choice in pancreatic imaging, MRI is rapidly evolving and may play an increasing role in the evaluation of the patient

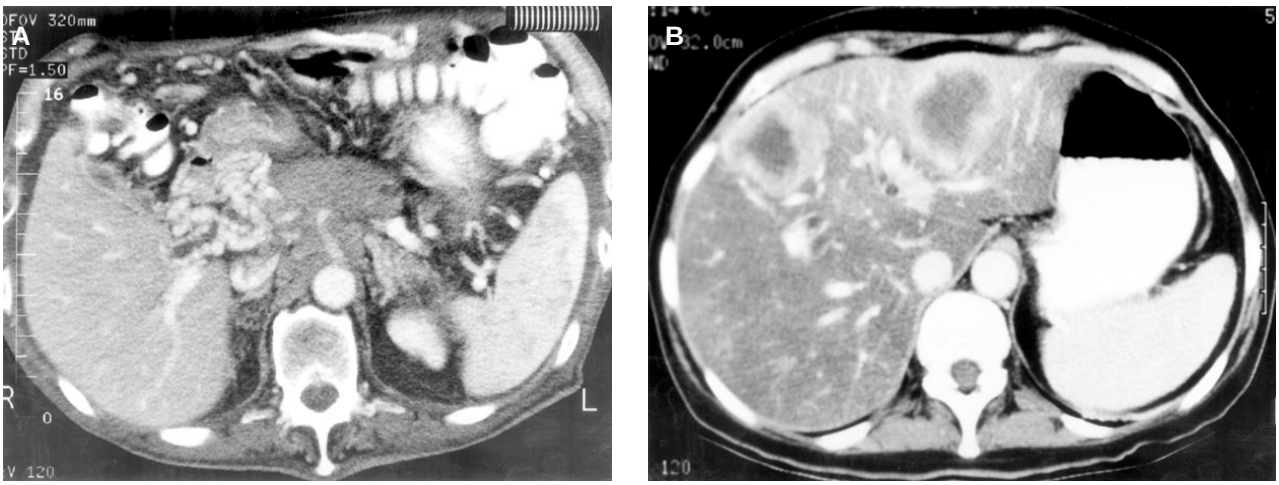


FIGURE 3 (A) CT scan of an unresectable adenocarcinoma of the pancreatic head with tumor encasement of the celiac artery. Note the extensive venous collateral secondary to occlusion of the superior mesenteric–portal vein confluence. (B) Large hypodense liver metastases in stage IV pancreatic adenocarcinoma.

with pancreatic cancer. Optimal MRI of the pancreas uses phased-array surface coils placed anteriorly on the patient's abdomen and posteriorly against the spine. Axial T1- and T2-weighted images are then obtained through the pancreas using a section thickness of 6 to 8 mm, with a 1 to 2 mm gap, respectively. On axial T1-weighted images, pancreatic abnormalities generally show decreased signal compared with the relatively high signal of the normal pancreas. Axial T2-weighted images are particularly useful in assessing the extent of inflammatory or cystic disease involving the pancreas. Duct dilation is demonstrated best on T2-weighted or magnetic resonance cholangiopancreatography (MRCP) images.

Endoscopic Retrograde Cholangiopancreatography

Until recently, ERCP played a major role in the diagnosis of pancreatic cancer. However, rapid technologic advances in newer imaging modalities, such as MRCP and endoscopic ultrasonography, are redefining the indications for ERCP. Furthermore, ERCP is an invasive test that carries a morbidity of 2 to 3%, most commonly from pancreatitis and cholangitis. Therefore, the role of ERCP in the diagnosis of pancreatic cancer is rapidly diminishing.

The utility of preoperative endoscopic stent drainage in the jaundiced patient has been controversial. Recent data suggest that preoperative biliary drainage is associated with increased morbidity and mortality rates in patients undergoing pancreaticoduodenectomy. At the Memorial Sloan Kettering Cancer Center, over half of patients undergoing a Whipple resection underwent preoperative biliary drainage (endoscopic stents, percutaneous drains/stents, or surgical drainage). Preoperative biliary drainage was determined to be the only statistically significant variable associated with infectious complications, intra-abdominal abscess, and postoperative death. These investigators concluded that stent drainage should be avoided whenever possible in patients with potentially resectable pancreatic and peripancreatic lesions.

ERCP will continue to be indicated in the patient who is not a surgical candidate. Stent drainage is widely considered a reasonable way of palliating malignant obstructive jaundice in patients who are not candidates for operative palliation. Other indications for ERCP include biliary obstruction that is not associated with a pancreatic mass. In such cases, diagnostic ERCP may reveal a nonmalignant cause of biliary obstruction such as choledocholithiasis, which can be effectively treated by sphincterotomy and stone extraction. However, diagnostic endoscopic ultrasound is emerging as the initial evaluative method, reserving ERCP for sphincterotomy and stone extraction.

Endoscopic Ultrasound

Endoscopic ultrasound produces high-frequency ultrasonographic images of the pancreas using the wall of the stomach and duodenum as an acoustical window. EUS provides an exceptional degree of anatomic detail that results in the ability to identify and stage lesions accurately. The pancreas, portal vein, celiac axis, common bile duct, gallbladder, and liver can all be visualized with EUS (Fig. 4). Although EUS itself does not provide a tissue diagnosis, the development of EUS-guided FNA makes diagnosis possible.

An abundance of data indicates that EUS is a sensitive method for the detection of pancreatic tumors. The larger series have demonstrated sensitivities of 90% or better. EUS is especially helpful for smaller lesions. In most studies, EUS is superior to transabdominal ultrasound and CT and is equivalent to ERCP for lesions that are 3 cm or smaller. Most tumors are relatively hypoechoic compared with pancreatic parenchyma, although the echo pattern becomes mixed and more variable the larger the size of the tumor. Cancers tend to have an irregular margin, sometimes with extending pseudopodia, but malignant lesions smaller than 3 cm may have a smooth margin. Unfortunately, areas of focal inflammation can also have a similar hypoechoic pattern and therefore the specificity for EUS for

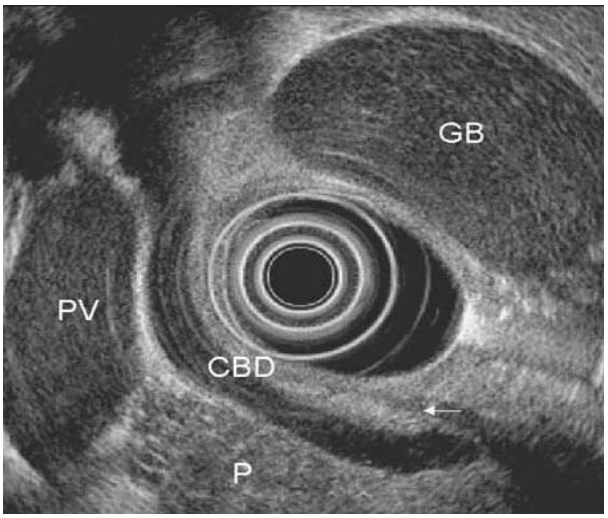


FIGURE 4 Endoscopic ultrasound of the pancreas. CBD, common bile duct; GB, gallbladder; P, pancreas; PV, portal vein.

differentiating pancreatic cancer from focal pancreatitis is approximately 70%.

Fine-Needle Aspiration Cytology

In patients with potentially resectable disease, percutaneous or endoscopic ultrasound-guided needle biopsy should be considered *only* in cases where pretreatment cytologic confirmation of the diagnosis is important, such as in patients being considered for preoperative multimodality therapy as part of a clinical trial. FNA is performed by passing a 21- to 23-gauge needle into the pancreatic mass using the appropriate imaging technique for guidance and then applying suction with a syringe. The specimen is then expressed onto a microscope slide, smeared, fixed in 95% alcohol, and stained by the Papanicolaou or other method. If an experienced cytopathologist is available, a diagnosis can generally be made within 20 minutes of obtaining the biopsy.

The typical findings of pancreatic adenocarcinoma include single or irregularly arranged clusters of cells exhibiting cellular pleomorphism, large vesicular nuclei, and prominent nucleoli. However, such cytologic appearances may sometimes fail to differentiate between an adenocarcinoma, a lymphoma, or an islet cell tumor. The main problem with fine-needle aspiration, however, is sampling error. A negative cytologic specimen does *not* exclude cancer.

FNA of the pancreas is useful for masses in the body and tail of the gland because cancers in this location are usually unresectable, especially if there is CT or EUS evidence of unresectability. It is especially valuable in the frail or elderly patient in whom one wishes to avoid a purely diagnostic laparotomy when surgical palliation is not indicated or warranted. The technique, however, should not be used for small, potentially resectable cancers because of sampling error and the theoretical possibility of seeding along the needle tract. The pancreas is a vascular organ with a rich lymphatic network. Common sense dictates that meddlesome needling can disseminate a cancer that already has a high propensity for local invasion and vascular permeation. When FNA is indicated, EUS may be the preferred approach, owing to a more direct route that does not violate the peritoneum.

Laparoscopy

Pancreatic cancer is notorious for its high predilection for liver and peritoneal metastases. Identifying small (1 to 2 mm) implants that are not definable on CT scan can now be achieved with laparoscopy. Laparoscopy is minimally invasive and can establish the diagnosis through visualization and biopsy of metastatic lesions. It can also serve as an avenue for obtaining peritoneal cytology, which can be positive in 20 to 30% of cases. With peritoneal lavage, cytologic exam of washings can be positive up to 30% of the time, indicating tumor unresectability and poor survival (median < 6 months).

The exact role of laparoscopy as a diagnostic tool and in the staging of pancreatic cancer remains to be defined. Studies by Cuschieri and Warshaw have established the value of laparoscopy in detecting liver and peritoneal metastases. However, many of their patients had locally advanced disease that was not amenable to surgical resection. When laparoscopy is limited to patients who fulfilled strict CT criteria for resectability, Fuhrman and colleagues could not confirm the diagnostic value of routine preoperative laparoscopic evaluation.

Sequence of Diagnostic Tests

The goal of preoperative diagnostic staging in pancreatic cancer is to differentiate the small group of patients who have a potentially resectable localized cancer from those with locally advanced disease and/or disseminated metastases. Figure 5 is an algorithm for the diagnosis and staging of pancreatic tumors. Patients with a clinical suspicion of pancreatic cancer should undergo physical examination and laboratory evaluation including CA 19–9 levels. Helical CT scan is the initial modality used to image the pancreas. MRI offers an alternative or adjunct to CT imaging of the pancreas. If the CT scan is equivocal, endoscopic ultrasound has been found to be useful, especially in the detection of small ampullary or periampullary neoplasms. If EUS reveals findings of unresectablity, such as superior mesenteric or celiac artery encasement, portal vein occlusion, or liver metastases, a tissue diagnosis can be established in the same session by EUS-guided FNA.

If the patient appears to have localized, surgically resectable disease by CT and EUS, it is appropriate to prepare the patient for an operation. Laparoscopy may be used as a preliminary test to detect small peritoneal or liver metastases that were not seen on initial imaging studies. The authors do not advocate routine biopsy of the mass or peripancreatic lymph nodes at surgery because these are excised *en bloc* with the specimen. On the other hand, if the tumor is found to be unresectable at operation, every attempt must be made to establish a tissue diagnosis by biopsy and frozen section histology prior to leaving the operating room. This removes any doubt about the exact diagnosis and the patient can be referred for palliative treatment.

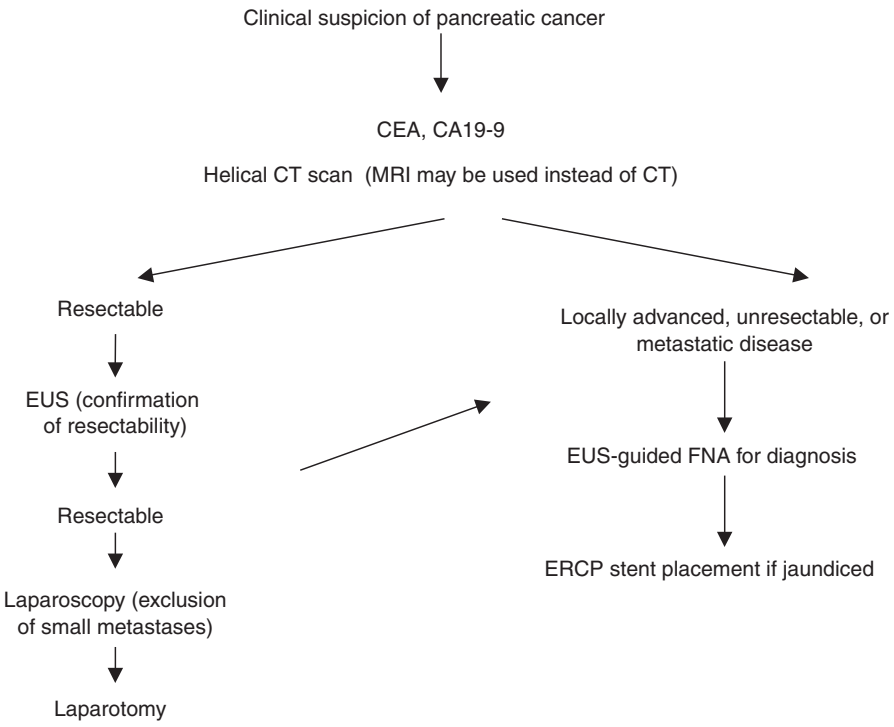


FIGURE 5 Algorithm for diagnosis and staging of pancreatic tumors.

CURRENT TREATMENT STRATEGIES AND PROGNOSIS

Emphasis must be placed on preoperative evaluation and adequate preparation of patient with pancreatic cancer. As mentioned earlier, CT scan, cytology, and EUS provide the surgeon with valuable preoperative information and obviate the need for time-consuming maneuvers on the operating table. It cannot be overemphasized that pancreatic exploration with a view to resection should not be performed by the occasional surgeon or resident in training or at institutions where there is not sufficient back-up expertise (endoscopy, radiology, cytology, and critical care) that is necessary for the care and management of these difficult problems.

Surgery

Pancreaticoduodenectomy (Whipple operation) is most commonly employed for tumors of the head of the pancreas (Fig. 6). The operation is optimal for malignant tumors that are confined to the duodenum, ampulla of Vater, or lower common bile duct. The neck of the gland is divided to the left of the superior mesenteric vein and the body and tail of the pancreas and spleen are left undisturbed. *En bloc* excision of the regional lymph nodes from the porta hepatis, aortocaval, and superior mesenteric regions again forms part of the operation. With the Whipple operation, endocrine function can be preserved. Although there is the possibility of an anastomotic leak from the pancreaticojejunostomy, this complication occurs in less than 10–15% of patients at centers experienced with pancreatic surgery. Also, more effective management of pancreatic anastomotic leakage with hyperalimentation, percutaneous drainage, and a somatostatin analogue has reduced the magnitude of this problem. Total pancreatectomy should be reserved for situations when there is tumor at the pancreatic margin on serial frozen sections or if the pancreas is not suitable for an anastomosis. A pylorus-preserving Whipple operation is a reasonable alternative but may result in transient gastric stasis.

A cancer of the pancreas is considered unresectable if there are distant (liver or peritoneal) metastases, invasion of major vessels (portal vein, hepatic artery, superior mesenteric vessels, or celiac axis), or any extension beyond the area of usual total pancreatectomy specimen. The possible exception is the case of isolated portal vein invasion provided the vein is patent. In these selected cases, portal vein resection with interposition graft placement has been described. Puckering of the

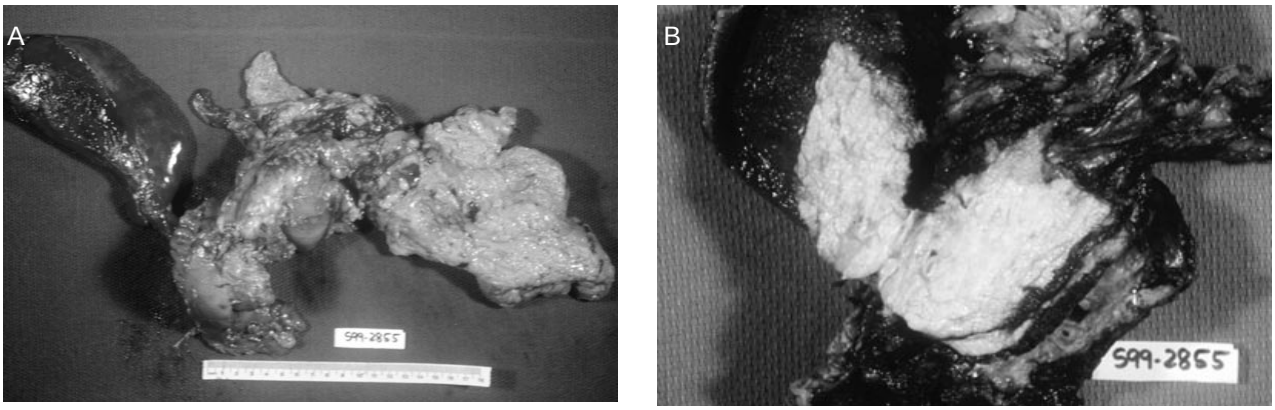


FIGURE 6 (A) Whipple specimen from a patient with pancreatic cancer. (B) The tumor is sectioned.

transverse mesocolon per se does not always indicate unresectability since the transverse mesocolon and, if necessary, the transverse colon can be excised in a total pancreatectomy specimen if there are no contraindications to a resection.

The mortality for major pancreatic resection is between 0 and 5% at specialized centers. Death as a result of operative complications usually occurs within the first 2 months of operation. After 2 months and up to 2 years later, death is usually due to metastatic pancreatic cancer although a few individuals can present as late as the end of the third year with metastatic disease. If the patient has survived 3 years, the cause of death is usually unrelated to pancreatic cancer. The estimated 5-year survival rate in patients with localized adenocarcinoma of the pancreas who undergo surgical resection of the primary tumor ranges from 6 to 24%. Survival after potentially curative resection for periampullary neoplasms including pancreatic cancer depends primarily on pathologic factors. Factors shown to influence survival include tumor size, lymph node metastases, tumor differentiation, margin status, and vascular and perineural invasion. In addition, tumors of the ampulla of Vater, distal common bile duct, and duodenum have a more favorable prognosis than that of pancreatic ductal carcinoma.

Chemotherapy and Radiation

Data from several randomized trials have shown a significant survival benefit in patients treated with infusional 5-fluorouracil (5-FU) and external beam radiation following pancreatectomy. Therefore, in patients who are well enough to tolerate it, adjuvant chemoradiation is recommended. Newer agents, such as gemcitabine, are currently being studied in combination with radiation as alternatives to 5-FU.

If diagnostic studies indicate that the patient has locally advanced, unresectable disease or metastatic disease, nonoperative palliation of jaundice can usually be achieved by endoscopically placed biliary stents. If the patient survives for more than a few months, recurrent cholangitis associated with stent blockage is a problem that necessitates regular endoscopic removal and replacement of the stent. Newer self-expandible metallic devices, such as the Wall stent, may provide improved patency rates compared to plastic stents. Percutaneous transhepatic placement of an internal expandable metal stent is being tried by interventional radiologists and offers yet another option for palliation of the jaundiced patient with malignant biliary obstruction.

If these nonoperative techniques are unsuccessful, biliary tract decompression can be performed either by cholecystojejunostomy or by hepaticojejunostomy (each with a diverting enterostomy), depending on whether the cystic duct is widely patent and is in full communication with the biliary tree proximal to obstructing cancer. When there is evidence of gastric outlet obstruction from tumor compression of the duodenum, gastrojejunostomy can be performed. Chemoradiation with 5-FU and external beam irradiation for locally advanced, unresectable tumors is indicated. If liver metastasis or carcinomatosis is found at laparotomy, single-agent chemotherapy with gemcitabine may provide palliation and improve the patient's quality of life.

CONCLUSION

Pancreatic cancer is a common cause of cancer-related mortality. Although the exact etiology for pancreatic cancer is unknown, the molecular events that take place involve mutations in oncogenes and tumor

suppressor genes. The diagnosis of pancreatic cancer continues to evolve with new technological advances in imaging, endoscopy, and biochemical tumor markers. The goal of such new diagnostic testing remains earlier tumor detection, thus providing the patient with the best chance for cure. The primary treatment for localized disease is surgical resection. Adjuvant chemoradiation has been shown by most, but not all, studies to improve local control and survival. In patients with advanced disease, palliation of jaundice, pain, and gastric outlet obstruction can be achieved by the use of stents and/or surgical bypass. Newer agents, such as gemcitabine, have been shown to improve the quality of life for patients with advanced pancreatic cancer.

See Also the Following Articles

Cancer, Overview • Computed Tomography (CT) • Endoscopic Ultrasonography • Laparoscopy • Magnetic Resonance Imaging (MRI) • Pancreatic Cancer • Pancreatic Tumors, Other

Further Reading

Bluemke, D. A., and Fishman, E. K. (1998). CT and MR evaluation of pancreatic cancer. *Surg. Oncol. Clin. North Am.* **7**, 103–124.

Bouvet, M., Binmoeller, K. F., and Moossa, A. R. (2001). The diagnosis of adenocarcinoma of the pancreas. *In* "Atlas of Clinical Oncology: Pancreatic Cancer" (J. L. Cameron, ed.), pp. 67–85. B. C. Decker, Hamilton, Ontario, Canada.

Bouvet, M., Gamagami, R. A., Gilpin, E. A., Romeo, O., Sasson, A., Easter, D. W., and Moossa, A. R. (2000). Factors influencing survival after resection for periampullary neoplasms. *Am. J. Surg.* **180**, 13–17.

Casey, G., Yamanaka, Y., Friess, H., Kobrin, M., Lopez, M., Buchler, M., Beger, H., and Korc, M. (1993). p53 mutations are common in pancreatic cancer and are absent in chronic pancreatitis. *Cancer Lett.* **69**, 151–160.

Cerny, W. L., Mangold, K. A., and Scarpelli, D. G. (1992). K-ras mutation is an early event in pancreatic duct carcinogenesis in the Syrian golden hamster. *Cancer Res.* **52**, 4507–4513.

Cuschieri, A. (1988). Laparoscopy for pancreatic cancer—Does it benefit the patient? *Eur. J. Surg. Oncol.* **14**, 41–44.

Evans, D. B., Abbruzzese, J. L., and Rich, T. A. (1997). Cancer of the pancreas. *In* "Cancer Principles and Practice of Oncology" (V. T. DeVita, S. Hellman, and S. A. Rosenberg, eds.), pp. 1054–1081. Lippincott–Raven Press, Philadelphia, PA.

Fernandez-del Castillo, C., Rattner, D. W., and Warshaw, A. L. (1995). Further experience with laparoscopy and peritoneal cytology in the staging of pancreatic cancer. *Br. J. Surg.* **82**, 1127–1129.

Fuhrman, G. M., Charnsangavej, C., Abbruzzese, J. L., Cleary, K. R., Martin, R. G., Fenoglio, C. J., and Evans, D. B. (1994). Thin-section contrast-enhanced computed tomography accurately predicts the resectability of malignant pancreatic neoplasms. *Am. J. Surg.* **167**, 104–111, discussion 111–113.

Hahn, S. A., Schutte, M., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., and Kern, S. E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353.

Moossa, A. R., and Gamagami, R. A. (1995). Diagnosis and staging of pancreatic neoplasms. *Surg. Clin. North Am.* **75**, 871–890.

Povoski, S. P., Karpeh, M. S., Jr., Conlon, K. C., Blumgart, L. H., and Brennan, M. F. (1999). Association of preoperative biliary drainage with postoperative outcome following pancreaticoduodenectomy [see comments]. *Ann. Surg.* **230**, 131–142.

Safi, F., Schlosser, W., Falkenreck, S., and Beger, H. G. (1998). Prognostic value of CA 19–9 serum course in pancreatic cancer. *Hepatogastroenterology* **45**, 253–259.

Spitz, F. R., Abbruzzese, J. L., Lee, J. E., Pisters, P. W., Lowy, A. M., Fenoglio, C. J., Cleary, K. R., Janjan, N. A., Goswitz, M. S., Rich, T. A., and Evans, D. B. (1997). Preoperative and postoperative chemoradiation strategies in patients treated with pancreaticoduodenectomy for adenocarcinoma of the pancreas [see comments]. *J. Clin. Oncol.* **15**, 928–937.

Yeo, C. J., Cameron, J. L., Sohn, T. A., Lillemoe, K. D., Pitt, H. A., Talamini, M. A., Hruban, R. H., Ord, S. E., Sauter, P. K., Coleman, J., Zahurak, M. L., Grochow, L. B., and Abrams, R. A. (1997). Six hundred fifty consecutive pancreaticoduodenectomies in the 1990s: Pathology, complications, and outcomes. *Ann. Surg.* **226**, 248–257; discussion 257–260.

Pancreatic Enzyme Secretion (Physiology)

CRAIG D. LOGSDON
University of Michigan, Ann Arbor

acetylcholine The major neurotransmitter of the parasympathetic nervous system.
cholecystokinin A gastrointestinal peptide hormone with multiple functions regulating pancreatic secretion and gastrointestinal function.
cholecystokinin-releasing peptide A trypsin-sensitive peptide whose presence in the lumen of the intestine leads to the release of cholecystokinin from intestinal enteroendocrine cells.
chyme The mixture of ingested foodstuffs with gastric secretions.
dorsal motor nucleus of the vagus A region of the brainstem that contains nuclei for vagal efferent neurons.
enteroendocrine cells Endocrine cells found throughout the epithelium of the gastrointestinal tract.
G-proteins Guanine nucleotide-binding proteins.
protein kinases Enzymes that add phosphate groups to proteins and thereby regulate their functions.
secretin A gastrointestinal peptide hormone that stimulates bicarbonate and fluid secretion from pancreatic ducts.
vagus nerve The 10th cranial nerve, which innervates much of the digestive system, including the pancreas.
zymogen granule The membrane-bound granules within pancreatic acinar cells that store digestive enzymes.

Pancreatic secretion occurs in various "phases" corresponding to the demand for pancreatic juice. The major phases are the (1) interdigestive, (2) cephalic, (3) gastric, (4) intestinal, and (5) humoral phases. The interdigestive phase is that period between meals when there is little need for pancreatic juice. The cephalic phase occurs when there is anticipation of food, perhaps due to smell, taste, or habitual behavior associated with feeding. The gastric phase begins with the ingestion of food and corresponds to that period of time when the food resides within the stomach, before entering the small intestine. The intestinal phase is the time during which the food resides within the intestine; this phase corresponds to the period when there is peak demand for pancreatic juice. The humoral phase describes the time after significant digestion has taken place and nutrient absorption has occurred, but while active digestive processes are still present. Specific regulatory mechanisms are responsible for regulating pancreatic secretion during these different phases and these regulatory mechanisms are the primary focus of this article.

INTRODUCTION

The exocrine pancreas produces a bicarbonate-rich secretion termed "pancreatic juice" that contains abundant digestive enzymes, proteins specialized for the molecular disassembly of complex organic constituents. Pancreatic juice is necessary for the proper digestion of ingested foodstuffs. Relative to its weight, the pancreas secretes more protein than any other organ. Without proper digestion, food cannot be absorbed and hence insufficient pancreatic secretion can lead to malnutrition. Furthermore, undigested food that reaches the colon promotes bloating, gas, and diarrhea. Thus, it is important that pancreatic secretion matches food intake.

Pancreatic juice originates from the exocrine pancreas, which is composed of numerous acini, cluster-like groups of acinar cells, and associated ducts. The acinar cells are responsible for the synthesis, storage within zymogen granules, and regulated release of the digestive enzymes and may also contribute to the fluid component of pancreatic juice. The duct cells provide a conduit for passage of pancreatic juice out of the pancreas and are also the primary site of pancreatic bicarbonate and fluid secretion. Because the major constituents of pancreatic juice, digestive enzymes and bicarbonate-rich fluid, originate in two different cell types, it is possible experimentally to separate the factors that influence each component. However, physiologically these components are not separated but occur in a coordinated manner with greater or lesser proportions of each being secreted depending on the demand. In addition to these two basic cell types that produce pancreatic juice, there are other important influences on pancreatic secretion. One such influence is the sphincter of Oddi. This sphincter, located at the entrance of the common bile/pancreatic duct into the intestine, regulates the flow of juice from the pancreas into the intestine and is influenced by nervous and hormonal inputs. Another influence on pancreatic secretion is pancreatic blood flow. Secretion is an energy-consuming process such that increased blood flow, supplying oxygen to the tissue, is required during

 Copyright 2004, Elsevier (USA). All rights reserved.

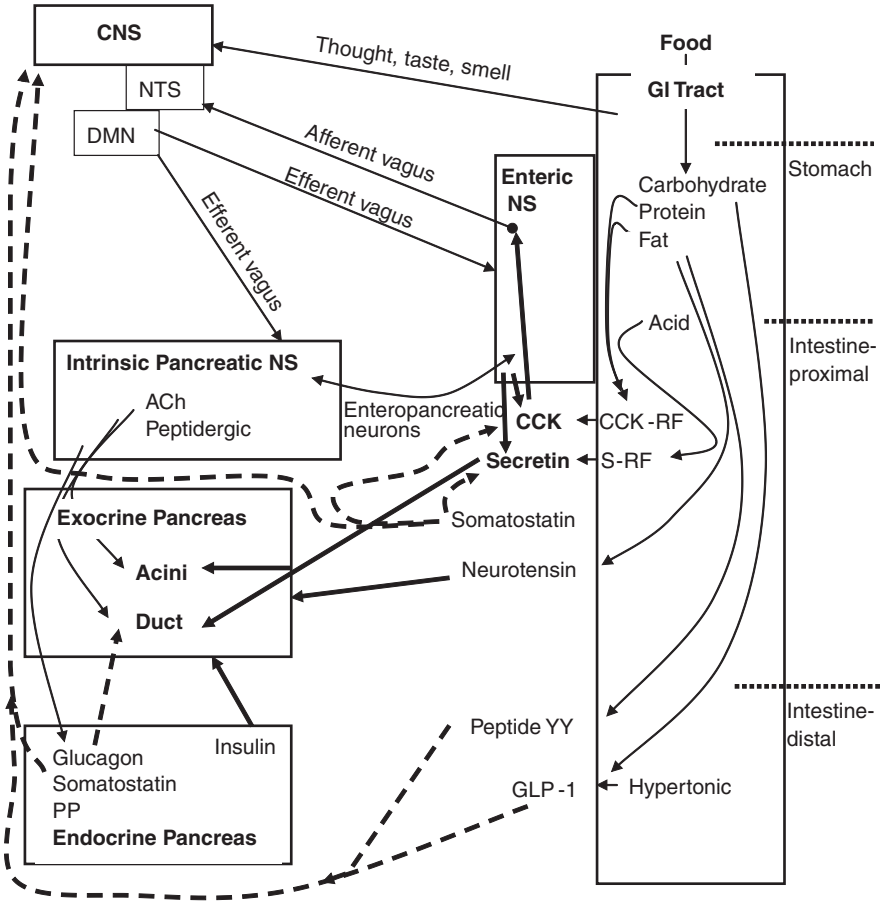


FIGURE 1 Regulation of exocrine pancreatic secretion. Pancreatic secretion of protein from acinar cells and of fluid and bicarbonate from duct cells is regulated by cholinergic (ACh) and peptidergic neurons within the intrinsic pancreatic neural network. Activity of the intrapancreatic nerves is modulated by extrinsic innervation through the vagus nerve originating in the brainstem. Vagal signals are influenced by information arising from sensory input both within the vagus and from other central nervous system (CNS) centers that are coordinated in the nucleus of the solitary tract (NTS). Pancreatic secretion is also modulated both positively (solid arrows) and negatively (dashed arrows) by hormones. Hormones arise both from the endocrine pancreas and from enteroendocrine cells within the epithelium of the intestine. Hormones can have direct effects or influence the nervous system. CCK and secretin are the major hormonal regulators of pancreatic secretion. CCK influences pancreatic secretion in humans by activating afferent neurons within the enteric nervous system. Secretin acts directly on pancreatic duct cells and synergizes with the effects of ACh on acinar cells. Somatostatin is an inhibitory substance that appears to act as a hormone on the CNS and also as a paracrine inhibitor of CCK and secretin release. Hormone-secreting cells and nerves within the gastrointestinal (GI) tract respond to the presence of specific components of chyme within specific compartments of the GI tract. The release of CCK and secretin appears to be regulated by releasing factors originating from cells within the GI tract. Signals from proximal portions of the GI tract tend to be stimulatory, whereas those from distal portions tend to be inhibitory.

active secretion and factors that diminish pancreatic blood flow reduce pancreatic secretion.

GENERAL ISSUES CONCERNING THE INTEGRATIVE PHYSIOLOGIC REGULATION OF PANCREATIC ENZYME SECRETION

Pancreatic secretion is influenced by many types of regulators, including hormones, neurotransmitters, and paracrine effectors. Figure 1 summarizes the roles of several of the major regulatory factors that will be discussed in this article. As noted in Fig. 1, the effects of regulatory substances on pancreatic secretion can occur either directly, by effects on acinar or duct cells, or indirectly, by effects on afferent neurons, the central nervous system (CNS), or enteroendocrine cells. Regulatory substances can also have interactions with one another that can be antagonistic, additive, or synergistic.

Many factors are able to influence pancreatic secretion experimentally. However, not all factors that are able to affect secretion do so in a normal physiological setting. Thus, it becomes important to distinguish between those that normally regulate pancreatic secretion, generally termed "physiological," and those that influence secretion only under experimental conditions, often termed "pharmacological." In order for a regulatory substance to be considered physiological, it must meet certain criteria. First, the concentration of the substance must increase in the physiologically relevant compartment during the appropriate phase of digestion. For a hormone, that means that plasma levels, which are relatively easy to monitor, must be elevated. For neurotransmitters or paracrine regulators, the relevant compartment is much less accessible and determination of relevant concentrations can be difficult. Second, concentrations of the candidate regulator that occur naturally must be able to influence pancreatic secretion. Again, many substances can affect pancreatic secretion at high concentrations that do not influence pancreatic secretion in normal circumstances. This criterion requires knowledge of the physiologic concentration, which is again easier for hormones than for neurotransmitters or paracrine substances. Furthermore, some regulators have important synergistic interactions with other regulators. In this case, administration of one without the other may not be sufficient to elicit a response. The third criterion is that inhibition of the regulator should influence pancreatic secretion during a normal meal. Experimental strategies to test this may include pharmacological, immunological, and genetic approaches. Each of these experimental approaches has its caveats. However, when several approaches verify the same role, then strong conclusions can be made. Similarly, it may not be possible to fulfill all three of these criteria for a specific molecule, but the more criteria that can be fulfilled, the stronger the argument for a physiologic role.

INTEGRATIVE REGULATORS OF PANCREATIC SECRETION

Neural Mechanisms

Intrinsic Pancreatic Neurons and the Enteropancreatic Neural Reflex

Intrapancreatic postganglionic cholinergic and noncholinergic neurons directly influence acinar and duct cell secretory activity. These neurons are activated by the central nervous system via the vagus nerve. Important vagovagal reflexes occur during gastric and intestinal phases of pancreatic secretion in which both afferent and efferent nerves are carried by the vagus. The vagal afferent nerve terminals in the stomach and duodenum are responsive to cholecystokinin (CCK), leptin, serotonin, interleukin-1B, and mechanical stimuli and communicate with the central nervous system. Direct neural connections have also been recently described between neurons in ganglia of the myenteric plexi of the stomach and duodenum and the intrapancreatic plexus. These enteropancreatic neural pathways have cholinergic and serotonergic components.

The Central Nervous System Controls Pancreatic Secretion through the Autonomic Nervous System

During the cephalic phase of pancreatic secretion, information from olfactory, visual and other inputs is integrated at the level of the nucleus of the solitary tract (NTS), which is located in the medulla or brainstem. Similarly, during later phases of pancreatic secretion, information arriving from vagal afferent neurons is processed in the NTS. The NTS projects to the efferent vagal neurons in the dorsal motor nucleus of the vagus. Central stimulants of vagally mediated pancreatic secretion remain to be clearly identified. However, central administration of orexin-A and thyrotropin-releasing hormone stimulates pancreatic juice flow. Central inhibitors of vagal efferent neurons include pancreastatin and pancreatic polypeptide.

The parasympathetic nervous system is the major controller of pancreatic secretion, with the vagus nerve being the primary source of parasympathetic

innervation. Activation of parasympathetic activity stimulates pancreatic fluid and enzyme secretion. In contrast, the sympathetic nervous system primarily plays an inhibitory role in regulating pancreatic secretion. Sympathetic innervation occurs chiefly through the splanchnic nerves. Splanchnic neural activation appears to inhibit pancreatic secretion largely through its vasoconstrictive effects on pancreatic blood flow.

Neurotransmitters Regulating Pancreatic Secretion

The most important stimulant of pancreatic secretion is acetylcholine released from the nerve termini of intrapancreatic neurons, which acts directly on both acinar and duct cells through the occupation of muscarinic cholinergic receptors (m3 receptors). Activation of these receptors leads to increased cellular signaling events, which result in the exocytotic release of digestive enzymes from the pancreatic acinar cell and increased bicarbonate and fluid secretion from the duct cells. Thus, inhibition of cholinergic nervous transmission, for example, by administration of atropine, inhibits both fluid and enzyme secretion.

There are also direct actions of noncholinergic, peptidergic neurotransmitters released from intrapancreatic neurons. Gastrin-releasing peptide, vasoactive intestinal peptide (VIP), and pituitary adenylate cyclase-activating polypeptide are stored in and released from intrapancreatic neurons and can act as stimulants of pancreatic secretion. These effects are more pronounced in certain animal species. Somatostatin, peptide tyrosine tyrosylamide (PYY), and enkephalins are examples of peptides that act as inhibitors of pancreatic secretion. Few of these, however, act directly on acinar cells and most inhibit the CNS or act at the level of intrapancreatic ganglia. The specific roles of these peptidergic neurotransmitters has been difficult to determine because of significant interactions between regulators and species differences in the activity of the regulators.

Hormonal Mechanisms

Pancreatic secretion is also regulated by the actions of a number of gastrointestinal hormones that originate from endocrine cells within the mucosa of the gastrointestinal tract and are therefore referred to as "enteroendocrine" cells. These hormones are released into the blood on ingestion of a meal. Individual hormones are regulated by specific components of the chyme, including both ingested foodstuffs and components secreted into the gut by mucosal cells and accessory organs. Hormones travel through the blood and can have direct effects on pancreatic acinar or duct cells via high-affinity receptors for the hormones or they may interact with afferent neurons to regulate local neural mechanisms or act on central neural sites.

Cholecystokinin

CCK is the major peptide hormone that regulates pancreatic protein secretion. CCK also influences the flow of secretions into the small intestine by relaxing the sphincter of Oddi and contracting the gallbladder. In addition, CCK reduces gastric emptying. Taken together, these actions of CCK regulate the movement of both food and digestive secretions into the small intestine. CCK also supports pancreatic secretion through its tropic effects on the pancreas. It stimulates the synthesis of new digestive enzymes and maintains the growth of the pancreas. In various species, the pancreas atrophies in the absence of intralumenal nutrients and will grow under conditions in which plasma CCK is chronically elevated, as occurs following administration of trypsin inhibitor.

CCK is produced and secreted by a specific class of intestinal enteroendocrine cells (I cells). It is released by hydrolytic products of digestion including amino acids and fatty acids. The mechanism by which these nutrients induce CCK release is not fully understood. The presence of active trypsin in the intestinal lumen inhibits CCK release in many species and ingestion of trypsin inhibitors causes high and sustained levels of plasma CCK. Diversion of the pancreatic juice from the duodenum also increases CCK plasma levels. Taken together, these observations suggest that feedback inhibition of CCK release occurs. The "feedback hypothesis" suggests that proteins bind or inhibit intralumenal endopeptidases, which would otherwise inactivate a CCK-releasing peptide. Several candidate molecules that have the ability to increase CCK release when infused into the small intestine have been identified. However, the physiological significance of these CCK-releasing peptides remains to be elucidated.

CCK plays a central role in the stimulation of pancreatic secretion during a meal. Under fasting conditions, the plasma CCK levels are very low ($\sim$1 pmol/liter in humans). After ingestion of a typical meal, the concentration increases to 6–8 pmol/liter within 10 to 30 min, followed by a gradual decline to basal levels over $\sim$3 h. Infusion of similar doses of CCK produces the same levels of pancreatic enzyme secretion. Furthermore, administration of CCK antagonists produces a 50–60% reduction of meal-stimulated pancreatic secretion.

The mechanism of CCK stimulation of pancreatic secretion appears to be somewhat determined by the species. In rodents, receptors exist on pancreatic acinar cells that can respond directly to CCK within physiologic concentrations. However, in humans, these receptors are not expressed on pancreatic acinar cells. Therefore, it appears that the major mechanism of CCK stimulation of pancreatic secretion is via an indirect mechanism that involves activation of receptors on vagal afferent neurons that are present within the enteric nervous plexus in the intestinal mucosa. These neurons communicate with the central nervous system and a stimulatory signal is conveyed back to the pancreas via vagal efferents to mediate secretion that is dependent on cholinergic signaling. The major evidence in support of this model is that the cholinergic antagonist atropine blocks pancreatic secretion in response to infusion of physiological doses of CCK. There may also be effects of CCK on enteropancreatic neurons that do not require an intact vagus nerve, as CCK can stimulate pancreatic secretion in patients after vagotomy. However, this remains unproven.

Secretin

Secretin is the intestinal peptide hormone that is the major regulator of pancreatic fluid and bicarbonate secretion. Secretin is released from intestinal enteroendocrine cells (S cells) in response to a reduction in duodenal pH. Acidic chyme entering from the stomach leads to secretin release and subsequent pancreatic bicarbonate secretion, which neutralizes gastric acid. The mechanisms responsible for secretin release remain under active investigation, but there is some evidence for a secretin-releasing factor that may be released by low-pH conditions. Nonacid factors may also play a role in secretin release. Bile, as well as fatty acids and other digestive products of fat, can increase plasma secretin levels. However, the physiologic importance of these nonacid factors is questionable, as plasma secretin does not increase in subjects in whom meal-induced acid secretion is neutralized.

Infusion of secretin at concentrations similar to those observed after a meal leads to the stimulation of pancreatic fluid and bicarbonate secretion, and neutralization of secretin, using antiserum, greatly reduces pancreatic bicarbonate release. Secretin activates specific receptors expressed on pancreatic duct and acinar cells. Thus, secretin appears to directly mediate its effects on pancreatic secretion. Secretin appears to act in synergy with CCK, as the combination of these hormones leads to much greater levels of pancreatic secretion than either one alone.

Other Stimulatory Hormones

A variety of other hormones have been reported to have stimulatory effects on pancreatic secretion. Insulin potentiates the secretory response to CCK and secretin, an observation that may explain why enzyme secretion is frequently reduced in human diabetics who otherwise do not exhibit overt pancreatic disease. The influences of insulin on exocrine secretion may also help to explain the location of the insulin-secreting cells that are within the islets of Langerhans, which are dispersed throughout the exocrine pancreas. Insulin is also critical for maintaining rates of acinar cell protein synthesis and adequate stores of digestive enzymes.

Other hormones that may be stimulatory for pancreatic secretion include gastrin, neurotensin, and motilin. Gastrin, which is structurally related to CCK, can also stimulate pancreatic secretion when infused at relatively high concentrations. However, it is unclear whether or not the serum levels of gastrin obtained after a meal are sufficient for pancreatic stimulation. Neurotensin is released from enteroendocrine cells by intestinal fatty acids and infusion of neurotensin can stimulate pancreatic secretion. However, the concentrations of neurotensin required to stimulate pancreatic secretion are higher than those normally observed after a meal. Therefore, it is unclear whether or not neurotensin is a physiological stimulant of pancreatic secretion. Motilin is a hormone that is known to regulate the cyclic interdigestive migrating motor complex that influences intestinal motility. Bolus injection of motilin also results in a transient increase in pancreatic secretion. Thus, it has been suggested that motilin may regulate the cyclic secretion of pancreatic juice during the interdigestive state.

Inhibitory Hormones

After a meal, pancreatic secretory levels decline and return to a low basal level. Much of this reduction in pancreatic secretion is due to a decrease in the levels of stimulatory signals. However, there is also evidence for inhibitory control. For example, infusion of glucose or amino acids to raise serum concentrations inhibits pancreatic secretory responses to a test meal. The release of inhibitory hormones from the islets of Langerhans, distal small intestine, and colon has been postulated to account for this observation. The most well-established inhibitory hormones include glucagon and its related peptides, as well as somatostatin, and PYY.

Glucagon and the glucagon-related molecules oxyntomodulin and glucagon-like peptide-1 (GLP-1) inhibit pancreatic secretion stimulated by secretin and CCK or by ingestion of a test meal. The inhibitory effect

includes a reduction of fluid and bicarbonate as well as enzyme secretion. Glucagon is released from pancreatic islets of Langerhans by the hyperaminoacidemia observed after a high-protein meal. Oxyntomodulin is a 37-amino-acid glucagon-containing peptide that is 10 times more potent than pancreatic glucagon in terms of its ability to reduce pancreatic secretion. Oxyntomodulin originates in enteroendocrine cells located in the distal intestine and its release is stimulated by hypertonic solutions. GLP-1 is another GI hormone derived from the glucagon precursor whose cells of origin (K cells) are located within the epithelium of the more distal small intestine. GLP-1 appears to be secreted in response to carbohydrates within the gut lumen. Glucagon, oxyntomodulin, and GLP-1 appear to inhibit pancreatic secretion via action on a central vagal site.

Somatostatin is another candidate as a physiological inhibitor of pancreatic secretion. Somatostatin is produced and secreted by both delta cells of the islets of Langerhans and enteric endocrine cells. Infusion of somatostatin inhibits CCK-stimulated pancreatic enzyme secretion. However, the concentrations required for this effect are higher than those observed in the serum under physiologic conditions. Furthermore, the mechanisms involved in the inhibitory effects of somatostatin remain uncertain. Somatostatin does not seem to exert effects on vagal afferent or efferent pathways but instead has effects on a central vagal site where it can reduce the effects of CCK. Somatostatin also has been reported to reduce the release of secretin and CCK from enteroendocrine cells and it is therefore likely that somatostatin also has a paracrine effect on CCK and secretin release.

PYY and pancreatic polypeptide are also inhibitory hormones. PYY is a 36-amino-acid peptide that is present in the distal small intestine, colon, and rectum and is released by fat and protein in the distal gut. Pancreatic polypeptide (PP) is closely related to PYY, but PP is localized in the islets of Langerhans. The only apparent physiologic actions of PP are the inhibition of pancreatic and biliary secretion. PP secretion is regulated by a cholinergic mechanism. Infusion of either PYY or PP inhibits meal-stimulated pancreatic secretions. These effects appear to be mediated by an influence on central vagal regulation of pancreatic secretion.

CELLULAR MECHANISMS REGULATING PANCREATIC SECRETION

Receptors

As described above, hormones and neurotransmitters that stimulate pancreatic secretion may do so by directly regulating acinar and duct cells or by regulating these cells indirectly through nerves or blood vessels. Determination of the physiologic regulatory pathways can be aided by analysis of effects on isolated acinar and duct cells and localization of receptors for each regulator on its target cell. For pancreatic acinar cells, the ability to stimulate amylase secretion *in vitro* is useful for determining the direct effects of agonists and antagonists. The presence of specific receptors has also been confirmed by binding studies with radiolabeled analogues and antagonists. Receptors can also be localized using electron microscopic autoradiography and confocal fluorescence microscopy. Preparations of isolated duct segments or cultured monolayers of duct cells that can be utilized to investigate direct interactions with regulatory molecules have also been developed.

Acinar cells have been found to bear receptors for CCK, bombesin, acetylcholine (m3 muscarinic), VIP, and secretin. However, important differences exist between species. For example, rodents, but not humans, express CCK_1 receptors. CCK interacts with both CCK_1 (previously CCK_A) receptors, which are highly specific for CCK, and CCK_2 (gastrin, previously CCK_B) receptors, which respond to both CCK and gastrin. Much is known about CCK_1 and m3 receptors and their signaling, because of their presence on rodent acinar cells. CCK receptors on afferent nerves appear to be of the CCK_1 type and have properties similar to those on acinar cells. It is more difficult to study the receptors regulating the pancreatic duct cells because of the relatively small number of these cells and the difficulty in studying their physiologic functions (e.g., ion transport) *in vitro*. However, studies have indicated the presence of receptors for secretin, ATP, CCK, VIP, and acetylcholine on these cells.

Transmembrane Signaling

All major secretagogue receptors are members of the seven hydrophobic transmembrane domain family of receptors that interact with guanine nucleotide-binding proteins (G-proteins) to activate intracellular signaling. The secretagogue receptors couple to the activation of heterotrimeric G-proteins that are composed of three subunits (α, β, and γ) and are usually defined on the basis of their α-subunit. The G-proteins primarily responsible for the stimulation of pancreatic acinar cell secretion are members of the G_q family, which includes G_q and G_{11}. These G-proteins regulate secretion through interactions with a phosphoinositide specific phospholipase C, leading to increases in diacylglycerol and inositol trisphosphate and ultimately protein kinase C and Ca^{2+}. Other secretagogues, such as secretin, activate G_s proteins. G_s is coupled to the activation of adenylate

cyclase and increases intracellular cyclic AMP (cAMP) and protein kinase A activity. G_s-coupled receptors are weak stimulators of acinar cell secretion but have synergistic effects when combined with receptors that activate the G_q/Ca^{2+} pathways.

It should also be noted that secretagogue receptors may have other important actions on pancreatic acinar cells, such as tropic effects or influences on enzyme gene expression, and these effects may involve other signaling pathways. CCK has been reported to activate a wide variety of effector proteins including phospholipase C, phospholipase A2, phospholipase D, protein kinase A, protein kinase C, phosphatidylinositol 3-kinase, focal adhesion kinase, tyrosine kinases, several mitogen-activated protein kinases, and stress-activated protein (SAP) kinase and the nuclear oncogenes c-fos, c-myc, and c-jun.

Mechanism of Action of Intracellular Messengers

The mechanisms by which increases in intracellular messengers act to induce protein and fluid secretion are not completely understood. It is generally held that second messengers exert their effects by changes in the phosphorylation of regulatory proteins. In addition to protein kinase C and protein kinase A (PKA), which were previously mentioned, several Ca^{2+}-activated kinases have been identified in pancreatic acinar cells including Ca^{2+}/calmodulin-activated type II and type III kinase and myosin light chain kinase. Pancreatic acinar cells also contain the major classes of serine/threonine phosphatases (i.e., PP1, PP2A, and PP2B), which may also be important for the regulation of pancreatic secretion. One of these, PP2B or calcineurin, is activated by the Ca^{2+}/calmodulin complex.

The role of intracellular messengers and effectors in pancreatic enzyme secretion is summarized in Fig. 2. Simulation of secretion normally involves synergistic interactions among intracellular messengers. In the case of acetylcholine and CCK, this includes interactions between Ca^{2+} and diacylglycerol-activated pathways. Agents such as VIP and secretin, which increase cAMP, add a further interaction at the intracellular effector level. Proteins localized on the granule and lumenal plasma membrane and several soluble and cytoskeletal proteins may be involved in exocytosis. In pancreatic duct cells, the same intracellular messengers and kinases regulate ion pumps, carriers, and channels involved in fluid and electrolyte secretion. In this case, the primary intracellular messenger is cAMP, which activates PKA, which phosphorylates cystic fibrosis transmembrane regulator, the ion channel involved in anion egress from duct cells.

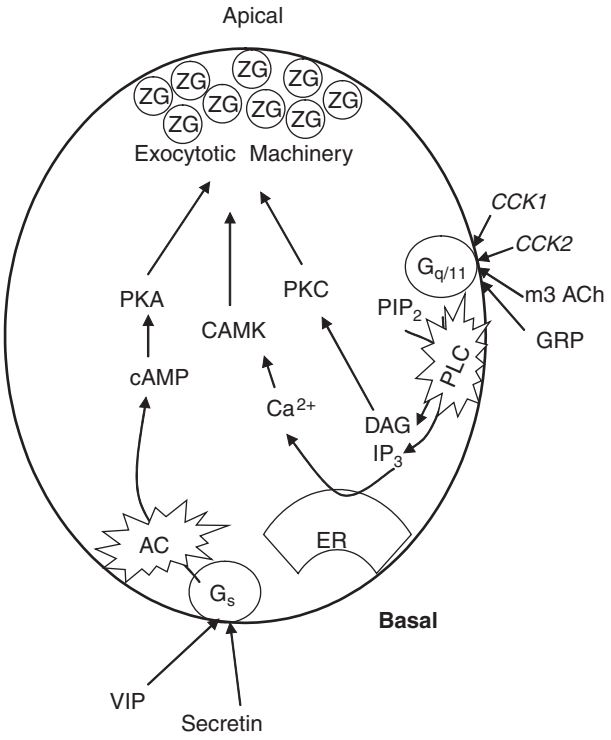


FIGURE 2 Pancreatic acinar cell stimulus–secretion coupling. The pancreatic acinar cell expresses receptors for a variety of molecules that stimulate secretion. Specific high-affinity receptors including muscarinic cholinergic (m3 ACh) and gastrin-related peptide (GRP) receptors, as well as in some species either CCK1 or CCK2 receptors or both, activate G-proteins of the G_q/G_{11} type to activate a phospholipase C (PLC), which converts phosphotidylinositide 1,4-bisphosphate (PIP_2) into diacyglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). IP_3 subsequently activates specific receptors on the endoplasmic reticulum (ER) intracellular Ca^{2+} stores, causing an increase in cytoplasmic Ca^{2+}. Increased Ca^{2+} activates Ca^{2+}-dependent kinases including calmodulin kinase (CAMK). DAG activates protein kinase C (PKC). Other stimulatory molecules, such as the neurotransmitter vasoactive intestinal peptide (VIP) and the hormone secretin, activate specific receptors that are coupled to G-proteins of the G_s type. Activation of G_s stimulates adenylate cyclase (AC) production of cyclic AMP (cAMP). Increased cAMP levels activate protein kinase A (PKA). Increased activity of the various kinases leads to activation of the exocytotic machinery, which includes membrane fusion-related molecules including Rab and soluble *N*-ethylmaleimide-sensitive attachment protein receptor (SNARE) proteins existing on zymogen granule (ZG) membranes and the apical plasma membrane, as well as cytoskeletal components including actin and microtubule networks.

Zymogen Granule Exocytosis

The terminal steps in protein secretion involve exocytotic release of enzymes from zymogen granules. This involves a membrane fusion event between zymogen granules and the apical plasma membrane. Membrane

fusion involves a number of specific membrane-associated proteins of the soluble *N*-ethylmaleimide-sensitive attachment protein receptor (SNARE) family and Rab families. The actin cytoskeleton is also involved in exocytosis, serving alternately as a barrier, preventing premature secretion, and as a potential motile element. Microtubules are also involved in the movement of zymogen granules to the apical portion of the cell, where they can interact with the exocytotic machinery.

SUMMARY

In summary, pancreatic exocrine secretion is regulated by a complex interplay of nervous, hormonal, and paracrine regulation that matches fluid, bicarbonate, and enzyme levels to meet physiologic demands. Complex interactions occur between multiple regulatory molecules and signaling pathways. The relative contribution of various components to the overall secretory function varies with the phase of secretion. Significant species differences occur with regard to receptor localization and the importance of different arms of the regulatory mechanisms.

See Also the Following Articles

Cholecystokinin (CCK) • Enteroglucagon • Exocytosis • Gastrin • Gastrin-Releasing Peptide • Pancreatic Bicarbonate Secretion • Pancreatic Digestive Enzymes • Pancreatic Polypeptide Family • Pituitary Adenylate Cyclase Activating Peptide (PACAP) • Secretin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Chey, W. Y., and Chang, T. (2001). Neural hormonal regulation of exocrine pancreatic secretion. *Pancreatology* **1**, 320–335.

Lee, M. G., Ahn, W., Lee, J. A., Kim, J. Y., Choi, J. Y., Moe, O. W., Milgram, S. L., Muallem, S., and Kim, K. H. (2001). Co-ordination of pancreatic HCO_3^- secretion by protein–protein interaction between membrane transporters. *J. Pancreas* **2**, 203–206.

Logsdon, C. D. (1999). Role of cholecystokinin in physiologic and pathophysiologic growth of the pancreas. *In* "Gastrointestinal Endocrinology" (G. H. Greeley, Jr., ed.), Chap. 15, pp. 393–422. Humana Press, Totowa, NJ.

Naruse, S., Kitagawa, M., Ishiguro, H., and Hayakawa, T. (2002). Feedback regulation of pancreatic secretion by peptide YY. *Peptides* **23**, 359–365.

Niebergall-Roth, E., and Singer, M. V. (2001). Central and peripheral neural control of pancreatic exocrine secretion. *J. Physiol. Pharmacol.* **52**, 523–538.

Owyang, C. (1996). Physiological mechanisms of cholecystokinin action on pancreatic secretion. *Am. J. Physiol.* **271**, G1–G7.

Owyang, C., and Williams, J. A. (1999). Pancreatic secretion. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), Chap. 16, pp. 355–379. Lippincott, Williams & Wilkins, Philadelphia, PA.

Solomon, T. E. (1994). Control of exocrine pancreatic secretion. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Chap. 41, pp. 1499–1530. Raven Press, New York.

Spannagel, A. W., and Green, G. M. (1999). Pancreatic hypersecretion in the jejunal-bypass rat. *Pancreas* **18**, 47–52.

Wasle, B., and Edwardson, J. M. (2002). The regulation of exocytosis in the pancreatic acinar cell. *Cell Signal.* **14**, 191–197.

Williams, J. A. (2001). Intracellular signaling mechanisms activated by cholecystokinin-regulating synthesis and secretion of digestive enzymes in pancreatic acinar cells. *Annu. Rev. Physiol.* **63**, 77–97.

Yule, D. I., and Williams, J. A. (1994). Stimulus-secretion coupling in the pancreatic acinus. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Chap. 39, pp. 1447–1472. Raven Press, New York.

Pancreatic Function Tests

MATTHEW J. DIMAGNO
University of Michigan Medical School, Ann Arbor

bentiromide test Tubeless pancreatic function test that indirectly measures pancreatic chymotrypsin output using bentiromide as a substrate.
cholecystokinin Gastrointestinal hormone that stimulates secretion of digestive enzymes by pancreatic acinar cells.
chronic pancreatitis Clinical diagnosis reflecting permanent or progressive deterioration in pancreatic function or structure or both.
fecal pancreatic elastase I Tubeless pancreatic function test that directly measures fecal pancreatic elastase I in stool.
malabsorption Impaired intestinal absorption of micronutrients due to either intestinal disease or maldigestion.
maldigestion Impaired digestion of macronutrients (i.e., fat), as in pancreatic insufficiency.
secretin Gastrointestinal hormone that stimulates secretion of bicarbonate and water by pancreatic ductal cells.

The exocrine function of the human pancreas is to secrete digestive enzymes that aid in absorption of dietary nutrients. Tests to measure pancreatic secretions can be invasive, requiring insertion of tubes to collect pancreatic fluids, or noninvasive, requiring oral administration of compounds and subsequent assays of stool, blood, urine, or breath components.

INTRODUCTION

Pancreatic function tests are used to measure pancreatic secretion in humans. The tests are clinically useful to discriminate between pancreatic and nonpancreatic causes of malabsorption. Pancreatic malabsorption occurs when residual pancreatic enzyme secretion is only 5–10% of normal levels, as occurs in severe chronic pancreatitis, when tests of pancreatic function are frequently abnormal. Both invasive and noninvasive tests exist. The most sensitive and specific tests, the secretin and cholecystokinin (CCK) stimulation tests, are invasive in that they require placement of a gastroduodenal tube; they are also time-consuming (2–3 hours) and are performed in only a few medical centers. "Tubeless" tests are more efficient, but, unfortunately, their clinical use is limited by their relative insensitivity to detecting mild impairments in pancreatic function, as in early chronic pancreatitis, when the secretin or CCK test may be abnormal.

TESTS REQUIRING A GASTRODUODENAL TUBE

Secretin and cholecystokinin are gastrointestinal hormones that may be given alone or together to stimulate and measure pancreatic secretion directly. This protocol, which is primarily used in research, requires placement of a double-lumen tube (Dreiling tube), which is anatomically positioned using fluoroscopy. The gastric aspiration port is positioned in the distal stomach for removal of contaminating gastric juice, while the duodenal aspiration port is positioned in the first, second, and third segments of the duodenum for removal of pancreatic juice under continuous suction. Juice removed from the duodenal port is used for analysis of pancreatic secretion, including measurements of pancreatic juice volume and bicarbonate concentration for the secretin test and trypsin and lipase output for the CCK test. For example, either in response to meals (the Lundh test) or to intravenous secretin administration, both pancreatic juice flow and juice bicarbonate concentration increase. A peak bicarbonate fluid concentration of less than 80 mEq/liter of bicarbonate indicates impaired pancreatic secretory function, suggestive of chronic pancreatitis. However, a positive test (and impaired pancreatic secretion) may occur in the absence of clinical or radiologic features of chronic pancreatitis, in conditions such as diabetes mellitus, protein-calorie malabsorption, gastric surgery, truncal vagotomy, celiac sprue, and hepatic cirrhosis.

TUBELESS TESTS

Noninvasive tubeless pancreatic function tests indirectly measure pancreatic function by relying on two bioassay principles. Patients with chronic pancreatitis may produce insufficient pancreatic digestive enzymes (e.g., lipase, chymotrypsin, amylase, elastase, and trypsin), causing maldigestion and poor absorption of foods (e.g., fat, protein, and carbohydrates), which may be measured in the stool. Alternatively, oral administration of a synthetic compound that is acted on by a specific pancreatic digestive enzyme will generate a product that may be measured in stool, blood, breath, or urine.

 Copyright 2004, Elsevier (USA). All rights reserved.

Diminished recovery of the product suggests impaired pancreatic enzymatic activity. Insensitivity and failure to detect mild disease is a problem common to all tubeless tests; bioassays in these protocols are able to distinguish reliably only among patients with moderate and severe pancreatic insufficiency. In addition, these tests are not always specific to pancreatic function; patients with normal pancreatic secretion may have a positive test result if they have hepatobiliary disease or small bowel diseases causing malabsorption (i.e., celiac sprue, Crohn's disease, or Whipple's disease) or have had prior gastric surgery.

As illustration of a tubeless test, the bentiromide test indirectly measures pancreatic chymotrypsin output. Endogenous chymotrypsin, a pancreatic digestive enzyme, cleaves orally administered bentiromide (*N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid), liberating *para*-aminobenzoic acid (PABA), which is absorbed, hepatically metabolized, and excreted in urine. Decreased recovery occurs in patients with inadequate pancreatic function, as in chronic pancreatitis. Many medications interfere with urinary measurements of free PABA and its metabolites and should be discontinued 3 days prior to the test. Because renal dysfunction lowers urinary recovery of PABA, measurement of PABA in serum rather than in urine has been developed with similar test sensitivity. An alternative but similar tubeless test is the fluorescein dilaurate (pancreolauryl) test, which measures nonspecific lipase (cholesterol esterase) activity and is slightly more sensitive than the bentiromide test.

Measurement of fecal pancreatic elastase I is the most commonly used tubeless test in the United States and Europe but is nonetheless controversial, because it has not been shown to be consistently sensitive or specific for mild or moderately impaired pancreatic secretion. The test is simple, relatively inexpensive, and involves measuring human fecal pancreatic elastase by enzyme-linked immunosorbent assay (ELISA). Unlike other pancreatic digestive enzymes, fecal pancreatic elastase levels do not change during intestinal transit, are stable in stool samples for up to 1 week, and are unaffected by concurrent use of oral pancreatic enzyme replacement therapy.

SUMMARY

Tubeless pancreatic function tests are useful for diagnosis of chronic pancreatitis in the setting of malabsorption, when the pancreas is severely damaged. In the setting of mild to moderate disease (when malabsorption is not present), the clinical utility of these tests is less well defined due to poor test sensitivity. In this situation, when a diagnosis is necessary to achieve patient management, the invasive secretin and CCK stimulation tests are preferred, if available.

See Also the Following Articles

Amylase • Cholecystokinin (CCK) • Pancreatic Triglyceride Lipase • Pancreatitis, Chronic • Secretin • Trypsin

Further Reading

Chey, W. D., and Chey, W. Y. (1999). Tests of gastric and exocrine pancreatic function and absorption. *In* "Textbook of Gastroenterology" (T. Yamada *et al.*, eds.), 2nd Ed., Vol. 1, pp. 2924–2937. Lippincott Williams & Wilkins, Philadelphia.

DiMagno, E. P. (1998). A perspective on the use of tubeless pancreatic function tests in diagnosis. *Gut* **43**(1), 2–3.

Owyang, C. (1999). Chronic pancreatitis. *In* "Textbook of Gastroenterology" (T. Yamada *et al.*, eds.), 2nd Ed., Vol. 1, pp. 2151–2177. Lippincott Williams & Wilkins, Philadelphia.

Pancreatic Polypeptide Family

PETER J. MANNON
National Institute of Allergy and Infectious Diseases, Bethesda, Maryland

enteroendocrine cell A specialized epithelial cell diffusely distributed throughout the gut epithelieum that releases a hormone after exposure to luminal contents and other stimuli.
enterogastrone Intestine-derived factor released in response to intestinal fat; inhibits gastric acid secretion when administered at physiologic levels.
interdigestive motility complex A cyclic, periodic pattern of gastric and intestinal contraction during the fasting state characterized by four phases: no activity (Phase 1), uncoordinated activity (Phase 2), a brief period of sweeping peristaltic contractions (Phase 3), and a period of uncoordinated activity preceeding Phase 1 (Phase 4).
prepro hormone The precursor peptide containing all the hormone gene-encoded amino acids (including the signal peptide) before processing to the mature peptide.

The pancreatic polypeptide family of peptides consists of pancreatic polypeptide, peptide YY, and neuropeptide Y. These peptides share significant features of gene organization, amino acid sequence, three-dimensional structure, and receptor binding specificity. However, pancreatic polypeptide and peptide YY are restricted to endocrine cells of the pancreas and distal gut epithelium and are released in response to meals. Neuropeptide Y is restricted to neural tissue and is released by nerve stimulation in the manner of a neurotransmitter. Pancreatic polypeptide and peptide YY generally exert inhibitory effects on gut function, such as gastric acid secretion and gastrointestinal motility, in order to regulate the postprandial state. Neuropeptide Y is better known for its effects on the cardiovascular, endocrine, and central nervous systems, but it also exerts inhibitory effects on gut secretion and motility.

INTRODUCTION

The pancreatic polypeptide family of peptides includes pancreatic polypeptide (PP), peptide YY (PYY), and neuropeptide Y (NPY). These peptides are related by a high degree of sequence homology among their 36 amino acids as well as by a shared hairpin-like tertiary structure. These similarities are thought to reflect origins from common ancestral genes, gene duplications resulting in new peptides, and peptide sequence conservation related to evolutionary pressures. However, there has been significant divergence of distribution, production, and actions of these peptides, with PP and PYY produced by endocrine cells of the pancreas and the gut epithelium, respectively, and NPY produced by neurons in the peripheral and central nervous systems.

Family Discovery

Pancreatic polypeptide was the first family member to be purified and identified by amino acid sequence following its recognition as a persistent contaminant of insulin preparations. In contrast, peptide YY and neuropeptide Y were both discovered using a strategy to isolate novel peptides that have been amidated at their C terminus, a biochemical modification often indicating innate bioactivity in peptides. Peptide YY and neuropeptide Y were isolated from extracts of porcine intestine and brain, respectively. The purification and amino acid sequence of all of these peptides were completed in advance of establishing their biological effects.

Peptide Structure

The three-dimensional structures of the peptides in this family, deduced from crystallography and molecular modeling, have helped to explain the stability of these molecules and the specificity of interaction with receptor proteins. The structure is best described as two helices running antiparallel to each other; there is a polyproline type II helix (amino acids 1–8) connected to an α helix (amino acids 14–32) connected by a type II β turn and stabilized by hydrophobic interactions between the helices. The C-terminal hexapeptide is flexible, projecting from the base of the molecule. Much of the receptor binding specificity is found in the close approximation of amino acid residues from the two ends of the molecule, as well as the terminal hexapeptide composition.

 Copyright 2004, Elsevier (USA). All rights reserved.

PANCREATIC POLYPEPTIDE

Tissue Distribution

Pancreatic polypeptide expression is restricted to endocrine cells found predominantly in the pancreas, where they can exist as part of the endocrine islets or as single or groups of cells in the exocrine pancreatic tissue, and less often in the epithelium of pancreatic ducts. The density of PP-containing cells is highest in the pancreatic head, and the tissue concentration of PP is five to eight times higher in the pancreatic head compared to the tail. The PP-immunoreactive cells have also been reported in the gastric mucosa of some adult mammals (dog, cat, opossum) and in the human colon.

Gene Organization

The human gene for pancreatic polypeptide is made up of four exons and three introns that encode the 95-amino-acid pancreatic polypeptide prepro hormone. Exon 1 encodes the 5′ untranslated region, exon 2 encodes a 29-amino-acid signal peptide and PP, exon 3 encodes a 23-amino-acid extension, and exon 4 encodes a further 7-amino-acid extension and the 3′ untranslated region. Posttranslational processing produces the 36-amino-acid peptide with an amidated C-terminal tyrosine residue.

Mechanisms of Pancreatic Polypeptide Release

Pancreatic polypeptide is primarily released following nutrient ingestion and requires an intact vagus nerve for full response. PP release occurs with all phases of feeding, including the cephalic, gastric, and intestinal phases. The cephalic phase, induced by sham feeding in animals or spit-and-chew techniques in humans, stimulates between 10 and 20% of the total meal-stimulated PP response. The cephalic-phase release of PP enhances further PP release during the gastric and intestinal phases that follow. The early phases of meal-stimulated PP release depend on vagal input. The cephalic-phase PP release is blocked by anticholinergic drugs and truncal vagotomy. Similarly, gastric distension resulting in PP release is blocked by anticholinergic drugs and vagotomy. Vagal activity, whether by direct electrical vagal stimulation or stimulation by insulin hypoglycemia, can increase PP serum levels.

In contrast, the intestinal phase of the feeding response, simulated by administering nutrients directly into the duodenum, does not depend on the neural reflexes mediating the cephalic phase and gastric distension responses. Although intestinal-phase PP release is optimized by intact vagal inputs, it still occurs after vagotomy, likely through remaining local enteric–pancreatic neural reflexes. Moreover, it is thought that intestinal-phase feeding also involves other nutrient-stimulated hormones—cholecystokinin (CCK), for instance—that contribute to the later increases in meal-induced PP release. The hormonal regulation of PP release during intestinal digestion is complex and not well defined.

Serum pancreatic polypeptide levels in the fasting state vary rhythmically with the hormonal and motility events that characterize the interdigestive motility complex. Basal PP levels peak through phases 1 to 3 and then return to low levels during phase 4. These basal PP level fluctuations are abolished by local ganglionic blockade and anesthetic use, supporting a role for neural reflexes in this response. Finally, average basal (fasting) pancreatic polypeptide levels increase with age.

Pancreatic Polypeptide Receptors

The Y4 receptor, the high-affinity receptor for pancreatic polypeptide, has been cloned and sequenced from many mammalian species, including rodents, pigs, and humans. The Y4 receptor is a member of the G protein-coupled receptor superfamily, and Y4 receptors have dissociation constants in the range of 20–40 p*M* for same-species cognate ligands (human PP for the human PP receptor, for instance). However, Y4 receptors are only PP preferring and not PP specific, allowing binding of PYY and NPY peptides at 100-fold lower affinities. In certain species, PP binds to the Y5 receptor (human and bovine PP have high binding affinity, and rat PP has low binding affinity, for the rat PP Y5 receptor), although with lower affinity than NPY or PYY. Last, lower vertebrates possess a PP receptor that also recognizes PYY.

Pancreatic polypeptide receptors have been reported in the gut, pancreas, brain (hypothalamus, hippocampus, and vagal nuclei of the brain stem), and kidney of various species.

Pancreatic Polypeptide Biological Actions

There is limited understanding of the physiologic role of pancreatic polypeptide. Two biologic effects that pancreatic polypeptide exerts in a number of species are inhibition of exocrine pancreatic secretion and inducing relaxation of the gallbladder. When PP is administered to produce serum levels attainable after meals, stimulated pancreatic enzyme secretion in dogs and humans is inhibited. Reduced bilirubin secretion into the duodenum after PP infusion is also noted in humans, and decreases in intraluminal gallbladder pressure (pigs) and increased gallbladder filling are seen in animal models. Because these effects occur at

nonpharmacologic serum levels, they are interpreted as being possibly physiologic responses. The effect on the pancreas could act as a negative feedback signal to turn off meal-stimulated pancreatic secretion and the gallbladder effect could preserve extrahepatic bile excretion between meals.

Other effects attributed to PP include inhibition of gastric secretion (at high doses in dogs; no effect at lower doses in humans), modulation of gastric emptying, and modest inhibition of feeding when administered peripherally (accompanied by decreased gastric emptying) and stimulation of feeding when given centrally (accompanied by increased gastric emptying). However, hypersecretion of PP in humans due to PP-producing tumors produces no symptoms, and elevated PP levels in rats transgenic for the PP gene are associated with only modest inhibition in food intake and gastric emptying. On the other hand, patients with Prader–Willi syndrome (congenital obesity, hyperphagia, hyperglycemia, and hyperinsulinemia) have abnormally low basal and meal-stimulated levels of PP; PP administration can diminish their hyperphagia through an unknown mechanism.

PEPTIDE YY

Tissue Distribution

Peptide YY is predominantly expressed in endocrine cells but is also produced by neurons in certain species. The primary source of serum PYY released into the bloodstream is enteroendocrine cells in the distal gut epithelium; PYY-immunoreactive cells can be found in the upper small intestine of many mammals, but the concentration of these cells increases in an aboral direction, whereby the colon and rectum have the highest density of PYY-producing cells. The PYY-containing cells have a flasklike appearance characteristic of open-type endocrine cells, with a wide base containing the secretory granules and a thin apical arm extending to the gut lumen. A large proportion of the PYY-containing granules also contain proglucagon-derived peptides such as glicentin and glucagon-like peptides 1 and 2.

PYY is also contained in endocrine cells in the gastric mucosa of a few adult animal species, but these are a minor proportion of total PYY-containing cells. PYY can colocalize with glucagon (rats and mice) and PP (dogs and pigs) in pancreatic endocrine cells and in specialized epithelium of alveolar ducts (Syrian golden hamsters). Peptide YY has been found in neural structures of the enteric nervous system, including myenteric ganglia and serosal ganglia. Central nervous system localization of PYY has also been made in the hypothalamus, brain stem, and spinal cord of the rat.

Gene Organization

The human gene for peptide YY is also made up of four exons and three introns that span approximately 1.2 kb and encode a 97-amino-acid prepro hormone. Exon 1 encodes the 5′ untranslated region, exon 2 encodes a 28-amino-acid signal peptide and most of PYY, exon 3 encodes the C-terminal PYY tyrosine and a 26-amino-acid extension, and exon 4 encodes a further 7-amino-acid extension and the 3′ untranslated region. Posttranslational processing produces the 36-amino-acid peptide with an amidated C-terminal tyrosine residue. The structure of this gene is so similar to that of the human PP gene that it is thought to be the result of a gene duplication event.

Peptide YY has also been discovered to exist as a 34-amino-acid peptide, PYY 3–36, in some species, including humans. This shortened form is likely due to proline dipeptidyl peptidase IV digestion of the full-length peptide. PYY 3–36 can account for up to 50% of total PYY activity in postprandial human serum, and has affinity for Y receptors that bind C-terminal fragments of PYY and NPY. Last, the prostate- and testis-restricted 33-amino-acid human seminal plasmin peptide (so-called PYY2) is homologous to the 12 amino-terminal residues of human PYY, but does not appear to have binding activity at Y receptors.

Mechanisms of Peptide YY Release

Peptide YY is released in response to ingested food, presumably to coordinate postprandial gastrointestinal reflexes. In humans, PYY serum levels begin to increase within 30 minutes of a meal and reach a plateau after 1–2 hours. The maximum level is maintained for up to 4 hours and then returns slowly to baseline. It is hypothesized that the early increase is due to signals from the duodenum because ingested nutrients have not had time to reach the distal gut for direct stimulation of PYY-containing cells. There are data to support a neurohumoral mechanism regulating PYY release from the distal gut by upper intestinal stimuli, but this is only found in certain species.

There is evidence that gastrin and CCK may be involved in the early postprandial PYY release. Gastric acid secretion stimulated by meals (in dogs) contributes to gastric-phase PYY release because early PYY release can be blocked by histamine receptor blockade and proton pump inhibitors. Gastrin has been found to inhibit PYY release from intestinal endocrine L cells, so the inhibition of gastrin release accompanying gastric acid secretion may remove this inhibitory input on PYY secretion. In rats treated with proton pump inhibitors, the resulting increases in gastrin levels are thought to

be responsible for the decrease in PYY content and mRNA levels in the colon because this effect is blocked by a gastrin/CCK-B receptor antagonist. Furthermore, in dogs, CCK infusion increases PYY levels, and this effect can be blocked by CCK-A receptor antagonists. It is not clear if CCK acts directly on PYY-secreting cells or via intermediate neural pathways.

Neural factors play a role in physiologic PYY release. For instance, there may be a baseline inhibitory effect of intact gut innervation on PYY release because truncal vagotomy, general anesthesia, and jejunoileal denervation enhance meal-stimulated PYY release in dogs. On the other hand, a PYY-releasing signal from the duodenum to the distal gut can be interrupted by atropine (in dogs, not rats), hexamethonium (dogs and rats), truncal vagotomy (rats), and CCK receptor antagonists (dogs, not rats). In addition, adrenergic, cholinergic, and certain peptidergic (bombesin, gastrin-releasing peptide) stimuli have been shown to stimulate PYY release, but the physiologic relevance of this is unknown.

The majority of postprandial PYY release is related to exposure of the distal small bowel and colon to the luminal remnants of a meal. In humans, peptide YY is released in proportion to the calorie content (higher levels with more calories) and nutrient composition of a meal (fat meals stimulate significantly more PYY release compared to isocaloric protein or carbohydrate meals). This is consistent with the effect on PYY release following direct instillation of specific nutrients into the distal gut. Administration of oleic acid to distal intestinal epithelium can result in PYY release (seen in dog, cat, rat, and human models). However, the most consistent effect is seen with combinations of oleic acid and bile salts, specifically taurocholate and deoxycholate; oleic acid by itself may have no effect on PYY release (in humans, for instance). Although bile salts enhance the effect of fatty acids on PYY release via direct epithelial contact, they can also have similar stimulatory effects on their own. Glucose and protein or amino acids can also stimulate PYY release after instillation into the distal gut, but whether this is of physiologic importance is not known. Last, adenylyl cyclase-coupled stimuli increasing intracellular cyclic AMP levels appear to be the predominant signals for directing PYY release.

Peptide YY Receptors

Peptide YY is a high-affinity cognate ligand for several Y receptor subtypes, including the Y1, Y2, and Y5 receptors. Y1 receptors have high affinity for PYY, neuropeptide Y, and specific substituted forms of these peptides, [Pro^{34}]PYY and [Pro^{34}]NPY, and low affinity for C-terminal fragments of these peptides as well as intact PP. In addition to widespread distribution in the central nervous system, Y1 receptors are also found in veins and arteries as well as in colonic epithelial cells and nerve fibers and ganglia of the enteric nervous system in humans. Y2 receptors display high affinity for PYY, NPY, and various of their C-terminal fragments; compared to the Y1 receptor, there is low-affinity binding of [Pro^{34}]-PYY and [Pro^{34}]NPY and no binding of PP. Y2 receptors occur in the brain, but also in the colonic epithelium of rats, in blood vessels (dog saphenous vein), and in nerve fibers of the autonomic nervous system. The Y5 receptor is described as "Y1-like" because not only do PYY and NPY bind with high affinity, but so do the Y1 analogues ([Pro^{34}]PYY and [Pro^{34}]NPY) in addition to long C-terminal fragments such as NPY 2–36 and PYY 3–36; however the C-terminal fragment NPY 13–36 binds with much lower affinity (PP binding affinity for the rat and human Y5 receptor was low for rat PP but high for human and bovine PP). The Y5 receptor subtype is found primarily in the brain.

Other reported receptors for PYY include the Y6 receptor and a PYY-preferring receptor. The Y6 receptor was originally found in rat brain and subsequently rabbit, monkey, and human homologues were described. The rat Y6 receptor has been variably shown to possess a high affinity for PYY with a Y1-like binding profile or a high affinity for PP and lower affinity for PYY. The Y6 receptor has not been shown to be a physiologically relevant receptor in humans. A PYY-preferring receptor has been suggested to exist because of a three to five times greater potency compared to NPY for some biological effects, but this purported receptor has not been characterized molecularly and remains theoretical at present.

Peptide YY Biological Actions

Peptide YY has been characterized as a candidate enterogastrone but additionally has been shown to inhibit intestinal motility, gut epithelial secretion, and pancreatic secretion. There are several aspects of interpretation of the large body of data pertaining to the biological effects of PYY that must be kept in mind; although similar effects for PYY are seen in many animal models, the ultimate effect of PYY is very dependent on (1) the species studied, (2) whether resting or stimulated conditions are tested for PYY inhibitory effects (and the type of stimulation used), and (3) what doses of PYY are required for effects (whether effects of PYY are observed at postprandial serum concentrations or only at higher, so-called pharmacologic, doses).

Whether PYY acts as a classic enterogastrone is not clear. On the one hand, it has been shown that

intravenous administration of PYY at doses reproducing higher postprandial levels and supraphysiologic levels can inhibit gastric acid production in humans, rats, and dogs. On the other hand, the inhibition of gastric acid production and emptying induced by introducing fat into the upper small intestine is not due to the accompanying PYY elevations (in humans and dogs) but is due to other factors, such as CCK. On its own, PYY is most potent at inhibiting cephalic-phase gastric acid production (nearly 90% inhibition in sham-fed dogs), and this requires vagal innervation. PYY can also inhibit gastric acid production by additive effects with glucagon-like peptide-1, secretin, and somatostatin or by directly blocking gastrin-stimulated histamine release from enterochromaffin-like (ECL) cells (in rats).

PYY can also act in the central nervous system (CNS) to modulate gastric acid secretion. However, the route of access of PYY to the CNS results in different effects on gastric acid production. It has been shown that intravenous PYY can cross the blood–brain barrier to bind specifically to regions of the dorsal vagal complex (DVC) [the nucleus tractus solitarii, the dorsal motor nucleus (DMN) of the vagus, and the area postrema (AP)], where Y receptor subtypes Y1, Y2, and Y4 have been found. The inhibitory activity of PYY on centrally stimulated gastric acid secretion (cephalic phase, for instance) may occur within the CNS because intracisternal injection of neutralizing antibodies against PYY can block its inhibitory effects. On the other hand, when PYY is directly microinjected into the DVC of rat brain stem, there is a dose-dependent increase in gastric acid secretion (that is dependent on an intact vagus nerve). These observations suggest that PYY has a complex role in gastric acid regulation that varies with the phase of gastric acid production.

One of the most consistently observed effects of PYY is the inhibition of intestinal motility. This effect is thought to enhance intraluminal nutrient digestion and absorption, earning PYY the rubric "the ileal brake." Delays in orocecal transit time are induced by PYY infusions that reproduce normal postprandial PYY levels. The decrease in intestinal motility following fat exposure in the lower small bowel can be blocked by neutralizing serum PYY with specific antisera. This effect on intestinal motility may be beneficial in conditions resulting in delivery of unabsorbed nutrients to the distal bowel and colon, such as pancreatic insufficiency, Crohn's disease, and acute diarrhea, which are associated with high basal and postprandial PYY levels.

In the pancreas, supraphysiologic levels of PYY can inhibit secretin-stimulated secretion in the anesthetized cat. Similarly, high doses of intravenous PYY inhibit meal- and secretin-stimulated pancreatic secretion in the dog, but porcine PYY has no effect on secretin- or CCK-stimulated pancreatic secretion in humans.

Other PYY biological effects in the gut include inhibition of stimulated apical chloride secretion in gut epithelium and an increase in intestinal absorption in intact animals. PYY has been shown to stimulate proliferation in gut epithelium (Y receptors are expressed by the gut epithelium of humans, rabbits, and rats) and may be involved in gut epithelial differentiation signals. PYY effects on feeding behavior are included in the discussion of neuropeptide Y.

NEUROPEPTIDE Y

Tissue Distribution

The majority of neuropeptide Y is found in the central and peripheral nervous systems (also seen in chromaffin cells of the adrenal medulla and pheochromocytomas), so that its distribution in the gut is within nerve cells and fibers. The NPY-immunoreactive (NPY-IR) nerves in the gut may be intrinsic or extrinsic to the gut. The intrinsic NPY-IR nerves are part of the nonadrenergic enteric nervous system, where NPY may colocalize with other neuropeptides, such as vasoactive intestinal peptide (VIP) and peptide histidine isoleucine. The extrinsic NPY-IR nerves generally are adrenergic fibers where NPY colocalizes with norepinephrine, and they innervate vascular structures. NPY is found in ganglia cell bodies, with higher frequency in submucosal plexi but also occurring in the myenteric plexi and extrinsically in the celiac ganglion. NPY-IR nerve fibers are found extending from plexi to the muscle layers of the muscularis propria, as well as within the muscularis mucosa, and into the mucosa, where they may closely invest the crypts. Perivascular NPY-IR fibers are largely of extrinsic adrenergic origin.

NPY Gene Expression and Receptors

The human gene for neuropeptide Y is also made up of four exons and three introns that span approximately 8 kb and encode 97-amino-acid residues of the NPY prepro hormone. Exon 1 encodes the 5′ untranslated region, exon 2 encodes residues 1–28 of the signal peptide and most of NPY, exon 3 encodes the C-terminal tyrosine, the 3-amino-acid cleavage site, and a 23-amino-acid extension, and exon 4 encodes a further 7-amino-acid extension and the 3′ untranslated region. Posttranslational processing produces the 36-amino-acid peptide with an amidated C-terminal tyrosine residue.

The full-length amidated 36-amino-acid peptide is the major form of NPY found in the bloodstream after

release from nerve endings following sympathetic stimulation or feeding. NPY binds to several Y receptor subtypes as previously discussed. An NPY-specific receptor subtype, Y3, that does not bind PYY has been described pharmacologically but has yet to be cloned. Given the extensive representation of NPY in nerves in the gut, it is important to note that Y receptors are found on ganglion cells and nerve fibers (in humans) and on epithelial cells (in rabbits, humans, rats, and mice), where NPY released from nerve endings (and possibly PYY released from enteroendocrine cells) could exert an effect.

NPY Biological Actions

Neuropeptide Y has a wide variety of activities, including vasoconstriction and regulation of cardiac function, stimulation of feeding, modulation of pituitary release of hormones, and effects on circadian rhythm and anxiety state. In the gut, administration of NPY potently inhibits mucosal fluid and electrolyte secretion stimulated by prostaglandin E2, VIP, and cholera toxin. NPY has effects on gut motility as well, mediating contraction and relaxation of the lower esophageal sphincter (cat), contraction of longitudinal muscle and inhibition of peristalsis and circular muscle reflex contraction (guinea pig), and contraction of longitudinal muscle (rat, dog). NPY appears to mediate its effects on motility not by direct actions on muscle but rather by inhibition of excitatory neurons of the enteric nervous system.

Neuropeptide Y has emerged as one of the key neuropeptides active in stimulating feeding behavior and coordinating peripheral signals involved in body weight regulation. Studies in rodents show that NPY injected into the brain (cerebral ventricles and hypothalamus) induces feeding, prolongs feeding, and causes feeding to resume in fed animals. Increases in endogenous NPY occur in the paraventricular nucleus of the hypothalamus prior to nocturnal feeding, during fasting, and accompanying the hyperphagia of drug-induced diabetes and genetically obese animals such as fatty Zucker rats and *ob/ob* mice. These latter observations suggest that NPY plays a physiologic role in feeding behavior. Current models of NPY-induced feeding show that the NPY released from fibers originating in the hypothalamic arcuate nucleus is responsible for stimulating feeding. The stimulation of feeding appears to be mediated through Y1 or Y5 receptor subtypes, likely located in other hypothalamic areas such as the paraventricular nucleus. Though rises in NPY levels in the hypothalamus precede feeding, it is not yet clear what the exact proximate mechanisms are that control and coordinate this increase.

Neuropeptide Y-stimulated feeding may also be part of a larger regulatory mechanism that is influenced by signals from outside the central nervous system. It has been shown that activation of Y2 receptor subtypes in the rat hypothalamus can decrease the release of NPY. In support of this activity, a Y2-preferring ligand PYY 3–36 can inhibit nocturnal onset feeding in rats (injected intraperitoneally) and decrease calorie intake in humans following intravenous infusions that mimic postprandial blood levels. This suggests that peripheral postprandial PYY release (recall the PYY 3–36 fragment can account for up to 50% of total PYY activity in postprandial human serum) may contribute to a satiety effect by attenuating the NPY effect in the hypothalamus. Whether this inhibitory effect is accomplished by direct effects on the hypothalamus or indirectly via other neural mechanisms has yet to be established.

Leptin is another peripheral signal related to regulation of body weight that can influence hypothalamic NPY activity. Leptin can decrease the production of NPY in the arcuate nucleus of the hypothalamus (mice); subpopulations of NPY-containing neurons in the rat hypothalamus involved in reactions to fasting also express the leptin receptor, supporting a role for leptin regulation of NPY in feeding response. It is thought that fluctuations in leptin levels related to body fat stores may contribute to enhanced NPY-induced feeding during low-body-fat conditions (low leptin levels) and blunted NPY-induced feeding during high-body-fat states (high leptin levels).

These data suggest that NPY has an important role in central control of feeding behavior and that peripheral signals, including postprandial PYY release and leptin levels related to body fat stores, modulate the feeding response to establish satiety and energy homeostasis.

See Also the Following Articles

Cholecystokinin (CCK) • Gastric Acid Secretion • Gastric Motility • Gastrin • Ileal Brake • Pancreatic Enzyme Secretion (Physiology)

Further Reading

Blomqvist, A. G., and Herzog, H. (1997). Y-Receptor subtypes—How many more? *Trends Neurosci.* **20**, 294–298.

Cerda-Reverter, J. M., and Larhammar, D. (2000). Neuropeptide Y family of peptides: Structure, anatomical expression, function, and molecular evolution. *Biochem. Cell Biol.* **78**, 371–352.

Mannon, P., and Taylor, I. L. (1994). The pancreatic polypeptide family. *In* "Gut Peptides: Biochemistry and Physiology" (J. H. Walsh and G. J. Dockray, eds.), pp. 341–370. Raven Press., New York.

Michel, M. C., Beck-Sickinger, A., Cox, H., Doods, H .N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T., and Westfall, T. (1998). XVI. International union of pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors. *Pharmacol. Rev.* **50**, 143–150.

Pancreatic Pseudocysts

THEODORE R. LIN AND MICHAEL W. MULHOLLAND
University of Michigan Medical Center, Ann Arbor

computerized tomography A radiographic technique allowing visualization of a plane or section through the body.
endoscopic retrograde cholangiopancreatography A procedure whereby pancreatic and biliary ducts are visualized by endoscopic injection of contrast medium.
pseudocyst A non-epithelial-lined collection of pancreatic secretions.

Pancreatic pseudocysts are non-epithelial-lined fluid collections that arise as a consequence of acute pancreatitis, trauma, or chronic pancreatitis. They are believed to result from disruption of the pancreatic duct with subsequent leakage of pancreatic exocrine secretions into surrounding peripancreatic tissue. The incidence of pancreatic pseudocysts varies widely, especially with the increased use of axial imaging in patients with pancreatitis and abdominal pain. Most pseudocysts are asymptomatic; the development of symptoms may herald significant complications, including perforation, obstruction, hemorrhage, and infection. Most pseudocysts are single, unilocular lesions. One-third of pseudocysts are located in the pancreatic head and the remainder are found in the body and tail regions. The fluid filling the pseudocyst cavity is usually watery in nature and rich in pancreatic enzymes, including amylase, lipase, and trypsin. Some endoscopic retrograde cholangiopancreatogram studies have shown that in 80% of cases, the pancreatic duct communicates with the pseudocyst.

PRESENTATION AND DIAGNOSIS

A lingering course of acute pancreatitis, persistent abdominal pain, or an epigastric mass may be potential indicators of pseudocyst development. Additional signs and symptoms may signify complications. Infection manifests with fever and abdominal pain. Patients with ruptured pseudocysts present with severe acute abdominal pain, peritoneal irritation, and shock. Gastrointestinal bleeding and vomiting occur when pseudocysts erode into the surrounding viscera and vascular structures.

The diagnosis of pseudocysts has been greatly facilitated by the use of axial imaging. Computerized tomography (CT) is the standard for diagnosis and detects small pseudocysts even less than 1 cm in diameter. Moreover, CT is helpful in delineating the features of pseudocysts, including anatomic location, fibrosis of the wall, presence of blood or infectious material, and viability of the surrounding pancreatic parenchyma (Fig. 1).

An assessment of the likelihood of spontaneous resolution and the risk of developing complications guides clinical decision-making. Management may consist of observation and surveillance or intervention. Pancreatic pseudocysts resolve spontaneously through absorption, rupture into neighboring viscus, or drainage through the pancreatic duct, with spontaneous resolution in 50% of cases. Factors that affect the likelihood of spontaneous resolution include size, chronicity, and multiplicity. Most pseudocysts will resolve without intervention within 6 weeks. Lesions that are less than 6 cm in size are more likely to heal than those that are larger than 6 cm. Those attributed to chronic pancreatitis are less likely to heal than pseudocysts secondary to acute pancreatitis. Multilocular lesions

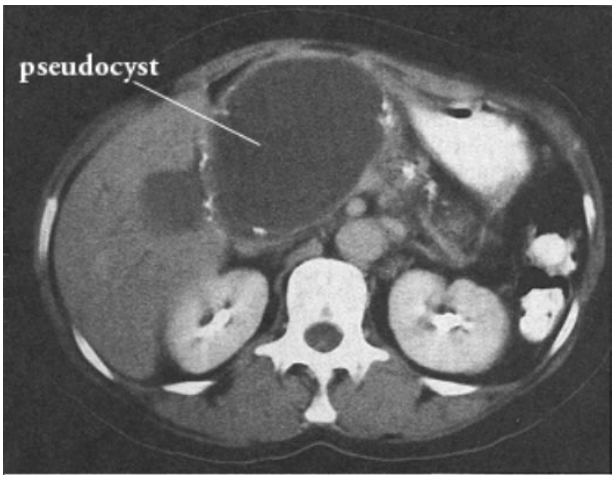


FIGURE 1 CT scan showing a large pancreatic pseudocyst. Reprinted from Guice, K. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., with permission. Copyright Lippincott-Raven Publishers, 1997.

 Copyright 2004, Elsevier (USA). All rights reserved.

are also less likely to heal spontaneously than unilocular lesions.

THERAPEUTIC APPROACHES

Endoscopic, percutaneous, and surgical approaches may be employed to treat complicated or symptomatic pancreatic pseudocysts. Endoscopy may be used to create a communication between the pseudocyst and either the stomach or small bowel. A broadly adherent, opposing wall of stomach or small bowel is absolutely required for this approach to be considered; preprocedural axial scanning is required to document this circumstance. A stent or catheter is then deployed, creating a conduit for internal drainage of the pseudocyst. Radiographically guided percutaneous placement of an indwelling catheter into the pseudocyst is another means of treating unresolving or complicated cases of pancreatic pseudocyst. External catheter drainage of the pseudocyst may be complicated by secondary infection of the pseudocyst, catheter obstruction from particulate debris, and the development of enterocutaneous or pancreaticocutaneous fistulas. Surgical approaches to complicated pancreatic pseudocysts are based on creating an internal drainage communication between the pseudocyst and either the stomach or small intestine (Fig. 2). As in the endoscopic approach, a broadly adherent pseudocyst wall is essential to the success of the drainage procedure. Moreover, an adequately mature and fibrous wall is required to ensure a stable anastamosis. Pseudocysts located in the tail of the pancreas may be amenable to resection by distal pancreatectomy. In these cases, splenectomy is warranted if the lesion is in close proximity to the splenic hilum.

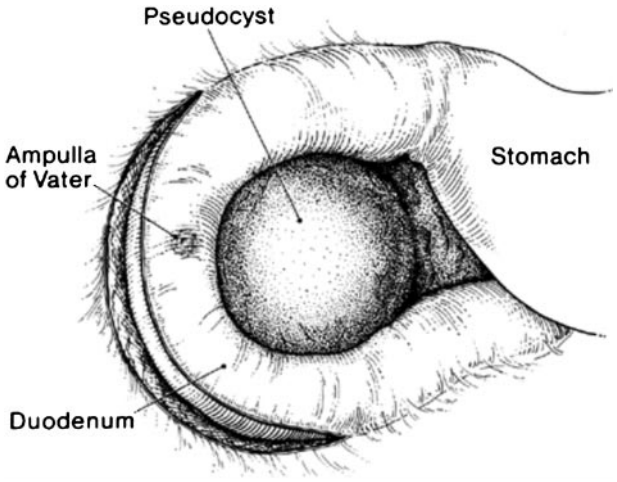


FIGURE 2 Pancreatic pseudocyst in close apposition to the duodenum, allowing for cyst drainage into the duodenum (cystoduodenostomy). Reprinted from Bradley, E. L. (1997). "Mastery of Surgery" (Nyhus, Baker, Fischer, eds.), 3rd Ed., p. 1228, with permission. Copyright Lippincott Williams & Wilkins.

See Also the Following Articles

Exocrine Pancreas • Pancreatic Tumors, Other • Pancreatitis, Acute • Pancreatitis Chronic

Further Reading

Cooperman, A. M. (2001). An overview of pancreatic pseudocysts: The emperor's new clothes revisited. *Surg. Clin. North Am.* **81**, 391–397.

Guice, K. S. (1997). Acute pancreatitis. *In* "Surgery: Scientific Principles and Practice" (L. Greenfield, M. Mulholland, K. Oldham, G. Zelenock, and K. Lillemoe, eds.), 2nd Ed., pp. 886–887. Lippincott-Raven, Philadelphia.

Yeo, C. J., and Cameron, J. L. (2001). Exocrine pancreas. *In* "Sabiston Textbook of Surgery" (C. Townsend, R. Beauchamp, B. Evers, and K. Mattox, eds.), 16th Ed., pp. 1129–1130. W. B. Saunders Co., Philadelphia, PA.

Pancreatic Transplantation

ROBERT M. MERION
University of Michigan, Ann Arbor

allograft A graft transplanted between genetically nonidentical individuals of the same species.
diabetes mellitus A chronic metabolic disorder in which utilization of carbohydrate is impaired and that of lipid and protein is enhanced; it is caused by an absolute or relative deficiency of insulin and is characterized, in more severe cases, by chronic hyperglycemia, glycosuria, water and electrolyte loss, ketoacidosis, and coma.
diabetic neuropathy A generic term for any diabetes mellitus-related disorder of the peripheral nervous system, autonomic nervous system, and some cranial nerves.
immunosuppression Prevention or interference with the development of an immunologic response; it may reflect natural immunologic unresponsiveness (tolerance), may be artificially induced by chemical, biological, or physical agents, or may be caused by disease.
pancreatic islet Cellular masses varying from a few cells to hundreds of cells lying in the interstitial tissue of the pancreas that constitute the endocrine portion of the pancreas and are the source of insulin and glucagon.
renal failure Loss of kidney function, either acute or chronic, that results in azotemia and syndrome of uremia.
retinopathy Retinal changes occurring in diabetes mellitus, marked by microaneurysms, exudates, and hemorrhages, and sometimes by neovascularization.

Pancreatic transplantation is a surgical treatment for severe diabetes mellitus and its complications. Most commonly, the entire gland is transplanted heterotopically in cases of longstanding type 1 diabetes. In a number of scattered reports, isolated pancreatic islets have been transplanted, with mainly unsatisfactory results. The two primary objectives of pancreatic transplantation are to restore normal glucose homeostasis and to prevent, delay, or reverse the end-organ complications associated with the underlying disease.

INTRODUCTION

Pancreatic transplantation has been performed since 1966. Results were poor during the first two decades for three main reasons. First, inability to successfully preserve pancreatic allografts during the required period of time between removal from the donor and implantation into the recipient led to early allograft failure in a significant proportion of cases. Second, fragility of the organ and injury related to ischemia and reperfusion resulted in a high incidence of allograft thrombosis. Third, immunosuppression was inadequate to prevent rejection and the treatment of pancreas transplant rejection was unsatisfactory.

Since the 1980s, results of pancreatic transplantation have steadily improved. Improved procurement techniques and more reliable preservation solutions have reduced the early failure rate and, combined with surgical innovations, have lessened the incidence of early allograft thrombosis. Immunosuppressive drug developments have significantly reduced the incidence, severity, and consequences of allograft rejection and now result in excellent long-term insulin-independent survival.

INDICATIONS AND CONTRAINDICATIONS

Patients with longstanding type 1 diabetes mellitus are candidates for pancreatic transplantation. A demonstrated absence of C peptide confirms the failure of endogenous insulin production. The development of end-organ complications such as diabetic proliferative retinopathy, peripheral and autonomic neuropathy, gastrointestinal dysmotility, and diabetic nephropathy signals advanced stages of the disease and warrants consideration for transplantation.

Three forms of pancreatic transplantation are offered to suitable patients depending upon the degree of renal impairment. These comprise pancreas transplant alone (PTA), simultaneous pancreas–kidney transplantation (SPK), and sequential pancreas after kidney transplant (PAK).

In patients without marked renal impairment, control of blood glucose via conventional means such as intermittent subcutaneous insulin injections or insulin pump is almost always successful in preventing life-threatening hypoglycemic episodes. However, a small number of patients have severe hypoglycemic

 Copyright 2004, Elsevier (USA). All rights reserved.

unawareness, multiple hospitalizations for diabetic coma or seizures, and inability to control blood sugar despite adequate compliance with an aggressive regimen. These patients may be candidates for PTA, provided that an adequate renal reserve is present.

Individuals who have already developed renal failure are also candidates for pancreatic transplantation and require renal replacement therapy as well, in the form of a kidney transplant. Two modalities are available for such patients and the choice between them depends on the availability of a kidney donor. In the first case, if a living donor is available, the kidney transplant is performed first and the patient may receive a subsequent cadaveric pancreas transplant (PAK), provided that the kidney transplant is functioning adequately. If a living donor transplant is not feasible, the patient may receive both organs from a single cadaveric donor (SPK).

RECIPIENT EVALUATION

Candidates up to 45 years of age are routinely considered, but older patients require more careful scrutiny due to demonstrated inferior patient and graft survival outcomes. The duration of diabetes mellitus should be assessed and absence of C-peptide secretion confirmed. Systematic assessment of end-organ complications is important. Diabetic retinopathy requiring laser photocoagulation, vitrectomy, or other measures should be assessed. Blindness is not an absolute contraindication to pancreatic transplantation, but its presence may suggest more advanced systemic disease. Peripheral neuropathy and associated musculoskeletal manifestations such as Charcot joints may be present. Autonomic neuropathy with symptomatic orthostatic hypotension and gastrointestinal dysmotility syndromes are common in this patient population. Meticulous investigation of correctable coronary artery disease is mandatory in light of the high incidence of both overt and asymptomatic lesions. The presence of severe peripheral vascular disease is a marker for systemic atherosclerosis and, particularly if associated with major amputation, is considered a relative contraindication. The need for immunosuppressive therapy, with its attendant risks of neoplasia and infection, means that preexisting malignancies (other than nonmelanoma skin cancers) and active or ongoing infections are also contraindications to pancreatic transplantation.

Unlike most surgical procedures, solid organ transplantation requires lifelong immunosuppression and cooperation with a complex medical regimen. Accordingly, patients must be capable of understanding the risks and benefits of the procedure and have a requisite system of social support.

DONOR CONSIDERATIONS

Virtually all pancreatic transplants are procured from brain-dead cadaveric donors. Donor assessment is a critical step as the pancreas is more susceptible to preprocurement ischemic injury than other solid organs. Donors over the age of 45 years and those whose cause of death was cerebrovascular accident are risk factors for allograft failure. High levels of vasopressor drugs and the use of antidiuretic drugs (vasopressin, desmopressin) to treat diabetes insipidus in the donor may also be associated with a higher risk of failure.

Most cadaveric pancreata are procured from multiorgan donors from whom kidneys, liver, heart, and lungs are also being retrieved. Coordination between several donor operative teams may therefore be necessary, although these procedures are now fairly standardized. The entire pancreas along with the associated donor duodenum is procured. The vascular anatomy of the liver and pancreas must be shared, since the portal vein is the venous outflow for the pancreatic allograft and is also required for the liver. Similarly, the arterial blood supply to the pancreas and liver arises from the celiac axis and arterial anomalies, which may be found in upward of 20% of donors, must be assiduously searched for and preserved.

RECIPIENT OPERATION

Technical details of the recipient pancreas transplant have evolved significantly. Placement of the organ into the iliac fossa within the peritoneal cavity (as opposed to the retroperitoneum) is preferred. Arterial blood supply to the pancreas, arising from branches of the celiac axis and from the superior mesenteric artery, dictate *ex vivo* reconstruction prior to implantation. A Y-graft of bifurcated donor iliac artery is generally used to enable a single arterial anastomosis to the common or external iliac artery in the recipient. The donor portal vein is anastomosed to the external iliac vein of the recipient.

The more common systemic drainage results in chronic hyperinsulinemia, although no specific adverse consequences have emerged and glucose homeostasis is excellent. Some groups have advocated portal venous drainage of the pancreas as a more physiological setup. First passage of pancreatic venous blood through the liver is associated with normal levels of circulating

insulin and a glucose tolerance that more closely approximates a normal pattern.

Exocrine secretion of the pancreas is the Achilles heel of pancreatic transplantation. The two major choices are to create an anastomosis between the donor duodenum and the recipient bladder or to create an anastomosis between the donor duodenum and the recipient small bowel. The former was used for many years, as it was perceived to be safer in the event of a postoperative leak. However, dehydration and acidosis from urinary losses of alkaline pancreatic secretions have led to a return to enteric drainage of the donor pancreas in over one-half of cases.

The pancreas allograft is placed contralateral to a kidney transplant. Most groups prefer to place the pancreas graft on the right side if possible. Since the left iliac vein passes under the iliac artery on that side, placement of the pancreas on the left may result in venous hypertension in the graft and a higher risk of allograft thrombosis.

IMMUNOSUPPRESSION

The majority of transplant centers use anti-lymphocyte preparations as part of an induction immunosuppression regimen following pancreatic transplantation. The agents available include polyclonal (equine and rabbit) and murine monoclonal antibodies to the T-cell CD3 receptor or the receptor for interleukin-2. Maintenance immunosuppression usually consists of a three-drug cocktail using a calcineurin inhibitor (cyclosporine or tacrolimus), an anti-metabolite (usually mycophenolate mofetil), and corticosteroids.

Rejection occurs in up to one-half of recipients. It must be diagnosed early and treated aggressively to avoid graft failure. The diagnosis increasingly relies on a tissue biopsy obtained percutaneously, laparoscopically, or via laparotomy. Pancreas allograft rejection is most commonly treated with a full course of anti-lymphocyte globulin or monoclonal anti-T-cell antibody.

OUTCOME

The success of pancreatic transplantation can be measured in terms of patient and graft survival as well as by the effects of the transplant on end-organ complications. Patient survival for all modalities of the procedure (PTA, SPK, PAK) exceeds 94% at 1 year and 83% at 5 years. Pancreas graft survival rates are dependent on several factors, the most important of which is the modality of therapy (Table I). SPK recipients have the highest pancreas graft survival rates, closely followed by recipients of PAK transplants. Rejection is easiest to diagnose in SPK recipients, where the kidney transplanted from the same cadaveric donor acts as a barometer for the immunologic response to the grafts. Those who receive a PTA have the poorest results, presumably because the diagnosis of rejection is more difficult.

TABLE I 1- and 3-Year Pancreas Transplant Graft Survival Rates by Modality of Transplantation

Pancreas transplant modality	1-year graft survival (%)	3-year graft survival (%)
Simultaneous pancreas–kidney transplant	84	77
Pancreas after kidney transplant	70	56
Pancreas transplant alone	64	50

Diabetic retinopathy may paradoxically worsen transiently after pancreatic transplantation, but by 2 years post-SPK transplantation a higher proportion of patients have stable eyes when compared to comparable diabetic patients treated with kidney transplant alone. Peripheral neuropathy may have multifactorial etiology, particularly in patients with established renal failure. Nevertheless, pancreas transplant recipients have been shown to have less severe and more slowly progressive neuropathy than their nontransplanted counterparts. A successful pancreatic transplant protects the transplanted kidney (or the native kidney in the case of a PTA) from the development or progression of diabetic nephropathy.

Quality of life is dramatically improved for recipients of pancreatic transplantation. Freedom from decades of dietary restrictions and abolition of the fear of hypoglycemic coma or seizures are obvious benefits. Objective measures of health status and quality of life have been repeatedly demonstrated to be greatly improved among pancreatic transplant recipients.

STATUS OF ISLET TRANSPLANTATION

Since the primary objective of pancreatic transplantation is the restoration of glucose homeostasis, the concept of isolated pancreatic islet transplantation has been attractive for many years. Unfortunately, despite several decades of research and modest numbers of human clinical investigations, long-term insulin independence has been achieved in fewer than 10% of cases. A recent effort using a corticosteroid-free immunosuppressive regimen in a small number of patients has provided promising results. Improved methods for large-scale islet isolation and more effective immunosuppressive drugs are

credited with these successes. However, long-term follow-up will be required before any conclusions can be drawn regarding the efficacy of this approach.

See Also the Following Articles

Diabetes Mellitus • Diabetic Neuropathies • Endocrine Pancreas • Exocrine Pancreas • Transplantation Immunology

Further Reading

Gruessner, A. C., and Sutherland, D. E. (2000). Pancreas transplant outcomes for United States (US) cases reported to the United Network for Organ Sharing (UNOS) and non-US cases reported to the International Pancreas Transplant Registry (IPTR) as of October, 2000. *Clin. Transplant.*, 45–72.

Hricik, D. E. (2000). Kidney–pancreas transplantation for diabetic nephropathy. *Semin. Nephrol.* **20**, 188–198.

Ojo, A. O., Meier-Kriesche, H. U., Hanson, J. A., *et al.* (2001). The impact of simultaneous pancreas–kidney transplantation on long-term patient survival. *Transplantation* **71**, 82–90.

Robertson, R. P., Davis, C., Larsen, J., Stratta, R., and Sutherland, D. E. (2000). Pancreas and islet transplantation for patients with diabetes. *Diabetes Care* **23**, 112–116.

Rosenberg, L. (1998). Clinical islet cell transplantation: Are we there yet? *Int. J. Pancreatol* **24**, 145–168.

Shapiro, A. M., Lakey, J. R., Ryan, E. A., *et al.* (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N. Engl. J. Med.* **343**, 230–238.

Sutherland, D. E., Gruessner, R. W., Dunn, D. L., *et al.* (2001). Lessons learned from more than 1,000 pancreas transplants at a single institution. *Ann. Surg.* **233**, 463–501.

Sutherland, D. E., Gruessner, R. W., and Gruessner, A. C. (2001). Pancreas transplantation for treatment of diabetes mellitus. *World J. Surg.* **25**, 487–496.

Pancreatic Triglyceride Lipase

MARK E. LOWE

Washington University School of Medicine, St. Louis, Missouri

dietary fats Predominantly triglycerides, which consist of acyl chains or fatty acids linked to glycerol through ester bonds.

lipase Enzyme that catalyzes the hydrolysis of ester bonds in dietary fats to release fatty acids.

pancreatic triglyceride lipase Specific enzyme produced and secreted by pancreatic acinar cells; primarily responsible for intestinal digestion of triglycerides.

Pancreatic triglyceride lipase is the major pancreatic lipase synthesized and secreted by pancreatic acinar cells. It accounts for most of the luminal digestion of triglycerides in the small intestine. This function is critical for the utilization of dietary triglycerides, because intestinal enterocytes cannot absorb long-chain triglycerides unless lipases convert them to fatty acids and monoacylglycerols.

INTRODUCTION

The digestion of dietary fats in humans begins in the stomach, where gastric lipase (a distinct protein) releases about 15% of the fatty acids from triglycerides. Lipases secreted by the pancreas complete dietary fat digestion in the small intestine. Of these lipases, pancreatic triglyceride lipase (PTL) predominates, as evidenced by patients with congenital deficiency of PTL who malabsorb 50–60% of dietary fats. Cholesterol esterase, another nonspecific lipase, mediates the hydrolysis of other dietary lipids. Other more distantly related lipases, such as lipoprotein lipase and hormone-sensitive lipase, mediate the uptake and release of fatty acids from various tissues.

PHYSIOLOGY

Only the pancreas synthesizes significant amounts of PTL, and this occurs in the acinar cells of the pancreas. In addition to this tissue-specific expression, mRNA encoding PTL shows temporal and cell-specific regulation. The fetal pancreas of humans and rodents does not express mRNA for PTL. After birth, the pancreas of suckling mouse or rat pups does not express PTL mRNA until the suckling–weanling transition.

Copyright 2004, Elsevier (USA). All rights reserved.

credited with these successes. However, long-term follow-up will be required before any conclusions can be drawn regarding the efficacy of this approach.

See Also the Following Articles

Diabetes Mellitus • Diabetic Neuropathies • Endocrine Pancreas • Exocrine Pancreas • Transplantation Immunology

Further Reading

Gruessner, A. C., and Sutherland, D. E. (2000). Pancreas transplant outcomes for United States (US) cases reported to the United Network for Organ Sharing (UNOS) and non-US cases reported to the International Pancreas Transplant Registry (IPTR) as of October, 2000. *Clin. Transplant.*, 45–72.

Hricik, D. E. (2000). Kidney–pancreas transplantation for diabetic nephropathy. *Semin. Nephrol.* **20**, 188–198.

Ojo, A. O., Meier-Kriesche, H. U., Hanson, J. A., *et al.* (2001). The impact of simultaneous pancreas–kidney transplantation on long-term patient survival. *Transplantation* **71**, 82–90.

Robertson, R. P., Davis, C., Larsen, J., Stratta, R., and Sutherland, D. E. (2000). Pancreas and islet transplantation for patients with diabetes. *Diabetes Care* **23**, 112–116.

Rosenberg, L. (1998). Clinical islet cell transplantation: Are we there yet? *Int. J. Pancreatol* **24**, 145–168.

Shapiro, A. M., Lakey, J. R., Ryan, E. A., *et al.* (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N. Engl. J. Med.* **343**, 230–238.

Sutherland, D. E., Gruessner, R. W., Dunn, D. L., *et al.* (2001). Lessons learned from more than 1,000 pancreas transplants at a single institution. *Ann. Surg.* **233**, 463–501.

Sutherland, D. E., Gruessner, R. W., and Gruessner, A. C. (2001). Pancreas transplantation for treatment of diabetes mellitus. *World J. Surg.* **25**, 487–496.

Pancreatic Triglyceride Lipase

MARK E. LOWE

Washington University School of Medicine, St. Louis, Missouri

dietary fats Predominantly triglycerides, which consist of acyl chains or fatty acids linked to glycerol through ester bonds.

lipase Enzyme that catalyzes the hydrolysis of ester bonds in dietary fats to release fatty acids.

pancreatic triglyceride lipase Specific enzyme produced and secreted by pancreatic acinar cells; primarily responsible for intestinal digestion of triglycerides.

Pancreatic triglyceride lipase is the major pancreatic lipase synthesized and secreted by pancreatic acinar cells. It accounts for most of the luminal digestion of triglycerides in the small intestine. This function is critical for the utilization of dietary triglycerides, because intestinal enterocytes cannot absorb long-chain triglycerides unless lipases convert them to fatty acids and monoacylglycerols.

INTRODUCTION

The digestion of dietary fats in humans begins in the stomach, where gastric lipase (a distinct protein) releases about 15% of the fatty acids from triglycerides. Lipases secreted by the pancreas complete dietary fat digestion in the small intestine. Of these lipases, pancreatic triglyceride lipase (PTL) predominates, as evidenced by patients with congenital deficiency of PTL who malabsorb 50–60% of dietary fats. Cholesterol esterase, another nonspecific lipase, mediates the hydrolysis of other dietary lipids. Other more distantly related lipases, such as lipoprotein lipase and hormone-sensitive lipase, mediate the uptake and release of fatty acids from various tissues.

PHYSIOLOGY

Only the pancreas synthesizes significant amounts of PTL, and this occurs in the acinar cells of the pancreas. In addition to this tissue-specific expression, mRNA encoding PTL shows temporal and cell-specific regulation. The fetal pancreas of humans and rodents does not express mRNA for PTL. After birth, the pancreas of suckling mouse or rat pups does not express PTL mRNA until the suckling–weanling transition.

Copyright 2004, Elsevier (USA). All rights reserved.

Human newborns have decreased lipase activity in pancreatic secretions and may also have decreased expression of PTL. Once expressed, mRNA encoding PTL appears in the pancreatic acinar cells and not in other cell types in the pancreas. Pancreatic acinar cells synthesize and secrete PTL into the pancreatic duct through both basal and regulated pathways. In the common duct, PTL mixes with biliary lipids and bile salts before entering the duodenum, where lipolysis occurs. PTL is active at the neutral pH normally present in the intestinal lumen. It can be inactivated by low pH, as occurs with gastric hypersecretion.

LIPOLYSIS

PTL is a carboxyl esterase that prefers acylglycerides to other dietary lipids, such as phospholipids, cholesterol esters, and galactolipids. PTL efficiently hydrolyzes a broad range of acyl chains of varying length and saturation from the 1 and 3 positions of tri- and diglycerides, producing fatty acids and 2-monoacylglycerides. *In vitro*, PTL cleaves acyl chains from C_{14} to C_{22} carbon chains with only a sixfold difference in rates between the best and worst substrates.

Several properties of PTL distinguish it from other enzymes. Like all lipases, PTL has low activity against water-soluble substrates and has much higher activity against water-insoluble substrates at oil–water interfaces, such as those presented by emulsions of dietary lipids. Paradoxically, many of the usual constituents of the duodenum, such as bile salts, phospholipids, proteins, and polysaccharides, inhibit PTL. Another pancreatic protein, colipase, restores activity to PTL under these conditions by anchoring it at the lipid–water interface.

PROTEIN STRUCTURE

The primary structures of PTL and of colipase from multiple species have been solved by chemical methods or predicted from their cDNA sequence. Both proteins are synthesized with 17-amino-acid signal peptides. Mature human PTL contains 449 amino acids (molecular mass, 49,558 Da) and human colipase contains 95 amino acids (molecular mass, 10,104 Da). In contrast to many other pancreatic exocrine proteins, PTL is not synthesized as an inactive proform or zymogen. Colipase is secreted as a proform, called procolipase, that contains a five-amino-acid propiece, which is cleaved to produce colipase. Unlike the exocrine proenzymes, procolipase does function and can restore activity to bile-salt-inhibited PTL.

The three-dimensional structure of PTL without colipase and two structures of the PTL–colipase complex have greatly increased our understanding of PTL function. PTL consists of two distinct domains, an N-terminal α/β hydrolase fold and a C-terminal β-sheet structure with homology to the C2 domains that are found in a wide range of proteins. Colipase binds to the C-terminal domain of PTL. The active site of PTL resides in the N-terminal domain and contains a glutamic acid–histidine–serine catalytic site. In the PTL structure and in one of the PTL–colipase structures, a surface loop covers the active site and prevents substrate from entering the active site. In the other PTL–colipase structure, the "lid" has moved into a position that opens and configures the active site and, together with colipase, presents a large hydrophobic surface that serves as the lipid binding site.

See Also the Following Articles

Amylase • Fat Digestion and Absorption • Pancreatic Digestive Enzymes • Pancreatic Enzyme Secretion (Physiology) • Trypsin

Further Reading

Brockman, H. (2001). Pancreatic lipase: Physiological studies. *In* "Intestinal Lipid Metabolism" (C. Mansbach II, P. Tso, and A. Kukis, eds.), pp. 37–54. Kluwer Academic/Plenum Publ., New York.

Lowe, M. (2001). Molecular mechanisms of pancreatic lipase and colipase. *In* "Intestinal Lipid Metabolism" (C. Mansbach II, P. Tso, and A. Kukis, eds.), pp. 37–54. Kluwer Academic/Plenum Publ., New York.

Pancreatic Tumors, Other

R. Todd Frederick and John M. Carethers
University of California, San Diego

pancreatic cystic neoplasms Pancreatic tumors that arise from parenchymal cells and often form cystic shapes seen in imaging studies.

pseudocyst A non-epithelial-lined collection of pancreatic secretions.

The vast majority of pancreatic tumors are adenocarcinomas arising from the ductal epithelium. Pancreatic tumors other than ductal adenocarcinoma are rare and can be broadly classified as cystic versus solid lesions. The primary cystic neoplasms of the pancreas represent the majority of these tumors; other rare solid tumors are occasionally seen.

INTRODUCTION AND CLASSIFICATION

Lesions arising from within the pancreas can generally be classified as either cystic or solid and as benign or malignant. However, the distinction between benign and malignant neoplasms can be difficult to make and several tumors represent a spectrum of disease with an indistinct transition point from benign to malignant. The most commonly encountered masses in the pancreas are not tumors or cysts but collections of fluid with a fibrous wall, termed pseudocysts. Pancreatic adenocarcinoma represents the most common neoplasm of the pancreas, accounting for perhaps 90–95% of tumors. The other tumors are all considered rare and can be classified based on appearance as cystic or solid (see Table I). The solid tumors include nonendocrine tumors, metastatic cancers, and endocrine tumors. This article will focus on the primary cystic neoplasms of the pancreas with a brief discussion of the assortment of other solid nonendocrine tumors.

TABLE I Classification of Pancreatic Tumors

A. Cystic
- a. Mucinous cystic neoplasms
  - i. Mucinous cystadenoma
  - ii. Mucinous cystic tumor with moderate dysplasia
  - iii. Mucinous cystadenocarcinoma
- b. Serous cystadenoma
- c. Serous cystadenocarcinoma (case reports)
- d. Intraductal papillary mucinous tumor
- e. Acinar cell cystadenocarcinoma
- f. Cystic choriocarcinoma
- g. Cystic teratoma
- h. Cystic islet cell tumor
- i. Cytic necrosis of adenocarcinoma/lymphoma
- j. Papillary cystic epithelial neoplasm
- k. Angimatous neoplasms
  - i. Angioma
  - ii. Lymphangioma
  - iii. Hemangioendothelioma

B. Solid
- a. Ductal adenocarcinoma
- b. Pancreatic lymphoma (primary vs systemic)
- c. Adenosquamous carcinoma
- d. Giant cell carcinoma (sarcomatoid or osteoclast-like carcinoma)
- e. Acinar cell carcinoma
- f. Pancreaticoblastoma

Cystic Tumors of the Pancreas

The cystic neoplasms were first recognized in 1824 but their classification has continued to evolve for the past 180 years. Initially, cystic tumors were recognized according to the sizes of the cysts, being either microcystic or macrocystic. In 1978, the distinction between two major types of cystic neoplasms was made according to the nature of the cyst contents: serous versus mucinous. A few years later in 1982, the third major tumor in this class was discovered. However, the terminology varied considerably and not until the late 1990s was the term intraductal papillary mucinous tumor (IPMT) widely accepted.

The list of cystic tumors within the pancreas is large but more than 90% of these tumors are composed of mucinous cystic neoplasms, serous cystadenomas, and IPMTs. The remainder are exceedingly rare and limited information is available regarding their nature. These tumors have many similar features in age of onset, clinical presentation, and treatment; however, prognosis varies considerably. The most important differential diagnosis to consider with all cystic tumors is the vastly more common pseudocyst for which detailed historical

 Copyright 2004, Elsevier (USA). All rights reserved.

information regarding pancreatitis or abdominal trauma is paramount. Other findings to suggest pseudocyst rather than cystic tumor include lack of septations, loculations, solid components, or calcifications on computed tomography (CT) or any communication between the cyst and ductal system on endoscopic retrograde cholangiopancreatography (ERCP). Perhaps the best distinguishing feature is the amylase content of the aspirated fluid, generally greater than 5000 IU/liter.

Mucinous Cystic Neoplasms

These tumors represent a spectrum of disease ranging from presumably benign to clearly malignant with local invasion and distant metastases occasionally seen. Previously classified as mucinous cystadenoma with variable degrees of dysplasia versus cystadenocarcinoma, the currently favored term is mucinous cystic neoplasms (MCNs) due to the difficulty in accurately diagnosing these lesions as benign versus malignant.

The MCNs are the most frequently encountered form of cystic neoplasm in the pancreas, representing 45–50% of these tumors. Women are more commonly affected, perhaps as high as 4 : 1, and the median age at diagnosis is 50–55 years. MCNs are usually composed of multiple cysts and range in size from 1 to 26 cm. Benign tumors tend to be smaller (5 to 6 cm), whereas malignant tumors average 8 to 11 cm. The mucinous tumors tend to be "macrocystic" in appearance, with fewer cysts and each cyst greater than 2 cm in size. Solid components and calcifications are occasionally found and usually connote a malignant change. The majority of MCNs are found in the body and tail of the pancreas and rarely connect with the ductal system (5%).

The majority of patients diagnosed with MCNs are symptomatic. Common presenting symptoms include vague abdominal pain, bloating or abdominal fullness, and early satiety. A few patients are able to feel an abdominal mass. Jaundice, pancreatitis, and weight loss are less common to rare. An increasing percentage of patients are asymptomatic and found to have incidental pancreatic masses after abdominal imaging for unrelated conditions. Diagnosis is then made after abdominal imaging is obtained, with ultrasound (US), CT, or magnetic resonance imaging. Although MCNs can be suspected based upon radiologic findings, a definitive diagnosis requires either aspiration of cyst fluid or surgical resection. Large tumors can be percutaneously aspirated safely; however, smaller tumors are increasingly being sampled via endoscopic US (EUS). Characteristic and diagnostic features of MCN fluid include high viscosity, positive mucin staining, low amylase, positive cytology (in 50%), and moderately elevated carcinoembryonic antigen (CEA) >200 ng/ml.

The lining of MCNs is composed of a mucin-producing tall columnar epithelium with a ductal origin. Up to 75% of cysts have been found to have regions of cyst wall denuded of epithelium, involving on average 40% of the wall. In addition, multiple reports of benign, dysplastic, and malignant epithelium coexisting within the same tumor have been described. For these reasons, once an MCN is suspected, surgical resection is strongly advised. This should entail either a distal pancreatectomy with splenectomy or a pancreaticoduodenectomy or Whipple procedure. Limited enucleation is not recommended. Long-term survival is good for patients with resectable disease, ranging from 50 to 75% at 5 years. Studies of adjuvant therapy including chemotherapy and radiation are ongoing and thus their effectiveness is not fully known.

Serous Cystadenoma

As opposed to its mucinous counterpart, the serous cystic tumor represents a more homogenous and predictable entity. Almost universally benign, only six case reports of malignant serous cystadenocarcinomas have been reported.

The serous cystadenomas are probably the second most common cystic tumor of the pancreas, representing 16 to 27% of cystic tumors in most series. Similar to MCNs, women are predominantly affected, again approaching 4:1 in incidence, and age at diagnosis is typically the sixth or seventh decade. These tumors can also be a component in 15% of patients with Von Hippel-Lindau syndrome. Concomitant renal, retinal, or cerebellar vascular tumors are usually found. The overall size of the serous tumors is similar to MCNs at 1 to 26 cm with an average of 6 to 10 cm, but the individual cysts tend to be much smaller, ranging from micrometers to 2 cm in diameter in a honeycomb fashion. The previously descriptive classification of "microcystic" adenoma is no longer used. Unlike the MCNs, serous tumors often have characteristic findings on radiologic imaging. The vascular, fibrous, calcified connective tissue of the tumor creates a central stellate scar with a characteristic "sunburst" pattern on plain films, CT, or angiography. Unfortunately, this pathognomonic finding is present in only 10 to 30% of tumors. Some authors believe that if it is present, further diagnostic testing, such as cyst aspiration, is unnecessary. Although found throughout the gland, these tumors tend to occur in the head of the pancreas more commonly than MCNs.

As with MCNs, the majority of patients diagnosed with serous tumors are symptomatic at presentation.

Symptoms are generally related to mass effect and include abdominal pain, fullness, and the feeling of a mass in up to 33% of patients. Biliary or pancreatic obstruction resulting in jaundice, cholangitis, or pancreatitis is unusual. Weight loss may be present, probably related to early satiety rather than to anorexia as is commonly found with malignancy. The proportion of patients who are asymptomatic at diagnosis will likely continue to increase with the rising use of abdominal imaging modalities for various complaints. As with MCNs, definitive diagnosis of serous cystadenomas usually requires additional testing after initial radiographic studies reveal a cystic mass. Aspiration of cyst fluid is helpful but may be more difficult due to smaller sizes of the individual cysts. The fluid can usually be distinguished from MCNs or pseudocysts by finding low viscosity, negative mucin staining, low CEA < 5 ng/ml, low amylase, and a unique characteristic of positive PAS (periodic acid-Schiff)-staining cells on cytologic analysis due to the abundant glycogen. PAS-positive cells are reported in approximately half of cases and if present, strongly support the diagnosis.

The epithelium of serous cystadenomas is composed of glycogen-rich low cuboidal cells. No mucin is produced. Unlike the MCNs, the progenitor cell is believed to be the centroacinar cell as opposed to ductal epithelial cells. The problem of frequently coexisting benign and dysplastic cells has not been reported for serous tumors. Due to the more benign nature of these tumors, patient management includes more options. Patients who are elderly or otherwise poor surgical candidates and asymptomatic patients may be safely followed with surveillance imaging. If the lesion continues to grow or symptoms develop or worsen, then surgical referral can be reconsidered. Most authors agree that young patients or symptomatic patients should undergo complete resection with either a distal pancreatectomy or a Whipple procedure to eliminate the chance of malignant degeneration as well as potential complications, such as hemorrhage or obstruction. Long-term survival is excellent.

Intraductal Papillary Mucinous Tumor

In the early 1980s, a new subset of the mucinous cystic tumor was noted that tended to involve the ductal system causing ductal dilation as well as acute and chronic pancreatitis. Various names were given to describe this subset of tumors, such as mucinous ductal ectasia, mucin-hypersecreting tumor, mucinous villous adenomatosis, and ductectatic mucinous cystadenoma, to name a few. By the mid-1990s, the consensus was for intraductal papillary mucinous tumor (IPMT). Similar to MCNs, IPMTs encompass a spectrum of disease from entirely benign to frankly malignant with many intermediates in between. Approximately 40% of tumors are malignant at time of presentation.

IPMTs are less common than MCNs, but are similar in incidence to serous tumors, representing the third most common cystic tumor of the pancreas (some authorities list IPMT as the second most common type). The incidence has risen over the past decade, likely due to increasing recognition of this distinct tumor type as well as improved radiographic techniques. Unlike both MCNs and serous tumors, males are more commonly affected with IPMTs. Age of onset is usually over 60 years. The gross and histologic findings are distinct from the other cystic tumors in that there is a papillary growth of tumor within the main or secondary pancreatic ducts, giving rise to main duct or branch duct types. The ducts then become obstructed and subsequently dilated, giving the appearance of a cystic tumor. "Mucin lakes" have been described in the surrounding pancreas and one series reported 20% to have concomitant lesions similar to MCNs, typically in the head of the pancreas. These likely represent the branch duct type with cystic dilations. The other characteristic finding, usually seen at ERCP, is protrusion of the major and/or minor papillae into the duodenum with extrusion of copious amounts of mucin.

As with the other cystic tumors of the pancreas, the majority of patients with IPMT are symptomatic at diagnosis. Unlike other cystic tumors, patients with IPMT frequently suffer from recurrent acute pancreatitis or chronic pancreatitis with episodic or chronic abdominal pain and symptoms related to obstruction of the bile ducts, pancreatic ducts, or small bowel. Weight loss, steatorrhea, and diabetes are not uncommon. Elevations of amylase or lipase are found in up to half of patients at presentation. The tumor markers CEA and CA19-9 are elevated in approximately 10 to 15% but are not felt to be useful in diagnosis due to the low specificity. Diagnosis usually requires advanced endoscopic techniques such as ERCP and EUS with fine-needle aspiration. Characteristic findings include a patulous ampulla with mucus emanating from a dilated orifice, cystic dilation of the main and secondary ducts, filling defects within the ducts, and pancreatic parenchymal atrophy. Strictures are notably absent. ERCP also allows for collection of ductal secretions and forceps biopsy for pathology and cytology. EUS is particularly helpful in distinguishing this tumor from chronic pancreatitis, which can have a similar appearance. These procedures are important for outlining the extent of disease to plan for surgical resection.

The epithelium of the IPMTs is composed of tall columnar mucin-producing cells, similar to MCNs.

These cells tend to form papillary projections, which can also be seen in MCNs. The epithelium frequently undergoes hyperplastic, dysplastic, and carcinomatous changes. Local invasion is not uncommon. The disease typically begins in the head of the pancreas with slow spread throughout the gland. Due to the high likelihood of finding malignant disease at diagnosis or on follow-up, complete resection is strongly recommended for all surgical candidates without evidence of local or metastatic spread. This typically requires a pancreaticoduodenectomy, distal pancreatectomy with splenectomy, or total pancreatectomy. The use of intraoperative frozen sections to outline the extent of malignancy is essential for curative resections. The survival for patients with curative resections is good, with estimates of 60 to 80% at 5 years. Patients with unresectable disease have outcomes similar to those for pancreatic adenocarcinoma.

Other Cystic Lesions

The vast majority of cyst-like lesions within the pancreas are not true cysts or tumors but actually pseudocysts. These are collections of pancreatic secretions related to ductal disruption after acute pancreatitis or pancreatic trauma.

Other cysts within the pancreas are exceedingly rare. These can be classified as congenital/inherited, infectious, and neoplastic. The congenital/inherited lesions include simple cysts, polycystic disease (with or without kidney involvement), Von Hippel-Lindau disease, dermoid cysts, and macrocysts secondary to cystic fibrosis. Infections with parasites such as *Echinococcus granulosis* and *Taenia solium* can also produce cystic lesions within the pancreas. Other neoplastic cystic lesions in addition to the three most common ones described above include acinar cell cystadenocarcinoma, cystic choriocarcinoma, cystic teratoma, cystic islet cell tumors/acinar cell cystadenocarcinoma (usually nonfunctioning), cystic necrosis of adenocarcinoma or lymphoma, papillary cystic epithelial neoplasms, and angiomatous lesions, such as angiomas, lymphangiomas, and hemangioendotheliomas. These tumors are described in the literature with case reports, and only minimal details regarding their clinical behavior are known.

Solid Tumors of the Pancreas

Though the overwhelming majority of solid tumors in the pancreas are adenocarcinomas, a few other types of tumors are known to occur and will be described here.

Lymphomas can involve the pancreas in up to one-third of patients. However, primary lymphoma of the pancreas is a rare tumor, accounting for less than 1% of all extranodal non-Hodgkin's lymphomas and only 0.3% of pancreatic tumors. Abdominal pain and weight loss are common complaints at presentation. Lactate dehydrogenase may be elevated. Due to the large size attained by these tumors, a palpable mass can be felt in over half of patients. Jaundice may be found in one-third. Pancreatitis is uncommon. These tumors are frequently misdiagnosed as adenocarcinoma prior to surgery. A clue to the diagnosis may be an unusually large solid tumor. Therapy consists of combination chemotherapy and radiation but survival is poor.

Two variants of ductal adenocarcinoma include adenosquamous carcinoma and giant cell carcinoma of the pancreas. Both arise from the ductal cell. Adenosquamous cancers have a transition point to squamous epithelium. They tend to be hypervascular. The giant cell carcinoma is also known as sarcomatoid and several subtypes have been described. Of note, these tumors tend to spread hematogenously and generate distant metastases unlike adenocarcinoma. Overall, the clinical presentation, response to therapy, and prognosis are very similar to those for adenocarcinomas.

Acinar cell carcinomas have also been described. These malignant tumors arising from the endocrine portion of the pancreas can present as cystic or solid tumors. Most are nonfunctioning and are discovered incidentally or due to mass effect. Serum lipase levels may be elevated and peripheral fat necrosis may be noted.

Pancreaticoblastoma is a rare tumor usually found in children at a mean age of 4 years. Girls are affected more often than boys. Rare cases can present in adulthood. The tumors tend to be large at the time of presentation and symptoms of pain or fullness are related to mass effect. Distant metastases can develop. Treatment involves wide excision.

See Also the Following Articles

Pancreatic Ductal Adenocarcinoma • Pancreatic Pseudocysts

Further Reading

Farrell, J., and Brugge, W. (2002). Intraductal papillary mucinous tumor of the pancreas. *Gastrointest. Endosc.* **55**, 701–714.

Fernandez-del Castillo, C., and Jimenez, R. (2002). Pancreatic cancer, cystic pancreatic neoplasms, and other nonendocrine pancreatic tumors. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., pp. 970–987. W. B. Saunders, Philadelphia, PA.

Goldschmiedt, M., and Gutta, K. (1998). Other tumors and diseases of the pancreas. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 6th Ed., pp. 895–901. W. B. Saunders, Philadelphia, PA.
Kimmey, M. (2002). Cystic neoplasms of the pancreas. *In* "The American College of Gastroenterology 2002 Postgraduate Course" (P. McNally and G. Zuccaro, eds.), pp. 481–488. American College of Gastroenterology, Arlington, VA.
Saar, M., and Yeo, C. (1994). Cystic and pseudocystic diseases of the pancreas. *Curr. Prob. Surg.* **31**, 167–243.

Pancreatitis, Acute

MICHAEL L. STEER
Tufts University of Medicine and Tufts–New England Medical Center

computed tomography scan Radiographic technique allowing visualization by computer reconstruction of a plane or section through the body.
endoscopic retrograde cholangiopancreatography Procedure whereby the pancreatic ducts are visualized by endoscopic injection of contrast media.
pancreatic pseudocyst Non-epithelial-lined collection of pancreatic secretions.
pancreatitis Inflammatory disease of the pancreas.
serum amylase Pancreatic amylase in the serum; serves as a marker of acinar cell damage, i.e., pancreatitis.
trypsin activation peptide Part of the molecule cleaved from trypsinogen to activate trypsin.
zymogens Proteolytic proenzymes that exist in cells as an inactive precursor; pancreatic zymogens are normally activated by cleavage in the gut lumen.

Pancreatitis is an inflammatory disease of the pancreas. In clinical terms, acute pancreatitis usually develops rapidly and without a prior history of repeated pancreatitis attacks. In contrast, chronic pancreatitis may be characterized by symptoms that slowly develop and a history of multiple previous pancreatitis attacks. Pathologically, acute pancreatitis is defined as pancreatitis that occurs in a previously healthy gland and that, after the acute attack resolves, leaves a pancreas that may be both morphologically and functionally normal. In contrast, the gland is usually abnormal before an attack of chronic pancreatitis becomes clinically apparent. That preexisting abnormality usually involves diffuse fibrosis of the gland with loss of both exocrine and endocrine elements, and those changes persist even after the attack of chronic pancreatitis has resolved.

PATHOLOGY

An attack of acute pancreatitis may be either mild or severe. Pathological changes of mild acute pancreatitis include edema, peripancreatic fat necrosis, and a mild intrapancreatic inflammatory reaction. In contrast, severe pancreatitis is pathologically characterized by varying degrees of parenchymal necrosis along with a severe intrapancreatic inflammatory reaction. Intrapancreatic hemorrhage as well as both intrapancreatic and peripancreatic fluid collections can be seen in severe pancreatitis along with extensive peripancreatic fat necrosis.

ETIOLOGY

Acute pancreatitis typically occurs as a consequence of some other process; collectively, those processes are referred to as the "etiologies" of acute pancreatitis. In developed countries, roughly 80% of patients with acute pancreatitis develop their disease as a result of either biliary tract stone disease or as a result of prolonged and excessive abuse of ethanol. In any particular population, the distribution between these two causes is related to the incidence of those inciting events in that population—i.e., in inner city populations, ethanol abuse is the most common cause, whereas in more affluent suburban groups, biliary tract stones account for most cases of pancreatitis. A number of drugs can also cause pancreatitis (see Table I). Duodenal or pancreatic tumors as well as other lesions causing pancreatic duct

Copyright 2004, Elsevier (USA). All rights reserved.

Goldschmiedt, M., and Gutta, K. (1998). Other tumors and diseases of the pancreas. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 6th Ed., pp. 895–901. W. B. Saunders, Philadelphia, PA.
Kimmey, M. (2002). Cystic neoplasms of the pancreas. *In* "The American College of Gastroenterology 2002 Postgraduate Course" (P. McNally and G. Zuccaro, eds.), pp. 481–488. American College of Gastroenterology, Arlington, VA.
Saar, M., and Yeo, C. (1994). Cystic and pseudocystic diseases of the pancreas. *Curr. Prob. Surg.* **31**, 167–243.

Pancreatitis, Acute

MICHAEL L. STEER
Tufts University of Medicine and Tufts–New England Medical Center

computed tomography scan Radiographic technique allowing visualization by computer reconstruction of a plane or section through the body.
endoscopic retrograde cholangiopancreatography Procedure whereby the pancreatic ducts are visualized by endoscopic injection of contrast media.
pancreatic pseudocyst Non-epithelial-lined collection of pancreatic secretions.
pancreatitis Inflammatory disease of the pancreas.
serum amylase Pancreatic amylase in the serum; serves as a marker of acinar cell damage, i.e., pancreatitis.
trypsin activation peptide Part of the molecule cleaved from trypsinogen to activate trypsin.
zymogens Proteolytic proenzymes that exist in cells as an inactive precursor; pancreatic zymogens are normally activated by cleavage in the gut lumen.

Pancreatitis is an inflammatory disease of the pancreas. In clinical terms, acute pancreatitis usually develops rapidly and without a prior history of repeated pancreatitis attacks. In contrast, chronic pancreatitis may be characterized by symptoms that slowly develop and a history of multiple previous pancreatitis attacks. Pathologically, acute pancreatitis is defined as pancreatitis that occurs in a previously healthy gland and that, after the acute attack resolves, leaves a pancreas that may be both morphologically and functionally normal. In contrast, the gland is usually abnormal before an attack of chronic pancreatitis becomes clinically apparent. That preexisting abnormality usually involves diffuse fibrosis of the gland with loss of both exocrine and endocrine elements, and those changes persist even after the attack of chronic pancreatitis has resolved.

PATHOLOGY

An attack of acute pancreatitis may be either mild or severe. Pathological changes of mild acute pancreatitis include edema, peripancreatic fat necrosis, and a mild intrapancreatic inflammatory reaction. In contrast, severe pancreatitis is pathologically characterized by varying degrees of parenchymal necrosis along with a severe intrapancreatic inflammatory reaction. Intrapancreatic hemorrhage as well as both intrapancreatic and peripancreatic fluid collections can be seen in severe pancreatitis along with extensive peripancreatic fat necrosis.

ETIOLOGY

Acute pancreatitis typically occurs as a consequence of some other process; collectively, those processes are referred to as the "etiologies" of acute pancreatitis. In developed countries, roughly 80% of patients with acute pancreatitis develop their disease as a result of either biliary tract stone disease or as a result of prolonged and excessive abuse of ethanol. In any particular population, the distribution between these two causes is related to the incidence of those inciting events in that population—i.e., in inner city populations, ethanol abuse is the most common cause, whereas in more affluent suburban groups, biliary tract stones account for most cases of pancreatitis. A number of drugs can also cause pancreatitis (see Table I). Duodenal or pancreatic tumors as well as other lesions causing pancreatic duct

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Drugs That May Cause Acute Pancreatitis

Definite causes	Probable causes	Equivocal causes
Furosemide	L-Asparaginase	Rifampin
Valproic acid	Phenformin	H2 blockers
Azathioprine	Acetominophen	Isoniazid
6-Mercaptopurine	Procainamide	Steroids
Methyl-DOPA	Chlorthalidone	Acetaminophen
Tetracycline	Erythromycin	Propoxyphene
Ethycrynic acid	Iatrogenic hypercalcemia	
Sulfonamides		
Metronidazole		
5-Aminosalicylate		
Pentamidine		
Estrogens		
Dideoxyinosine		
Thiazides		

obstruction (e.g., duodenal Crohn's disease, duodenal ulcers, or periampullary diverticulitis) can also precipitate an attack of acute pancreatitis. Some patients have acute pancreatitis on a hereditary basis. Recent studies have identified a number of genetic mutations that are associated with pancreatitis in those individuals. In some people, the mutation results in an abnormal cationic trypsinogen that, once activated, may be resistant to inactivation by trypsin inhibitors in the pancreas. In other individuals, hereditary pancreatitis may be due to genetic defects that result in production of ineffective trypsin inhibitors in the pancreas. Pancreatitis can also be triggered by a number of miscellaneous events, including hyperlipidemia, shock, hypothermia, trauma, scorpion bites, and endoscopic retrograde cholangiopancreatography (ERCP). The anatomic variant pancreas divisum has also been associated with episodes of pancreatitis, but whether it should be considered a cause of acute pancreatitis is a controversial issue. Roughly 15% of patients develop pancreatitis for no identifiable reason. In those cases, the pancreatitis is considered to be "idiopathic." Recent reports have suggested that many of those patients with idiopathic pancreatitis actually have pancreatitis as a result of overlooked biliary tract disease (microlithiasis), an autoimmune process, or cystic fibrosis.

The mechanisms by which a biliary tract stone might trigger acute pancreatitis have been the subject of considerable study and speculation. Opie suggested in 1901 that a stone might migrate into or through the terminal biliopancreatic ductal system, causing obstruction and creating, behind that obstruction, a common bile–pancreatic channel that would permit bile to reflux, retrogradely, into the pancreatic duct. More recent studies, however, have indicated that stone-induced obstruction of the pancreatic duct, rather than reflux of bile into the pancreatic duct, is the event that probably triggers acute pancreatitis. Presumably, the stone, or inflammation caused by passage of a stone, causes obstruction of the distal duct, and continued secretion above that obstruction leads to pancreatic ductal hypertension. By mechanisms that have not been defined, pancreatic ductal hypertension then causes changes within pancreatic acinar cells that result in acinar cell injury and pancreatitis (see below).

The mechanisms by which ethanol abuse might cause acute pancreatitis are not known with certainty, but it is generally believed that ethanol or one of its metabolites triggers pancreatitis by exerting a direct toxic effect on the pancreas, much like a drug. Experimental evidence to support this hypothesis is, however, limited. The chronic pancreatitis that is associated with prolonged and excessive ethanol abuse may reflect changes resulting from repeated subclinical episodes of acute ethanol-induced pancreatitis.

PATHOPHYSIOLOGY

The earliest changes of acute pancreatitis appear to occur within pancreatic acinar cells, and perhaps the earliest of those changes involves intraacinar cell activation of digestive enzyme zymogens. The pancreatic acinar cell synthesizes and secretes a large number and amount of digestive enzymes, and those enzymes play a critical role in digestion. Under normal conditions, the potentially harmful digestive enzymes are synthesized and secreted as inactive proenzymes or zymogens, which become activated only after they traverse the pancreatic ductal system and enter the duodenum. There, the brush border enzyme enterokinase (enteropeptidase) catalyzes the activation of trypsinogen and trypsin activates the other zymogens, including chymotrypsinogen, proelastase, and the procarboxypeptidases. During the earliest stages of pancreatitis, activation of digestive enzyme zymogens, including trypsinogen, occurs within acinar cells. The mechanisms responsible for that intraacinar cell activation of trypsinogen are not entirely clear, but one widely accepted hypothesis suggests that it is catalyzed by lysosomal hydrolases such as cathepsin B and that the activation occurs because lysosomal hydrolases and digestive enzymes become colocalized within membrane-bounded intracellular organelles during the very early stages of pancreatitis. Presumably, intraacinar cell zymogen activation leads to acinar cell injury and that injury then leads to pancreatitis.

CLINICAL PRESENTATION

Symptoms

The classical symptoms of acute pancreatitis are abdominal pain, nausea, and vomiting. The pain is usually experienced in the upper abdomen and usually radiates straight through to the back. It is usually constant and gradually increases in severity. It may be relieved by sitting upright and/or leaning forward. Nausea and vomiting usually follow the onset of pain but the pain is not relieved even by repeated episodes of vomiting. Fever is common but rigors, when they occur, should suggest that biliary obstruction and cholangitis are also present. Clinical jaundice can be present either as a result of biliary tract obstruction or as a result of cholestasis caused by pancreatitis.

Physical Examination

Acute pancreatitis is usually associated with diffuse abdominal tenderness, and both voluntary and involuntary guarding are typically noted on physical examination. An abdominal mass, particularly in the upper abdomen, can sometimes be felt. The patient is frequently described as "writhing about" on the stretcher in search of a position of comfort. Physical changes indicative of severe dehydration may be present either as the result of repeated vomiting or because of loss of fluid into the extravascular spaces. Tachycardia, fever, tachypnea, and hypotension are commonly seen when pancreatitis is severe. When retroperitoneal hemorrhage has occurred, ecchymoses in the flank (Grey–Turner sign) or periumbilically (Cullen's sign) may be observed. Jaundice is not uncommon, but during the early stages of pancreatitis, the jaundice is usually mild.

Blood Tests

Routine blood tests usually reveal an elevated white blood cell count and a so-called shift to the left manifested by an increase primarily in neutrophils. The hematocrit is usually elevated as a result of fluid loss, either externally as the result of vomiting or internally as a result of fluid shifts to the extravascular compartment. Measurement of serum electrolytes in the former case usually reveals changes indicative of a hypochloremic alkalosis whereas serum electrolytes may be normal in spite of severe dehydration if most of the fluid has been lost into the extravascular space. In either case, hypoalbuminemia is common. The bilirubin and transaminases may be elevated, and the alkaline phosphatase may be elevated with these changes due either to extrahepatic biliary obstruction or to the cholestasis of severe illness. Patients with hyperlipidemia-induced pancreatitis typically have very high serum triglycerides during an attack and, not infrequently, lactescent serum can be seen. The classical change in blood chemistry associated with acute pancreatitis is a rise in the serum amylase, lipase, or both. The magnitude of these rises does not indicate the severity of the attack and a substantial fraction of patients may be examined in the hospital setting after these changes have resolved and their amylase or lipase levels have returned to normal. It is important to recognize that an elevated amylase can be noted in a variety of disease states other than pancreatitis (see Table II) and that a rise in serum amylase is therefore not necessarily diagnostic of pancreatitis. Similarly, a normal or near-normal serum amylase value does not exclude the diagnosis of acute pancreatitis.

A number of other tests have been proposed for the diagnosis of acute pancreatitis. These include measurement of trypsinogen activation peptide (TAP) in blood or serum, measurement of serum immunoreactive trypsin, measurement of urine amylase levels, and measurement of the clearance ratio of amylase to creatinine. Unfortunately, none of these tests is without false positives or false negatives and, at present, they would appear to have little to offer for the diagnosis of acute pancreatitis. Recent studies have suggested that a

TABLE II Causes of Elevated Serum Amylase Activity

Pancreatic causes	Intraabdominal nonpancreatic causes	Extraabdominal causes
Acute pancreatitis	Perforated hollow viscus	Salivary gland pathology
Pancreatic ascites	Cholangitis	Burns
Pancreatic trauma	Renal failure	Lung tumors
Chronic pancreatitis	Mesenteric ischemia/infarction	Diabetic ketoacidosis
Pancreatic cancer	Cholecystitis	Pneumonia
Endoscopic retrograde cholangiopancreatography	Ruptured ectopic pregnancy	
	Bowel obstruction	
Pseudocysts	Ovarian cyst	
Duct obstruction		

number of blood tests can be used to predict the severity of pancreatitis. These include measurement of serum interleukins (IL-1 and IL-6), tumor necrosis factor α (TNFα), and C-reactive protein. Changes in the levels of these elements are not specific to pancreatitis and these tests should therefore not be used for the diagnosis of pancreatitis.

Imaging Studies

Plain films of the abdomen usually reveal an ileus pattern. Occasionally, a prominent loop of air-filled jejunum (sentinel loop) near the pancreas is seen in the left upper quadrant of the abdomen, but this is not uniformly the case. On chest X-ray, pleural effusions and lower lobe atelectasis are common. Occasionally, ultrasound examination of the abdomen may be helpful in the diagnosis by revealing pancreatic swelling and the presence of gallbladder and/or bile duct stones. However, ultrasound examination is often incomplete because the dilated, gas-filled bowel loops that are present as the result of pancreatitis-induced ileus may preclude a comprehensive ultrasound exam.

Perhaps the most useful imaging study for acute pancreatitis is the computed tomography (CT) scan, particularly when it is combined with intravenous bolus administration of contrast material. This exam usually detects all but the mildest changes of pancreatitis, including pancreatic swelling, peripancreatic inflammation, the presence of acute peripancreatic fluid collections, and the development of pancreatic necrosis. The latter change is manifested by areas of non-contrast-perfused pancreatic tissue. The contrast-enhanced CT scan can be particularly useful by excluding other causes of severe abdominal pain and peritonitis, including perforated viscus and ischemic bowel. The finding of a normal pancreas on contrast-enhanced CT examination in the presence of clinical changes suggestive of severe pancreatitis should alert the clinician to the high likelihood that the patient does not, indeed, have pancreatitis.

Prognosis

Several scoring systems have been used in an attempt to determine, early in the course of pancreatitis, whether the attack is mild or severe. These include the Ranson system (see Table III), the Glasgow system, the CT scoring system, and the Acute Physiology and Chronic Health Evaluation (APACHE) scoring system. Obesity has also been identified as an independent risk factor. It is also likely that an experienced clinician using simple "good judgment" can also discriminate between mild and severe pancreatitis with similar accuracy. Patients with severe pancreatitis, for the most part, experience a complex disease with secondary failure of other organs (lungs, kidney, etc.), which often requires prolonged intensive care, may necessitate repeated operations and may be associated with mortality rates of 20–40%. On the other hand, mild pancreatitis is truly a mild disease with recovery in several days, little morbidity, and negligible mortality.

TREATMENT

Initial Treatment

The initial treatment of acute pancreatitis includes the firm establishment of a diagnosis. A number of other serious problems, including perforated viscus, ischemic or infarcted bowel, and bowel obstruction, can masquerade as pancreatitis, and making the firm diagnosis of acute pancreatitis may not be a trivial matter. Usually, patients present with classical symptoms, physical findings, and laboratory changes, including hyperamylasemia. An early CT scan can be used to confirm the diagnosis. However, when doubt persists, a diagnostic laparotomy, to exclude other causes of abdominal pain, may be required.

Once the diagnosis of acute pancreatitis is established, attention should be directed at relieving pain and providing fluid resuscitation. The pain of pancreatitis may be severe and difficult to control. It is generally

TABLE III Ranson's Signs for the Prognosis of Acute Pancreatitis[a]

On admission	During initial 48 hours
Age Over 70 Years	Hematocrit decrease over 10% with rehydration
White blood cell count over 18,000/mm^3	Blood urea nitrogen rise greater than 2 mg/dl
Blood glucose over 20 mg/dl	Serum calcium below 8 mg/dl
Lactate dehydrogenase over 400 IU/liter	Arterial oxygen below 60 mmHg
Glutamic aspartate aminotransferase above 250 U/liter	Base deficit over 5 mEq/liter
	Fluid sequestration over 4 liters

[a]Adapted from Ranson (1979). The total number of positive items correlates with morbidity and mortality.

believed that meperidine, rather than morphine, is the opiate of choice for pain control in pancreatitis. Fluid requirements, particularly in severe pancreatitis, can be enormous, and, in many ways, resemble those of a severe burn. The hematocrit elevation can be used to estimate the intravascular fluid deficit, and that deficit should be replaced with an isotonic solution containing albumin. Usually, the standard serum electrolytes are normal and fluid replacement can be achieved with either lactated Ringer's solution or normal saline. Hypocalcemia can be present and, if ionized serum calcium levels are depressed, calcium administration may be indicated. Hypomagnesemia, particularly in patients with a history of ethanol abuse, may also be present and require treatment.

Adequate fluid resuscitation is perhaps the single most important element in the early treatment of severe pancreatitis, and there is evidence that inadequate fluid repletion can markedly worsen the severity of a pancreatitis attack. Proper administration of fluid in this setting may require placement of a central venous or pulmonary artery pressure monitor. This is especially true in patients with comorbidities such as cardiac disease and in those with adult respiratory distress syndrome (ARDS), because overhydration in these patients may also create serious problems.

Prophylactic antibiotics may be helpful in the early treatment of patients with severe acute pancreatitis. Several reports have indicated that early administration of broad-spectrum antibiotics that penetrate the pancreas may decrease either the mortality or the morbidity of severe pancreatitis by reducing the incidence of later septic complications. This is a controversial issue, however, because the administration of prophylactic antibiotics may favor the emergence of resistant organisms, particularly fungi, in the inflamed pancreas, actually worsening the clinical course.

Patients with severe pancreatitis may suffer from a prolonged illness, and provision of adequate nutrition is therefore essential. The early institution of total parenteral nutrition in these patients has gained wide acceptance, although, at present, there is considerable interest in providing nutrition via an enteral route. Early enteral nutrition, provided via a nasoenteric or nasogastric tube, may also be beneficial in preventing bacterial translocation from the bowel to the pancreatic bed, thus, in this manner, reducing the incidence of pancreatic infection. Randomized studies to evaluate this possibility are currently underway.

A number of other therapies for pancreatitis have been proposed, frequently on the basis of preclinical animal experiments, but none of these interventions has been shown to be of benefit. Included among these unproved therapies are peritoneal lavage, plasma ultrafiltration, nasogastric suction, and administration of variety of agents, including atropine, glucagon, calcitonin, somatostatin, steroids, protease inhibitors, heparin, and platelet-activating factor antagonists. There is currently great interest in the possibility that agents that modify the inflammatory response by interfering with cytokine/chemokine action may be useful in the early treatment of pancreatitis, and it is likely that prospective trials of such agents will be forthcoming. However, at present, no agents directed at altering the inflammatory response have been of proved benefit in the treatment of severe acute pancreatitis.

The role of early (i.e., within 48–72 hours of the onset of an attack) endoscopic biliary stone clearance, achieved by ERCP, with or without endoscopic sphincterotomy, has been the subject of several prospective randomized trials. Each of those studies has shown that early stone clearance is of no benefit in mild pancreatitis. On the other hand, in severe pancreatitis, early stone clearance has been shown to be of benefit in two studies but of no benefit in a third study. Thus, at present, the role of early stone clearance by ERCP and endoscopic sphincterotomy is controversial.

Late Treatment of Complications

Acute pancreatitis may be complicated by the late development of pancreatic infection or by the evolution of a pancreatic pseudocyst. Pancreatic infection, when it involves infection of either necrotic pancreatic tissue or necrotic peripancreatic tissue, usually requires surgical debridement. For the most part, this is accomplished by repeated operations, during which the infected necrotic tissue is removed and drains are placed. Recently, attempts to accomplish this by percutaneous or even endoscopic methods have also been described, but the role of those methods remains to be established.

Pancreatic pseudocysts are collections of fluid that are contained within the boundaries created by adjacent structures. Most communicate with the pancreatic ductal system and contain fluid with high concentrations of pancreatic digestive enzymes. They may be colonized by bacteria, but clinical infection, with signs of sepsis, is usually absent. When clinical infection is present, a pancreatic abscess (i.e., infected pseudocyst) should be suspected. Pancreatic pseudocysts that are asymptomatic and of stable size do not require treatment. When infected, they should be drained, usually by percutaneous methods externally. It is generally believed that, in the absence of infection, pseudocysts that are symptomatic and/or enlarging after 6 weeks of observation should be

treated, but there is no agreement as to the ideal method of treating such pseudocysts. Successful treatment has been reported using percutaneous drainage (i.e., with CT or ultrasound guidance), endoscopic internal drainage (i.e., with endoscopic ultrasound guidance), or by surgical internal drainage (i.e., by cyst gastrostomy or cyst Roux-en-Y jejunostomy). In practice, the approach is usually determined by the local expertise available at the treating institution.

See Also the Following Articles

Amylase • Computed Tomography (CT) • Pancreatic Digestive Enzymes • Pancreatic Function Tests • Pancreatic Pseudocysts • Pancreatic Triglyceride Lipase • Pancreatitis, Experimental Models • Trypsin • Ultrasonography

Further Reading

Banks, P. A. (1993). Medical management of acute pancreatitis and complications. *In* "The Pancreas: Biology, Pathobiology and Disease" (V. L. W. Go, E. P. DiMagno, J. D. Gardner, E. Leventhal, H. A. Reber, and G. A. Scheele, eds.), pp. 593–612. Raven Press, New York.

Beger, H. G., Buchler, M., Bittner, R., Block, S., Nevalainen, T., and Roscher, R. (1988). Necrosectomy and postoperative local lavage in necrotizing pancreatitis. *Br. J. Surg.* **75**, 207–212.

DiMagno, E. (1993). Treatment of mild pancreatitis. *In* "Acute Pancreatitis: Diagnosis and Therapy" (E. Bradley III, ed.), pp. 261–264. Raven Press, New York.

Kloppel, G. (1999). Progression from acute to chronic pancreatitis. A Pathologist's view. *Surg. Clin. North Am.* **79**, 801–814.

Leavitt, M. D., and Eckfeldt, J. H. (1993). Diagnosis of acute pancreatitis. *In* "The Pancreas: Biology, Pathobiology, and Disease" (V. L. W. Go, E. P. DiMagno, J. D. Gardner, E. Leventhal, H. A. Reber, and G. A. Scheele, eds.), pp. 613–636. Raven Press, New York.

Ranson, J. H. C. (1979). The timing of biliary surgery in acute pancreatitis. *Ann. Surg.* **189**, 654–663.

Russell, C. (1998). Long-term prognosis after acute pancreatitis. *In* "The Pancreas" (H. G. Beger, A. L. Warshaw, M. Buchler, D. L. Carr-Locke, J. P. Neoptolemos, C. Russell, and M. G. Sarr, eds.), pp. 656–664. Blackwell Science, London.

Sarr, M. G., Nagorney, D. M., Mucha, P. J. R., Farnell, N. B., and Johnson, C. D. (1991). Acute necrotizing pancreatitis: Management by planned, staged pancreatic necrosectomy/debridement and delayed primary wound closure over drains. *Br. J. Surg.* **78**, 576–581.

Sigurdsson, G. I. H. (1998). Necrotizing pancreatitis: Intensive-care measures. *In* "The Pancreas" (H. G. Beger, A. L. Warshaw, M. W. Buchler, D. L. Carr-Locke, J. P. Neoptolemos, and C. Russell, eds.), pp. 508–520. Blackwell Science, London.

Steinberg W., and Tenner, S. (1994). Acute pancreatitis. *N. Engl. J. Med.* **330**, 1198–1210.

Vitas, G. J., and Sarr, M. G. (1992). Selected management of pancreatic pseudocysts: Operative versus expectant management. *Surgery* **111**, 123–130.

Pancreatitis, Chronic

CHRIS E. FORSMARK
University of Florida, Gainesville

cholecystokinin A gastrointestinal hormone that stimulates pancreatic secretion of digestive enzymes as well as gallbladder contraction.

endoscopic retrograde cholangiopancreatography A procedure utilizing an endoscope to inject radiographic contrast into the pancreatic and bile ducts. Treatment of certain problems including ductal strictures and removal of ductal stones is possible with this technique.

endoscopic ultrasonography A procedure utilizing an endoscope with an ultrasound probe mounted on the tip, which allows highly detailed examination of the pancreas and surrounding structures.

neurolysis Destruction of nerves to produce pain relief

pseudoaneurysm Aneurysmal dilation of a peripancreatic artery produced by the local injurious effects of a pseudocyst.

pseudocyst A collection of fluid produced by leakage of pancreatic juice and surrounded by a capsule of fibrous and granulation tissue.

secretin A gastrointestinal hormone that stimulates pancreatic secretion of fluid and bicarbonate.

steatorrhea The presence of significant undigested fat in the stool. A marker of inadequate or incomplete digestion of fat.

trypsinogen A pancreatic digestive enzyme that, once activated, has the capacity to activate all the other pancreatic digestive enzymes.

trypsin inhibitor protein (also known as SPINK-1) A pancreatic secretory protein that inhibits the action of activate trypsin and functions as a protective mechanism to prevent inadvertent activation of digestive enzymes within the pancreas.

Chronic pancreatitis is characterized by inflammation, fibrosis, and eventually destruction of both exocrine and endocrine tissue. By definition, the damage is permanent and irreversible. This histological definition, although still widely accepted, is difficult to apply in practice, as pancreatic tissue is rarely available for evaluation. Over a number of years, a series of international panels have attempted to draw up definitions for chronic pancreatitis but these have also been based on histology. More recent attempts to define and categorize chronic pancreatitis have focused on the results of diagnostic tests or on etiology, bypassing the difficult issue of defining the disease. For many clinicians, the definition has been based on abnormalities of the pancreas visible on imaging studies such as computed tomography, ultrasound, or endoscopic retrograde cholangiopancreatography. Unfortunately, these abnormalities (such as diffuse pancreatic calcifications or a dilated pancreatic duct) may take years to develop and may not even develop in some patients. Defining the disease based on imaging studies will therefore miss many patients with the disease. More recent attempts to categorize chronic pancreatitis have focused more on etiology, but no single, widely accepted method for defining and categorizing chronic pancreatitis exists.

DEMOGRAPHIC FINDINGS

Estimates of the incidence and prevalence of chronic pancreatitis vary widely. Autopsy studies suggest a prevalence of 0.04–5%. Autopsy studies may overestimate true prevalence as histologic changes of chronic pancreatitis can be found in certain subgroups (e.g., longstanding alcohol use, very elderly) in whom no symptoms of chronic pancreatitis were present in life. Incidence studies range from less than 1 to more than 10 new cases per 100,000 population. The single prospective study that was performed found a prevalence of 27.4 cases per 100,000 population and an incidence of 8.2 new cases per 100,000 population. In the United States, this accounts for more than 60,000 hospital admissions yearly in which chronic pancreatitis is one of the discharge diagnoses.

Most of these demographic data are based on patients with alcohol-induced chronic pancreatitis, as these are the easiest to diagnose using routinely available imaging studies. Patients with other forms of pancreatitis (e.g., idiopathic chronic pancreatitis) and those with less advanced cases of alcoholic chronic pancreatitis may lack these easily identifiable abnormalities and so may not be included in epidemiological studies. The true incidence and prevalence are therefore probably higher than the available estimates.

Chronic pancreatitis produces significant morbidity, mortality, and health care costs. A recent analysis of both acute and chronic pancreatitis estimated total

 Copyright 2004, Elsevier (USA). All rights reserved.

costs at 2.5 billion dollars yearly. Morbidity and mortality occur as a consequence of the disease itself, the therapy used to treat it (e.g., surgery), or ongoing alcohol abuse (in those with alcohol-induced chronic pancreatitis). Abdominal pain, maldigestion with weight loss, diabetes mellitus, secondary cancers, and continued alcohol abuse are the major negative influences on quality and length of life. Overall, 10-year mortality is approximately 30% and 20-year mortality is approximately 50% or greater. Death is often not due to chronic pancreatitis but rather to other medical conditions (cirrhosis, malignancy, and vascular disease), continued alcohol abuse, or postoperative complications.

ETIOLOGY AND PATHOPHYSIOLOGY

Despite intense effort, no clear unifying pathophysiologic mechanisms have been identified. Although all etiologies of chronic pancreatitis may produce similar histologic damage, the pathophysiologic mechanism may vary from etiology to etiology. The ultimate effect is damage to the pancreatic acini, ducts, nerves, and islet cells and the development of the cardinal manifestations of abdominal pain, maldigestion, and diabetes mellitus. The etiologies of chronic pancreatitis are listed in Table I and are discussed below.

Alcohol Consumption

Alcohol consumption is the major cause of chronic pancreatitis. Although there is no "safe" level of intake below which chronic pancreatitis does not occur, prolonged alcohol intake is usually required (e.g., 4 pints of beer or 800 ml of wine per day for 6–12 years). Only a minority of alcoholics (approximately 15%) drinking this much will actually develop chronic pancreatitis, suggesting that an important cofactor (such as diet or genetics) exists. In Westernized societies, 70% of cases of chronic pancreatitis are due to alcohol consumption, with the remaining 30% due to other causes or idiopathic disease. The mechanism by which alcohol produces pancreatic injury and chronic pancreatitis is unknown.

By the time the first clinical attack of alcoholic pancreatitis occurs, most patients have already developed histologic changes of chronic pancreatitis. Continued alcohol abuse leads to ongoing pancreatic damage although damage may continue even with complete abstinence, albeit at a slower rate. The prognosis of alcoholic chronic pancreatitis is poor, with the frequent development of exocrine or endocrine insufficiency and increased mortality due to the consequence of continued alcohol abuse.

TABLE I Causes of Chronic Pancreatitis

Alcohol abuse
Tropical pancreatitis
Obstruction of pancreatic duct
Trauma to pancreatic duct
Benign or malignant ductal stricture
Consequences of pancreatic duct stent
Pancreas divisum
Genetic causes
Hereditary pancreatitis (trypsinogen gene mutations and others)
Cystic fibrosis
Autoimmune chronic pancreatitis
Recurrent or severe acute pancreatitis
Idiopathic pancreatitis
Early-onset
Late-onset

Tropical Pancreatitis

Tropical pancreatitis is the most common form of chronic pancreatitis in certain areas of Indonesia, India, and Africa. The most classic form includes abdominal pain beginning in childhood, with subsequent diabetes, malnutrition, and diffuse pancreatic calcifications. Less severe forms also seem to exist. Most people ultimately die from complications of the disease. Malnutrition is felt to be important in the development of tropical chronic pancreatitis, perhaps due to deficiencies in trace elements or antioxidants. Toxic products contained in the diet (e.g., cassava or sorghum) or the environment or genetics may also play a role in pancreatic injury.

Pancreatic Duct Obstruction

Any condition that chronically obstructs the main pancreatic duct can lead to chronic pancreatitis in the gland "upstream" from the obstruction. Obstruction may be due to benign strictures, ampullary stenosis or neoplasms, pancreatic tumors or pseudocysts, congenital variants, or endoscopically placed pancreatic duct stents. Long-standing obstruction leads to irreversible chronic pancreatitis, but both functional and structural improvement can be seen if the obstruction is discovered and relieved.

One form of pancreatic duct obstruction deserves specific mention. Pancreas divisum, a congenital failure of fusion of the dorsal and ventral pancreatic ducts, may lead to obstruction of the pancreatic duct and chronic pancreatitis. The small ventral pancreas drains through the larger major papilla and the larger dorsal pancreas drains through the smaller accessory papilla. Pancreas

divisum occurs in up to 7% of the population. Although the vast majority of patients with this congenital abnormality have no symptoms, a small subset of patients will develop acute or chronic pancreatitis.

Genetic Forms of Chronic Pancreatitis

Hereditary Pancreatitis

Several kindreds from around the world with acute and chronic pancreatitis have been described. The pattern of inheritance is autosomal dominant with incomplete penetrance and the typical clinical features are recurrent acute pancreatitis beginning at an early age, culminating in advanced chronic pancreatitis. Pancreatic cancer is a common complication, occurring in up to 40% of these patients. There are a number of genetic defects that have been identified in these kindreds with hereditary pancreatitis. Functional genomic studies initially identified a single point mutation in the cationic trypsinogen gene. Trypsinogen is a proenzyme that, once activated, has the ability to activate not only itself but other digestive enzymes as well. In the normal state, trypsinogen is activated only within the duodenum, safely away from the pancreas. Trypsinogen may, however, undergo a very slow autoactivation within the pancreas. A number of mechanisms exist to prevent this activated trypsin from activating other digestive enzymes within the pancreas and causing pancreatitis. Based on molecular modeling, the mutation was thought to convey a gain-of-function mutation, in which the mutated trypsinogen was resistant to inactivation once activated. Trypsin, the activated form of trypsinogen, could then activate all of the other pancreatic digestive enzymes within the pancreas. This continual low-grade activation of digestive enzymes within the pancreas is believed to produce ongoing damage, which ultimately produces severe chronic pancreatitis. Since this initial discovery, several other gene mutations have been identified in the cationic trypsinogen gene as well as within an enzyme meant to inactivate trypsin, the trypsin inhibitor protein (serine protease inhibitor kazal type 1 or SPINK-1). Studies have not yet identified these gene mutations as being important in patients with other forms of chronic pancreatitis (idiopathic and alcoholic chronic pancreatitis).

Cystic Fibrosis

In children, cystic fibrosis is the most common cause of pancreatic insufficiency. Precipitation of protein plugs and inspissated mucus occur in the pancreatic duct, much as in the bronchioles, leading to damage to the gland due to chronic obstruction. Recent studies have noted cystic fibrosis gene mutations in patients with acute and chronic pancreatitis who have no pulmonary or sinus conditions commonly associated with cystic fibrosis. In some studies, at least half of patients initially diagnosed with "idiopathic" chronic pancreatitis have at least one cystic fibrosis allelic mutation. Many of these mutations are not measured on commercially available screens and complete genetic analysis is required to identify the mutations.

Rare Forms of Chronic Pancreatitis

Autoimmune Chronic Pancreatitis

Chronic pancreatitis may rarely be seen in association with autoantibodies, elevated levels of immunoglobulins, and dense lymphocytic infiltrate within the pancreas (if histology is available). In over half of these cases, other autoimmune diseases coexist including primary sclerosing cholangitis, primary biliary cirrhosis, autoimmune hepatitis, and Sjogren's syndrome.

Recurrent or Severe Acute Pancreatitis

A very severe attack of acute pancreatitis or multiple less severe attacks may produce enough damage to the gland to result in chronic pancreatitis. Most commonly, this occurs after the development of a stricture in the pancreatic duct that produces continuing damage to the gland upstream of the stricture.

Hypertriglyceridemia

Triglyceride levels above 1000 mg/dl can initiate acute pancreatitis. Recurrent attacks of hyperlipidemic pancreatitis may ultimately produce chronic pancreatitis. This occurs most commonly as a result of familial hyperlipidemias (types IV and V) exacerbated by estrogen use or poorly controlled diabetes. This cause of both acute and chronic pancreatitis should not be forgotten, as effective therapy is available.

Idiopathic Chronic Pancreatitis

Between 10 and 30% of patients will not have an easily definable cause of chronic pancreatitis. Idiopathic chronic pancreatitis is the second most common cause of chronic pancreatitis in adults, behind alcohol abuse. Some of these patients may actually be alcoholics and may be missed if this history cannot be elucidated. Some may have cystic fibrosis gene mutations, as noted above. Two different forms of idiopathic chronic pancreatitis have been described, a late-onset form and an early-onset form. The late-onset form occurs at a mean age of 56 years, commonly presenting with steatorrhea or diabetes and less commonly with abdominal pain. The early-onset form begins at approximately age 20 and is

characterized by severe pain in essentially all patients but very infrequently by steatorrhea or diabetes. Many of these patients with early-onset idiopathic chronic pancreatitis do not have easily identifiable abnormalities of the pancreas on imaging studies [computed tomography (CT), ultrasound (US), or endoscopic retrograde cholangiopancreatography (ERCP)] and are commonly misdiagnosed.

CLINICAL FEATURES

Abdominal Pain

Abdominal pain is the predominant symptom of chronic pancreatitis and the one that most adversely affects quality of life. The most frequent type of pain is dull, located in the epigastrium with associated back pain, and made worse by eating or lying in the supine position. Food may be avoided, leading to weight loss and malnutrition. Pain may be episodic (lasting from hours to weeks) or more constant or continuous. The pain is generally moderate to severe and narcotic analgesics are required in many patients. Pain is not universal, however. Pain never develops in up to 15% of patients with alcoholic chronic pancreatitis and up to 25% (or more) of patients with late-onset idiopathic chronic pancreatitis. In some patients, the pain may "burn out" after many years of chronic pancreatitis, although this is unpredictable.

There are many potential sources of pain. The most common contributing factors are considered to be inflammation and damage to pancreatic and peripancreatic visceral afferent nerves, pancreatic tissue ischemia, hyperstimulation of the pancreas due to interruption of normal feedback control, and complications of chronic pancreatitis (pseudocyst, common bile duct or duodenal obstruction, superimposed pancreatic carcinoma).

Maldigestion

Exocrine insufficiency (steatorrhea) due to chronic pancreatitis does not occur until the capacity of the pancreas to secrete lipase and colipase is reduced to less than 10% of normal. Maldigestion is thus a marker of far-advanced chronic pancreatitis. Fat maldigestion is more common and more clinically important than protein or carbohydrate maldigestion. Maldigestion is due to diminished secretion of pancreatic enzymes as well as reduced secretion of bicarbonate from the pancreatic ductal system; with reduced bicarbonate, the lower duodenal pH inactivates many pancreatic digestive enzymes. Weight loss is not invariable with maldigestion as intake may be increased, but can occur if food is avoided due to pain or if intake is inadequate due to chronic alcoholism. Folate deficiency may be seen, particularly in chronic alcoholics. Osteopenia and osteoporosis may also develop as a consequence of vitamin D deficiency.

Diabetes Mellitus

Endocrine insufficiency (diabetes mellitus) develops when enough pancreatic islet cells have been destroyed to significantly impair insulin production. The islet cells are more resistant to damage than the acinar and ductal cells, so diabetes is also a marker of far-advanced chronic pancreatitis. This form of diabetes is rarely associated with ketoacidosis but frequent treatment-associated episodes of hypoglycemia can occur (due to inadequate glucagon reserves). Microvascular complications of diabetes (retinopathy or neuropathy) occur as frequently as in other diabetics, if corrected for the duration of disease.

DIAGNOSIS

Clinical features such as abdominal pain, steatorrhea, or diabetes usually suggest the possibility of chronic pancreatitis, and this is confirmed by one of a wide variety of diagnostic tests. The true gold standard diagnostic test, pancreatic histology, is rarely available and alternative tests serve as imperfect substitutes. Most of the diagnostic tests currently in use are imaging studies that detect structural abnormalities within the pancreas. These can include such features as a dilated or irregular pancreatic duct, pancreatic gland atrophy, or diffuse pancreatic calcifications. Chronic pancreatitis can be a slowly progressive disease and these changes may take years to develop. Similarly, functional problems such as exocrine or endocrine insufficiency may not develop for years. In advanced and long-standing disease, easily visible structural and functional abnormalities may be seen and this makes the diagnosis straightforward. In less advanced disease or disease of shorter duration, however, these structural and functional changes may not be present and the diagnosis can be challenging. This has led to a general clinical differentiation: "big-duct disease" and "small-duct disease." Big-duct disease is the presence of significant structural abnormalities of the pancreas (dilation of the main pancreatic duct, diffuse calcifications) and is often associated with exocrine or endocrine insufficiency. Small-duct disease implies the absence of these advanced structural abnormalities and usually the absence of exocrine or endocrine insufficiency. This differentiation is clinically useful; those with big-duct disease have advanced or long-standing disease, are easiest to diagnose, are usually alcoholic,

TABLE II Diagnostic Tests and Studies for Chronic Pancreatitis (Listed in Order of Decreasing Sensitivity)

Tests of function	Procedures examining structure
Secretin or secretin–cholecystokinin test	EUS[a]
Fecal elastase or chymotrypsin	ERCP
Serum trypsin	MRI/MRCP[a]
Fecal fat	CT
Blood glucose	Ultrasound
	Plain abdominal radiograph

[a] Estimated. The sensitivity of EUS may be better than that of ERCP, MRI, and CT. The specificity of EUS is, however, suboptimal and limits its overall accuracy. The sensitivity of MRI/MRCP is not known but is still probably less than that of ERCP using current image technology.

and are most suitable for endoscopic or surgical therapy. Those with small-duct disease are much more difficult to diagnose, are more likely idiopathic, and are most suitable for medical therapy.

Very few diagnostic tests measure abnormalities of pancreatic function. With the exception of direct hormonal stimulation testing, these tests of function are abnormal only in far-advanced disease when exocrine insufficiency has already developed. Diagnostic tests used for chronic pancreatitis are listed in Table II and are discussed below.

Tests of Pancreatic Structure

Plain Abdominal X Rays

The finding of diffuse pancreatic calcification on a plain abdominal radiograph is specific for chronic pancreatitis but is seen only in far-advanced disease. Focal pancreatic calcification may be due to other conditions such as trauma, islet-cell tumors, or hypercalcemia.

Abdominal Ultrasonography and Computed Tomography

A transabdominal ultrasound may note pancreatic parenchymal calcifications, pancreatic atrophy, or a dilated pancreatic duct. Overlying bowel gas can interfere with visualization of the pancreas, making the sensitivity of ultrasound greater than that of plain X rays but less than that of CT. CT findings of chronic pancreatitis include atrophy of the gland, irregular contour of the pancreas, dilation or irregularity of the pancreatic duct, and calcified pancreatic calculi (Fig. 1). CT is also useful in looking for complications of chronic pancreatitis (e.g., pancreatic carcinoma or pancreatic pseudocyst).

Magnetic Resonance Imaging and Magnetic Resonance Cholangiopancreatography

The quality of standard magnetic resonance imaging (MRI) images of the pancreas continues to improve but has not yet reached the overall quality of CT. Magnetic resonance cholangiopancreatography (MRCP) utilizes an anatomic reconstruction of the biliary and pancreatic ductal systems. MRCP is most accurate in advanced chronic pancreatitis but the accuracy with less-advanced chronic pancreatitis is substantially reduced.

Endoscopic Ultrasonography

The endoscopic ultrasonography (EUS) instrument is a high-frequency ultrasound probe mounted on an endoscope. High-resolution images can be acquired of the pancreatic duct, parenchyma, and surrounding structures. EUS is highly accurate in advanced chronic pancreatitis (Fig. 2). The sensitivity and especially the specificity of the test in patients without these advanced abnormalities are not yet known.

Endoscopic Retrograde Cholangiopancreatography

ERCP is one of the most commonly used imaging techniques in the evaluation of patients with presumed chronic pancreatitis. The procedure involves injection of the pancreatic duct with radiographic contrast. The diagnosis of chronic pancreatitis by ERCP is based on changes in the main pancreatic duct and the duct side branches. These changes include duct dilation, stricture formation, irregular duct contour, associated filling of

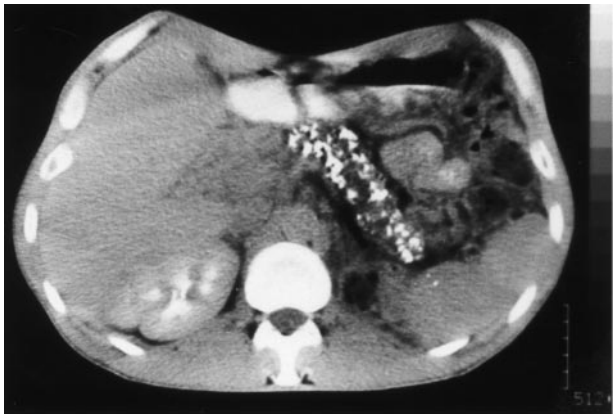


FIGURE 1 An abdominal CT scan demonstrates diffuse and dense calcification within the pancreas of a patient with long-standing alcohol-induced chronic pancreatitis.

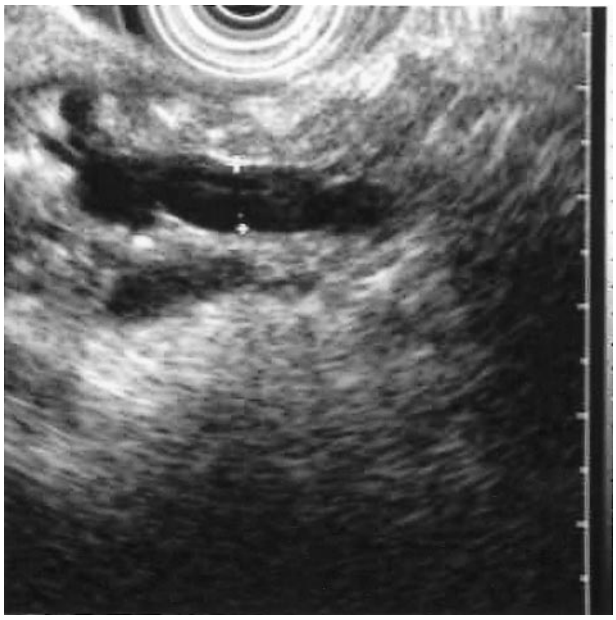


FIGURE 2 An EUS image of chronic pancreatitis. The instrument is seen at the top of the image. The dilated pancreatic duct (measurement markers) is seen at the bottom. The parenchyma of the gland between the scope and dilated duct has irregular echotexture, also consistent with chronic pancreatitis.

cavities or pseudocysts, and filling defects (i.e., pancreatic ductal calculi). ERCP is highly accurate in advanced disease and reasonably accurate in less severe disease (Fig. 3). Some patients, however, can have chronic pancreatitis with a normal ERCP. ERCP is also limited by a number of clinical considerations. In up to one-third of procedures, the ERCP image obtained is of inadequate quality to allow a definitive conclusion. Other conditions can mimic the ERCP changes seen in chronic pancreatitis (pancreatic carcinoma, acute pancreatitis, pancreatic duct stenting, and aging). The procedure is expensive and requires substantial experience and skill and procedure-related complications can occur in up to 10% of patients. Given these factors, ERCP is usually a late step in the evaluation of patients with suspected chronic pancreatitis. ERCP does have one advantage in that therapy can also be accomplished in some patients. This therapeutic, rather than diagnostic, role of ERCP is discussed below.

Test of Pancreatic Function

Laboratory Tests

Serum amylase or lipase levels are often normal and also not of diagnostic importance. Serum trypsin (another pancreatic enzyme) is more useful as a diagnostic test. Levels below 20 mg/dl are seen in advanced chronic pancreatitis (i.e., steatorrhea is present). Pancreatic enzymes may also be measured in stool. Since these enzymes are not reabsorbed, diminished levels in stool are an indirect measure of output of pancreatic enzymes. Both fecal elastase and fecal chymotrypsin have been measured, but only fecal elastase is commercially available. Overall, the sensitivity approaches 90% in advanced disease but is only approximately 50–60% in less advanced disease. This is equivalent to the accuracy of serum trypsin. Measurement of serum glucose or 72 h fecal fat output can document endocrine or exocrine insufficiency but are not useful for diagnostic purposes.

Indirect Tests of Pancreatic Function

There are a number of methods to indirectly measure pancreatic function. Usually, a substrate that requires the presence of pancreatic digestive enzymes within the gut lumen for metabolism is given. Metabolic products are then measured; their relative production rate indirectly reflects pancreatic enzyme secretion. Maldigestion of these substrates does not occur until overall exocrine insufficiency has developed; hence, these tests will be accurate in advanced or end-stage disease but inaccurate in earlier disease. Many variations of this type of test have been developed using a

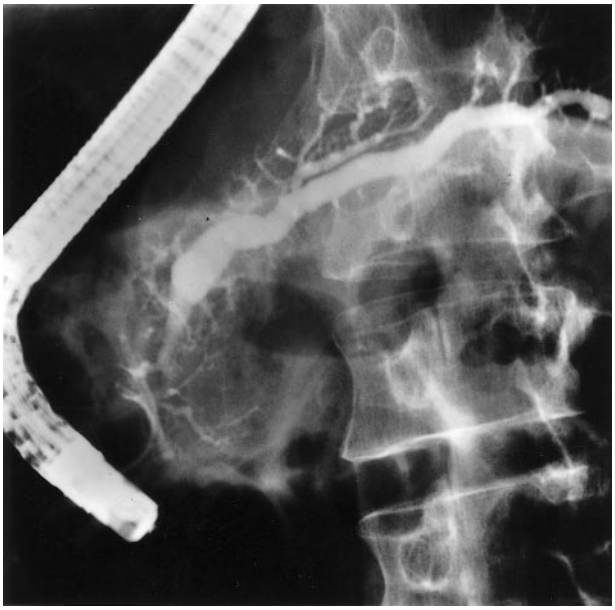


FIGURE 3 An ERCP of a patient with chronic pancreatitis. The scope can be seen, along with a pancreatic duct filled with radiographic contrast. The pancreatic duct is dilated and the side branches of the duct are also dilated and irregular. The findings are consistent with chronic pancreatitis and demonstrate big-duct disease (see text for explanation).

variety of substrates (bentiromide test, pancrealauryl test, Lundh test meal, amino acid consumption test, and dual-label Schilling test). None of these tests is available for routine clinical use.

Direct Tests of Pancreatic Function

These tests measure the actual output from the pancreas after stimulation with a secretagogue. The secretagogues used are secretin, cholecystokinin, or both. Pancreatic secretions are collected with a tube placed in the duodenum. The sensitivity of these tests, like all diagnostic tests, depends on the severity of the disease. The overall sensitivity (74–90%) and specificity (80–90%) are however, superior to those of any other currently available diagnostic test. There are substantial data indicating that direct hormonal stimulation tests can also diagnose chronic pancreatitis at a somewhat earlier stage than any other available test. Unfortunately, these tests are available at only a few referral centers and are unavailable to many clinicians.

Diagnostic Strategy

The ideal test would have high sensitivity (in both small-duct and big-duct disease) and specificity and be inexpensive, safe, and widely available. No ideal diagnostic test exists. The appropriate use of the variety of available tests requires an understanding of their sensitivity, cost, and risk. Initially, tests that are safe, simple, and inexpensive are used. These tests generally are able to identify only patients with advanced or big-duct chronic pancreatitis. Such tests could include serum trypsin, fecal elastase, and US. When these first-echelon tests are not diagnostic, more invasive, risky, or costly tests would be considered. Second-echelon tests include hormonal stimulation tests, CT scan, MRI/MRCP, or EUS. The most invasive and risky test, ERCP, is most commonly used when the diagnosis remains unclear or when therapy, rather than diagnosis, is required.

TREATMENT

Treatment of Exocrine and Endocrine Insufficiency

Steatorrhea

Steatorrhea occurs only when 90% of pancreatic output has been lost, due to inadequate delivery of both lipase and bicarbonate. The acidic environment further inactivates what lipase may be present and precipitates bile salts, worsening fat absorption. The replacement of endogenous digestive enzymes with exogenous enzymes is the goal of therapy. The use of appropriate dosages of these enzymes leads to resolution of diarrhea and weight loss, although steatorrhea is rarely completely corrected.

The various commercially available preparations and their lipase content are listed in Table III. At least 30,000 IU of lipase needs to be delivered to the duodenum during each meal. Patients using the lower-potency preparations must therefore take three to four pills with each meal. Fewer pills must be taken with the more potent formulations. The enteric-coated preparations are protected from gastric acid. The non-enteric-coated preparations will be destroyed by gastric acid and so must be used in conjunction with an agent to reduce gastric acid output (histamine-2 receptor antagonists or proton pump inhibitors). The most common reason for failure of these enzymes to correct steatorrhea is inadequate dose, usually due to the patient's unwillingness to take the number of pills required. Another common reason is inactivation of lipase by gastric acid in those receiving non-enteric-coated preparations. Dietary manipulations may be needed in the management of malabsorption and malnutrition. The diet should usually contain a moderate percentage of fat (30%), a high percentage of protein (30%), and a low percentage of carbohydrates (40%) and should be supplemented with a high-quality multivitamin, calcium, and vitamin D.

TABLE III Commercially Available Pancreatic Enzymes for the Treatment of Steatorrhea[a] or Pain[b]

Brand name	Units of lipase per pill
Non-enteric-coated enzyme preparations	
Viokase, Viokase 16	8,000; 16,000
Ku-Zyme HP	8,000
Generic pancrealipase	8,000
Enteric-coated enzyme preparations	
Creon 5, 10, 20	5,000; 10,000; 20,000
Pancrease MT 4, 10, 16, 20	4,000; 10,000; 16,000; 20,000
Ultrase 6, 12, 18, 20	6,000; 12,000; 18,000; 20,000

[a] For the treatment of steatorrhea, both non-enteric-coated and enteric-coated preparations can be used. The dosage depends on the lipase content; 30,000 units of lipase should be delivered with each meal. Non-enteric-coated enzymes require cotreatment with agents to suppress gastric acid.

[b] For the treatment of pain, only non-enteric-coated enzyme preparations are used: four to eight pills (depending on potency) before meals and at night. As above, an adjuvant agent to reduce gastric acid is required.

Diabetes Mellitus

Diabetes is late complication of chronic pancreatitis. Microangiopathic complications occur with regularity, including retinopathy, nephropathy, neuropathy, and more rapid atherosclerosis. Another common complication is treatment-induced hypoglycemia. These patients can have inadequate glucagon as well as insulin reserves and cannot respond to hypoglycemia with a glucagon surge and subsequent increase in blood glucose levels. Overly vigorous attempts to control blood sugar can be associated with disastrous complications of treatment-induced hypoglycemia.

Some patients will respond to oral hypoglycemics, but many require insulin. The goal of insulin therapy is usually to control urinary losses of glucose rather than attempt tight control of blood sugar. Tight control of blood sugar is usually indicated in only one subgroup, those with hyperlipidemic pancreatitis, where tight control of blood sugar is needed to allow control of triglyceride levels.

Pain

Medical Treatment

There are many causes of pain and no single treatment is effective in all patients. Abstinence from alcohol reduces the risk of other alcohol-related complications such as cirrhosis, prolongs life, slows the rate of progression of chronic pancreatitis, and may reduce pain. Analgesics are routinely required. Nonnarcotic analgesics should be used first and if narcotics are required, the least potent formulation should be tried first (e.g., propopxyphene with acetaminophen, tramadol). More potent narcotics are required in many patients and addiction occurs in up to 20% of these patients. In patients who require narcotics, the addition of an antidepressant (tricyclic antidepressants or selective serotonin reuptake inhibitors) can be helpful as these can potentiate the effect of narcotics.

Several small studies have suggested that non-enteric-coated pancreatic enzymes can reduce abdominal pain in some patients with chronic pancreatitis. The theoretical basis of their effect is to reduce pancreatic stimulation (or hyperstimulation) by cholecystokinin. This effect, of reestablishing normal negative feedback control of pancreatic secretion, is operative only in the duodenum. Enteric-coated preparations typically release their enzymes in the jejunum; hence, non-enteric-coated preparations are selected if the intent is to treat pain. The response in these trials is mixed. Patients who seemed to respond best to the use of conventional enzymes are those with mild to moderate chronic pancreatitis (without steatorrhea or small-duct disease). Patients with advanced disease (steatorrhea or big-duct disease) did not seem to respond. If tried, this therapy should be considered only in select patients (small-duct disease, no steatorrhea) and should use the correct enzyme and the correct dose (non-enteric-coated preparations at high dosage, coupled with an agent to reduce gastric acid).

A few small studies of octreotide (Sandostatin) have suggested that the use of this agent may reduce the pain of chronic pancreatitis, perhaps by reducing pancreatic stimulation or perhaps by a direct anti-nociceptive effect. A few small studies have also suggested that the use of antioxidants (mixtures of vitamins E and C, selenium, methionine, and β-carotene) may reduce pain. Neither treatment is currently recommended outside clinical trials.

Neurolysis

Interruption of nociceptive visceral afferents from the pancreas can be accomplished by ablation of the celiac nerve plexus. This has been tried by both CT-guided and EUS-guided methods, using injection of anesthetics, steroids, or alcohol. The therapy seems most effective for the pain of pancreatic cancer but is usually too short-lived for patients with chronic pancreatitis. EUS-guided techniques appear to work better and for somewhat longer periods of time than CT-guided techniques. Destruction of visceral afferents by destruction of the splanchnic nerves (thoracoscopic splanchnicectomy) is also being investigated but the efficacy remains to be established.

Endoscopic Therapy

Available endoscopic therapies include pancreatic duct sphincterotomy, pancreatic duct stenting, dilation of ductal strictures, removal of pancreatic stones, and treatment of complications such as pseudocyst or biliary obstruction. The primary goal of endoscopic therapy is to remove or bypass any obstruction within the pancreatic duct. Careful selection of patients for this therapy is critical. Appropriate candidates are those with a significant pancreatic ductal stricture in the head of the pancreas or those with a few obstructing stones in the head of the gland. Approximately one-third to one-half of patients with big-duct disease satisfy these criteria. In these carefully selected groups, up to three-fourths of patients will experience pain relief after endoscopic therapy. Complications of endoscopic treatment of chronic pancreatitis occur in 15–20% of patients (pancreatitis, bleeding, perforation, and sepsis). Pancreatic stents, used for therapy in these patients, can in and of themselves also injure the pancreatic duct and

parenchyma in up to half of patients treated, and these changes may not resolve. Endoscopic therapy is a reasonable alternative in appropriately selected patients at centers with appropriate expertise.

Surgical Treatment

Surgical therapy is considered for pain and for complications of chronic pancreatitis (pseudocyst, bile duct obstruction, and duodenal obstruction). The most commonly performed operation for pain is the lateral pancreaticojejunostomy (modified Puestow procedure). The dilated pancreatic duct is longitudinally incised along its length and overlaid with a defunctionalized Roux limb. The procedure carries low rates of complications (5%) and mortality (2%). The pancreatic duct must usually be dilated to greater than 5 mm; hence, this type of surgery is considered in those with big-duct disease. After a Puestow procedure, immediate pain relief occurs in 70–90% of patients. Pain is controlled in only 50% after 1–3 years of follow-up. In some centers, resection of all or part of the head of the pancreas is combined with a ductal drainage procedure. Morbidity and mortality from these operations is higher than that seen with a Puestow procedure, but long-term pain relief is also higher. Subtotal or total pancreatic resections are rarely indicated and are currently associated with unacceptable postoperative complications, particularly brittle diabetes mellitus.

Treatment for Pain

No single therapy is effective in all patients and the choice of a treatment depends on a variety of factors. The most important include pancreatic ductal anatomy, presence of complications, and local expertise. The first step is to make sure the diagnosis is correct. It is inappropriate to consider therapies with potential side effects or complications unless the diagnosis is secure. Second, it is worthwhile to look for specific complications that have specific therapy, such as pancreatic pseudocyst, duodenal obstruction, common bile duct obstruction, peptic ulcer disease, or pancreatic carcinoma (these are discussed below). Medical therapy is appropriate in all patients. This should include encouragement of abstinence from alcohol (if applicable), a low-fat diet, and analgesics. The choice of subsequent therapy depends in large part whether the patient has small-duct or big-duct disease. In those with small-duct disease, a trial of high-dose non-enteric-coated enzymes coupled with acid suppression is appropriate. Treatment options for those with big-duct disease are largely mechanical, with either endoscopic or surgical attempts to decompress the enlarged pancreatic duct. Patients who fail the above therapies may be considered for more experimental therapies such as celiac plexus block, splanchnicectomy, or octreotide. Significant pancreatic resections are considered an option of last resort in both groups of patients.

COMPLICATIONS

Pancreatic Pseudocyst

Approximately one-quarter of patients with chronic pancreatitis develop a pseudocyst. Many pseudocysts remain asymptomatic. Symptomatic pseudocysts may obstruct a surrounding hollow viscus (duodenum producing nausea and vomiting or common bile duct producing jaundice), may cause pain, may bleed (internally or into the gut lumen), or may rupture (into the peritoneal cavity producing pancreatic ascites or into the pleural space producing a pancreatic pleural effusion). Overall, these complications occur in approximately one-third of patients. The diagnosis is best made with CT. The diagnosis is usually suspected due to a worsening pattern of pain, persistent elevations in amylase or lipase, or symptoms pointing toward one of the above complications.

Not all pseudocysts require treatment. Those that are less than 6 cm in diameter, that are causing no symptoms, and that occur in a reliable patient can be safely observed. Symptomatic pseudocysts can be managed with percutaneous, endoscopic, or surgical decompression, depending on location and local expertise.

Two complications deserve specific mention. Bleeding is rare but can be life-threatening. Bleeding can occur from a pseudoaneurysm associated with a pseudocyst. Evidence of gastrointestinal bleeding in a patient with a pseudocyst is usually treated as an emergency. If initial evaluations do not reveal a source, an emergent CT is performed to look for a pseudoaneurysm. Confirmation of the pseudoaneurysm with CT should be followed with emergent angiography and embolization. Rupture of a pseudocyst or leak from the pancreatic duct can produce pancreatic ascites of a pancreatic pleural effusion. These should be suspected when a very high level of amylase (typically >4000 IU/liter) is found in ascites or a pleural effusion. Endoscopic treatment at ERCP, with sphincterotomy and stenting, is usually curative. Surgical therapy is reserved for failure of endoscopic therapy.

Pseudocysts should not be mistaken for cystic neoplasms. These neoplasms usually have a thick wall or internal nodules along the wall, often with internal sepatations. They most frequently occur in middle-aged women without previous pancreatitis and without risk factors for pancreatitis. These neoplasms are

frequently misdiagnosed; the therapy is resection, not drainage.

Gastrointestinal Bleeding

Bleeding may occur not only as a consequence of a pseudoaneurysm (as noted above) but also from thrombosis of the splenic vein as it travels behind the pancreas. This produces a "left-sided" portal hypertension with gastric varices out of proportion to esophageal varices. Splenectomy is curative.

Pancreatic Carcinoma

Chronic pancreatitis is a risk factor for pancreatic adenocarcinoma and the two may coexist. The overall risk is approximately 4% lifetime, although this may increase to 40% for those with hereditary pancreatitis. There is no currently effective screening method, but superimposed carcinoma is usually suspected due to weight loss or a worsening in a previously stable pain pattern.

Duodenal or Common Bile Duct Obstruction

Obstruction of the duodenum or intrapancreatic common bile duct may occur due to fibrosis and inflammation within the head of the pancreas or due to an associated pseudocyst. Duodenal obstruction produces nausea, vomiting, and early satiety and is best confirmed by a barium upper gastrointestinal series. Common bile duct obstruction can produce jaundice, biliary pain, cholangitis, or asymptomatic elevations in liver chemistries. These patients may also have intrinsic liver disease, so a liver biopsy is usually indicated before deciding on treatment for asymptomatic elevations in liver chemistries. All other symptomatic presentations require therapy. Endoscopic biliary stenting is an appropriate temporizing measure but these patients generally require surgical biliary bypass.

Others

Small bowel bacterial overgrowth and gastroparesis may complicate chronic pancreatitis. Both can interfere with effective treatment of steatorrhea and gastroparesis may cause pain of its own accord.

See Also the Following Articles

Amylase • Computed Tomography (CT) • Cystic Fibrosis • Diabetes Mellitus • Magnetic Resonance Imaging (MRI) • Malabsorption • Pancreatic Bicarbonate Secretion • Pancreatic Digestive Enzymes • Pancreatic Function Tests • Pancreatic Pseudocysts • Pancreatic Triglyceride Lipase • Trypsin • Ultrasonography

Further Reading

Ammann, R. W., Heitz, P. U., and Kloppel, G. (1996). Course of alcoholic chronic pancreatitis: A prospective clinicomorphological long-term study. *Gastroenterology* **111**, 224–231.

Ammann, R. W., and Muellhaupt, B. (1999). The natural history of pain in alcoholic chronic pancreatitis. *Gastroenterology* **116**, 1132–1140.

Cohn, J. A., Bornstein, J. D., and Jowell, P. S. (2000). Cystic fibrosis mutations and genetic predispositions to idiopathic chronic pancreatitis. *Med. Clin. North Am.* **84**, 621–631.

Eisen, G. M., Chutkan, R., Goldstein, J. L., *et al*. (2000). Endoscopic therapy of chronic pancreatitis. ASGE position paper. *Gastrointest. Endosc.* **52**, 843–848.

Etemad, B., and Whitcomb, D. C. (2001). Chronic pancreatitis: Diagnosis, classification, and new genetic developments. *Gastroenterology* **120**, 682–707.

Forsmark, C. E. (2000). The diagnosis of chronic pancreatitis. *Gastrointest. Endosc.* **52**, 293–298.

Forsmark, C. E., and Grendell, J. H. (1991). Complications of pancreatitis. *Semin. Gastrointest. Dis.* **2**, 165–176.

Forsmark, C. E., and Toskes, P. P. (1995). What does an abnormal pancreatogram mean? *Gastrointest. Endosc. Clin. North Am.* **5**, 105–124.

Layer, P., Yamamoto, H., Kalthoff, L., *et al*. (1994). The different courses of early- and late-onset idiopathic and alcoholic chronic pancreatitis. *Gastroenterology* **107**, 1481–1487.

Lowenfels, A. B., Maisonneuve, P., Cavallini, G., *et al*. (1993). Pancreatitis and the risk of pancreatic cancer. *N. Engl. J. Med.* **328**, 1433–1437.

Saforkas, G. H., Farnell, M. B. Farley, D. R., *et al*. (2000). Long-term results after surgery for chronic pancreatitis. *Int. J. Pancreatol.* **27**, 131–142.

Vitas, G. J., and Sarr, M. G. (1992). Selected management of pancreatic pseudocysts: Operative versus expectant management. *Surgery* **111**, 123–130.

Wallace, M. B., and Hawes, R. H. (2001). Endoscopic ultrasound in the evaluation and management of chronic pancreatitis. *Pancreas* **23**, 26–35.

Warshaw, A. L., Banks, P. A., and Fernandez-del Castillo, C. (1998). AGA technical review: Treatment of pain in chronic pancreatitis. *Gastroenterology* **115**, 763–776.

Pancreatitis, Experimental Models

ASHOK K. SALUJA AND LAKSHMI BHAGAT
University of Massachusetts Medical School

acute pancreatitis Acute inflammatory disease of the pancreas.
caerulein An analogue of cholecystokinin that, because it is more stable, is used for secretagogue-induced pancreatitis.
CDE diet Choline-deficient diet supplemented with ethionine; this diet induces pancreatitis in female mice.
sodium taurocholate A bile salt that induces pancreatitis when injected retrograde into the pancreatic duct.
trypsin The pancreatic digestive enzyme whose premature activation within the pancreas is a common feature of various models of pancreatitis.

Acute pancreatitis is an inflammatory disease of the pancreas that is associated with little or no fibrosis. It can be initiated by several factors including gallstones, alcohol, trauma, and infections and in some cases it may be hereditary. Very often, the patients develop additional complications, such as sepsis, shock, and respiratory and renal failure, resulting in considerable morbidity and mortality. Approximately 300,000 cases occur in the United States each year; some 20% of cases are severe. Despite considerable work being carried out in this area, the pathophysiological mechanisms associated with this disease still remain incompletely understood. This can be attributed in part to the relative inaccessibility of clinical material for experimental studies and has therefore led to the development of several models of experimental pancreatitis to explore its etiology, pathophysiology, and treatment regimens. This article will focus on the experimental models used to study acute pancreatitis. For the sake of simplicity, these models have been divided into two broad categories depending on the approach used to induce pancreatitis, namely, noninvasive and invasive models.

NONINVASIVE MODELS OF PANCREATITIS

Secretagogue-Induced Pancreatitis

More than a century ago, Mouret observed that excessive neural stimulation results in injury of the exocrine pancreas. Subsequent studies have not only confirmed these findings but also have shown that in addition to stimulation with the cholinergic agonists (carbamylcholine), excessive stimulation with cholecystokinin (CCK) and its analogue, caerulein, also results in acute pancreatitis. In fact, the caerulein-induced model has become the most widely used model with which to study acute pancreatitis in rodents.

CCK receptors present on the pancreatic acinar cells exist in two affinity states—high and low. Occupancy of high-affinity receptors by low concentrations of CCK (or caerulein) results in stimulation of digestive enzyme secretion, whereas interaction of CCK with low-affinity receptors results in inhibition of digestive enzyme secretion. Acute pancreatitis is induced by occupancy of these low-affinity receptors by caerulein and it can be prevented by administration of CCK-receptor antagonists (e.g., L-364-718 or CCK-JMV-180, an antagonist of low-affinity CCK receptors in rats).

Administration of caerulein to rodents at doses that are 10- to 100-fold higher than those that evoke maximal secretion of digestive enzymes from pancreatic acinar cells results in a rapid development of pancreatitis. Stimulation with supramaximal concentrations of CCK and carbamylcholine can also induce pancreatitis; however, caerulein is used most frequently. Caerulein can be given intravenously at doses sufficient to deliver 5–10 μg/kg/h in a continuous infusion. Alternatively and perhaps more conveniently, pancreatitis can also be induced by hourly intraperitoneal or subcutaneous injections, although this approach is usually not as reliable as intravenous infusions. Pancreatic injury in rats develops rapidly, increases with the duration of infusion, and is maximal within 12 h. This relatively mild form of pancreatitis is characterized by massive pancreatic edema (which can be visualized macroscopically), increased serum levels of pancreatic enzymes (such as amylase and lipase), cytoskeletal (actin) reorganization, acinar cell vacuolization, necrosis, and inflammation. Strikingly, intra-acinar cell activation of trypsinogen can be observed within minutes of initiation of caerulein infusion.

The caerulein-induced model has become the model of choice for studying pancreatitis in mice, particularly in genetically altered strains. For these studies,

 Copyright 2004, Elsevier (USA). All rights reserved.

pancreatitis is usually induced by administration of 8 to 12 hourly intraperitoneal injections of caerulein (50 μg/kg). The resulting pancreatitis is more severe in mice than in rats and after 12 injections of caerulein, the pancreas develops extensive necrosis and inflammation. This noninvasive model has also proved to be a good model with which to study pancreatitis-associated lung injury.

Due to its ease of use and reproducibility, the caerulein-induced model has been extensively used to study the pathophysiological mechanisms involved in pancreatitis. Several recent studies using this model have shown that one of the earliest events in the onset of pancreatitis is the colocalization of digestive enzyme zymogens and lysosomal hydrolases in large vacuoles observed during pancreatitis. This colocalization could result in premature intra-acinar activation of trypsinogen and perhaps other digestive zymogens, leading to acinar cell injury. This model has also been extensively used to elucidate the role of heat shock proteins and inflammatory mediators in pancreatitis and associated lung injury.

Overall, the caerulein-induced model is probably the most popular experimental model of pancreatitis because it offers several advantages: it is very easy and inexpensive to use, is noninvasive, develops rapidly and reproducibly in mice and rats, which are easy to handle and in addition, the severity of the resulting disease can be regulated and manipulated. However, the main drawback of this model is its questionable clinical relevance.

Diet-Induced Pancreatitis

Administration of ethionine, the ethyl analogue of methionine, was initially shown to induce mild edematous, nonlethal pancreatitis in rats, cats, dogs, and monkeys. Subsequently, Lombardi and co-workers in 1975 developed a model of acute necrotizing hemorrhagic pancreatitis by feeding a choline-deficient ethionine-supplemented (CDE) diet to young female mice. However, this disease is lethal and all the mice die within 5 days if they are fed the CDE diet *ad libitum*. Later, this protocol was modified to reduce the mortality rate to 50 to 70% by limiting diet consumption and feeding the CDE diet for only 24 h. The mice are fasted for 24 h before and after being given the CDE diet. Estrogen and a reduced capacity to neutralize acute pancreatitis enzymes probably mediate the sex difference in response to the dietary regimen.

Like the secretagogue model, this model has also been shown to be associated with a defect in stimulus–secretion coupling and the digestive enzyme zymogens have been shown to be colocalized with the lysosomal enzymes within 1 day of the start of the CDE diet. This model shares many histological and biochemical features with clinical pancreatitis in humans and is also associated with acute lung injury.

Arginine-Induced Pancreatitis

This infrequently used model, originally described by Tani and co-workers in 1990, involves administration of a single large intraperitoneal injection of L-arginine to rats and results in necrotizing pancreatitis that is less severe than the CDE diet-induced pancreatitis. The disease evolves over a period of 72 h with the initial appearance of small intracellular vesicles within 6 h, a significant elevation in serum amylase and lipase and acinar cell necrosis within 24 h, and a marked inhibition of protein synthesis within 72 h after administration of the injection. Changes in the actin cytoskeleton are an early component in this model of pancreatitis. It is also accompanied by a stress response with a large increase in heat shock proteins 27 and 70. The pancreas starts recovering by 7 days and regains its normal function by 14 days. The mechanism by which L-arginine induces pancreatitis is not well understood although it is likely that it acts by inducing an intracellular stress response.

Immune-Induced Pancreatitis

Immune-induced pancreatitis can be initiated by a Schwartzmann or an Arthus type of reaction, by injection of foreign serum intraperitoneally or into the pancreatic duct, and by the intraductal injection of anti-pancreatic basement membrane antibodies.

INVASIVE MODELS OF PANCREATITIS

Observations made as early as 1856 by Claude Bernard and in 1901 by Opie laid the foundation for the development of invasive models of acute pancreatitis. Opie proposed the "common channel" theory on the basis of the autopsy results of two patients who died from acute necrotizing pancreatitis and were found to have gallstones impacted in the ampulla of Vater. According to this theory, the stones obstructed the terminal common channel of the bile and pancreatic ducts so that the bile could now reflux into the pancreatic ductal system and initiate pancreatitis. Although several objections have been raised and alternative hypotheses suggested, this hypothesis is still accepted by many as a valid explanation for the triggering event in gallstone pancreatitis. Alternatively, a stone on passage into the duodenum stretches the sphincter of Oddi, thus permitting reflux of duodenal contents into the pancreatic duct.

Yet another widely accepted theory is based on the premise that obstruction of the pancreatic duct leads to retention of the pancreatic juice within the pancreatic ductal system and that rupture of the intrapancreatic ducts because of the increased ductal pressure leads to spilling of the pancreatic juice containing the digestive enzymes into the gland itself, thereby initiating a cascade of events resulting in pancreatitis.

For the most part, these experimental models have been designed to test the validity of the aforementioned theories and thus include perfusion or retrograde injection of bile and other agents into the pancreatic duct, construction of a closed duodenal loop to facilitate reflux of duodenal contents into the pancreatic ductal system, and temporary obstruction/ligation of the pancreatic or biliopancreatic ductal system.

Closed Duodenal Loop

This model, one of the first experimental models of acute pancreatitis when it was described by Pfeffer *et al*. in 1957, has undergone a series of modifications over the years and is used mainly to establish the etiology of pancreatitis. In their initial studies, Pfeffer and coworkers, using fasted mongrel dogs, isolated 10 cm of the duodenum just beyond the pylorus and ligated the bile duct so that the closed duodenal segment communicated with the pancreatic ductal system. Gastric outflow was reestablished by construction of a gastrojejunostomy. Under these conditions, edematous changes were noted after 4 h and parenchymal necrosis developed between 9 and 11 h after the surgery. However, fat necrosis and inflammation occurred infrequently. The possible mechanisms of pancreatitis induced using this approach include pancreatic ischemia, overdistension of the duodenal loop, and reflux of the duodenal contents into the pancreatic ductal system since pancreatic duct ligation or pancreatic duct cannulation ameliorates the changes associated with pancreatitis in this model.

The model has undergone several changes including placement of an intralumenal tube in the ligated area to maintain intestinal continuity, performing a gastrojejunostomy and insertion of a bypass cannula (Herrera fistula), and injection of infected bile or bile salt–trypsin mixture into the closed loop.

Since the severity of pancreatitis associated with this model is highly variable and is associated with little or no fat necrosis and inflammation, it is not universally accepted as a model of acute pancreatitis. Furthermore, its clinical relevance still remains to be established.

Retrograde Ductal Injection and Prograde Perfusion

Acute pancreatitis can be induced in relatively large animals (dogs, cats, pigs, rats) by cannulating the pancreatic duct and retrogradely injecting agents such as activated digestive enzymes (trypsin, elastase, lipase, phospholipase A), bile, bile plus trypsin or the purified bile salt, sodium taurocholate alone or with trypsin, and fatty acids. The critical features of this model are the pressure that is used for injecting these agents and the finding that the severity of pancreatitis is directly correlated with the pressure and the volume of the injected material.

Perhaps the earliest model of acute pancreatitis was developed in 1856 by Claude Bernard, who injected olive oil and bile retrogradely into the pancreatic duct of the dog and observed the subsequent development of pancreatitis. This model is suitable to study the late but not the early events associated with severe pancreatitis. In an attempt to study the early events and to reduce the severity of pancreatitis, Aho and co-workers modified the protocol by injecting small volumes of sodium taurocholate (3 to 5%) over a period of 1 min in rats. The pancreatitis induced in this case evolves slowly over a period of 72 h and the severity of pancreatitis correlates directly with the injected concentration of sodium taurocholate.

Pancreatitis of severity ranging from mild edematous to necrotizing can also be initiated in large animals such as cats, pigs, dogs, and primates by cannulating the pancreatic duct and prograde perfusing it with a permeability-increasing agent at low pressure followed by infusion of infected bile, hydrochloric acid, aspirin, bile salts, or even activated digestive enzymes. Perfusion of the last agent in combination with prostaglandin E2 has been shown to induce necrotizing pancreatitis.

Duct Ligation/Obstruction Models

Pancreatic duct ligation or obstruction models were designed to recapitulate the events that occur during gallstone pancreatitis in humans. Ligation or temporary obstruction of the pancreatic duct in most animals results in pancreatic edema, mild inflammation, and acinar cell apoptosis, ultimately leading to atrophy of the gland. However, the severity of the pancreatitis can be enhanced by combining duct ligation with stimulation of secretion and pancreatic ischemia. Most of these studies have been performed on rabbits and rats.

On the other hand, ligation of the pancreatic duct or of the common biliopancreatic duct of the American opossum results in severe hemorrhagic pancreatitis, which is associated with acute lung injury and a

14-day mortality rate of 100%. The opossum is a very useful animal for studies mimicking gallstone pancreatitis because its biliopancreatic anatomy closely resembles that of humans. However, unlike in humans where necrotizing pancreatitis is focal and restricted, in the opossum it is diffuse and uniformly distributed throughout the gland. In the opossum, pancreatic duct obstruction appears to be the sole cause of pancreatitis and neither biliary duct obstruction nor bile reflux is essential for triggering or worsening the severity of pancreatitis.

IN VITRO MODEL OF PANCREATITIS

Although pancreatitis is an inflammatory disease, many of the early changes take place in pancreatic acinar cells. For example, in all models of pancreatitis, intra-acinar cell activation of trypsinogen is one of the earliest events that precede overt pancreatitis. In order to study the mechanisms of pathobiological events initiated in pancreatic acinar cells, an *in vitro* model has been developed and has been extensively used in the past few years. In this model, stimulation of rat pancreatic acini with supramaximal concentrations of caerulein results in intra-acinar cell activation of trypsinogen.

For these studies, dispersed pancreatic acini are prepared from freshly harvested pancreas by collagenase digestion and gentle shearing. Acini thus prepared are more than 95% viable and remain healthy for up to 24 h. These acini can be incubated with high concentrations of caerulein and the resulting acinar cell changes can be monitored. Such an *in vitro* reductionist approach has been useful in studies designed to elucidate the mechanisms involved in the activation of trypsinogen and the ensuing acinar cell injury that is observed during pancreatitis. Furthermore, studies using this *in vitro* model have shown that a sustained rise in intracellular calcium is required for activation of trypsinogen and that inhibition of cathepsin B activity prevents this activation.

CONCLUSIONS

Although a number of models, both invasive and noninvasive, have been developed to study the different aspects of acute pancreatitis, there is still no "perfect model" that recapitulates the events occurring during clinical acute pancreatitis in their entirety. Notwithstanding the shortcomings associated with the models described above, these experimental approaches have helped greatly in unraveling the signaling events and the complex interaction of acinar, nonacinar, and inflammatory cells during the initiation and progression of this disease.

See Also the Following Article

Pancreatitis, Acute

Further Reading

Aho, H. J., Koskensalo, S. M., and Nevalainen, T. J. (1980). Experimental pancreatitis in the rat: Sodium taurocholate-induced acute haemorrhagic pancreatitis. *Scand. J. Gastroenterol.* **15**, 411–414.

Gorelick, F. S., Adler, G., and Kern, H. F. (1993). Cerulein-induced pancreatitis. *In* "The Pancreas: Biology, Pathobiology and Disease" (V. L. W. Go, E. P. DiMagno, J. D. Gardner, E. Lebenthal, H. A. Reber, and G. A. Scheele, eds.), 2nd Ed., pp. 501–526. Raven Press, New York.

Lampel, M., and Kern, H. F. (1977). Acute interstitial pancreatitis in the rat induced by excessive doses of a pancreatic secretagogue. *Virch. Arch. A Pathol. Anat. Histol.* **373**, 97–117.

Lerch, M. M., Saluja, A. K., Runzi, M., Dawra, R., Saluja, M., and Steer, M. L. (1993). Pancreatic duct obstruction triggers acute necrotizing pancreatitis in the opossum. *Gastroenterology* **104**, 853–861.

Lombardi, B., Estes, L. W., and Longnecker, D. S. (1975). Acute hemorrhagic pancreatitis (massive necrosis) with fat necrosis induced in mice by DL-ethionine fed with a choline-deficient diet. *Am. J. Pathol.* **79**, 465–480.

National Digestive Disease Clearinghouse Web Site. Pancreatitis. Available at http://www.niddk.nih.gov/health/digest/pubs/pancreas/pancreas.htm.

Opie, E. L. (1901). The etiology of acute hemorrhagic pancreatitis. *Bull. Johns Hopkins Hosp.* **12**, 182–192.

Pfeffer, R. B., Stasior, O., and Hinton, J. W. (1957). The clinical picture of the sequential development of acute hemorrhagic pancreatitis in the dog. *Surg. Forum* **8**, 248–251.

Saluja, A. K., Donovan, E. A., Yamanaka, K., Yamaguchi, Y., Hofbauer, B., and Steer, M. L. (1997). Cerulein-induced *in vitro* activation of trypsinogen in rat pancreatic acini is mediated by cathepsin B. *Gastroenterology* **113**, 304–310.

Steer, M. L., and Saluja, A. K. (1993). Experimental acute pancreatitis: Studies of the early events that lead to cell injury. *In* "The Pancreas: Biology, Pathobiology and Disease" (V. L. W. Go, E. P. DiMagno, J. D. Gardner, E. Lebenthal, H. A. Reber, and G. A. Scheele, eds.), 2nd Ed., pp. 489–500. Raven Press, New York.

Tani, S., Itoh, H., Okabayashi, Y., Nakamura, T., Fujii, M., Fujisawa, T., Koide, M., and Otsuki, M. (1990). New model of acute necrotizing pancreatitis induced by excessive doses of arginine in rats. *Digest. Dis. Sci.* **35**, 367–374.

Williams, J. A. (2001). Intracellular signaling mechanisms activated by cholecystokinin-regulating synthesis and secretion of digestive enzymes in pancreatic acinar cells. *Annu. Rev. Physiol.* **63**, 77–97.

Pancreatitis, Hereditary

DAVID C. WHITCOMB
University of Pittsburgh

cationic trypsinogen Precursor form of trypsin defined by its cationic (net positive) charge.
hereditary pancreatitis Form of the disease that appears to be acquired via an inherited gene.
pancreatitis Inflammation of the pancreas.

Hereditary pancreatitis refers specifically to otherwise unexplained acute or chronic pancreatitis in an individual from a family in which pancreatitis appears to be passed down from generation to generation through an inherited gene. Familial pancreatitis is a broader term that refers to pancreatitis of any cause in an individual from a family whose members develop pancreatitis at a higher rate than would be expected by chance alone. Familial pancreatitis may or may not be caused by a genetic mutation. Hereditary pancreatitis therefore defines a narrow, specific term whereas familial pancreatitis is a broad, general term that also encompasses inherited types of pancreatitis.

CLINICAL FEATURES

Pancreatitis is inflammation of the pancreas. Acute pancreatitis occurs when normal pancreatic tissue suddenly becomes inflamed. Inflammation is occasionally caused by a bacterial or viral infection or by direct trauma to the abdomen. However, in most cases, acute pancreatitis develops when the inactive pancreatic digestive enzymes become active inside the pancreas rather than inside the intestine. In this case, the pancreas literally begins to digest itself, leading to severe inflammation. Chronic pancreatitis describes the effects of prolonged inflammation in which the normal pancreatic tissue is partially or completely destroyed and replaced with scar tissue.

A typical description of hereditary pancreatitis includes a family history in which the older generations describe lifelong histories of attacks of abdominal pain beginning in childhood. The pain often becomes continuous by adulthood or disappears. In more severe cases, chronic pancreatitis or diabetes mellitus develops, and pancreatic cancer (or an adenocarcinoma of unknown origin) may be seen after age 50 years. In the younger generations, the childhood attacks would be diagnosed as acute pancreatitis. However, only about half of the offspring in each generation develops the typical symptoms, and the other half of the family tree never has symptoms. In some cases, the symptoms skip a generation, demonstrating that about one out of five people with the pancreatitis-causing gene mutation are silent carriers. This pattern of inheritance is typical of autosomal dominant genetic disorders that are passed from generation to generation through a single defective gene.

The average person inheriting a gene mutation causing hereditary pancreatitis has several symptoms. At about age 10 years they begin developing attacks of severe abdominal pain caused by acute pancreatitis. The exact age of first attack in an individual can vary from about 1 year old to adulthood, but usually occurs in childhood. The pain is centered in the epigastrium, which is about a third of the way from the rib cage to the navel. The pain often increases, beginning as a poorly described ache with nausea and becoming a very severe, steady pain with nausea and vomiting. The pain may extend to the back, to the sides (usually left), or to the entire abdomen. It is unusual to have fever or diarrhea, and during the attack the person usually tries to lie in a curled up position and to remain motionless. The attacks may last for 2–3 days (this is also variable), with attacks usually occurring a couple of times a year.

Later in life, about 5–10 years after the first attack of acute pancreatitis, up to half of the individuals with hereditary pancreatitis will develop chronic pancreatitis. The symptoms of chronic pancreatitis include continuous abdominal pain, difficulty digesting food, bulky stools or diarrhea-like symptoms, and (later) diabetes mellitus. Patients with chronic pancreatitis for several decades are also susceptible to pancreatic cancer, especially if they smoke cigarettes.

CAUSES

The first mutation causing hereditary pancreatitis was discovered in 1996 on chromosome 7 within the *cationic trypsinogen* gene, or *PRSS1*. Trypsinogen is the inactive form of the protein-digesting enzyme trypsin, which

 Copyright 2004, Elsevier (USA). All rights reserved.

cuts protein chains at arginine or lysine amino acids. Trypsinogen normally becomes activated to trypsin after passing into the intestine, and trypsin activates the other inactive digestive enzymes inside the intestine, where ingested food is digested. The hereditary pancreatitis mutation changes the DNA code for a key arginine amino acid (symbol R) at position 122 of the trypsin molecule to code for the amino acid histidine (symbol H), resulting in a "R122H" mutation. Arginine 122 acts an emergency "fail-safe" self-destruction site on the back side of the trypsin molecule, allowing this digestive enzyme to be eliminated by a second trypsin molecule if the trypsinogen becomes active trypsin inside the pancreas. This site is regulated by calcium, and in low calcium concentrations the trypsin will destroy itself (e.g., within the acinar cell). When this site is mutated to histidine, the pancreatic acinar cells cannot eliminate prematurely activated trypsin, which then activates other digestive enzymes. The prematurely active digestive enzymes begin digesting the pancreas, leading to acute pancreatitis. Repeated episodes of acute pancreatitis lead to severe scarring of the pancreas, or chronic pancreatitis. About two-thirds of the families with hereditary pancreatitis have one of several known cationic trypsin mutations, and the other third has unknown mutations.

DIAGNOSIS AND TREATMENT

The diagnosis of hereditary pancreatitis is currently made by diagnosing pancreatitis in two or more members of a family in two generations in which other causes have been excluded (for example, pancreatitis from gallstones). In addition, hereditary pancreatitis can be diagnosed in individuals with pancreatitis through genetic testing. The two common mutations for which testing is recommended are the cationic trypsinogen R122H and N21I mutations. Mutations in the trypsin inhibitor gene (*SPINK1*) and some mutations in the cystic fibrosis gene (*CFTR*) are also seen in pancreatitis, but usually not in families with hereditary pancreatitis. Mutations in these genes are common, and a single mutation does not cause the pancreatitis seen in hereditary pancreatitis.

Currently, there are no specific, proved therapies to prevent or treat hereditary pancreatitis. Smaller meals, diets low in fat and protein, and vitamin and antioxidant supplements have been recommended by some experts, and are usually not discouraged. Smoking cigarettes and alcohol consumption are strongly discouraged because these factors promote pancreatitis and contribute to pancreatic cancer. Medications to reduce stomach acid, supplements containing pancreatic enzymes, and even insulin injections may be necessary if chronic pancreatitis develops. Surgery may be necessary to treat some complications.

See Also the Following Articles

Diabetes Mellitus • Pancreatic Cancer • Pancreatitis, Acute • Pancreatitis, Chronic • Trypsin

Further Reading

Ellis, I., Lerch, M. M., and Whitcomb, D. C. (2001). Consensus committees. Genetic testing for hereditary pancreatitis: Guidelines for indications, counseling, consent and privacy issues. *Pancreatology* **1**(5), 405–418.

Etemad, B., and Whitcomb, D. C. (2001). Chronic pancreatitis: Diagnosis, classification, and new genetic developments. *Gastroenterology* **120**, 682–707.

Lowenfels, A. B., Maisonneuve, P., Whitcomb, D. C., Lerch, M. M., and DiMagno, E. P. (2001). Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis. *J. Am. Med. Assoc.* **286**(2), 169–170.

Whitcomb, D. C. (2002). Hereditary diseases of the pancreas. *In* "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, L. Laine, C., Kaplowitz, N., Owyang, and D. W. Powell, eds.), 4th Ed., pp. 2147–2165. Lippincott Williams & Wilkins, Philadelphia.

Whitcomb, D. C., Gorry, M. C., Preston, R. A., Furey, W., Sossenheimer, M. J., Ulrich, C. D., *et al.* (1996). Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat. Genet.* **14**(2), 141–145.

Pancreatitis, Pediatric

M. JAMES LOPEZ
University of Michigan Medical Center, Ann Arbor

acute pancreatitis An acute inflammatory condition of the pancreas associated with a broad variety of etiologies in children.

chronic pancreatitis An inflammatory condition of the pancreas with evidence of permanent and progressive morphological damage.

hereditary pancreatitis An autosomal dominant disease that accounts for 1% of cases of both chronic and recurrent pancreatitis; genetic mutations have been identified in cationic trypsinogen for many patients.

metabolic pancreatitis Pancreatic disorder related to abnormalities such as hyperlipidemia or hypercalcemia, which can cause acute or recurrent pancreatitis but which is extremely rare in pediatrics.

pseudocyst A complication of pancreatitis with a cystic structure not lined with epithelium, which distinguishes it from a true cyst. Most commonly, the walls of the cyst are formed in part by the stomach, spleen, and pancreas.

sphincterotomy Incision of the sphincter of Oddi (with reference to pancreatitis).

Pancreatic disorders in childhood include congenital anatomic and genetic disorders as well as acquired disorders. Most of these disorders are rare. This article will focus on the various types of pancreatitis and briefly describe their clinical characteristics.

ACUTE PANCREATITIS

Acute pancreatitis is an acute inflammatory condition of the pancreas. It is associated with a broad variety of causes in childhood, in contrast to the two main causes of gallstones and alcohol in adults (Table I). Nearly 20% of pediatric cases of pancreatitis are associated with trauma, approximately 10% are associated with gallstones or congenital structural abnormalities of the pancreatico-biliary ductular system, and less than 5% are associated with medications; only rarely are cases associated with metabolic abnormalities and between 15 and 20% have no explanation even after exhaustive evaluation (idiopathic). Over half of all children with pancreatitis have an associated systemic illness. Of those with multisystem disease, between one-quarter and one-third present with sepsis or shock, approximately 5% have an associated viral illness, and over half have a systemic illness, such as a collagen vascular disease or heart disease. In children under 4

TABLE I Reported Causes of Acquired Pancreatitis in Children

Trauma
Biliary tract disease
Gallstones
Structural abnormalities (e.g., choledochal cyst)
Drug-induced
Azathioprine
Valproic acid
L-Asparaginase
Thiazides
Tetracyclines
Sulfonamides
Furosemide
Estrogen
Infection
Mumps virus (uncommon now)
Enterovirus
Epstein–Barr virus
Hepatitis A virus
Coxsackie virus B
Influenza A
Measles (uncommon now)
Leptospirosis
Mycoplasmosis
Typhoid fever
Ascariasis
Malaria
Rubella (uncommon now)
Metabolic
Protein-calorie malnutrition
Hypercalcemia
Reye's syndrome (rare now)
Hypertriglyceridemia
Cystic fibrosis (pancreatic sufficiency)
Miscellaneous
Henoch-Schönlein purpura
Systemic lupus erythematosus
Perforated duodenal ulcer
Kawasaki disease
Congenital partial lipodystrophy
Juvenile tropical pancreatitis
Following endoscopic retrograde cholangiopancreatography

 Copyright 2004, Elsevier (USA). All rights reserved.

years of age, the final diagnosis is rarely idiopathic, and in children under 3 years of age, all patients will have an associated systemic illness, such as sepsis or shock.

The incidence of pancreatitis in childhood has been difficult to determine. Acute pancreatitis has been an uncommon diagnosis at most pediatric institutions. However, a recent single institution review suggested that there may be an increasing incidence with between 70 and 100 new cases of pancreatitis now identified each year. The symptoms at presentation are similar to those in adults. Typically, children present with abdominal pain and/or vomiting. The combination of elevated serum lipase and amylase levels is reliable for the diagnosis of pancreatitis in children. Pancreatitis appears to be less severe in children with a low mortality compared to adults (no prospective studies available). However, severe metabolic and physical complications, such as hypoalbuminemia, hypocalcemia, pseudocysts, necrosis, pleural effusions, and hypoxia, can result.

Hemorrhagic pancreatitis is said to occur in approximately 13% of cases, but in recent experience it appears to be less frequent. Use of Ranson's or the Glasgow criteria to assess severity has not been validated for the pediatric population, but these criteria appear to be useful in providing evidence of complications.

Pseudocysts are reported to occur in approximately 15% of children with pancreatitis but limited information about this complication is available. Usually, the amylase and lipase levels are elevated and associated with abdominal pain, emesis, and/or fever.

The mainstay of treatment is symptomatic and similar to that in adults.

CHRONIC PANCREATITIS

Chronic pancreatitis is an inflammatory condition of the pancreas with evidence of permanent and progressive morphological damage observed on biopsy, autopsy, or radiological studies. Clinically, it is characterized by recurrent or persistent abdominal pain (80% of patients). Steatorrhea and weight loss may also be present. The incidence and prevalence of this condition in children are uncertain but chronic pancreatitis is uncommon. However, estimates in adults suggest that the prevalence is between 1 and 5% with an incidence of close to 8 cases per 100,000 population.

In children, chronic pancreatitis appears to result in two primary forms: (1) a calcifying form that can be associated with hereditary pancreatitis, hypercalcemia, hyperlipidemia, cystic fibrosis (with pancreatic sufficiency), juvenile tropical pancreatitis, and idiopathic pancreatitis and (2) an obstructive form (unusual in children) that is more commonly associated with trauma, sclerosing cholangitis, sphincter of Oddi dysfunction, congenital anomalies, idiopathic fibrosing pancreatitis, and chronic renal failure. Autoimmune pancreatitis resulting in chronic pancreatitis has been described in adults but, as of this writing, it has not been recognized in children.

Genetic susceptibility to chronic pancreatitis appears to be related to mutations in the cationic trypsinogen gene (PRSS1; see Hereditary Pancreatitis below), cystic fibrosis transmembrane conductance regulator (CFTR), and polymorphisms in the serine protease inhibitor, Kazal type 1 (SPINK1). These mutations do not account for all cases of chronic pancreatitis.

Treatment is symptomatic and similar to that used in adults.

HEREDITARY PANCREATITIS

Hereditary pancreatitis is an autosomal dominant disease that accounts for 1% of cases of both chronic and recurrent pancreatitis. The disease has been associated with mutations in the cationic trypsinogen gene. In addition, polymorphisms in SPINK 1 appear to increase the risk for the development of both familial (hereditary) and chronic pancreatitis but not primarily cause it. Clinically, there is incomplete penetrance with variability in the age at presentation and the severity of symptoms. Symptoms can begin as early as the first decade and are present in up to 80% of affected individuals by 20 years of age (range between 11 months and old age). Initial attacks resemble those of acute pancreatitis. Over time, patients develop signs and symptoms more typical of chronic pancreatitis although some patients will have calcification and some evidence of atrophy at the time of first presentation.

If hereditary pancreatitis evolves to chronic pancreatitis, some patients will develop complications more typical of this chronic disorder, including fibrosis, ductal abnormalities, calcifications, diabetes mellitus (reported in 10–25% of cases), steatorrhea and malabsorption (reported in 5 to 45% of cases), pseudocysts, and portal and splenic vein thrombosis. A more serious complication is that of pancreatic cancer, which may have a greater than 50-fold increased incidence in patients with hereditary pancreatitis.

Therapeutic interventions remain symptomatic. Thus, the approaches used are the same as those for other forms of pancreatic inflammation. Additionally, antioxidant cocktails are commonly prescribed. There are limited data supporting their effectiveness in this condition, although they are reported to be effective in recurrent pancreatitis.

METABOLIC PANCREATITIS

Hyperlipidemia

Hyperlipidemia is a reported cause of pancreatitis but appears to be exceedingly rare in pediatrics. The reported incidence of pancreatitis for each type of hyperlipidemia is as follows: type 1, 35%; type IV, 15%; and type V, 30–40%. The pancreatitis appears to be acute and recurrent. Use of an antioxidant cocktail has been reported to stop recurrent episodes of pancreatitis related to hyperlipidemia. If triglycerides are elevated during an episode of acute pancreatitis, they should be reevaluated after resolution of the pancreatitis to determine whether an underlying error in lipid metabolism is present.

Hypercalcemia

Hypercalcemia is a rare cause of pancreatitis. It can result from a variety of disorders, including total parenteral nutrition, vitamin D poisoning, sarcoidosis, metastatic bone disease, and the infusion (iatrogenic) of high doses of calcium. The role of primary hyperparathyroidism in either acute or chronic pancreatitis is unclear but certainly is a rare cause. It appears to occur in 7–15% of all patients with hyperparathyroidism and accounts for less than 1% of all cases of pancreatitis in adults or children.

See Also the Following Articles

Hyperlipidemia • Pancreatic Disease, Pediatric • Pancreatitis, Acute • Pancreatitis, Chronic • Pancreatitis, Hereditary

Further Reading

Digmagno, E. P., and Chari, S. (2002). Acute pancreatitis. *In* "Slesinger & Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 913–941. W. B. Saunders, New York.

Etemad, B., and Whitcomb, D. C. (2001). Chronic pancreatitis: Diagnosis, classification, and new genetic developments. *Gastroenterology* **120**, 682–707.

Forsmark, C. E. (2002). Chronic pancreatitis. *In* "Slesinger & Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 943–969. W. B. Saunders, New York.

Lopez, M. J. (2002). An increasing incidence of acute pancreatitis: A single institution perspective. *J. Pediatr.* **140**, 622–624.

On-line Mendelian Inheritance in Man web site. Available at http://www.ncbi.nlm.nih.gov/Omim/. [See entries for No. 167800, Pancreatitis, Hereditary; No. 276000, Protease, Serine, 1 (cationic trypsinogen); and No. 167790, Serine Protease Inhibitor, Kazal-Type, 1; Spink1].

Whitcomb, D. C. (2002). Hereditary and childhood disorders of the pancreas, including cystic fibrosis. *In* "Slesinger & Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 881–912. W. B. Saunders, New York.

Paraneoplastic Syndrome

ROBERTO DE GIORGIO*, MONICA BOVARA*, GIOVANNI BARBARA*, VINCENZO STANGHELLINI*, STEFANIA GUERRINI*, FABRIZIO DE PONTI*, MARCELLO TONINI†, AND ROBERTO CORINALDESI*

University of Bologna and †University of Pavia, Italy

chronic intestinal pseudo-obstruction A syndrome due to a failure of intestinal propulsion and characterized by a clinical picture resembling mechanical obstruction in the absence of any lesion occluding the lumen of the gut.

enteric nervous system The collection of neurons embedded in the wall of the gastrointestinal tract, organized in two major ganglionated (i.e., the myenteric or Auerbach's and submucosal or Meissner's) plexuses and able to control all digestive functions, including motility, absorption/secretion, sensitivity, and microcirculation.

onco-neural antigens Antigens shared by tumor cells and neural tissues that are likely capable of triggering the immune response involved in the neuropathophysiology of paraneoplastic syndrome.

Paraneoplastic syndromes encompass a wide array of clinical manifestations that may affect virtually all organs of the human body as a result of a systemic response evoked by a remote tumor. The gastrointestinal tract may be the target of a paraneoplastic syndrome evoked most commonly by small-cell lung carcinoma, thymoma, and gynecological and breast tumors. Symptoms arising in this context include diarrhea, constipation, nausea, vomiting, and abdominal pain. These symptoms may precede the diagnosis of the underlying neoplastic disease; thus, the differential diagnosis of gastrointestinal manifestations may also take into account the occurrence of a remote cancer. Symptoms are the result of altered gastrointestinal secretory motor function related to an impairment of the intrinsic innervation supplying the gastrointestinal tract (i.e., the enteric nervous system).

INTRODUCTION AND GENERAL FEATURES

Certain types of tumors express molecules, namely, onco-neural antigens, against which the immune system reacts as part of the immune-mediated anti-neoplastic response. This humoral and cell-mediated immune response cross-reacts with onco-neural antigens physiologically expressed by the enteric nervous system, thus leading to dysfunction and structural damage. As a result, a severe perturbation of gastrointestinal motility, ranging from achalasia, gastroparesis, and megacolon to chronic intestinal pseudo-obstruction, occurs. The histopathological hallmark of these motor disorders is an inflammatory lympho-plasmacellular infiltrate within the ganglionated plexuses of the enteric nervous system, mainly in the myenteric plexus (or Auerbach's plexus). This histopathological picture, referred to as myenteric ganglionitis, is associated with degeneration and loss of neurons, likely reflecting the immune-mediated injury within the myenteric plexus, throughout the gastrointestinal tract.

TYPES OF ANTI-NEURONAL ANTIBODIES IN PARANEOPLASTIC SYNDROME

In addition to a cellular immune response, several distinctive anti-neuronal antibodies can be found in the serum of patients with paraneoplastic dysmotility. Depending on their molecular target, they are referred to as anti-Hu (also known as type 1 anti-neuronal nuclear antibodies), anti-Yo (anti-Purkinje cell cytoplasmic antibodies), P/Q- and N-type Ca^{2+} channel antibodies, and ganglionic-type nicotinic acetylcholine receptor antibodies. Although the pathogenetic role played by anti-neuronal antibodies in enteric neuronal damage is still under investigation, their detection is of great importance in the diagnosis of an underlying, often occult, tumor.

Anti-Hu antibodies, usually found in the serum of patients with paraneoplastic encephalomyelitis/subacute sensory neuropathy, can be also detected in paraneoplastic gastrointestinal dysmotility. The Hu antigens are four nervous system-specific RNA-binding proteins, with a crucial role in neuronal development and survival. It is plausible that binding of autoantibodies to the enteric neuronal Hu proteins may play a role in the neuronal dysfunction underlying gastrointestinal paraneoplastic syndrome. Another class of anti-neuronal antibodies includes the anti-voltage-gated

 Copyright 2004, Elsevier (USA). All rights reserved.

Ca^{2+} channels, which are often identified in the serum of patients with Lambert-Eaton myasthenic syndrome associated with small-cell lung carcinoma. These autoantibodies, including the P/Q- and N-type Ca^{2+} channels (which control acetylcholine release), may also be involved in central as well as autonomic nervous system dysfunction related to paraneoplastic syndromes. Together with anti-Hu antibodies, the N-type anti-voltage-gated Ca^{2+} channel antibodies are the most common autoantibodies described in patients with paraneoplastic dysmotility.

Anti-Yo antibodies, found in the serum of patients with paraneoplastic cerebellar degeneration as a manifestation of gynecologic or breast cancer, are also detected in rare cases of paraneoplastic gastrointestinal dysmotility related to ovarian carcinoma. These antibodies target the Yo (recently redefined as cerebellar degeneration-related) antigens, which exert a functional inhibition of c-*myc* transcriptional activity, likely inducing neuronal damage through the activation of apoptosis.

In severe gastrointestinal dysmotility, the suspicion of a paraneoplastic syndrome should be always taken into account. Testing for neuronal autoantibodies is helpful for accurate diagnosis of patients with paraneoplastic syndrome. Further research is needed to better clarify the mechanisms underlying paraneoplastic-related enteric neuropathy.

See Also the Following Articles

Enteric Nervous System • Gastric Motility • Intestinal Pseudoobstruction

Further Reading

De Giorgio, R., Stanghellini, V., Barbara, G., Corinaldesi, R., De Ponti, F., Tonini, M., Bassotti, G., and Sternini, C. (2000). Primary enteric neuropathies underlying gastrointestinal motor dysfunction. *Scand. J. Gastroenterol.* **35**, 114–122.

Krishnamurthy, S., and Schuffler, M. D. (1987). Pathology of neuromuscular disorders of the small intestine and colon. *Gastroenterology* **93**, 610–639.

Lennon, V. A. (1996). Calcium channel and related paraneoplastic disease autoantibodies. *In* "Textbook of Autoantibodies" (J. B. Peter, and Y. Shoenfeld, eds.), pp. 139–147. Elsevier Science, Amsterdam, The Netherlands.

Musunuru, K., and Darnell, R. B. (2001). Paraneoplastic neurologic disease antigens: RNA-binding proteins and signaling proteins in neuronal degeneration. *Annu. Rev. Neurosci.* **24**, 239–262.

Sutton, I., and Winer, J. B. (2002). The immunopathogenesis of paraneoplastic neurological syndromes. *Clin. Sci.* **102**, 475–486.

Parasitic Diseases, Overview

RODNEY D. ADAM
University of Arizona College of Medicine, Tucson

cyst An environmentally stable form of the life cycle of a protozoan that reproduces asexually. The cyst has a hard outer wall and is involved in the transmission of the organism from one host to another.

definitive host The host in which a parasite achieves sexual maturity.

intermediate host For a parasite with sexual reproduction, this is an additional host(s) in which a parasite develops to some extent, but not to sexual maturity.

oocyst Cystic stage of the Apicomplexa, resulting from sporogony (e.g., *Plasmodium*, *Cryptosporidium*, *Toxoplasma*, *Cyclospora*).

parasite An organism that is physiologically dependent on another organism and extracts its nutrients from this host. More specifically, when the term is used for organisms colonizing humans, it frequently refers to protozoans and helminths and sometimes ectoparasites (e.g., ticks, lice).

trophozoite The vegetative or growing form of an organism, used to describe an asexual stage of a protozoan organism.

All the noncommensal (and perhaps commensal) bacteria, fungi, protozoa (or protists), and helminths are parasitic organisms when they colonize their human hosts. However, this article will be limited to the more commonly used, narrow definition that includes only the protozoa and helminths. The parasitic diseases are major causes of morbidity and mortality throughout the world, but strike the populations of the tropical and developing regions disproportionately. *Plasmodium falciparum,* the species that causes most cases of fatal malaria, along with *Mycobacterium tuberculosis*, is one of the two organisms most likely to cause fatal infection. Substantial disease burden also results from numerous other protozoan and helminthic organisms. This article will focus primarily on those parasitic organisms that cause diseases with significant gastrointestinal manifestations.

INTRODUCTION

In addition to the readily apparent differences in biology and size of the protozoa in comparison to the helminths, there are a number of important generalizations that can be made in regard to the differences in human infections and clinical manifestations. All of the protozoa that are pathogenic for humans are microscopic unicellular organisms that have the ability to reproduce asexually within the human host and thus have the potential to cause a very high disease burden from a single inoculating organism. In contrast, the pathogenic helminths are multicellular organisms that replicate sexually in the definitive host and many require at least one intermediate host to complete their life cycles. Therefore, they do not multiply within the human host, so the disease manifestations are determined by the initial inoculum size, the chronicity, and the host response to the infection. Most of the helminths do not live in the human host for more than several years, so late clinical manifestations are generally the residual of earlier infection and inflammatory response rather than ongoing infection. A corollary of this principle is that acquired immunodeficiency syndrome (AIDS) and other forms of immunocompromise do not result in the markedly increased disease burden that can be seen with protozoa, fungi, or bacteria. The major exception to this general rule is *Strongyloides stercoralis*, which has an autoinfection cycle and is able to replicate to very high numbers in immunocompromised human hosts, especially those receiving high doses of corticosteroids.

Blood eosinophilia is common with invasive helminth infections, frequently reaching very high levels (e.g., 30 to 50% of cases). The degree of eosinophilia is relatively mild in helminths that reside in the gut lumen. Protozoan pathogens typically do not evoke significant blood eosinophilia, but *Entamoeba histolytica* infections do frequently cause tissue eosinophilia. The most common (and not so common) medically important protozoa are listed in Table I and the helminths are listed in Table II. The subsequent portion of this article will give a brief description of the parasites with significant gastrointestinal manifestations. Even for those organisms without significant gastrointestinal manifestations, the diagnosis may be established by identification of the organism or its eggs in intestinal or fecal samples. The organisms for which the diagnosis may be made by gastrointestinal samples are listed in Table III.

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Protozoa Pathogenic for Humans

Organism	Classification	Geographic distribution	Means of human acquisition	Other mammalian hosts	Tissue tropism	Method of diagnosis	Clinical manifestations
Plasmodium sp.	Apicocomplexa (Sporozoa)	Tropical and subtropical regions	Anopheline mosquito bite	None (for human *Plasmodium* species)	Erythrocytes	Microscopic examination of blood smear	Febrile illness
Babesia spp.	Apicocomplexa	Worldwide	Tick bite	Rodent (*B. microti*); cattle (*B. bovis* and *B. divergens*)	Erythrocytes	Microscopic examination of blood smear	Febrile illess
Toxoplasma gondii	Apicocomplexa	Worldwide	Ingestion of pseudocysts in contaminated meat; ingestion of oocysts in material contaminated by cat fees; vertical transmission	Cat is definitive host; numerous mammals are incidental hosts	Nucleated cells, especially in heart and brain	Histology; serology	Multiple
Cryptosporidium parvum	Apicocomplexa	Worldwide	Ingestion of oocysts, especially in contaminated water	Cattle and other mammals	Small intestinal epithelial cells	Stool microscopy; intestinal histology	Diarrhea; rare extraintestinal disease
Cyclospora cayetanensis	Apicocomplexa	Worldwide	Ingestion of oocysts in contaminated food or water	None known	Small intestine	Stool microscopy	Diarrhea
Sarcocystis sp.	Apicocomplexa	Worldwide, especially Southeast Asia	Ingestion of sporocysts from feces or meat (human is definitive host); ingestion of fecal oocysts (human is intermediate host)	Carnivore is definitive host; herbivore is intermediate host	Muscle cells	Histology (muscle)	Usually asympto-matic; myositis; diarrhea
Isospora belli	Apicocomplexa	Tropical and subtropical regions	Ingestion of oocyst	None known	Small intestine	Stool microscopy	Diarrhea
Leishmania sp.	Kinetoplastid flagellate	Different locales for different species in tropical and subtropical regions	Bite of sandfly	Numerous	Histology; serology (for visceral)	Visceral infection; mucocutaneous; cutaneous	
Trypanosoma cruzi	Kinetoplastid flagellate	South and Central America	Bite of Reduviid bug	Numerous	Muscle, especially cardiac and GI	Serology	Acute (meningitis or carditis); chronic (cardiomyopathy, mega-GI disease)
Trypanosoma brucei sp.	Kinetoplastid flagellate	Equatorial Africa	Bite of Tsetse fly	Game animals (East African); none (West African)	Lymphatics, blood; CNS	Microscopy of blood or CSF Histology	Febrile illness with lymphadenopathy; meningo-encephalitis

(*continues*)

TABLE I Protozoa Pathogenic for Humans (*continued*)

Organism	Classification	Geographic distribution	Means of human acquisition	Other mammalian hosts	Tissue tropism	Method of diagnosis	Clinical manifestations
Giardia lamblia	Diplomonad flagellate	Worldwide	Ingestion, most commonly through contaminated water	Numerous	Small intestine	Stool microscopy; stool antigen detection	Diarrhea; malabsorption
Entamoeba histolytica	Rhizopod (ameba)	Developing regions	Ingestion, most commonly through contaminted food	None	Stool microscopy; serology	Colitis; hepatitic "abscesses"	
Dientamoeba fragilis	Trichomonad	Worldwide	Water; fecal–oral		Intestine	Stool microscopy	Usually asymp-tomatic; diarrhea
Balantidium coli	Ciliate	Worldwide	Ingestion of cysts	Numerous, especially pigs	Colon	Stool microscopy; intestinal histology	Usually asymptomatic; colitis
Blastocystis hominis	Unknown	Worldwide, especially developing regions	Probably fecal–oral	Unknown	Intestine	Stool microscopy	Usually asymptomatic; may cause diarrhea
Naegleria fowleri	Rhizopod (ameba)	Warm climate; freshwater	Nasal invasion from swimming in warm freshwater	Free-living	CNS	Microscopy of CSF	Meningitis
Acanthamoeba spp.	Rhizopod (ameba)	Worldwide	Eye (direct contact)	Free-living	CNS; cornea	CNS (brain histology); eye (histology or culture)	Meningoencephalitis; keratitis
Trichomonas vaginalis	Trichomonad	Worldwide	Sexual	None	Vaginal epithelium	Microscopy of vaginal fluid	Vaginitis
Enterocytozoon bieneusi	Microsporidia	Worldwide	Ingestion of contaminated water	Numerous	Gastrointestinal tract	Stool microscopy	Diarrhea (chronic in AIDS)

TABLE II Helminths Pathogenic for Humans

Organism	Geographic distribution	Means of human acquisition	Definitive host	Intermediate host(s)	Tissue tropism	Method of diagnosis
Nematodes						
Ascaris lumbricoides	Worldwide, especially tropics	Egg ingestion	Human, possibly other mammals	None	Small intestine	Identify ova in stool
Ancylostoma duodenale (Old World hookworm)	Eastern hemisphere	Penetration of skin by larvae	Human	None	Small intestine	Identify ova in stool
Necator americanus (New World hookworm)	Americas	Penetration of skin by larvae	Human	None	Small intestine	Identify ova in stool
Enterobius vermicularis	Worldwide	Egg ingestion	Human	None	Cecum	Identify worms in perianal area
Trichuris trichiura	Worldwide, especially developing regions	Egg ingestion	Human, other mammals	None	Large intestine	Identify ova in stool
Trichinella spiralis	Worldwide	Ingestion of larvae from muscle	Numerous carnivores	Same as definitive	Small intestine, muscle	Serology, identify larvae in muscle
Wuchereria bancrofti	Tropics and subtropics	Larvae injected by mosquito	Human	Mosquito	Blood, lymphatics	Identify microfilariae in blood, serology
Brugia malayi	South and Southeast Asia	Larvae injected by mosquito	Human	Mosquito	Blood, lymphatics	Identify microfilariae in blood, serology
Onchocerca volvulus	Africa, Central and South America	Larvae injected by blackfly	Human	Blackfly	Eye, skin	Identify microfilariae in skin, serology
Strongyloides stercoralis	Worldwide, especially tropics and subtropics	Penetration of skin by larvae; Penetration of colon by larvae (autoinfective cycle)	Human (also free-living)	None	Small intestine	Identify larvae in concentrated fecal samples, serology
Trematodes						
Schistosoma mansoni	Africa, South America	Penetration of skin by cercariae	Human	Snail	Inferior mesenteric veins	Identify eggs in stool, serology
Schistosoma hematobium	Africa, Middle East	Penetration of skin by cercariae	Human	Snail	Vesical plexus	Identify eggs in urine
Schistosoma japonicum	Southeast Asia	Penetration of skin by cercariae	Human	Snail	Superior mesenteric veins	Identify eggs in stool, serology
Paragonimus westermani	Asia, Africa, South America	Ingestion of metacercariae	Human	Snail, crab	Lungs	Identify eggs in stool
Clonorchis sinensis	Southeast Asia	Ingestion of metacercariae	Human	Snail, fish	Biliary system	Identify eggs in concentrated stool
Fasciola hepatica	Sheep-raising areas	Ingestion of metacercariae	Human	Snail, watercress	Biliary system	Identify eggs in stool
Cestodes						
Taenia solium	Scattered worldwide	Ingestion of cysticerci in meat (to be definitive host); ingestion of eggs (to be intermediate host)	Human	Pigs (occasionally humans)	Intestine (muscle and brain as intermediate host)	Identify eggs or proglottids in stool, serology
Taenia saginata	Cattle-raising areas	Ingestion of cysticerci in meat	Human	Cattle	Intestine	Identify eggs or proglottids in stool
Echinococcus spp.	Scattered worldwide	Egg ingestion	Canines	Numerous mammals	Liver, other organs	Serology, identify cysts in lesions
Diphyllobothrium latum	Scattered worldwide	Cyst ingestion from uncooked fish	Human, other mammals	Freshwater fish	Intestine	Identify eggs or proglottids in stool

TABLE III Important Parasitic Infections Diagnosed from Gastrointestinal Samples

Organism	Type of sample
Protozoa	
Entamoeba histolytica	Identification of cysts or trophozoites in feces or colon biopsy specimens (also do serology)
Giardia lamblia	Identification of cysts or trophozoites in feces
Cryptosporidium parvum	Identification of oocysts in feces or merozoites in intestinal biopsies
Cyclospora cayetanensis	Identification of oocysts in feces
Isospora belli	Identification of oocysts in feces
Enterocytozoon bieneusi	Identification of organism in intestinal biopsy or fecal sample
Helminths	
Trichuris trichiura	Identification of ova in microscopic examination of stool
Enterobius vermicularis	Identification of pinworms in perianal region
Ascaris lumbricoides	Identification of ova in microscopic examination of stool
Strongyloides stercoralis	Identification of larvae in feces or duodenal contents; poor sensitivity, so serologic studies should also be performed
Hookworm (*N. americanus* and *A. duodenale*)	Identification of ova in microscopic examination of stool
Schistosoma mansoni	Identification of ova in feces or rectal biopsy; serologic studies are helpful
Clonorchis sinensis	Identification of ova in feces using concentration techniques
Fasciola hepatica	Identification of ova or adult worms in feces
Paragonimus westermani	Identification of ova in microscopic examination of stool
Taenia solium	Identification of parasite proglottids or ova in feces (ova appear identical to *T. saginata* ova)
Taenia saginata	Identification of parasite proglottids or ova in feces (ova appear identical to *T. solium* ova)
Diphyllobothrium latum	Identification of parasite proglottids or ova in feces

PROTOZOA

The protozoa have generally been classified according to morphologic criteria into four major groups, the Sporozoa, the flagellates, the amebas, and the ciliates. The Sporozoa form a diverse group of organisms in terms of their hosts, transmission cycles, tissue tropism, and clinical manifestations, but clearly fall into one clade or genetic group. In contrast, the flagellates form several genetic groupings: the kinetoplasts (*Trypanosoma* spp., *Leishmania* spp.), the diplomonads (*Giardia lamblia*), and the trichomonads (*Trichomonas vaginalis*). *En. histolytica* is the most important pathogen of the amebas. *Balantidium coli* is rarely a pathogen, but is frequently listed since it is the only medically important ciliate. In addition to these four groups, the microsporidia have emerged as important pathogens, especially as complications of human immunodeficiency virus (HIV) infection. The microsporidia may actually be fungi, but will be addressed here since they are commonly considered to be protozoa. In addition to the pathogenic organisms, a number of protozoa are frequently identified in fecal specimens, but are seldom or never associated with disease. *Blastocystis hominis* may occasionally be pathogenic, although this remains controversial. *Entamoeba coli*, *Entamoeba hartmanni*, and *Endolimax nana*, as well as a number of other organisms, may be identified in fecal specimens, but are nonpathogenic.

En. histolytica is the most frequent parasitic cause of life-threatening gastrointestinal disease, resulting in an estimated 100,000 deaths per year throughout the world. Human infection is initiated by ingestion of *En. histolytica* cysts via contaminated food or water or by direct contact with infected individuals. There are no known nonhuman hosts, so zoonotic transmission does not occur. An accumulation of literature since 1978 has provided convincing evidence of pathogenic and nonpathogenic strains of *En. histolytica*, eventually resulting in the division of these organisms into two species, the pathogenic *En. histolytica* and the nonpathogenic *En. dispar*. The two organisms appear the same morphologically, so routine fecal specimens will not distinguish between the two.

After excystation in the small intestine, *En. histolytica* trophozoites replicate in the large intestine, where they ingest host immune cells and erythrocytes and cause tissue necrosis. Infected patients have locally invasive disease with abdominal pain, bloody diarrhea, and fever. Occasionally, the trophozoites disseminate beyond the large intestine, most commonly to the

liver. Patients with liver involvement have right upper quadrant abdominal pain and fever and discrete lesions that can be identified by ultrasound or computed tomography (CT) scanning. The finding of liver abscesses in a patient with a compatible clinical presentation who has anti-amebic antibodies is nearly diagnostic of amebic liver abscess, but pyogenic liver abscesses should be ruled out in patients with negative serologic tests and those who fail to respond to appropriate medical therapy (usually metronidazole).

Although the observation of *Entamoeba* cysts in fecal specimens in fecal specimens will not distinguish between *En. histolytica* and *En. dispar* infections, the discovery of trophozoites engulfing red blood cells (RBCs) is diagnostic of *En. histolytica* infection. RBCs are commonly present in stool specimens, but white blood cells are uncommon, probably because the trophozoites ingest the neutrophils. In invasive colonic disease, an antibody response is found in over 90% of patients. Although these are not protective antibodies, their measurement is very useful in the diagnosis of amebiasis. Up to 99% of patients with extraintestinal amebiasis will demonstrate an antibody response.

G. lamblia is a common cause of diarrheal disease with malnutrition and weight loss throughout the world. Human infection is initiated when the environmentally stable cysts are ingested, most commonly via contaminated water or by direct fecal–oral contact, such as may occur in a day-care center. Although there has been considerable controversy regarding the potential role of zoonotic transmission, recent molecular studies allow the conclusion that most dog and livestock *G. lamblia* isolates are genetically different from human isolates, indicating that these organisms are not major sources of human infection. In contrast, cats and beavers harbor genotypes that are the same as the human isolates, leaving open the possibility that human infection can be acquired from these animals.

Excystation occurs in the proximal small intestine, yielding the disease-causing trophozoites that replicate in the small intestine. Trophozoites adhere to the intestinal wall by mechanical suction generated by the ventral disk. The mechanism of diarrhea is unknown, since invasion does not occur and no toxins have been identified. Many people remain asymptomatic, but those with symptoms have diarrhea with loose foul-smelling stools and bloating. Weight loss is common and symptoms frequently last weeks or months in the absence of treatment. In the United States, giardiasis causes approximately the same number of hospitalizations as shigellosis. The diagnosis is usually established by the identification of cysts or trophozoites in fecal specimens or by detection of *Giardia* antigens in fecal specimens, either by indirect fluorescent antibody (IFA) or by enzyme-linked immunosorbent assay. The diagnosis can sometimes be made using a string test in which patients swallow a capsule on a string. After several hours, the capsule is examined microscopically for the presence of *Giardia* trophozoites. In patients with persistent diarrhea and malabsorption who have negative fecal examinations, upper endoscopy with small intestinal biopsy can be used to look for evidence of giardiasis or other small intestinal disease.

Cryptosporidium parvum is an apicocomplexan organism that was recognized as an important cause of diarrhea after the advent of the AIDS epidemic. Human infection results when oocysts are ingested, typically from contaminated water. The infections may be acquired from other humans or from livestock. An outbreak involving an estimated 400,000 persons in Milwaukee, Wisconsin, in 1993 may have resulted from human or livestock contamination of the water. In immunocompetent individuals, diarrhea with loose stools typically lasts approximately 1 week. However, in patients with advanced HIV infection, profuse watery diarrhea may persist for months, leading to dehydration. In addition, *C. parvum* may rarely disseminate to the gallbladder, causing symptomatic biliary disease. The asexual replication of merozoites in an intracellular, but extracytoplasmic space in small intestinal endothelial cells results in the symptoms. The diagnosis of intestinal disease is established when the oocysts are identified in fecal samples, either by routine staining or by acid-fast or fluorescence antibody staining. *Candida* spp. can be confused with *C. parvum* oocysts on routine stains, but the acid-fast stain will easily distinguish the two, since *C. parvum* oocysts are acid-fast. The diagnosis can also be established by identifying the merozoites in small intestinal endothelial cells obtained by intestinal biopsy. There is no known effective therapy for cryptosporidiosis, so treatment is supportive.

Cyclospora cayetanensis is a recently identified cause of diarrheal illness in humans. It was initially thought to be a cyanobacteria-like organism or a blue green alga, but in 1993, was described as an apicocomplexan, which is closely related to the poultry pathogens, the *Eimeria* spp. Cyclosporiasis gained prominence in the United States when it was associated with outbreaks of diarrheal disease acquired from imported raspberries in the summer of 1996 and subsequent summers. Human infections are initiated when sporulated oocysts are ingested followed by replication in the small intestine. Diarrhea with loose stools begins an average of 1 week after exposure and may last several weeks in the absence of therapy. The diagnosis is established by identifying the oocysts in fecal specimens from

infected persons after routine preparation or after acid-fast staining or by the detection of autofluorescing oocysts. To date, no nonhuman reservoirs have been identified.

Trypanosoma cruzi is a zoonotic organism that infects a variety of mammals throughout much of South and Central America and is transmitted from one host to another by the bite of a Reduviid (kissing) bug. Humans are accidental hosts and not part of the usual life cycle of the organism. Epimastigotes replicate in the Reduviid bug, followed by differentiation into metacyclic trypomastigotes. The infection results when the metacyclic trypomastigotes from the insect feces gain access through a mucous membrane surface or through compromised skin. The bite of the insect is intensely pruritic and the scratching that occurs after the bite may allow the *Tr. cruzi* epimastigotes access. The organisms then invade host cells, where they differentiate into amastigotes, replicate, and then differentiate into trypomastigotes followed by cell rupture and invasion of the next cell.

An acute illness with fever, local inflammation, and sometimes cardiac or central nervous system (CNS) manifestations may result. Most of these cases resolve spontaneously. Subsequently, whether or not symptoms are present acutely, some patients go on to a prolonged asymptomatic parasitemia (indeterminate phase). Some of those patients go on to develop chronic Chagas' disease, sometimes known as mega-gastrointestinal disease. The most common cause of morbidity and mortality is cardiomyopathy with concomitant ventricular arrhythmias. The other major cause of morbidity is hypertrophy and dilation of the gastrointestinal tract, most commonly the colon or esophagus. Patients may thus present with severe constipation or megacolon or with severe dysphagia and regurgitation. When the diagnosis is suspected in someone with an appropriate exposure history and clinical syndrome, serologic testing will usually support the clinical diagnosis. Endogenous cases in the United States have been documented rarely in Texas, but cases have been acquired through organ transplants from infected individuals and may also occur by blood transfusion. Anti-parasitic treatment is sometimes effective in acute infections, but is ineffective in chronic infections. Therefore, treatment of chronic infection is supportive.

The microsporidia are currently classified with the protozoa, but their classification is uncertain and they may belong to the fungi. They are obligate intracellular organisms and became recognized as significant human pathogens after the advent of the AIDS epidemic. Many species from multiple genera have been associated with human infection. *Enterocytozoon bieneusi* is the most common microsporidial infection in AIDS patients and infects the small intestine, biliary tract, and liver. Patients have diarrhea and malabsorption, with the severity and duration of disease depending on the level of immunosuppression. The diagnosis can be established by the identification of organisms in intestinal biopsy samples, by electron microscopy, or by routine histochemical methods. Sometimes the organisms can be detected in biliary, intestinal, or fecal fluids. Other microsporidia cause primarily disseminated or corneal disease.

HELMINTHS

The helminths that cause human infections are classified as nematodes (roundworms), cestodes (flatworms), and trematodes (flukes). The nematodes are frequently divided intestine and tissue-dwelling organisms.

Ascaris lumbricoides (roundworm) infection is initiated when the ova are ingested from an environmental source. When larvae hatch from the ova, they migrate through the lungs, sometimes resulting in cough and pulmonary infiltrates. Blood eosinophilia is common at this time. The larvae are then swallowed and return to the small intestine where they become adult worms. Most infections result in no gastrointestinal symptoms, but heavy infections can result in intestinal obstruction due to a mass of adult worms. Biliary tract invasion occurs rarely, resulting in acute abdominal pain and biliary obstruction. In roundworm infections with no overt symptoms, malabsorption is common, especially in developing areas.

Hookworm infections are established when larvae from the soil penetrate the skin and migrate through the lungs. They are then swallowed to make their way to the small intestine. The initial phase of the infection may result in pruritis at the site of larval penetration or pulmonary symptoms similar to those of roundworm infections. The majority of infections do not result in gastrointestinal symptoms, but abdominal pain, diarrhea, and malabsorption may be seen, especially with heavy infections. The most important clinical consequence of infection is iron deficiency anemia, which is proportional to the worm burden and is more severe with Old World hookworms (*Ancylostoma duodenale*) than with New World hookworms (*Necator americanus*).

Trichuris trichiura (whipworm) infections are initiated by the ingestion of ova, followed by hatching into larvae in the small intestine and migration of the larvae to the large intestine, where the adult worm lays eggs. Most infections are probably asymptomatic, but heavy

infections can result in iron deficiency anemia, dysentery, or rectal prolapse.

Trichinella spiralis infections (trichinosis) are established when meat that contains larvae from infected mammals (e.g., pigs, polar bears) is ingested. Cooking and prolonged freezing kill the larvae, so the risk is greatest from fresh raw meat. Mild gastrointestinal symptoms may ensue within a week after infection, but most patients remain asymptomatic at this stage. After approximately 2 to 6 weeks, larval migration results in fever, headache, periorbital or facial edema, conjunctivitis, and neurologic symptoms. The severity of illness generally correlates with the degree of the initial inoculum. Eosinophilia is pronounced. Death may occur from heart failure or CNS involvement. A suspected diagnosis may be supported by serologic testing and confirmed by the detection of larvae in muscle biopsy specimens. Treatment consists of corticosteroids in addition to anti-helminthic therapy.

St. stercoralis third-stage (L3) larvae in soil infect humans by direct penetration of intact skin. They pass through the lungs and mature in the intestine into parthogenetic adult worms, which lay eggs that hatch in the intestine to form L1 larvae. The L1 larvae may then complete their life cycle in the soil. However, as part of the autoinfective cycle, the larvae can mature to the L3 stage in the intestine and migrate to a location in the intestine or elsewhere. The number of larvae in the typical *Strongyloides* infection is small, but in immunocompromised hosts (especially those on high doses of corticosteroids for prolonged periods of time), large numbers of larvae are produced through the autoinfective cycle. These larvae may invade extraintestinal sites such as the lungs and CNS, carrying bacteria with them as they go. These bacteria are actually the major cause of morbidity and mortality in overwhelming strongyloidiasis.

Schistosomiasis refers to disease caused by several species of the genus, *Schistosoma*. The *Schistosoma* species that infect humans cause most of their disease when the adult worms lay eggs in the venous system for which they are tropic, causing scarring and venous obstruction. *Schistosoma hematobium* infects the bladder and its venous system, whereas the other *Schistosoma* species infect the gastrointestinal system and are tropic for parts of the mesenteric venous system. The most common cause of gastrointestinal and mesenteric infection is *Schistosoma mansoni*; the others have similar clinical manifestations, but vary in their geographic distributions. *Sc. mansoni* infection is initiated when cercariae released by the intermediate host snail, which resides in freshwater, penetrate the intact skin. Pruritis (swimmer's itch) may develop at the time of cercarial penetration, but is more commonly associated with schistosomes that do not infect humans; these latter schistosomes may be encountered in the United States. Approximately 1 to 2 months after a massive *Sc. mansoni* or *Schistosoma japonicum* infection, patients may develop a syndrome with acute fever with systemic and pulmonary symptoms, called Katayama fever. They have hepatosplenomegaly, lymphadenopathy, and eosinophilia. Most people recover spontaneously, but deaths may occur, especially with *Sc. japonicum*.

However, most of the morbidity associated with schistosome infections is due to the chronic scarring that results. The degree of scarring is generally proportional to the worm burden and egg output of the infecting organisms. The cercariae transform into schistosomulae, which first migrate to the liver and lungs and then to the inferior mesenteric plexus, where they mature into adults and lay their eggs. The scarring results in elevated splenic and portal pressures with splenomegaly and portal hypertension. The diagnosis can be established by the identification of *Sc. mansoni* eggs in fecal specimens or in rectal biopsies. In endemic areas, liver ultrasound is very effective in evaluating the extent of hepatic fibrosis resulting from the chronic infection. Screening is important in these areas, since anti-helminthic treatment may eliminate the organisms and prevent scarring, but does not reverse scarring that has already occurred.

Clonorchis senensis is a liver fluke found in eastern Asia that infects fish-eating mammals, whereas *Fasciola hepatica* is found in sheep-raising areas throughout the world. Human infections with these parasites may result in symptomatic biliary disease.

The human is the definitive host for *Taenia saginata*, the beef tapeworm, and cattle are the intermediate hosts. Human infections are generally asymptomatic, although mild abdominal pain or diarrhea may result. Infections may be diagnosed by finding the eggs or the proglottids in fecal specimens. When only the eggs are found, they cannot be distinguished morphologically from those of *Taenia solium*.

Humans are normally the definitive hosts for *Ta. solium*, the pork tapeworm. The human infection results when inadequately cooked meat containing cysticerci (the larval form) is ingested. The tapeworm then matures in the intestine and lays eggs, which are passed in the feces to be ingested by the porcine intermediate host. This form of infection is relatively inconsequential. However, humans may develop cysticercosis when they become an accidental intermediate host by ingesting eggs either from their own infection or from another source. The larvae then migrate to muscles or brain, where they eventually calcify. The inflammatory

response and calcifications of cysterci in the brain may result in seizures or other CNS manifestations. In endemic areas, cysticercosis is the most common cause of adult-onset seizures. Cysticercosis can be diagnosed by the classic clinical findings as well as by serologic testing. Patients may have simultaneous intestinal infection, so fecal specimens should be examined in patients with a new diagnosis of cysticercosis.

Canines are the definitive hosts of the *Echinococcus* species and the intermediate hosts vary with the species. *Echinococcus granulosus* causes hydatid disease and is the most common species to infect humans. Domestic dogs are the usual definitive hosts and sheep and other livestock animals are the normal intermediate hosts, so the organism is endemic in many grazing areas. In the United States, most cases are found in sheep-grazing areas of southern Utah and northern Arizona. Human infections are initiated when people ingest eggs from canine feces, becoming accidental intermediate hosts. The eggs hatch to form oncospheres, which migrate to their final destination to form larval cysts. The most common location is the liver, but the lungs or other organs may also be involved. Symptoms usually result from the space-occupying effect of the cysts, but an abrupt onset of fever with hypotension or even anaphylaxis can result from accidental rupture of a cyst. Imaging with ultrasound or CT scanning yields characteristic results and the suspected diagnosis can be confirmed by serologic testing. Therapy may consist of medical (albendazole) and/or surgical therapy and should be individualized. When needle aspiration or surgical excision of a possible echinococcal lesion is performed, care must be taken to avoid spilling the contents into the peritoneum, which could lead to life-threatening anaphylaxis or spreading of the infection. *Echinococcus multilocularis* causes a more aggressive disease (alveolar hydatid disease), since the cysts produce buds that metastasize throughout the body. Although liver lesions are the most common lesions, life-threatening cerebral lesions may also result. The usual definitive hosts of *Ec. multilocularis* are northern foxes and the intermediate hosts are the rodents on which the foxes feed; therefore, the organism is endemic to parts of the Arctic and sub-Arctic regions.

See Also the Following Articles

Amebiasis • Cestodes • Chagas' Disease• *Cryptosporidium* • Giardiasis • Helminth Infections • Nematodes • Trematodes • *Trichinella*

Further Reading

Adam, R. D. (2001). Biology of *Giardia lamblia*. *Clin. Microbiol. Rev.* **14**, 447–475.

Bhatia, G. (1997). Echinococcus. Semin. *Respir. Infect.* **12**, 171–186.

Capo, V., and Despommier, D. D. (1996). Clinical aspects of infection with *Trichinella* spp. *Clin. Microbiol. Rev.* **9**, 47–54.

Chen, X. M., Keithly, J. S., Paya, C. V., and LaRusso, N. F. (2002). Cryptosporidiosis. *N. Engl. J. Med.* **346**, 1723–1731.

Clark, D. P. (1999). New insights into human cryptosporidiosis. *Clin. Microbiol. Rev.* **12**, 554–563.

Didier, E. S. (1998). Microsporidiosis. *Clin. Infect. Dis.* **27**, 1–7.

Espinosa-Cantellano, M., and Martinez-Palomo, A. (2000). Pathogenesis of intestinal amebiasis: From molecules to disease. *Clin. Microbiol. Rev.* **13**, 318–331.

Franzen, C., and Muller, A. (1999). Molecular techniques for detection, species differentiation, and phylogenetic analysis of microsporidia. *Clin. Microbiol. Rev.* **12**, 243–285.

Herwaldt, B. L. (2000). *Cyclospora cayetanensis*: A review, focusing on the outbreaks of cyclosporiasis in the 1990s. *Clin. Infect. Dis.* **31**, 1040–1057.

Katz, D. E., and Taylor, D. N. (2001). Parasitic infections of the gastrointestinal tract. *Gastroenterol. Clin. North Am.* **30**, 797–815.

MacDonald, A. S., Araujo, M. I., and Pearce, E. J. (2000). Immunology of parasitic helminth infections. *Infect. Immunity* **70**, 427–433.

Mandell, G. L., Bennett, J. E., and Dolin, R. (eds.) (2000). "Principles and Practice of Infectious Diseases." Churchill Livingstone, Philadelphia, PA.

Marshall, M. M., Naumovitz, D., Ortega, Y., and Sterling, C. R. (1997). Waterborne protozoan pathogens. *Clin. Microbiol. Rev.* **10**, 67–85.

Nichols, G. L. (2000). Food-borne protozoa. *Br. Med. Bull.* **56**, 209–235.

Ortega, Y. R., and Adam, R. D. (1997). *Giardia*: Overview and update. *Clin. Infect. Dis.* **25**, 545–549.

Petri, W. A. Jr., and Singh, U. (1999). Diagnosis and management of amebiasis. *Clin. Infect. Dis.* **29**, 1117–1125.

Ross, A. G., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y., Williams, G. M., and McManus, D. P. (2002). Schistosomiasis. *N. Engl. J. Med.* **346**, 1212–1220.

Siddiqui, A. A., and Berk, S. L. (2001). Diagnosis of *Strongyloides stercoralis* infection. *Clin. Infect. Dis.* **33**, 1040–1047.

Umezawa, E. S., Stolf, A. M., Corbett, C. E., and Shikanai-Yasuda, M. A. (2001). Chagas' disease. *Lancet* **357**, 797–799.

Parasympathetic Innervation

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

dorsal vagal complex The parasympathetic center in the brainstem (medulla oblongata) where sensory signals from the gut and motor outflow to the gut are integrated.

efferent neuron A neuron that transmits signals from the brain or spinal cord to the gut.

Three divisions of the autonomic nervous system innervate the digestive tract. One is the parasympathetic division; the other two are the sympathetic and enteric divisions. The parasympathetic division is subdivided anatomically into the cranial and sacral divisions due to the neuroanatomic organization in which neurons that send fibers to the gut are located both in the brainstem and in the sacral region of the spinal cord. Projections to the digestive tract from these regions of the central nervous system are termed preganglionic efferents.

PARASYMPATHETIC NEURONS

Neuronal cell bodies of the parasympathetic cranial division reside in the medulla oblongata and project out of the brain in the vagus nerves (see Fig. 1). Cell bodies of the sacral division are located in the sacral region of the spinal cord and project in the pelvic nerves to the large intestine. Efferent fibers in the pelvic nerves make synaptic contact with neurons in ganglia located on the serosal surface of the colon and in ganglia of the enteric nervous system deeper within the large intestinal wall. Efferent vagal fibers form synaptic connections with neurons of the enteric nervous system in the esophagus, stomach, small intestine, and colon, as well as in the gallbladder and pancreas.

DORSAL VAGAL COMPLEX

Cell bodies of efferent vagal neurons are in the dorsal motor nucleus of the medulla oblongata. They are part of the dorsal vagal complex, which consists of the dorsal motor nucleus of the vagus, nucleus tractus solitarius, area postrema, and nucleus ambiguous. The nucleus tractus solitarius handles the sensory information entering the brain from the gut. The area postrema is a chemical sensor that signals the presence of agents in the blood and the nucleus ambiguous consists of motor neurons, some of which project to innervate the esophagus.

The dorsal vagal complex is the central integrative center for the outflow of signals from the brain to the gut. This center in the brain is more directly involved in the control of the specialized digestive functions of the esophagus, the stomach, and the functional cluster of the duodenum, gallbladder, and pancreas than in the distal small bowel and large intestine. The circuits in the dorsal vagal complex and their interactions with higher centers are responsible for the rapid and more precise control required for adjustments to rapidly changing conditions in the upper digestive tract during anticipation, ingestion, and digestion of meals of varied composition.

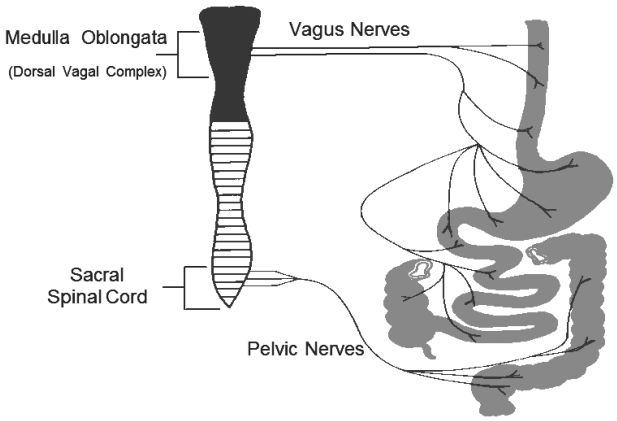


FIGURE 1 The parasympathetic division of the autonomic nervous system has cranial and sacral divisions. The cranial division consists of nerve cell bodies located in the medulla oblongata and the projections of the vagus nerves to the digestive tract. Cell bodies of neurons of the sacral division are positioned in the sacral region of the spinal cord. They project from the pelvic nerves to the large intestine.

 Copyright 2004, Elsevier (USA). All rights reserved.

FUNCTION

Efferent nerves of the cranial parasympathetic division transmit signals to the enteric innervation of the gastrointestinal musculature to control digestive processes both in anticipation of food intake and following the meal. This involves both stimulation and inhibition of contractile behavior in the stomach, stimulation of gastric acid secretion, stimulation of pancreatic secretion, and contraction of the gallbladder. Stimulation of inhibition or contraction results from activation of the enteric circuits that excite inhibitory or excitatory motor neurons, respectively.

Parasympathetic sacral efferents to the small and large intestinal musculature are predominantly stimulatory due to their input to the enteric microcircuits that control the activity of excitatory motor neurons. Signals transmitted by the sacral efferents are involved mainly in the initiation of defecation.

See Also the Following Articles

Autonomic Innervation • Brain–Gut Axis • Defecation • Enteric Nervous System • Gastric Motility • Vagus Nerve

Further Reading

Wood, J. (1999). "Neurogastroenterology and Motility." Milner-Fenwick, Timonium, MD.

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480. Lippincott Williams & Wilkins, Baltimore, MD.

Parenteral Nutrition

DOUGLAS W. WILMORE

Brigham and Women's Hospital and Harvard University

enteral Gastrointestinal feedings, either by mouth or through a tube placed in the stomach or intestinal tract.

parenteral nutrition The infusion of all essential nutrients by the intravenous route, bypassing the gastrointestinal tract. Also referred to as PN, total parenteral nutrition, artificial nutrition, intravenous feedings, or hyperalimentation.

tonicity A measure of the number of osmotic particles within a fluid. Isotonic refers to a similar number of particles in an intravenous solution and in the blood plasma. A hypertonic solution has a higher concentration of a solute than is found in blood.

Parenteral nutrition is a method of feeding patients by infusing a mixture of all necessary nutrients into the circulatory system, thus bypassing the gastrointestinal tract. This approach is also referred to as intravenous nutrition, total parenteral nutrition, artificial nutrition, and hyperalimentation. This method of nutritional support is utilized in patients who cannot eat or take nutrients into the intestinal tract by tube feedings. In other situations, this method of feeding is indicated in patients who cannot absorb adequate nutrients due to gastrointestinal disease, injury, or loss of the intestinal tract and therefore can receive nutrients into the body only by intravenous infusion.

BACKGROUND

Intravenous fluids administered to hospitalized patients usually contain only sugar (glucose, also referred to as dextrose) and some minerals (usually only salt, sodium chloride). These solutions are similar in concentration (tonicity) to blood plasma and therefore this infusate does not injure the lining of the veins in the arms through which they are infused (this injury to veins is a process called phlebitis). Additionally, the isotonic solutions do not damage the circulating red blood cells. This means, however, that the

Copyright 2004, Elsevier (USA). All rights reserved.

FUNCTION

Efferent nerves of the cranial parasympathetic division transmit signals to the enteric innervation of the gastrointestinal musculature to control digestive processes both in anticipation of food intake and following the meal. This involves both stimulation and inhibition of contractile behavior in the stomach, stimulation of gastric acid secretion, stimulation of pancreatic secretion, and contraction of the gallbladder. Stimulation of inhibition or contraction results from activation of the enteric circuits that excite inhibitory or excitatory motor neurons, respectively.

Parasympathetic sacral efferents to the small and large intestinal musculature are predominantly stimulatory due to their input to the enteric microcircuits that control the activity of excitatory motor neurons. Signals transmitted by the sacral efferents are involved mainly in the initiation of defecation.

See Also the Following Articles

Autonomic Innervation • Brain–Gut Axis • Defecation • Enteric Nervous System • Gastric Motility • Vagus Nerve

Further Reading

Wood, J. (1999). "Neurogastroenterology and Motility." Milner-Fenwick, Timonium, MD.

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480. Lippincott Williams & Wilkins, Baltimore, MD.

Parenteral Nutrition

DOUGLAS W. WILMORE

Brigham and Women's Hospital and Harvard University

enteral Gastrointestinal feedings, either by mouth or through a tube placed in the stomach or intestinal tract.

parenteral nutrition The infusion of all essential nutrients by the intravenous route, bypassing the gastrointestinal tract. Also referred to as PN, total parenteral nutrition, artificial nutrition, intravenous feedings, or hyperalimentation.

tonicity A measure of the number of osmotic particles within a fluid. Isotonic refers to a similar number of particles in an intravenous solution and in the blood plasma. A hypertonic solution has a higher concentration of a solute than is found in blood.

Parenteral nutrition is a method of feeding patients by infusing a mixture of all necessary nutrients into the circulatory system, thus bypassing the gastrointestinal tract. This approach is also referred to as intravenous nutrition, total parenteral nutrition, artificial nutrition, and hyperalimentation. This method of nutritional support is utilized in patients who cannot eat or take nutrients into the intestinal tract by tube feedings. In other situations, this method of feeding is indicated in patients who cannot absorb adequate nutrients due to gastrointestinal disease, injury, or loss of the intestinal tract and therefore can receive nutrients into the body only by intravenous infusion.

BACKGROUND

Intravenous fluids administered to hospitalized patients usually contain only sugar (glucose, also referred to as dextrose) and some minerals (usually only salt, sodium chloride). These solutions are similar in concentration (tonicity) to blood plasma and therefore this infusate does not injure the lining of the veins in the arms through which they are infused (this injury to veins is a process called phlebitis). Additionally, the isotonic solutions do not damage the circulating red blood cells. This means, however, that the

Copyright 2004, Elsevier (USA). All rights reserved.

solutions can contain only 50 g of glucose per liter of solution. Because the usual adult patient requires only 2–3 liters of solution per day to maintain normal hydration, only 100–150 g of glucose will be delivered daily, providing approximately 400–600 cal/day. This amount of energy is inadequate to meet the daily energy requirements of the usual adult, which usually range from 1500 to 2500 kcal/day, and these needs may be further increased due to disease and weight loss.

The breakthrough in the development of parenteral nutrition came on two fronts. First, dietary fat (soybean oil) was emulsified and provided as lipid particles, which were safe and well tolerated when administered intravenously. Next, all other essential nutrients (glucose, amino acids, electrolytes, vitamins, and minerals) were mixed together to make a highly concentrated solution. This nutrient mix was combined with the fat emulsion and infused slowly into the superior vena cava (the large vein that drains blood from the head, upper chest, and arms into the heart) via a central venous infusion catheter. The concentrated solution was rapidly diluted by the high level of blood flow through the heart and the nutrients were delivered by the bloodstream to all organs of the body. Infusion of this concentrated solution into an arm vein causes severe chemical phlebitis, resulting in extensive local inflammation and clotting at the infusion site. Infusion through the central venous catheter, however, allowed adequate nutrients to be delivered in a reasonable volume of fluid without this complication occurring.

Approximately 35 years ago, researchers demonstrated the efficacy of this approach by feeding beagle puppies for up to 1 year entirely by the intravenous route. Normal growth and development were observed. Next, the mixture was infused into an infant who was born with only a short length of small intestine, inadequate to maintain sufficient absorption. The child was sustained by intravenous feedings for up to 2 years while surgical correction of the bowel was undertaken. During this time, the baby also demonstrated normal growth and development. These were the first demonstrations that by infusing all necessary nutrients by the intravenous route and bypassing the gastrointestinal tract of an organism normal growth and development could be achieved.

INDICATIONS

In adults, the general indications for parenteral feedings include hospitalized patients who cannot take food or tube feeding formulas by the enteral route (e.g., via the intestinal tract). In well-nourished individuals, 7–10 days of conventional intravenous support (using 5% dextrose solutions) is generally provided, but if the period of partial starvation is to extend beyond this time, parenteral nutrition is indicated to prevent the potential complications associated with malnutrition. If the duration of illness is known to extend beyond 10 days, such as occurs in patients with severe pancreatitis, major trauma, or burn injuries, and enteral feedings are impossible, parenteral nutrition is initiated at an earlier stage in the hospital course. If the patient is malnourished as indicated by weight loss, the feedings will also be started as soon as the patient has stabilized. Preoperative patients with weight loss (greater than 15% normal body weight) who cannot take enteral feedings may also benefit from a course of parenteral nutrition for 7–10 days before surgery.

Individuals who have undergone massive intestinal resection may also require parenteral nutrition as a form of long-term nutritional support. Others with inflammatory bowel disease, severe malabsorption, and motility and obstructive disorders also frequently require such intravenous support.

Premature infants have minimal body stores of adipose and lean tissue and often require several weeks of intravenous feedings before full enteral feeding is possible. Therefore, parenteral nutrition is often utilized in these infants and those requiring surgical correction of major gastrointestinal anomalies.

NUTRITIONAL REQUIREMENTS

In order to provide adequate nutrients to an individual patient, the specific nutritional requirements of the individual must be determined. These needs may differ from requirements that were present during health; malnutrition is associated with deficits that need repair and many diseases are associated with infection and inflammation that impose increased requirements for many nutrients.

Energy

Basal energy requirements are a function of the individual's weight, age, gender, and activity level and the disease process. In general, hospitalized adults require approximately 25–30 kcal/kg/day but these requirements may be greater in patients with injury or infection (see Table I).

Protein

Protein (or amino acids, the building blocks of proteins) is the functional and structural component of the

TABLE I Energy Requirements

Patient condition	Basal metabolic rate	Approximate energy requirement (kcal/kg/day)
No postoperative complications, gastrointestinal fistula without infection	Normal	25–30
Mild peritonitis, long-bone fracture or mild to moderate injury	25% above normal	30–35
Severe injury or infection	50% above normal	35–45
Burn 40–100% of total body surface	Up to 100% above normal	35–60

body. With disease, poor food intake, and inactivity, body protein is lost and individuals become weak and waste muscle mass. Protein requirements for most healthy individuals are 0.8 g/kg/day (approximately 40–70 g of protein/day). Critically ill patients may need 1.5–2.0 g protein/kg/day (approximately 60–150 g/day) depending on the disease process, but this amount is often reduced in patients with kidney or liver disease.

Vitamins and Minerals

These requirements are usually met when standard volumes of a nutrient mix are provided. Increased amounts of vitamins are usually provided to severely ill patients (Table II) and blood levels are periodically determined to adjust the infusion levels of minerals such as sodium, potassium, chloride, phosphorous, magnesium, and zinc. Trace elements are usually added daily (Table III).

APPLICATION

The Solution

The patient is assessed and the nutritional requirements are determined by a physician, dietician, or other skilled provider. A prescription is submitted to the pharmacy or nutrition-mixing service for a mixture of parenteral nutrients to be composed for the specific patient. The nutrients are usually placed in 2- to 3-liter plastic bags and the contents are infused daily (Table IV). Alternatively, premixed solutions can be taken off the shelf and utilized but the fixed nutrient concentration limits the versatility of this approach in a heterogeneous patient population.

The Central Venous Catheter

To administer the highly concentrated nutrient solutions, a catheter must be placed into the central venous system with its tip in the superior vena cava.

TABLE II Vitamin Requirements

Vitamin	Units	Recommended dietary allowance (RDA) for daily oral intake	Daily requirement of the moderately injured	Daily amount provided by standard intervenous preparations
Vitamin A (retinol)	IU	1760 (females)–3300 (males)	5000	3300
Vitamin D (ergocalciferol)	IU	200	400	200
Vitamin E (tocopherol)	mg	8–10	Unknown	10
Vitamin K (phylloquinone)	μg	20–40	20	0
Vitamin C (ascorbic acid)	mg	60	75	100
Thiamine (vitamin B_1)	mg	1.0–1.5	2	3
Riboflavin (vitamin B_2)	mg	1.2–1.7	2	3.6
Niacin	mg	13–19	20	40
Pyridoxine (vitamin B_6)	mg	2.0–2.2	2	4
Pantothenic acid	mg	4–7	18	15
Folic acid	mg	0.4	1.5	0.4
Vitamin B_{12}	μg	3.0	2	5
Biotin	μg	100–200	Unknown	60

TABLE III Trace Mineral Requirements

Mineral	Recommended dietary allowance (RDA) for daily oral intake (mg)	Suggested daily intravenous intake (mg)
Zinc	15	2.5–5.0
Copper	2–3	0.5–1.5
Manganese	2.5–5.0	0.15–0.8
Chromium	0.05–0.2	0.01–0.015
Iron	10 (males)–18 (females)	3

Such a catheter can be placed via the subclavian vein, through a neck vein (the jugular vein approach is less desirable because of the high rate of associated infection), or by using a long catheter placed in an arm vein and then threaded into the central venous system (a peripherally inserted central catheter line) (Fig. 1). Once the correct position of the catheter has been established (usually by X ray), the infusion can begin.

Administration

To ensure that the solution is administered at a continuous rate, an infusion pump is utilized to administer the solution. In hospitalized patients, infusion usually occurs over 22–24 h/day. In ambulatory home patients, administration usually occurs overnight (12–16 h).

Monitoring

With solution infusion, a variety of determinations are made to ensure that the individual is responding appropriately to the infusate. This step involves primarily the measurement of glucose in the blood or urine (using techniques similar to those utilized by diabetic patients when monitoring sugar levels) but monitoring may be much more frequent and complex in critically ill patients. Urine output is measured over each 24 h period. In hospitalized patients, these measurements are usually performed by a highly trained nurse.

TABLE IV The Composition of a Standard Liter of Adult Parenteral Nutrient Solution

Glucose	150 g
Amino acids	42 g
Sodium	50 mEq
Chloride	60 mEq
Potassium	40 mEq
Phosphorous	15 mg
Acetate	75 mEq
Magnesium	12 mEq
Sulfate	12 mEq
Calcium	5 mEq
Gluconate	5 mEq

Variations Due to Age and Disease

Fluid volume and nutrient concentration must vary in pediatric patients depending on their age and weight. In adults with abnormal fluid loss, renal or liver failure, or other metabolic disorders, special mixtures are compounded to meet specialized fluid and nutrient requirements.

PERIPHERAL VEIN NUTRITION

Slightly hypertonic nutrition solutions can be prepared from commercially available amino acid mixtures (5%) dextrose solutions (10%), and fat emulsions (20%). These nutrient mixtures have a low caloric density (approximately 0.3 to 0.6 kcal/ml) and thus provide only 1200 to 2300 kcal/day in 2000 to 3500 ml of solution. The advantage to using these dilute nutrient mixtures is that they can be infused through a plastic cannula placed in a large-bore arm vein, thus avoiding the central catheter. This approach may be useful in the short term or, combined with minimal feedings administered via the gastrointestinal tract, the two methods of nutrient administration together may provide adequate nutrient requirements to a critically ill patient. These solutions tend to cause phlebitis (inflammation) of the arm veins and the infusion site must be inspected frequently and the infusion site changed every 48–72 h. Thus, this approach is acceptable only as a short-term solution to delivery of the parenteral nutrients.

COMPLICATIONS AND MONITORING

Catheter Care and Catheter Sepsis

Although it is not the most common problem related to intravenous feedings, catheter infection is probably the most common, serious problem related to this technique. The catheter is a plastic or Silastic tube that passes through a skin puncture site and then enters

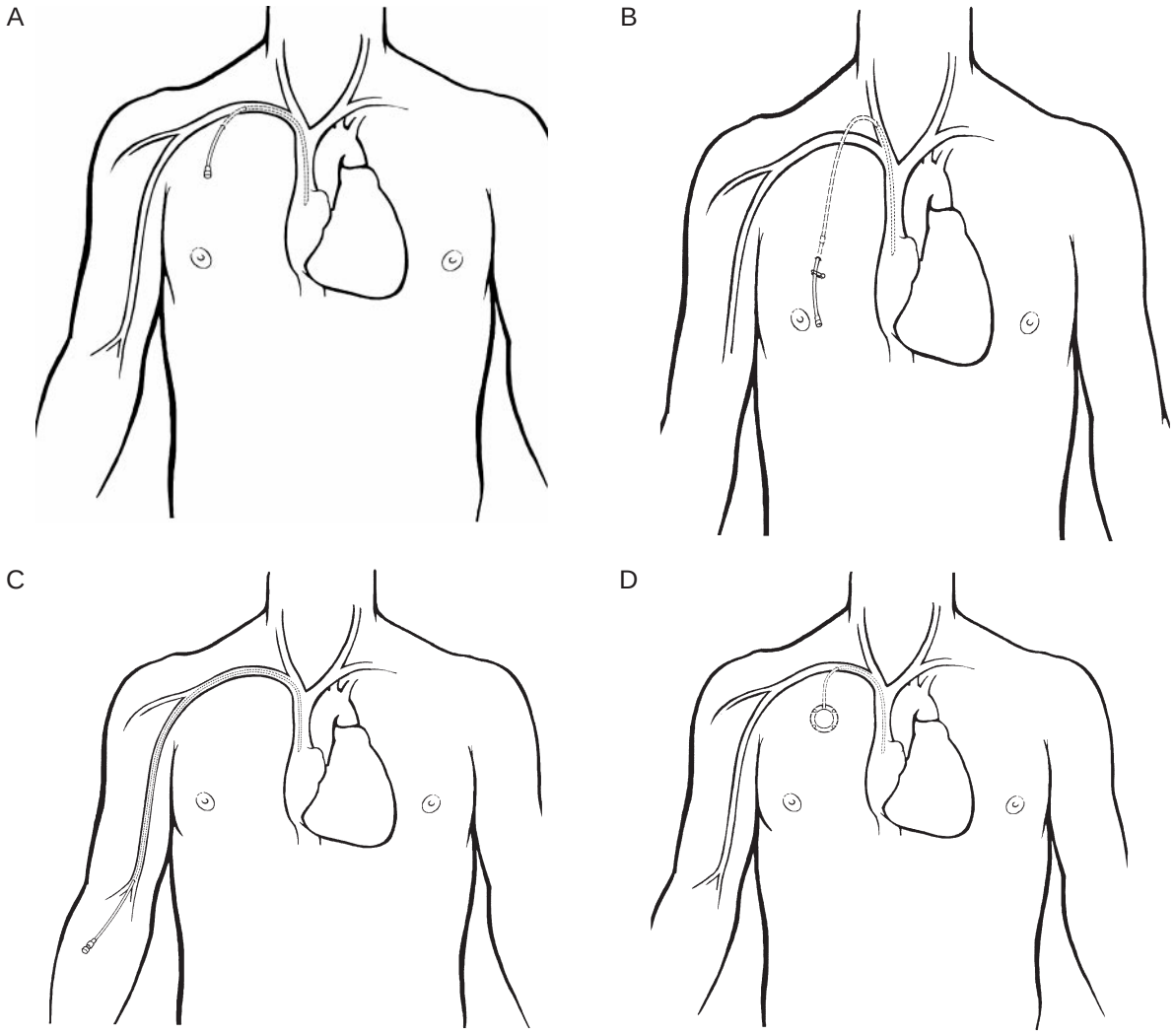


FIGURE 1 (A) This is a central venous catheter placed in the subclavian vein with its tip in the superior vena cava. (B) In the operating room, a catheter is inserted into the jugular vein in the neck and is directed into the superior vena cava. The distal portion of the catheter is tunneled under the skin to provide a subcutaneous tract that will resist infections. This device, which is inserted in patients requiring long-term (home) therapy, is often referred to as a Hickman catheter. (C) This fine plastic catheter is placed in an arm vein and is directed into the superior vena cava. This procedure is performed at the bedside, usually by a highly trained nurse. This approach is commonly referred to as a PICC (peripherally inserted central catheter) line. (D) This catheter is buried entirely below the skin and has a circular port covered by a rubber diaphragm, which is placed just below the skin. The patient intermittently pierces the skin and underlying diaphragm with the infusion needle in preparation for the nutrient infusion. The needle is removed at the termination of the therapy, so that no external catheter is visible.

the circulatory (venous) system. Thus, the catheter entrance site may serve as a portal for the entry of bacteria and other microorganisms into the body. In addition, many individuals requiring this therapy have inadequate mechanisms to fight infection, making these malnourished or critically ill patients all the more vulnerable to this completion of catheter sepsis (infection).

Catheter sepsis is characterized by the classic signs of infection: chills, fever, and, on occasion, drainage around the catheter entrance site. The white blood cell count is usually elevated and frequently

microorganisms are cultured from the bloodstream or the catheter tip. Fortunately, with removal of the infusion catheter, the symptoms usually abate in most patients. A short course of antibiotics is then usually adequate to clear the infection. However, many critically ill patients have fever from other sources (pneumonia or wound infections, for example) and the presence of this symptom complex associated with fever in a seriously ill patient often makes accurate diagnosis of catheter infection very difficult.

To prevent this complication, a rigorous program of catheter care is followed. Only intravenous nutrition solutions are administered through the catheter and no blood may be withdrawn from the catheter. Two to three times weekly, the dressing around the catheter is removed and, using sterile technique, the skin around the entrance site is scrubbed to reduce the number of microorganisms on the skin, thus decreasing the chance of catheter infection. The entrance site is inspected for signs of inflammation and drainage and if present, cultures are usually taken or the catheter is removed. Sterile technique is also utilized while attaching a new bag of nutrient solutions to the catheter for infusion. Catheter care and administration of the nutrient solution are generally carried out by a specially trained member of the nursing staff or a nurse with expertise in this field assigned to the nutritional support service.

Metabolic Complications

Although there is wide potential for problems (Table V) because of the variability in the patient's clinical status, metabolic complications are generally minimized by the adherence to a strict monitoring protocol (Table VI).

The major concern is the occurrence of hyperglycemia (an elevated blood sugar), which is associated with the infusion of excess glucose in the feeding solution or the diabetic-like state in the patient associated with many critical illnesses. Hyperglycemia can result in an osmotic diuresis (abnormal loss of fluid via the kidney), dehydration, and hyperosmotic coma. A diagnosis is made by the presence of elevated blood glucose levels and the detection of sugar in the urine. When this complication occurs, the infusion solution may be reformulated to decrease the amount of infused

TABLE V Some Metabolic Complications of Parenteral Nutrition

Problems	Possible causes	Solution
Glucose		
Hyperglycemia, glycosuria, osmotic diuresis, hyperosmola nonketotic dehydration, and coma	Excess quantity of glucose infused; inadequate endogenous insulin; increased glucocorticoids; sepsis	Reduce quanity of glucose infused
Ketaocidosis in diabetes mellitus	Inadequate endogenous insulin response; inadequate exogenous insulin therapy	Administer exogenous insulin; reduce glucose
Fat		
Altered coagulation	Hyperlipidemia	Decrease administration rate
Hypertriglyceridemia	Rapid infusion; decreased clearance	Decrease administration rate
Impaired liver function	May be caused by fat emulsion or by an underlying disease process	Consider infusions for only 16–18 h/day
Essential fatty acid deficiency	Inadequate essential fatty acid administration	Administer fat emulsion
Amino acids		
Serum amino acid imbalance	Unphysiologic amino acid profile of the nutrient solution; ?? amino acid utilization with various disorders	Change type of amino acid mixture administered
Hyperammonemia	Excessive ammonia in protein hydrolysate solutions; deficiency of specific amino acids; primary hepatic disorder	
Prerenal asotemia	Excessive amino acid infusion with inadequate calorie administration; inadequate free water intake; dehydration	Reduce amino acid intake
Elevated or subnormal blood levels of electrolytes and minerals	Inadequate or excess administration	Adjust administration rate; evaluate underlying pathophysiology

TABLE VI Monitoring Patients Receiving Parenteral Nutrition

Variables	Suggested monitoring frequency	
	First week	Later
Energy balance	**Daily**	**Daily**
Weight		
Metabolic variables		
Blood measurements	Daily	1–2 × weekly
Plasma electrolytes (Na^+, K^+, Cl^-)	3 × weekly	2 × weekly
Blood urea nitrogen	3 × weekly	2 × weekly
Plasma total calcium and inorganic phosphorus	Daily	3 × weekly
Blood glucose	3 × weekly	2 × weekly
Plasma transaminases	2 × weekly	Weekly
Plasma total protein and fractions	As indicated	As indicated
Blood acid–base status	Weekly	Weekly
Hemoglobin	2 × weekly	Weekly
Magnesium	Weekly	Weekly
Triglycerides		
Urine measurements	Daily	Daily
Glucose	Daily	Daily
Specific gravity or osmolarity		
General measurements	Daily	Daily
Volume of infusate	Daily	Daily
Oral intake (if any)	Daily	Daily
Urinary output		
Prevention and detection of infection		
Clinical observations (activity, temperature, symptoms)	Daily	Daily
WBC and differential counts	As indicated	As indicated
Cultures	As indicated	As indicated

glucose (usually the glucose infused should not exceed 4 mg/kg/min). Alternatively, insulin can be administered (either administered by subcutaneous injection or placed in the infusion bag).

The patient may also accumulate triglycerides in the bloodstream with infusion of the fat emulsion. Infusion of both glucose and fat emulsion in excess may result in pulmonary insufficiency. With excess glucose infusion, excess carbon dioxide (CO_2) production occurs as CO_2 accumulates in the bloodstream, a result of glucose metabolism. With lipid infusion, the lipid particles may accumulate in the lungs and reduce the diffusion capacity of respiratory gases.

In addition to these metabolic problems, the concentrations of a variety of electrolytes and minerals may vary and these levels are monitored by frequently determining concentrations in the bloodstream. Adjustments in the concentrations of the nutrients in the infusates are then made.

Mechanical Complications

As with any device, catheters and tubing may become clotted or twist and obstruct. Pumps may also fail or operate improperly. Of more concern is the infrequent complication of superior vena cava thrombosis, which is related to the catheter being placed in this large vein. This complication is associated with clotting of this large vessel draining the upper body and results in swelling of the arms and face. This occurs more commonly in children than adults but usually prevents further intravenous support and thus poses a serious life-threatening problem to many patients when this complication occurs.

HOME PARENTERAL NUTRITION

Patients who are unable to eat and absorb adequate nutrients for maintenance over the long term may be candidates for home parenteral nutrition. Many such individuals suffer from short-bowel syndrome caused by loss of the intestinal tract due to (1) extensive Crohn's disease, (2) mesenteric infarction, or (3) severe abdominal trauma. Pseudo-obstruction, radiation eneteritis, carcinomatosis, necrotizing enterocolitis, and intestinal fistulas are other reasons for poor bowel function and indications for home nutritional support.

To be eligible for this home parenteral nutritional therapy, patients must be able to master the techniques associated with this support system, be motivated, and have adequate social support in the home. Patients receive extensive evaluation, teaching, and training during a period of hospitalization that covers basic principles of parenteral nutrition and provides guidelines for catheter care, the maintenance of asepsis, and the use of infusion pumps.

A patient who is judged to be a candidate for home parenteral nutrition requires an indwelling Silastic catheter designed for long-term permanent use (see Fig. 1).

The nutrient solutions are prepared weekly and delivered to the patient's home. The patient sets up the infusion system and attaches the catheter to the delivery tubing in the evening for infusion over the next 12–16 h. The intravenous nutrition is terminated by the patient the next morning. Home care nurses see the patient at regular intervals and monitor infusion techniques, evaluate the patient's response to therapy, and continue the educational process for both the patient and family.

Home nutrition evolved because patients required this type of nutritional support for months or years, but costs of hospitalization made such therapy impossible in the hospital setting. This resulted in the development of private home care companies who have a support staff (including home nurses), mix and provide the solutions, maintain an inventory of supplies, respond to emergencies, and facilitate billing and collection for these services from insurers.

EFFICACY

Though the need for nutrition support in critically ill patients may seem obvious, this method of care has only recently became subject to critical review and to the objective assessment of outcome and cost benefit. A summary of these findings follows.

In one report, a total of 13 prospective randomized trials that evaluated the use of preoperative parenteral nutrition in surgical patients were identified. The patients were considered by their physicians to be moderately malnourished. A pooled analysis of the data showed that patients who received preoperative TPN had 10% fewer postoperative complications than the control group. The analysis found no significant difference in mortality between the parenterally fed groups and the controls.

Use of parenteral nutrition in the immediate postoperative period was also evaluated. In contrast to the analysis of the preoperative data, this study of postoperative feeding concluded that patients who received TPN after an operation had an approximately 10% higher risk of ensuing complications, with no associated benefit.

Trials addressing perioperative nutritional support have also been conducted in patients undergoing surgery in the upper abdomen for gastroenterological malignancies. In one such study, patients undergoing major pancreatic resection were randomly assigned either to a group that received TPN on postoperative day 1 or to a non-TPN group. No significant benefit from the use of adjuvant TPN could be demonstrated and the incidence of complications (primarily those associated with infection) was significantly greater in the TPN group.

Finally, a large meta-analysis evaluated the effects of providing parenteral nutrition versus no feeding in 2211 critically ill patients. The use of intravenous feedings did not influence mortality but may have reduced the complication rate, especially in malnourished patients. The authors concluded that further studies were necessary to firmly establish the latter conclusion.

All of these studies indicate the need to identify specific targeted patient groups who will benefit from this complex and expensive technique. Broad-scale use of this nutritional support technique is not indicated.

THE FUTURE

Because of costs, complications, and the ability to provide nutritional support by enteral tube feedings, the use of parenteral nutrition is decreasing in the United States. However, in specific patient groups it remains cost-effective and life-saving.

Several new components or products have become available to improve the type of substrate administered. New fat emulsions that include omega-3 fatty acids (fish oil) have been developed and these compounds are believed to be able to enhance the immune system in critically ill patients. In addition, some amino acids have not been included in the parenteral mixtures because of solubility problems or instability related to sterilization or prolonged shelf-life. However, these compounds have recently been combined with other amino acids to form stable dipeptides, which are now being included in the parenteral amino acid solutions. One such substance that is now available is the amino acid glutamine, which is thought to enhance immunological and bowel function in selected populations. This may translate into a method of nutritional support that improves recovery following major surgery and chemotherapy and improves resistance to infection in individuals with other life-threatening illnesses.

See Also the Following Articles

Enteral Nutrition • Nutritional Assessment • Short Bowel Syndrome

Further Reading

Baker, R. D., Jr., Baker, S. S., and David, A. M. (eds.) (1997). "Pediatric Parenteral Nutrition." Chapman and Hill, New York.

Dudrick, S. J., Wilmore, D. W., Vars, H. M., and Rhoads, J. E. (1968). Long-term total parenteral nutrition with growth, development, and positive nitrogen balance. *Surgery* **64**, 134–143.

Klein, S., Kinney, J., Jeejeebhoy, K., *et al.* (1997). Nutritional support in clinical practice: Review of published data and recommendations for further research directions. *J. Parenteral Enteral Nutr.* **21**, 133–156.

Rombeau, J. L., and Rolandelli, R. H. (eds.) (2000). "Clinical Nutrition: Parenteral Nutrition." W. B. Saunders, Philadelphia, PA.

Wilmore, D. W., and Dudrick, S. J. (1968). Growth and development of an infant receiving all nutrients exclusively by vein. *J. Am. Med. Assoc.* **203**, 860–864.

Parietal Cells

CATHERINE S. CHEW
Medical College of Georgia, Augusta

cotransporter An ion transporter that carries more than one ion in the same direction across the outer cell membrane.

endocytosis The process by which a cell membrane folds inward to internalize substances.

exchanger An ion transporter that transports ions in opposite directions across the cell membrane.

exocytosis The release of substances in a vesicle by a process in which the membrane surrounding the vesicle fuses with the membrane surrounding the outside of the cell.

H^+,K^+-ATPase The hydrogen, potassium ATPase protein that pumps protons across the cell membrane in a 1:1 exchange for potassium ions.

microvilli Microscopic, hair-shaped cellular projections that contain F-actin in their cores.

secretagogue A substance, such as a hormone or paracrine signal, that stimulates secretion.

Parietal cells are located within glands in the gastric mucosa. They have the unique ability to secrete a massive amount of hydrochloric acid (HCl), which sustains the highly acidic environment within the lumen of the stomach. The unusual morphology of the parietal cell has fascinated researchers for over 100 years. When parietal cells are stimulated by the appropriate secretagogues, they also undergo spectacular morphological transformations that are correlated with the activation of the enzyme H^+,K^+-ATPase. HCl is generated by the H^+,K^+-ATPase, acting in concert with ion channels and other ion transporters. The H^+,K^+-ATPase is composed of a catalytically active α-subunit and a β-subunit that targets the enzyme to the appropriate membrane compartment. The three major secretagogues, histamine, acetylcholine, and gastrin, stimulate HCl secretion by activating different signaling pathways that interact at the level of the parietal cell. Because of the common occurrence of peptic ulcer disease, the acid secretory function of the parietal cell has been studied extensively. However, the parietal cell has other functions. In human, monkey, cat, dog, sheep, rabbit, and guinea pig, parietal cells secrete intrinsic factor, which is essential for the efficient absorption of vitamin B12 or cyanocobalamin in the distal ileum. Parietal cells may also secrete prostaglandins as well as growth factors including transforming growth factor-α, amphiregulin, and heparin-binding epidermal growth factor. Prostaglandins exert a cytoprotective effect on the gastric mucosa and the local release of growth factors appears to play an important role in regulating the growth and differentiation of epithelial cells within the gland units.

LOCATION AND STRUCTURE OF THE PARIETAL CELL

Parietal cells are present in glands within the fundus and body of the stomach and are the largest cells in these

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Enteral Nutrition • Nutritional Assessment • Short Bowel Syndrome

Further Reading

Baker, R. D., Jr., Baker, S. S., and David, A. M. (eds.) (1997). "Pediatric Parenteral Nutrition." Chapman and Hill, New York.

Dudrick, S. J., Wilmore, D. W., Vars, H. M., and Rhoads, J. E. (1968). Long-term total parenteral nutrition with growth, development, and positive nitrogen balance. *Surgery* **64**, 134–143.

Klein, S., Kinney, J., Jeejeebhoy, K., *et al.* (1997). Nutritional support in clinical practice: Review of published data and recommendations for further research directions. *J. Parenteral Enteral Nutr.* **21**, 133–156.

Rombeau, J. L., and Rolandelli, R. H. (eds.) (2000). "Clinical Nutrition: Parenteral Nutrition." W. B. Saunders, Philadelphia, PA.

Wilmore, D. W., and Dudrick, S. J. (1968). Growth and development of an infant receiving all nutrients exclusively by vein. *J. Am. Med. Assoc.* **203**, 860–864.

Parietal Cells

CATHERINE S. CHEW
Medical College of Georgia, Augusta

cotransporter An ion transporter that carries more than one ion in the same direction across the outer cell membrane.

endocytosis The process by which a cell membrane folds inward to internalize substances.

exchanger An ion transporter that transports ions in opposite directions across the cell membrane.

exocytosis The release of substances in a vesicle by a process in which the membrane surrounding the vesicle fuses with the membrane surrounding the outside of the cell.

H^+,K^+-ATPase The hydrogen, potassium ATPase protein that pumps protons across the cell membrane in a 1:1 exchange for potassium ions.

microvilli Microscopic, hair-shaped cellular projections that contain F-actin in their cores.

secretagogue A substance, such as a hormone or paracrine signal, that stimulates secretion.

Parietal cells are located within glands in the gastric mucosa. They have the unique ability to secrete a massive amount of hydrochloric acid (HCl), which sustains the highly acidic environment within the lumen of the stomach. The unusual morphology of the parietal cell has fascinated researchers for over 100 years. When parietal cells are stimulated by the appropriate secretagogues, they also undergo spectacular morphological transformations that are correlated with the activation of the enzyme H^+,K^+-ATPase. HCl is generated by the H^+,K^+-ATPase, acting in concert with ion channels and other ion transporters. The H^+,K^+-ATPase is composed of a catalytically active α-subunit and a β-subunit that targets the enzyme to the appropriate membrane compartment. The three major secretagogues, histamine, acetylcholine, and gastrin, stimulate HCl secretion by activating different signaling pathways that interact at the level of the parietal cell. Because of the common occurrence of peptic ulcer disease, the acid secretory function of the parietal cell has been studied extensively. However, the parietal cell has other functions. In human, monkey, cat, dog, sheep, rabbit, and guinea pig, parietal cells secrete intrinsic factor, which is essential for the efficient absorption of vitamin B12 or cyanocobalamin in the distal ileum. Parietal cells may also secrete prostaglandins as well as growth factors including transforming growth factor-α, amphiregulin, and heparin-binding epidermal growth factor. Prostaglandins exert a cytoprotective effect on the gastric mucosa and the local release of growth factors appears to play an important role in regulating the growth and differentiation of epithelial cells within the gland units.

LOCATION AND STRUCTURE OF THE PARIETAL CELL

Parietal cells are present in glands within the fundus and body of the stomach and are the largest cells in these

Copyright 2004, Elsevier (USA). All rights reserved.

glands. They originate from immature progenitor cells in the gland isthmus and then migrate upward toward the pit region and downward toward the base of the gland. Parietal cells near the base of glands are smaller and frequently have dense accumulations of F-actin within their intracellular canaliculi. In contrast, parietal cells in the upper region of the gland have a more rounded appearance and contain more well-defined intracellular canaliculi. The designation "parietal" arose from the location of this cell within the gastric gland (bulging out along the wall; parietal being defined as relating to the walls of any hollow part of a plant or animal). Parietal cells are also referred to as oxyntic cells, based on the Greek word *oxyntos* (to generate an acidic substance). The "typical" parietal cell is usually depicted in a triangular shape with the apical region of the cell forming the apex of the triangle, which borders the lumen of the gastric gland. In stimulated parietal cells, the canaliculi are stylistically represented as simple bifurcations of microvillar membrane that extend partway into the cell. In reality, however, the intracellular canaliculi are complex, interconnected tubular systems that extend from the lumenally facing side of the cell to the basolateral membrane region (Fig. 1). The extensive array of intracellular membranes immediately beneath the intracellular canaliculi has historically been referred to as "tubulovesicles." This term was a compromise, because it was not possible to determine whether the membrane-rich structures detected with standard electron microscopy (EM) techniques were elongated tubules or round vesicles. Recent work, based on rapid freeze fixation and scanning EM as well as 3D reconstructions of transmission EM-based serial thin sections, suggests that the tubulovesicular system contains tightly packed cisternae that bear some resemblance to classical Golgi stacks.

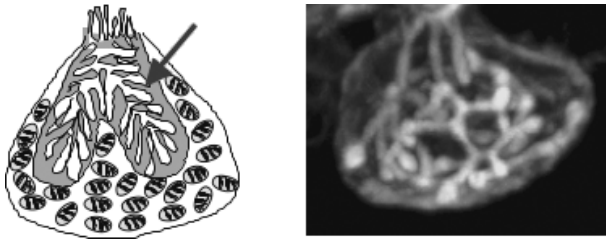


FIGURE 1 The complexity of the intracellular canaliculus is not conveyed in simple drawings. (Left) Typical depiction of an actively secreting parietal cell with a simple canaliculus containing elongated microvilli (arrow). (Right) 3D reconstruction of confocal microscopic images of a parietal cell within a gastric gland. The gland was fixed and stained with fluorescently tagged phalloidin, which specifically labels F-actin. Reprinted from Chew, C. S., Parente, J. A., Jr., Chen, X., and Chaponnier, C. (2000). The LIM and SH3 domain-containing protein, lasp-1, may link the cAMP signaling pathway with dynamic membrane restructuring activities in ion transporting epithelia. *J. Cell Sci.* **113**, 2035–2045, with permission.

MORPHOLOGICAL CHANGES AND ION TRANSPORT ACTIVITIES ASSOCIATED WITH HYDROCHLORIC ACID SECRETION

In addition to complex internal membrane structures, parietal cells are unusually rich in F-actin and possess a large number of energy-generating mitochondria. When parietal cells are stimulated to secrete hydrochloric acid (HCl), the canalicular membrane compartment expands at the expense of the intracellular tubulocisternal membrane compartment and this expansion is correlated with the appearance of numerous elongated microvilli. The mechanism driving these changes is controversial. One view is that the two compartments are continuous. The other, which forms the basis of the membrane recruitment hypothesis originally proposed by John Forte and colleagues in 1977, is that the canalicular and tubulocisternal membranes are distinct intracellular compartments. The membrane recruitment hypothesis proposes that the activation of the acid secretory response induces an exocytotic-like insertion of tubulocisternal membrane containing the H^+,K^+-ATPase into the canalicular membrane. Removal of the stimulus results in an endocytotic retrieval of membrane plus the H^+,K^+-ATPase back into the tubulocisternal compartment. Currently, this latter hypothesis is favored and is supported by the localization of a number of proteins associated with vesicle recycling within the parietal cell, as well as by observations that the H^+,K^+-ATPase translocates to a biochemically distinct compartment following stimulation.

Parietal cell proteins associated with vesicular trafficking include SNAP-25 (25 kDa synaptosomal-associated protein), syntaxin 3, and VAMP-2 (vesicle-associated membrane 2), which form a ternary complex that is implicated in the docking and fusion of vesicles. Proteins associated with endocytosis that have been localized in the parietal cell include clathrin, clathrin adapters, and dynamin-2. Also present are SNAREs, which are also implicated in vesicle fusion; SCAMPs, which are transmembrane proteins found in vesicles involved in membrane recycling; myosin 5b; and several rab proteins including rab11a, rab11b, and rab25. Evidence from other cellular systems suggests a role for rab proteins in regulating vesicular trafficking

between membrane compartments. The expression of a dominant negative form of rab11a in gastric parietal cells also inhibits the recruitment of the H^+,K^+-ATPase to the canalicular membrane.

Although the relationship between changes in F-actin polymerization and microvillar elongation has not yet been established, changes in the plasticity of the actin cytoskeleton are implicated in this process. At least two different actin-binding proteins, ezrin and lasp-1, are highly expressed in the parietal cell and are regulated by the cyclic AMP (cAMP) signaling pathway. Parietal cells also contain two different pools of actin: γ-actin, which is present at the cortical cell membrane, and β-actin, which is present mainly within the canalicular region.

The precise biochemical steps associated with the generation of HCl are unknown. Protons are thought to be derived from H_2O by the reaction, $HOH \rightarrow H^+ + OH^-$. Once generated, protons are actively extruded into the apically directed intracellular canaliculus by the H^+,K^+-ATPase in a neutral exchange for K^+. The activity of the H^+,K^+-ATPase is balanced by (1) the movement of Cl^- through a Cl^- channel and (2) the reuptake, or recycling, of K^+ through a K^+ channel. Although there are currently several candidates, the identity of these channels remains to be determined. The hydroxyl (OH^-) ions produced during the generation of protons combine with CO_2 in a reaction catalyzed by carbonic anhydrase to form bicarbonate (HCO_3^-), which is extruded from the cell by the action of Cl^-/HCO_3^- exchangers (AE2 isoforms) on the basolateral membrane. When this bicarbonate enters the general circulation, the pH of the venous blood exiting the stomach rises above that of the arterial blood entering the stomach, a phenomenon that has been labeled the "alkaline tide."

In addition to Cl^-/HCO_3^- exchangers, a $Na^+/K^+/2Cl^-$ cotransporter (NKCCl) resides on the basolateral membrane. Recent evidence suggests that there is a reciprocal relationship between the expression of the $Na^+/K^+/2Cl^-$ cotransporter and AE2. In rats, parietal cells above the neck of the glands express AE2 at high levels, whereas NKCCl expression is not detectable. This distribution is reversed in the neck and the base of the glands. The differences in the distribution of these transporters suggest that there are at least two different populations of parietal cells with different Cl^- entry mechanisms.

Na^+/H^+ exchangers (NHE2 isoforms) are also present on the basolateral membrane of the parietal cell. Although not directly coupled with AE2, these exchangers may act in concert with this Cl^-/HCO_3^- exchanger in a pH-dependent manner, thereby indirectly mediating the transport of NaCl into the cell. The activities of NHE2 and AE2, coupled with the basolateral Na^+,K^+-ATPase, or sodium pump, and a basolateral K^+ conductance, may facilitate the uptake of Cl^- into the cell against its electrochemical gradient. Based on targeted gene disruption studies, however, it appears that neither NHE2 nor NKCC1 is required for the acute acid secretory response.

SIGNALING PATHWAYS INVOLVED IN THE REGULATION OF HYDROCHLORIC ACID SECRETION

Parietal cells possess receptors for histamine (H2-receptor subtype), acetylcholine (muscarinic M3 receptor subtype), and gastrin [cholecystokinin B (CCK-B) or CCK_2 receptor subtype] (Fig. 2). When histamine binds to H2 receptors, the enzyme adenylyl cyclase is activated through a stimulatory heterotrimeric G-protein, G_s, catalyzing the generation of cAMP from ATP. Once elevated, cAMP activates a cAMP-dependent protein kinase(s) that phosphorylates two F-actin-associated proteins, ezrin and lasp-1. The functions of these proteins in the activation of the acid secretory response have not yet been defined; however, their localization within F-actin-rich compartments supports a role in the regulation of changes in the actin cytoskeleton that accompany active secretion. The response to acetylcholine and gastrin involves the elevation of intracellular calcium concentrations ($[Ca^{2+}]_i$), presumably by coupling of the receptors with another heterotrimeric G-protein, G_q, which activates phospholipase Cβ. Once activated, this phospholipase catalyzes the breakdown of phosphatidylinositol 4,5-bisphosphate to form diacylglycerol (DAG), which activates protein kinase C, and inositol 1,4,5-trisphosphate (IP_3). IP_3 activates an intracellular calcium channel, the IP_3 receptor, allowing for calcium release from intracellular stores. Intracellular calcium release is coordinated with the influx of extracellular calcium through unidentified storage-operated calcium channels in the plasma membrane (Fig. 2). In isolated parietal cells, histamine weakly elevates $[Ca^{2+}]_i$ in a subpopulation of parietal cells, but this response is not coupled to the activation of HCl secretion. In contrast, the presence of extracellular calcium has been shown to be essential for cholinergic activation of secretion.

Although the intracellular signaling pathways that are activated by acetylcholine and gastrin are similar with respect to calcium, only acetylcholine elicits a significant acid secretory response in the absence of a histamine background. The reason for the divergence in

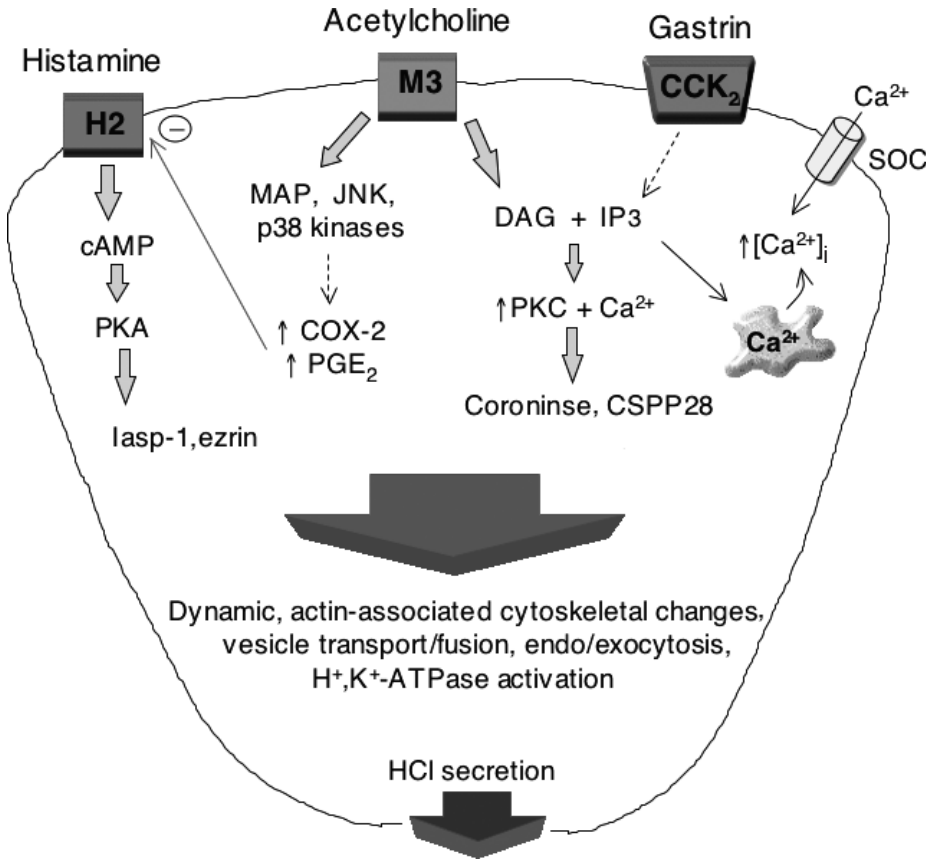


FIGURE 2 Overview of the intracellular signaling pathways involved in the activation of parietal cell HCl secretion. When histamine binds to H2-type receptors, the enzyme adenylyl cyclase is activated and catalyzes the breakdown of ATP to form cAMP, leading to the activation of a cAMP-dependent protein kinase(s). These kinases phosphorylate lasp-1 and ezrin, which are F-actin-associated proteins. Acetylcholine and gastrin activate a calcium-dependent signaling pathway(s) by binding to muscarinic M3 and CCK2 receptors, respectively. Gastrin is a weak agonist at the level of the parietal cell and little is known about its actions beyond elevation of $[Ca^{2+}]_i$. It is also unknown how elevated cAMP levels potentiate the responses to acetylcholine and gastrin. Acetylcholine and gastrin induce the formation of IP_3 and DAG, presumably by activating phospholipase Cβ. IP_3 induces the release of $[Ca^{2+}]_i$ from intracellular stores. Calcium also enters the cell through an unidentified storage-operated calcium channel(s) (SOC) in the plasma membrane. Protein kinase C (PKC) and calcium-dependent protein kinase(s) are subsequently activated, leading to increases in phosphorylation of coroninse and CSPP28. Acetylcholine also activates ERK/JNK/p38 kinases and may increase the production of PGE2. This latter response could serve as a delayed autocrine negative feedback loop to regulate histamine-stimulated HCl secretion. The link between changes in protein phosphorylation and agonist-induced changes in the actin-based cytoskeletal has not yet been defined. However, recent evidence suggests that protein phosphorylation may be involved in regulating dynamic cytoskeletal changes and vesicle transport events involved in the insertion and retrieval of the H^+,K^+-ATPase from the canalicular membrane. Dashed lines indicate less established pathways.

the secretory response to acetylcholine and gastrin is unclear, but may be associated with differences in their signaling response patterns. Acetylcholine induces a greater rise in $[Ca^{2+}]_i$ than does gastrin. In addition, acetylcholine elicits a uniform response from parietal cells in isolation, whereas gastrin stimulates only ~30% of the same cells. Moreover, although both agonists appear to activate protein kinase C, they do not appear to increase the phosphorylation of the same downstream proteins. Acetylcholine also activates the mitogen-activated protein kinases [extracellular signal-related kinase variants 1 and 2 (ERKs 1 and 2)], JUN

N-terminal kinase (JNK), and possibly the p38 kinase signaling pathway. Recent evidence suggests that the activation of these latter pathways increases cyclooxygenase 2 (COX-2) gene expression and prostaglandin E2 (PGE2) production by the parietal cell. Increased PGE2 production may serve as a delayed negative feedback control pathway to regulate histamine-stimulated HCl secretion (see below).

As with histamine, the intracellular signaling events that that are modulated by cholinergic stimulation have just begun to be defined. Thus far, two proteins have been found to undergo increased phosphoryation in parietal cells stimulated with acetylcholine: coroninse and CSPP28. Coroninse is an actin-binding protein that has been implicated in the regulation of endocytosis. CSPP28 is phosphorylated by a calcium-dependent mechanism and is enriched in light membrane fractions. Thus, like coroninse, CSPP28 may be involved in controlling vesicle movement and/or fusion.

In most, but probably not all species, histamine and cAMP are the most powerful stimulators of the acid secretory response at the level of the parietal cell. There is an acute potentiating interaction between the cAMP and calcium signaling pathways such that a low dose of acetylcholine administered along with a submaximal dose of histamine induces a greater than additive response. There is a similar potentiating interaction between histamine and gastrin. The mechanisms responsible for these potentiating interactions are not yet established.

Several factors have been found to suppress the acute acid secretory response to histamine in isolated parietal cells, including prostaglandins of the E series, somatostatin, epidermal growth factor (EGF), and transforming growth factor-α (TGF-α). All of these factors appear to couple with an inhibitory heterotrimeric G-protein, G_i. Longer exposure of parietal cells in primary culture to EGF/TGFα enhances the acid secretory response. This long-term response may be initiated by activation of the serine/threonine protein kinase, Akt, and the induction of H^+,K^+-ATPase gene expression.

See Also the Following Articles

Exocytosis • Gastric Acid Secretion • Gastric H^+,K^+-ATPase • Gastrin • Histamine

Further Reading

Forte, J. G., and Yao, X. (1996). The membrane-recruitment and -recycling hypothesis of gastric HCl secretion. *Trends Cell Biol.* **6**, 45–48.

McDaniel, N., and Lytle, C. (1999). Parietal cells express high levels of Na-K-2Cl cotransporter on migrating into the gastric gland neck. *Am. J. Physiol.* **276**, G1273–G1278.

Okamoto, C. T., and Forte, J. G. (2001). Vesicular trafficking machinery, the actin cytoskeleton, and H^+,K^+-ATPase recycling in the gastric parietal cell. *J. Physiol.* **532**, 287–296.

Schultheis, P. J., Clarke, L. L., Meneton, P., Harline, M., Boivin, G. P., Stemmermann, G., Duffy, J. J., Doetschman, T., Miller, M. L., and Shull, G. E. (1998). Targeted disruption of the murine Na^+/H^+ exchanger isoform 2 gene causes reduced viability of gastric parietal cells and loss of net acid secretion. *J. Clin. Invest.* **101**, 1243–1253.

Soll, A. H., and Berglindh, T. (1994). Receptors that regulate gastric acid secretory function. *In* "Physiology of the Gastrointestinal Tract" (L.R. Johnson, ed.), 3rd Ed., Vol. 2, 1139–1169, Raven Press, New York.

Pathologic and Paralytic Ileus

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

ileus Absence of propulsive motility in the gastrointestinal tract.
laparotomy Surgical incision through the abdominal wall.
physiologic ileus Normal absence of gastrointestinal contractile activity.

The term "ileus" is derived from the Greek word *eileos*, which is translated as "to block or to twist." Ileus is generally defined as absence of propulsive motility in the small or large intestine. Absence of propulsive motility implies stasis of the luminal contents. The presence of ileus can reflect normal intestinal physiology or it can be pathologic.

INTRODUCTION

Physiologic ileus refers to the absence of motility that normally occurs for short periods of time during specialized motility patterns, such as the inactivity found during the interdigestive state in the small intestine. Pathologic ileus refers to the prolonged absence of propulsive motility that is associated with clinical symptoms of an intestinal obstruction. Various names have been applied to this form of obstruction. It may be known as adynamic ileus, inhibitory ileus, or paralytic ileus. The latter term is not strictly accurate in a pathophysiologic sense because the intestinal musculature is not paralyzed in the same sense as in skeletal muscle paralysis.

CLINICAL SYMPTOMS

Pathologic ileus can result from a variety of diverse conditions that include disease changes in the intestine and disease activity in the peritoneal cavity. Pathologic ileus may also be induced as a neurally mediated reflex response to conditions outside the peritoneum and may be associated with conditions such as general infections, spinal cord injury, and irritation of the kidneys. This form of obstructive intestinal inactivity is most commonly seen during and for prolonged periods following laparotomy. Exposure and handing of the bowel is a strong stimulus for the production of ileus. Nevertheless, the most clinically important cause of pathologic ileus is peritonitis.

The entire intestine in pathologic ileus becomes dilated with large collections of both fluid and gas. The gaseous composition is mostly air. Unlike mechanical obstruction whereby the intestine above the point of occlusion becomes distended gradually, dilatation of the entire bowel occurs quickly, with thinning of the wall as the radius increases. This underlies the extreme degree of abdominal distension that is one of the primary signs of the condition.

The clinical features of pathologic ileus do not parallel those found in mechanical obstruction. Pain is usually not a complaint, or, if present, is a dull continuous ache rather than the intermittent sharp cramping pain that occurs in structural blockage. Nausea and vomiting can be prominent and fluids given by mouth are regurgitated. Defecation and the passage of gas usually stop completely with the onset of the ileus.

PATHOPHYSIOLOGY

Paralytic ileus is not entirely applicable as a synonym for pathologic ileus, because the intestinal musculature is not paralyzed. The ability of the smooth muscle to generate maximal force of contraction is unchanged in pathologic ileus. Neural influences actively inhibit the intestinal musculature and account for the prolonged absence of muscle contraction and motility seen in pathologic ileus. The way it starts and the prolonged nature of pathologic ileus are reminiscent of skeletal muscle reflex paralysis that is acutely associated with spinal cord injury and ensuing spinal shock. In spinal shock, the neural circuits of the spinal cord that normally mediate motor reflexes become nonfunctional for a transient period of time. In the case of pathologic ileus, the neural circuits of the enteric nervous system that normally initiate and organize propulsive motor behavior in the intestine become nonfunctional for a transient period. Although neural mechanisms are involved in both cases, the details of the neurophysiologic changes cannot be explained fully for either case.

 Copyright 2004, Elsevier (USA). All rights reserved.

Activation of sympathetic innervation is implicated in the initiation of pathologic ileus. Virtually all of the sympathetic nerve fibers, which enter the intestine outside of the sphincters, end in the ganglia of the enteric nervous system. Norepinephrine released from the sympathetic nerves acts to suppress the release of acetylcholine at the millions of nicotinic synapses that are part of the integrative neural networks of the enteric nervous system. This inactivates the neural circuits and thereby prevents them from activating propulsive motility. Subpopulations of enteric inhibitory motor neurons remain active in this circumstance and act to continuously suppress contraction of the inherently myogenic musculature.

Although sympathetic activation and the release and action of norepinephrine in the enteric nervous system are an attractive explanation for pathologic ileus, some observations do not support the hypothesis. For example, pretreatment with drugs that block the receptors for norepinephrine or that suppress its release from sympathetic nerves is ineffective for prevention of the ileus associated with laparotomy. Infiltration of the splanchnic innervation with a local anesthetic to block sympathetic nerve traffic to the intestine does not reverse the ileus seen during laparotomy in dogs. On the other hand, injection of a local anesthetic into the blood entering the intestine immediately evokes rigorous contractile activity. Two conclusions emerge from these observations. One is that activation of sympathetic reflexes may release substances other than norepinephrine that initially "lock" the integrated microcircuits of the intestinal enteric nervous system in a state of prolonged paralysis; this ensues for prolonged periods of time without further sympathetic input. The second conclusion is that ongoing activity of enteric inhibitory motor neurons accounts for the absence of contractile behavior. Blockade of the inhibitory neuronal activity by injection of a local anesthetic releases the muscle from the inhibition. Uncoordinated contractile activity appears after neural blockade due to the autogenic properties of the musculature.

See Also the Following Articles

Autonomic Innervation • Basic Electrical Rhythm • Colonic Obstruction • Disinhibitory Motor Disorder • Enteric Nervous System • Hirschsprung's Disease (Congenital Megacolon) • Intestinal Pseudoobstruction • Toxic Megacolon

Further Reading

Cullen, J., Caropreso, D., Hemann, L., Hinkhouse, M., Conklin, J., and Ephgrave, K. (1997). Pathophysiology of adynamic ileus. *Dig. Dis. Sci.* **42**, 731–737.

Summers, R., and Liu, C. (1995). Approach to the patient with ileus and obstruction. *In* "Textbook of Gastroenterology" (T. Yamada, D. Alpers, C. Owyang, D. Powell, and F. Silverstein, eds.), 2nd Ed., Vol. 1, pp. 796–798. J. B. Lippincott, Philadelphia.

Wood, J. (1999). "Neurogastroenterology and Motility." Milner-Fenwick, Timonium, Maryland.

Pepsin

JEAN-PIERRE RAUFMAN
University of Maryland School of Medicine, Baltimore

pepsin Aspartic proteinase (36 kDa), derived from pepsinogen under acidic conditions; assists the digestion of dietary protein.

pepsinogen Proenzyme (40 kDa) synthesized and stored in zymogen granules and secreted by chief cells of the stomach.

Pepsin, a protease present in the gastric lumen, is secreted by the chief cells of the gastric mucosa as an inactive precursor, pepsinogen; pepsinogen is activated by acid present in the gastric lumen, which initiates digestion of protein. Multiple genes code for immunologically distinct forms (isoenzymes) of pepsin, but all forms appear to have functional similarity. Peptic ulcers and acid-peptic disease derive their names from pepsin.

HISTORICAL OVERVIEW

In groundbreaking experiments using human gastric juice, the American physician/physiologist William Beaumont (1785–1853) hypothesized that, in addition to acid, an unknown factor contributed to protein digestion. The German physiologist Theodor Schwann (1810–1882) identified this factor from its ability to digest egg white and designated it "pepsin" (derived from the Greek word *pepsis*, for digestion). In England, using amphibian peptic cells as a model, the Cambridge physiologist John Langley (1852–1925) elucidated the formation of pepsinogen in chief cells, its storage in cytoplasmic zymogen granules, and, after secretion, its conversion to the active acid protease, pepsin. In 1930, John Northrop crystallized pepsin. Recent advances in structural biology have revealed how the three-dimensional structure of these molecules explains their activation and actions.

STRUCTURE AND ACTIVATION OF PEPSIN

Like other aspartic proteinases (EC 3.4.23.X), pepsin (approximate molecular mass, 36 kDa) is synthesized as a proenzyme, pepsinogen (approximate molecular mass, 40 kDa), which is stable at neutral and alkaline pH (>6) and is converted to active pepsin at acid pH by proteolytic cleavage of an N-terminal prosegment (inhibitory piece). Studies of the crystal structure of pepsinogen indicate that the inhibitory piece shields the substrate-binding portion of the active protease, with six basic amino acids in the prosegment forming electrostatic interactions with acidic amino acids in pepsin. Thus, at neutral pH, the inhibitory piece maintains the enzyme in its inactive form by sterically blocking access to the active site and neutralizing negative charges in pepsin, thereby stabilizing the conformation of the proenzyme.

Exposure to acid results in protonation of carboxylate groups and repulsion of the net positive charges that disrupt the electrostatic interactions, unblocking the active site and activating the enzyme. Moreover, pepsinogen is subject to the proteolytic action of activated pepsin (autocatalysis). By these mechanisms, exposure to pH <6 (as expected in the gastric lumen) activates a rapid (2 sec at pH 5–6; 5 msec as the pH approaches 2) cascade of pepsin activation. Returning ambient pH to neutrality can arrest or reverse these conformational changes. Increasing the pH to >7.2 (as expected in the normal small intestine) or the temperature to $>65°C$ irreversibly denatures pepsin, whereas pepsinogen is stable to pH 10 and 100°C.

The molecular structure of human pepsin (Fig. 1) is very similar to that of other members of the aspartic proteinase family. The central hydrophobic core of pepsin (catalytic aspartic acid residues at position 32 and 215) comprises the active site of the enzyme. This site can accommodate an approximately 8-amino-acid portion of protein substrate.

MEASUREMENT OF PEPTIC ACTIVITY

Measurement of peptic activity was first standardized by Anson and Mirsky in 1932; they defined 1 peptic unit as the activity of pepsin (pH 2, 37°C) that results in the release over 10 minutes of 0.1 μmol of tyrosine from 5 ml of 2% hemoglobin. To increase sensitivity and facilitate the analysis of multiple samples, investigators have modified this assay by using radiolabeled substrate (hemoglobin and albumin) and automated assays.

 Copyright 2004, Elsevier (USA). All rights reserved.

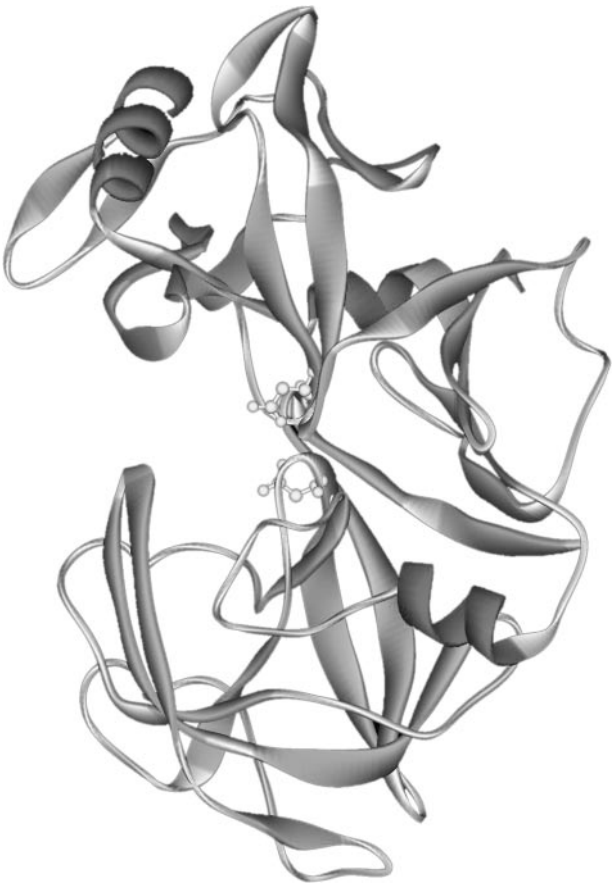


FIGURE 1 Schematic structure of human pepsin (EC 3.4.23.1), drawn according to coordinates deposited in the Protein Data Bank (identification number 1PSN). Light gray indicates β sheets; dark gray indicates α helices. The aspartic acid residues at positions 32 and 215 delineate the active site of the enzyme (in a ball-and-stick configuration).

Rather than using a defined absolute unit of peptic activity, many investigators express results as a percentage of peptic activity in experimental samples compared to control or, in secretory studies, as a percentage of total peptic activity in the sample.

ACTIONS IN NORMAL PHYSIOLOGY AND DISEASE

The major function of pepsin is to initiate digestion of ingested proteins. The greater activity of pepsin for meat proteins, such as collagen, indicates that the enzyme is less important for hydrolysis of vegetable proteins. In addition, in neonates, pepsin may assist in milk clotting. The pH optimum for peptic hydrolysis of proteins depends on the substrate (pH 1.5–2.5 for hemoglobin, pH 3 for albumin, and pH 5.5 for milk clotting).

Peptic activity releases free amino acids and small peptides that can be absorbed by enterocytes lining the small intestine. Relatively large peptides resulting from incomplete pepsin digestion also enter the duodenum and are further degraded by pancreatic proteinases. Some products of peptic digestion also serve a signaling function because they stimulate the release from the distal stomach (gastrin) and duodenum of hormones (cholecystokinin-releasing peptide and others) that play additional roles in regulating digestion.

Pepsin appears to play a crucial, if not necessary, role in ulceration of the stomach and duodenum. In the absence of pepsin, gastric acid does not cause ulceration. Hence, major benefits of antacid therapy in the treatment of ulcer disease may be inhibition of the conversion of pepsinogen to pepsin and the maintenance of a gastric luminal pH greater than the optimum for the enzyme. Changes in the level or distribution of pepsinogen in blood and urine that are observed in peptic ulcer disease and gastric cancer have not proved clinically useful.

See Also the Following Articles

Chief Cells • Duodenal Ulcer • Gastric Acid Secretion • Gastric Ulcer • Protein Digestion and Absorption of Amino Acids and Peptides

Further Reading

Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C., and James, M. N. (1995). Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**, 960–972.

Hershey, S. J. (1994). Gastric secretion of pepsins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1227–1238. Raven Press, New York.

Raufman, J.-P. (1996). Peptic activity and gastroduodenal mucosal damage. *Yale J. Biol. Med.* **69**, 85–90.

Sielecki, A. R., Fujinaga, M., Read, R. J., and James, M. N. (1991). Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* **219**, 671–692.

Percutaneous Drainage

DAVID B. LAUTZ AND RODNEY CHAN
Brigham and Women's Hospital and Harvard Medical School

abscess Collection of pus formed by tissue destruction in an inflamed area of a localized infection.
Seldinger technique Used for the insertion of a guidewire into a vessel.
tandem-trocar technique Used to cannulate into a vessel.

Even though there has never been a debate about whether abscesses should be drained; the available options for drainage have changed. Traditionally, an operation has been required, whether a thoracotomy for empyema, or a laparotomy for diverticular abscess. The modern technique of image-guided percutaneous drainage has now become the first choice for many clinicians and patients. Surgery is increasingly reserved only for difficult-to-reach locations, for cases in which percutaneous drainage fails, or for cases in which urgent surgical exploration is needed in view of impending sepsis.

INDICATIONS, APPROACHES, AND GUIDANCE

Patients harboring intraabdominal abscesses usually present with signs of systemic inflammation, including fever, leukocytosis, and pain. Postoperative fluid collections represent the most common cause of intraabdominal abscess and are often localized to the dependent pelvis. Other common causes of pelvic abscesses include appendicitis, diverticulitis, and pelvic inflammatory diseases. Although most abscesses are drained transabdominally when possible, the pelvis represents a challenge due to obscuring bowel loops from ileus and the abundance of vascular structures. Alternative approaches of transrectal and transvaginal drainage guided by ultrasound as well as transgluteal drainage guided by computed tomography (CT) are employed routinely.

Liver abscesses are common and they can be either single or multiple. Single pyogenic abscesses are frequently located in the right lobe, whereas multiple abscesses occupy both lobes. Drainage of single abscesses can often be guided by ultrasound alone, but multiple abscesses typically require multiple drains and are CT guided. Subdiaphragmatic abscesses represent a challenging location and require a combination of imaging procedures for guidance.

Pancreatic fluid collections occur following episodes of acute pancreatitis. They can present early as pancreatic necrosis or late as symptomatic pseudocysts. CT is the usual method of drainage in these cases. Aspiration without catheter placement is not appropriate as the chance of recurrence approaches 70%.

PROCEDURES

Before starting the procedure, it is important to ensure that drainage is clinically indicated, coagulopathy is corrected, and the imaging modality and the approach are chosen. Broad-spectrum antibiotics are usually given 1 hour preprocedure. The procedure is usually performed with a combination of intravenous conscious sedation (fentanyl, versed) and local anesthesia. A diagnostic scan is typically obtained with a radio-opaque skin marker.

Diagnostic aspiration with a 22-gauge needle will confirm the depth of the cavity as well as the path for catheter placement. The aspirate should be examined by gram stain and culture, although organisms may not be found if patients have been receiving antibiotics. White cells can be absent if patients are immunocompromised. In addition, the presence of creatinine suggests a urinoma, bilirubin suggests biloma, fat globules suggest lymphocele, and amylase suggests pseudocysts. Care is taken not to decompress the abscess cavity completely. The exchange of the aspiration needle for the catheter is done using the Seldinger technique or the tandem-trocar technique and a large catheter of 12- to 14-gauge lumen is left for drainage. There are many types of catheters. Pigtails are the most common types, and the drainage site is left in the dependent position if possible. After catheterization, the cavity is irrigated many times and, most importantly, the drain is secured externally. Finally, an additional scan is completed to ensure that there are no undrained collections.

Daily drain care and inspection are paramount. The usual recommendation for flushing is 5 ml, three or four times a day, to prevent clogging. When the catheter

 Copyright 2004, Elsevier (USA). All rights reserved.

drains less than 10 ml for a 24-hour period, it is usually reasonable to remove it, but it is important to confirm that the decrease in drainage is not secondary to blockage. Sudden decrease in drainage can be secondary to catheter obstruction. Imaging with either CT or a sinogram is important to demonstrate complete drainage.

CLINICAL TRIALS

There is a lack of randomized clinical trials comparing surgical drainage to percutaneous drainage of abscesses. In a retrospective review of 32 patients with Crohn's disease who underwent percutaneous drainage from 1985 to 1999, there was a 95% success rate, with 50% of patients avoiding surgery in the short term (< 60 days). There was not any significant increase in need for surgery long term among those patients treated by drainage. The abscess recurrence rate of 22% was comparable to surgical drainage.

Although most retrospective studies of percutaneous drainage have shown promising results, one review of 160 unselected patients who underwent drainage of pancreatic pseudocysts revealed lower success (88 vs. 42%), higher mortality (16 vs. 0%), more complications (64 vs. 27%), and longer hospital stay (45 vs. 18 days) in the percutaneous drainage group.

CATHETER COMPLICATIONS

In an older retrospective study on the use of percutaneous drainage in 118 patients from 1979 to 1984, there was a reported 4.2% major complication rate, which includes septic shock, hemorrhage, subphrenic abscess, and formation of an arteriovenous fistula, and a mortality rate of 2.5%. Other complications include the formation of an enteric fistula, which presents as a change in character and volume of the drainage.

CONCLUSION

There is still some debate among authorities on surgical approaches and percutaneous drainage of abscesses, but the majority of patients who develop intraabdominal abscesses are now treated noninvasively. This approach has radically changed the surgical management of many patients and has allowed them to avoid an open surgical approach, which may increase their chances for other postoperative complications. The increasing use of and demand for guided percutaneous drainage will revolutionize the treatment of many surgical disorders as well as offer clinicians and their patients another less invasive option for drainage.

See Also the Following Articles

Computed Tomography (CT) • Liver Abscess • Pancreatic Pseudocysts

Further Reading

Gervais, D. A., Hahn, P. F., O'Neill, M. J., and Mueller, P. R. (2002). Percutaneous abscess drainage in Crohn disease: Technical success and short- and long-term outcomes during 14 years. *Radiology* **222**(3), 645–651.

Hamlin, J. A, Friedman, M., Stein, M. G., and Bray, J. F. (1986). Percutaneous biliary drainage: Complications of 118 consecutive catheterizations. *Radiology* **158**(1), 199–202.

Harisinghani, M. G., Gervais, D. A., Hahn, P. F., Cho, C. H., Jhaveri, K., Varghese, J., *et al.* (2002). CT-guided transgluteal drainage of deep pelvic abscesses: Indications, technique, procedure-related complications, and clinical outcome. *Radiographics* **22**(6), 1353–1367.

Heider, R., Meyer, A. A., Galanko, J. A., and Behrns, K. E. (1999). Percutaneous drainage of pancreatic pseudocysts is associated with a higher failure rate than surgical treatment in unselected patients. *Ann. Surg.* **229**(6), 781–787; discussion, 787–789.

Men, S., Akhan, O., and Koroglu, M. (2002). Percutaneous drainage of abdominal abscess. *Eur. J. Radiol.* **43**(3), 204–218.

Percutaneous Endoscopic Gastrostomy (PEG)

JEREMY SPECTOR
Duke University Medical Center

enteral By way of the gastrointestinal tract.
fluoroscopy Examination of the tissues using an X-ray fluoroscope.
gastrostomy An opening into the stomach.
ligament of Treitz Suspensory muscle of the duodenum that anatomically divides the duodenum from the jejunum.
parenteral By means other than the gastrointestinal tract, such as via intravenous injection.
volvulus Twisting of the intestine, resulting in obstruction.

The placement of a percutaneous endoscopic gastrostomy tube is a widely used method of delivering enteral nutrition to those unable to eat. The process involves inserting a thin tube through the wall of the abdomen into the stomach; the tube allows nutrient augmentation, fluid administration, or medication delivery. In properly selected patients, the procedure is safe, well tolerated, and successful in over 90% of individuals.

INTRODUCTION

As the population ages and life-sustaining technologies are advanced, many circumstances arise in which patients are unable to eat or tolerate oral intake. In this setting, additional means exist to provide nutrition, fluids, and medications, either by enteral or parenteral methods. When possible, enteral feeding is the preferred route to provide nutrition to patients with a functioning gastrointestinal (GI) system. By maintaining even a minimal exposure of nutrients to the enteral GI tract, there are significant benefits for the patient, including preserving the function of the enteral immune system, preventing mucosal atrophy, maintaining the normal gut flora, and reducing the incidence of sepsis. In addition, enteral nutrition is generally safer and less costly than parenteral nutrition. For short-term needs, nasoenteric feeding tubes are safe, effective, and easy to place and subsequently to remove. However, when oral intake will be compromised for an extended period of time, percutaneous endoscopic gastrostomy (PEG) tubes have become the delivery systems of choice. Introduced in the early 1980s, PEG tubes have several advantages over surgical gastrostomies. PEG tubes can be placed using only mild sedation and topical anesthesia and require less time compared to surgical gastrostomies, which are generally performed under general anesthesia. In addition, PEG tubes are less expensive and have fewer complications compared to surgical gastrostomies.

INDICATIONS AND CONTRAINDICATIONS

In the adult population, the most common conditions that lead to PEG placement are neurologic disorders such as strokes, amyotrophic lateral sclerosis, neurologic trauma, and dementia. Other illnesses include head and neck cancers, facial trauma, and the short bowel syndrome. PEG tubes are most often placed for the purpose of providing long-term enteral nutrition. Other goals may include decreasing the risk of aspiration, improving survival, decreasing the risk of infection, and healing pressure ulcers. At times, PEG tubes are also placed to allow decompression of the stomach, such as in the setting of malignant intestinal obstruction or gastric outlet obstruction, often referred to as a "venting" PEG. Less common indications include treatment of gastric volvulus via fixation of the stomach to the anterior abdominal wall, and to allow access to the stomach for transgastric surgical instrumentation.

Absolute contraindications to PEG placement are the inability to advance the gastroscope, a limited life expectancy, or the inability to bring the gastric wall and anterior abdominal wall in apposition (which is required for proper healing and tract formation). The inability to transilluminate the anterior abdominal wall is considered by many to also be a contraindication to placement. However, there are data to suggest that with definite localization, with external palpation, and in the absence of air aspiration prior to needle visualization (see below), the technique can be performed safely. There are relative contraindications as well, including compromised wound healing, morbid obesity, massive ascites, coagulopathy, gastric varices, history of subtotal gastrectomy, peritoneal dialysis, portal hypertension, large hiatal hernia, hepatomegaly, and neoplastic or infiltrative disease involving the gastric wall.

 Copyright 2004, Elsevier (USA). All rights reserved.

PEG INSERTION

The following description is a standard technique for placing PEG tubes. Prior to the procedure, intravenous antibiotics are typically administered to minimize infectious complications. An endoscope is passed through the mouth and esophagus and into the stomach. The stomach is insufflated with air to fully distend it. The room lights are lowered and through the use of external palpation of the anterior abdominal wall, combined with internal transillumination through the gastric wall, the site for PEG tube insertion is identified. This site is marked and then prepared with a cleansing solution and a sterile field is created. Local anesthetic is administered to this site. A small-caliber needle is inserted along the projected tube insertion tract for confirmation. The needle is slowly advanced in the direction of the stomach lumen while continuous negative pressure is being applied to the plunger. Endoscopic visualization of the needle tip as it pierces the gastric wall must coincide with aspiration of air back into the syringe. Aspiration of air bubbles prior to endoscopic visualization of the needle suggests the penetration of an interposed hollow viscus (e.g., colon) and should prompt immediate removal of the needle. Once the direct tract of the PEG tube has been confirmed in this manner, a skin incision approximately 1 cm in length is made to allow passage of a trochar. The trochar is passed through the skin incision into the stomach and a flexible guidewire is advanced through the trochar. A snare is then passed through the endoscope into the stomach. The snare is opened and manipulated to secure the end of the guidewire. The scope, snare, and guidewire are then withdrawn through the mouth as a single unit.

At this point in the procedure either a "pull" or "push" technique can be performed. Both methods have similar success rates, and choice is based on operator preference and experience. Using the pull technique, a loop on the most proximal end of PEG tube is tied to the end of the guidewire that was withdrawn through the mouth. The tube is generously lubricated and the guidewire with the attached PEG is "pulled" back through the mouth into the stomach from the site of the skin incision. The tapered tube passes through the stomach wall and emerges through the premade skin incision. Traction is applied until gentle resistance, caused by an internal bumper attached to the distal end of the tube, is felt. Once mild resistance is met, the PEG tube is secured in place with an external bumper. Tube markings (in centimeters) indicate the distance between the internal and external bumpers. This distance, which is typically between 3 and 6 cm, is recorded. Usually, the endoscope is reinserted to visualize the internal bumper, in order to verify that there is the appropriate amount of traction on the tube, although this step is not required. A topical antibiotic ointment is applied to the skin site and the area is dressed and bandaged. The superfluous external portion of the PEG tube is trimmed to a workable length. An abdominal binder is applied to help keep the tube in place and minimize the risk of inadvertent tube removal while the tract matures. The push technique differs in that the tube has a central lumen, through which the guidewire is passed. The tube is "pushed" over the guidewire and advanced until it emerges through the skin surface. It is then secured in a manner similar to that used in the pull technique. A less commonly performed alternative to the push and pull techniques is the direct placement method, which uses a trochar with a peel-away surface that delivers the PEG tube and internal bumper into the stomach lumen, obviating the need to use a guidewire.

PERCUTANEOUS ENDOSCOPIC JEJUNOSTOMY

Clinical scenarios occur when enteral nutrition delivered directly to the small bowel is indicated, and a PEG tube will not suffice. Direct percutaneous endoscopic jejunostomy (PEJ) tubes are an available option for enteral access in these situations. Indications for PEJ tubes include patients who have undergone gastric resection or have gastric outlet obstruction, gastric dysmotility, or a nonfunctioning gastrojejunostomy, or patients at significant risk for aspiration events. PEJ tubes are placed in a manner similar to that of PEG tubes. A push enteroscope, which is longer than the standard upper endoscope, is used and passed into the jejunum beyond the ligament of Treitz. As with PEG insertion, transillumination of a loop of small bowel must be achieved followed by the insertion of a feeding tube into the jejunum, using the technique described for PEG placement. The tubes used for PEJs are those found in standard PEG kits. Fluoroscopy or a combined approach with a gastroenterologist and radiologist is often used to facilitate the placement of PEJ tubes. Although the procedure is more technically demanding, the success rate is comparable to that of PEG placements. Direct PEJ tubes should be differentiated from jejunal tubes that are placed through gastrostomy or PEG tubes (the so-called JET-PEG). These tubes are often used for indications similar to those for direct PEJ tubes, but it has been shown that the jejunal portions of these tubes tend to migrate back into the stomach, essentially rendering them a PEG tube.

COMPLICATIONS

PEG placement is generally a safe and well-tolerated procedure. The procedure has an associated mortality rate of approximately 1%, with major complications occurring in 2.5–3% of cases. Complications can be divided into major and minor categories. Major complications include perforation, peritonitis, aspiration, gastrocolocutaneous fistula, necrotizing fasciitis, and premature device dislodgment. Minor complications include peristomal wound infection, which is the most common complication. Antibiotic prophylaxis with 1 g of cephazolin 30 minutes prior to the procedure is recommended to reduce the incidence of this complication. Additional minor complications include tube malfunction or clogging, leakage around the tube site, and migration of the tube. Less common complications also include implantation or seeding of tumor cells at the PEG tube site, aortogastric fistula, and gastric volvulus.

ETHICAL CONSIDERATIONS

It is unfortunate that most patients in need of a PEG tube are severely debilitated and often near the end of life. Because of this, PEG use has often been the subject of charged ethical debate. In a difficult and emotional setting such as this, it is essential that the goals of the intervention are clear, and the risks and benefits thoroughly considered. Although there are many currently accepted indications for PEG placement, very few data are available to directly support its use, despite being a commonly performed procedure for over 20 years. A keen example is the use of PEG tubes to reduce the risk of aspiration pneumonia. Although direct instillation of nutrients into the stomach obviates problems with swallowing food, to date, there have been no trials that conclude either PEG or PEJ tubes reduce the risk of aspiration. Enteral feeding tubes will not eliminate aspiration of oropharyngeal secretions, a common and recurrent cause of aspiration. Further, it has been suggested that PEG tubes can actually increase the amount of gastroesophageal reflux, a risk factor for aspiration, by altering the gastroesophageal angle. Similar to the case for aspiration, there are few data to support PEG use to decrease mortality, improve wound healing, or improve markers of nutrition. These areas deserve further study.

When the benefit of a procedure is in question, the risks of the procedure must be carefully examined. PEG, although generally accepted as a safe procedure, does carry risks of morbidity and even mortality, as previously described. In several large studies, the 1-year mortality rate after PEG placement was greater than 50%. In this severely ill population, alternative noninvasive approaches should always be presented. Postprocedural ethical issues also need to be considered. These include the fact that often physical and chemical restraints need to be used in order to prevent confused or agitated patients from pulling out the tube, with the attendant complications. Additionally, when PEG feedings are to be used on a trial basis, the end points for discontinuing PEG use, such as duration and improvement in clinical parameters, should be clearly set prior to placement. All patients or their surrogates should be fully informed of the potential benefits, risks, and alternatives to PEG placement.

CONCLUSION

Since its description as an alternative to open gastrostomy, PEG tube placement has become a common endoscopic procedure. A pragmatic understanding of the realized benefits of PEG placement and enteral feeding, as well as the risks, by both the health care team and patient, is an essential first step in the process. In the future, technological advances, combined with responsible patient selection, will optimize the use of the PEG procedure for optimizing delivery of enteral feeding in various clinical settings.

See Also the Following Articles

Enteral Nutrition • Gastric Outlet Obstruction • Gastric Surgery • Gastrostomy • Parenteral Nutrition • Volvulus

Further Reading

Feldman, M. (1998). "Sleisenger & Fordtran's Gastrointestinal and Liver Disease." W. B. Saunders, Philadelphia, PA.

Finucane, T., Christmas, C., and Travis, K. (1999). Tube feeding in patients with advanced dementia: A review of the evidence (1999). *JAMA* **282**(14), 1365–1370.

Ginsberg, G., Bosco, J., Barkun, A., *et al.* (2002) Endoscopic enteral nutritional access devices. *Gastrointest. Endosc.* **56**, 796–802.

Safadi, B., Marks, J., and Ponsky, J. (1998). Percutaneous endoscopic gastrostomy. *Gastrointest. Endosc. Clin. North Am.* **8**, 551–563.

Schapiro, G., and Edmundowicz, S. (1996). Complications of percutaneous endoscopic gastrostomy. *Gastrointest. Endosc. Clin. North Am.* **6**, 409–422.

Percutaneous Transhepatic Cholangiography (PTC)

PAUL V. SUHOCKI
Duke University Medical Center

choledochocele Cystic biliary dilatation within the wall of the duodenum.

choledochoenterostomy Anastomosis created between bile duct and intestine. The most commonly performed anastomosis is between bile duct and jejunostomy (Roux-en-Y) because it eliminates biliary reflux of food material.

endoscopic retrograde cholangiopancreatography Visualization of the bile ducts and pancreatic duct, achieved by injecting iodinated contrast medium through an endoscope into the biliary and pancreatic ducts in a retrograde manner.

magnetic resonance cholangiopancreatography Visualization of the bile ducts and pancreatic duct, achieved during magnetic resonance imaging; does not require contrast medium to be injected into the ducts. The ducts are visible because of differences between the magnetic properties of the biliary/pancreatic fluid and surrounding tissues.

percutaneous transhepatic cholangiography Visualization of the intrahepatic and extrahepatic bile ducts, achieved by injecting iodinated contrast medium into the bile ducts. Injection is done by passing a needle and catheter through the abdominal wall, through the liver capsule, and into the bile ducts. Percutaneous access to the bile ducts also allows the operator to perform biliary tract interventions such as percutaneous biliary drainage.

portoenterostomy Surgical treatment (Kasai) procedure for extrahepatic biliary artesia. A Roux-en-Y limb is anastomosed to the dissected fibrous biliary plate on the undersurface of the liver.

Percutaneous transhepatic cholangiography (direct injection of iodinated contrast medium into the bile ducts under fluoroscopy) is a method of imaging the bile ducts with contrast medium. Prior to the emergence of endoscopic retrograde cholangiopancreatography (ERCP), which involves injection of contrast medium into the biliary tree endoscopically through the ampulla of Vater, percutaneous transhepatic cholangiography (PTC) was the sole means of imaging the bile ducts with contrast medium. Currently, because ERCP is less invasive than PTC for obtaining a cholangiogram, it is the method of choice in medical centers that are staffed by endoscopists who are adept at the procedure. Cholangiography obtained by injection of iodinated contrast medium into the bile ducts during ERCP or PTC is considered to be the gold standard for bile duct imaging.

INDICATIONS FOR PTC

Computed tomography (CT), ultrasound (US), and magnetic resonance cholangiopancreatography (MRCP) have emerged over the past two decades as noninvasive means of viewing the bile ducts. These modalities do not require injection of iodinated contrast into the bile ducts. However, current technical limitations make the bile duct detail obtained with these imaging methods inferior to that obtained with PTC and ERCP. This is particularly true for visualization of the subsegmental bile ducts.

PTC is indicated for the following reasons:

1. To determine the level of biliary obstruction and to attempt to categorize the lesion as benign or malignant.
2. For preoperative delineation of bile duct anatomy. In this case, it is important to demonstrate the amount of normal bile duct above an obstruction for planning a choledochoenterostomy and also for excluding the presence of variant biliary anatomy.
3. To determine if a biliary stricture is present in the patient with normal-caliber bile ducts on cross-sectional imaging and to assess clinical and laboratory evidence of biliary obstruction. This scenario is most often encountered in the postoperative patient (biliary bypass, partial liver resection, orthotopic liver transplant) who has undergone choledochoenterostomy.
4. To provide a percutaneous transhepatic tract for future percutaneous interventions, such as bile duct biopsy, stone removal, balloon dilatation of a biliary stricture, or placement of a permanent metallic stent across a nonoperable malignant biliary occlusion.

 Copyright 2004, Elsevier (USA). All rights reserved.

PREPROCEDURE WORKUP

The patient has usually had a CT scan, abdominal magnetic resonance imaging (MRI), MRCP, or liver ultrasound prior to PTC. When obstructive cholangitis is suspected, these studies will have supplied information about the presence or absence of bile duct dilatation, obstructing mass, gallstones, biliary strictures, collections of bile, and ascites. Bile collections related to bile leak are drained percutaneously under CT or ultrasound guidance before PTC. When the presence of biliary pathology is suggested on these studies, ERCP is usually performed next to provide information about the bile ducts with an iodinated contrast cholangiogram. ERCP is associated with less risk than PTC and obviates the need for PTC if technically successful. ERCP can be unsuccessful because of: postoperative anatomy (Roux-en-Y), difficulty passing the endoscope through the duodenum because of a duodenal stricture related to pancreatitis or tumor, inability to visualize or cannulate the ampulla of Vater because of duodenal stricture or tumor, and inability to achieve an adequate level of sedation.

When PTC becomes necessary, informed consent is obtained from the patient or guardian. General anesthesia is arranged for the pediatric patient or the adult patient in whom an adequate level of conscious sedation cannot be achieved. Antibiotics are administered intravenously. Coagulation and platelet count abnormalities are corrected to as near normal as possible. Vital signs are monitored throughout the case. The patient is given corticosteroids prior to the procedure if an allergy to iodinated contrast exists.

TECHNIQUE

A radiologist performs PTC in the interventional suite of the department of radiology. A dedicated interventional radiology nurse administers conscious sedation and monitors vital signs throughout the procedure. The patient is placed prone on a fluoroscopy table in a room outfitted with a multiangle image intensifier. The abdomen is prepped and draped in sterile fashion from the level of the nipples to the umbilicus. A 22-gauge Chiba needle is passed into the liver under fluoroscopic guidance. The stylet is withdrawn and contrast is injected through the needle while the needle is retracted slowly. The operator stops retracting the needle when contrast flows into a bile duct. Several milliliters of contrast are injected into the ductal system. A 0.018-inch-diameter guidewire is advanced through the needle into the duct. The needle is replaced with a #3 French catheter with multiple side holes. The potentially infected bile is aspirated and sent to the microbiology laboratory for culture. Contrast is then injected into the ducts to obtain a diagnostic cholangiogram. Radiographic images of the opacified bile ducts are exposed from anterior–posterior, in the right-anterior oblique and left-anterior oblique directions, in order that the total extent of each duct can be evaluated.

For percutaneous access to the right hepatic biliary tree, the needle is passed into the right hepatic lobe at the junction of the lower and middle thirds of the liver. The needle is directed toward the left shoulder. For percutaneous access to the left hepatic lobe, the needle is passed into the ventral duct of the left biliary tree under sonographic guidance. Entry to the left biliary tree is usually made through a window between the left costal margin and the xiphoid process.

Unusual means of accessing the bile ducts sometimes become necessary. The gallbladder can be injected with contrast to fill the biliary tree in a retrograde fashion via the cystic duct. This method can only be used if the patient has a gallbladder and a patent cystic duct with free communication of the cystic duct with the extrahepatic and intrahepatic ducts. In patients with massive ascites and dilated bile ducts, the ductal system can be accessed by passing a needle system through the right internal jugular vein, through the right atrium, into inferior vena cava, and into the middle hepatic vein. The needle is then passed through the wall of the middle hepatic vein, through liver parenchyma, and into the biliary tree. Iodinated contrast medium is then injected through the needle system.

After diagnostic cholangiography is performed, the decision whether to intervene percutaneously is made. Interventions include bile duct biopsy, biliary drainage, balloon dilation of a benign stricture (Fig. 1), and placing an endobiliary stent (Fig. 2). This decision is often made in conjunction with the surgical and medical services.

COMPLICATIONS

Complications occurring during PTC vary with the degree of technical difficulty experienced during the procedure. Patients with obesity, coagulopathy, comorbidities, sepsis, presence of ascites, and nondilated bile ducts are at increased risk. Complications include biliary sepsis, bile ascites, hemobilia, hemorrhage, and pneumothorax.

INTERPRETATION

Biliary Obstruction

Most cases of biliary obstruction are related to calculi or neoplasm. The most common tumors are

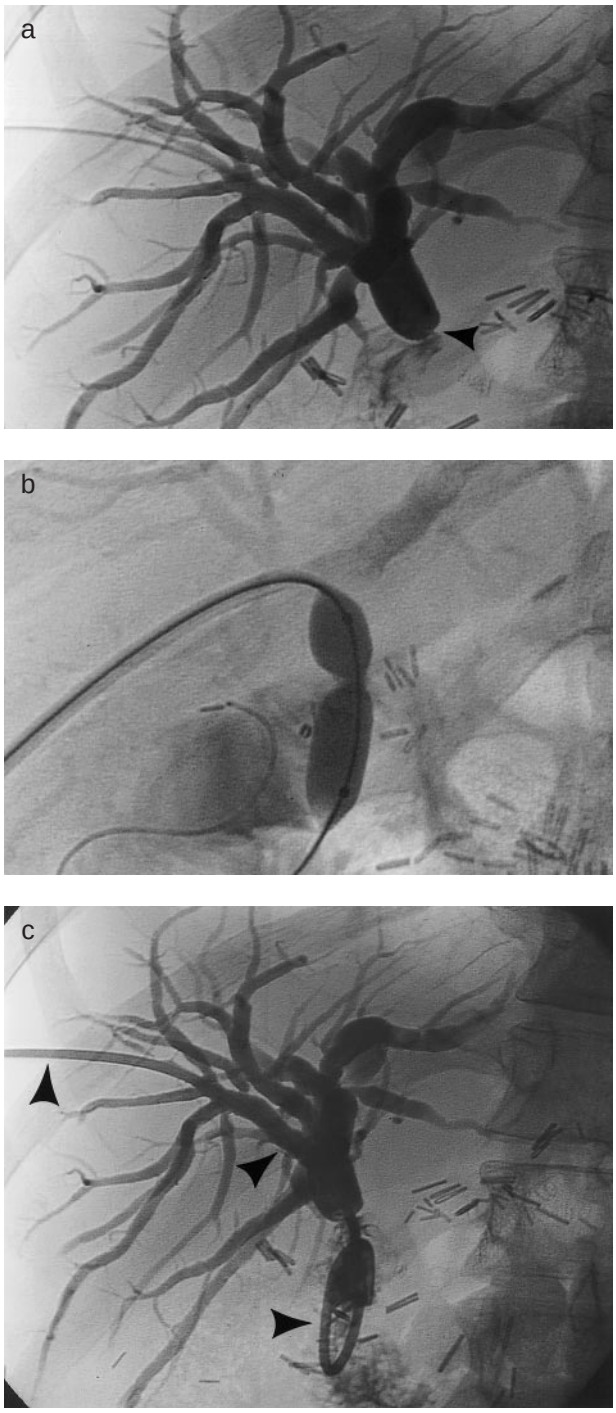


FIGURE 1 (a) Benign anastomotic stricture (arrowhead) developed 13 months following creation of an anastomosis between the common hepatic duct and the jejunum. (b) A waist can be seen while dilating the stricture percutaneously with a 10-mm-diameter balloon. (c) A biliary drainage catheter (#8 French; arrowheads) with multiple side holes is kept in place for 6 weeks postprocedure, allowing drainage of bile across the anastomosis.

pancreatic cancer (Fig. 3) and metastatic disease (Fig. 4), followed by cancer of the bile ducts (Fig. 5), ampulla, and gall bladder. Less common causes of biliary obstruction include congenital biliary obstruction, bile duct injury, pancreatitis, sclerosing cholangitis, biliary atresia, choledochal cyst, duodenal diverticulum, infection, and parasitic infestation.

Biliary atresia is the most common cause of obstructive jaundice in infancy. It can involve the intrahepatic or extrahepatic bile ducts, the latter being more commonly involved and found in 1/10,000–1/15,000 births. Hepatic biliary scintigraphy suggests the diagnosis if there is a delay of radiotracer excretion into the gastrointestinal tract. Treatment for extrahepatic biliary atresia is surgical. A portoenterostomy (Kasai procedure) is performed. Liver transplantation is considered for the patient who has a failed Kasai procedure or for whom the diagnosis of biliary atresia was made late.

Choledochal cysts are seen in 1/13,000 hospital admissions in the United States. Three-quarters of affected patients are female. Theories for the pathogenesis of choledochal cysts include (a) anomalous pancreatobiliary duct junction, (b) abnormal canalization of the bile ducts, and (c) abnormal autonomic innervation of the extrahepatic bile duct. Todani's classification of choledochal cyst disease is most widely used. Type I cysts are extrahepatic and are seen in 80–90% of all cases. These can be focal or fusiform. A type II cyst is a diverticulum of the extrahepatic duct and is seen in 3% of cases. Type III is a choledochocele, seen in 5% of cases. Type IV cysts account for 10% of cases and are subdivided into two groups; type IVA is characterized by dilatation of the intrahepatic and extrahepatic biliary tree, and type IVB cysts are multiple extrahepatic cysts. Type V cysts, known as Caroli's disease, are intrahepatic and rare. Complications of choledochal cysts include biliary obstruction, cholangitis, rupture, hepatic abscess, and cancer of the bile ducts or gallbladder. Complete excision of the cyst is recommended because of the associated risk of cancer. In patients with pancreatitis, long segment narrowing of the common bile duct (CBD) can be seen, with gentle tapering distally. The entire CBD may be involved and the ampullary segment may be normal.

Primary sclerosing cholangitis (PSC) is a progressive disease of the biliary tree characterized by inflammatory and fibrotic bile duct lesions. The etiology of the disease is unclear. Patients often present with recurrent episodes of fever, chills, weight loss, pruritus, jaundice, and mild right upper quadrant discomfort. Laboratory tests usually reveal an elevated white blood cell count and a cholestatic biochemical profile. Histologic

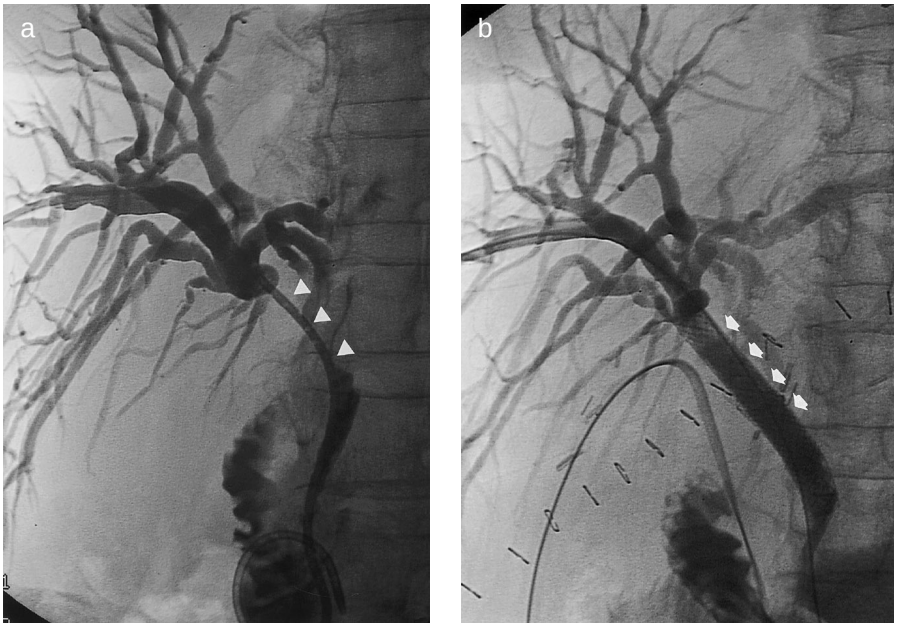


FIGURE 2 (a) A biliary drainage catheter (#8 French) with multiple side holes (arrowheads) traverses a long-segment malignant biliary occlusion. (b) A 10-mm-diameter stent (arrows) was positioned across the malignant occlusion to allow internal drainage of bile across the lesion. The percutaneous access was removed.

findings on liver biopsy are often nonspecific. The disease course is usually progressive, resulting in cholestatic liver disease. Cholangiography is a key component in making the diagnosis of PSC. Abnormalities of intrahepatic and extrahepatic bile ducts are seen in 75–80% of the patients. Isolated involvement of either the intrahepatic or extrahepatic ducts occurs in 10–20% of cases. Cholangiographic findings include focal narrowings, beading, pseudodiverticula, and pruning of intrahepatic duct branches. Long segment bile duct narrowing is not uncommon. However, brush cytology of dominant bile duct strictures should be performed because of the development of cholangiocarcinoma in 40% of the patients.

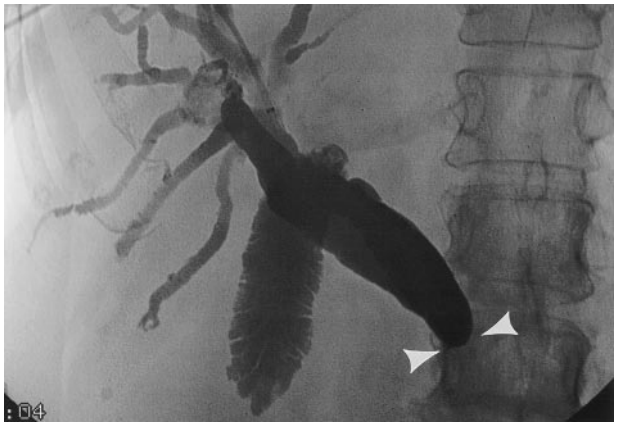


FIGURE 3 Abrupt termination (arrowheads) of the common bile duct is seen in a patient with pancreatic carcinoma.

Several disease processes can mimic PSC on cholangiography. These include metastatic disease to the liver, lymphoma, cholangiocarcinoma, cirrhosis, hepatic arterial chemotherapy, graft-versus-host disease, chronic allograft rejection, opportunistic infection of the bile ducts, multiple liver abscesses, systemic fungal disease, and ischemic bile duct strictures related to hepatic arterial compromise following hepatobiliary surgery or transplantation. Chronic biliary obstruction from gallstones or biliary strictures or chronic reflux of intestinal contents across biliary–enteric anastomoses can cause bacterial infection of the biliary tree, resulting in secondary cholangitis. These entities can be differentiated from PSC after careful history taking and review of hepatobiliary imaging and laboratory blood tests.

Common duct stones (Fig. 6) are seen in approximately 15% of patients with cystic duct stones. Retained stones in the biliary tree are present in 4% of patients who undergo cholecystectomy for gallbladder calculi. Bile duct calculi can also be seen in the setting of cholestasis related to intrahepatic or extrahepatic bile duct strictures (Fig. 7). Stones in the biliary tree are usually removed through an endoscopic approach. They are removed by a percutaneous transhepatic approach

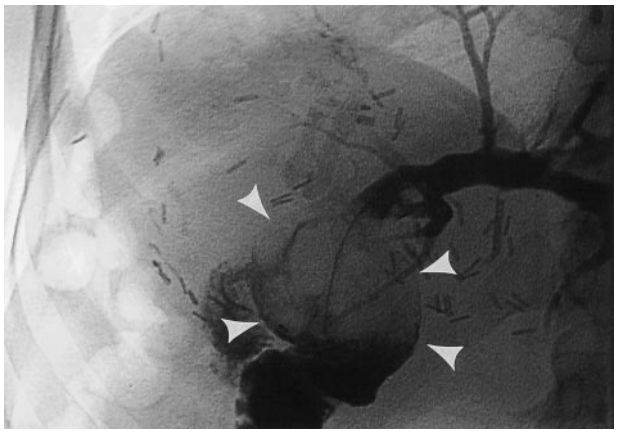


FIGURE 4 Colon carcinoma involving the portal lymph nodes creates a mass defect (arrowheads) on the common hepatic duct and the anastomosis of the common hepatic duct with the jejunum.

when the endoscopist is unable to access the biliary tree, usually in the setting of postoperative biliary anatomy. The diagnosis of biliary calculi is usually made by a noninvasive imaging method such as ultrasound or MRCP. Fifteen percent of gallstones are radio-opaque. On cholangiography, radio-opaque gallstones appear as well-defined filling defects. They are differentiated from tumors by their mobility within the ducts. Occasionally, an inflammatory reaction in the adjacent bile duct wall may cause them to be fixed, simulating a tumor. Contraction of the sphincter of Oddi can mimic the presence of a gallstone in the distal CBD. It can be difficult to differentiate gallstones from air bubbles and blood clots.

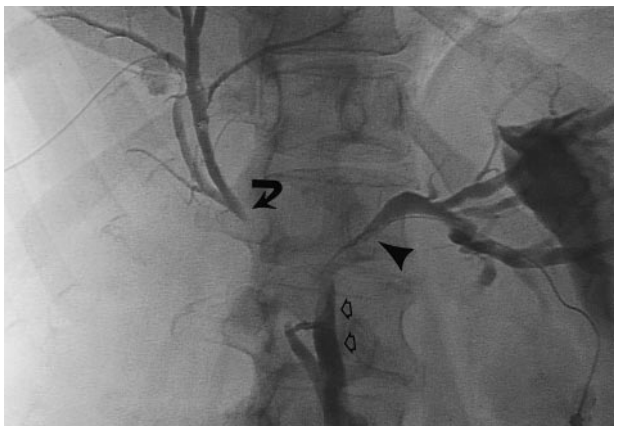


FIGURE 5 Cholangiocarcinoma (Klatskin tumor) involves the convergence of the right anterior sectoral duct (curved arrow), the left main hepatic duct (arrowhead), and the common hepatic duct (open arrows). The right posterior sectoral duct is completely occluded and does not fill with contrast.

Air bubbles coalesce, can be aspirated, and collect in anterior bile ducts. Blood clots usually clear after several days of biliary drainage.

Benign bile duct strictures are usually short. They may develop following hepatobiliary surgery, either as an iatrogenic injury to the duct or as anastomotic stricture following surgical creation of a choledochoenteric communication. When increased in length, benign biliary strictures are usually smooth and taper gradually. Cholangiographic features suggestive of cancer of the bile ducts, pancreas, or ampulla include abrupt termination of the bile duct, "apple core" concentric narrowing of the bile duct, and changes in caliber along the length of the stricture. Tumors can extend from the wall into the lumen of the bile duct in an exophytic fashion. Cross-sectional imaging often helps in making the diagnosis of malignancy. Brush biopsy of the biliary stricture or surgical exploration may be necessary to differentiate between benign and malignant lesions.

Bile Duct Leak

Bile duct injury can develop during cholecystectomy, partial liver resection, slippage of clip or ligature from the cystic duct following cholecystectomy, PTC, liver biopsy, and blunt or sharp abdominal trauma. Drainage catheters are usually placed into

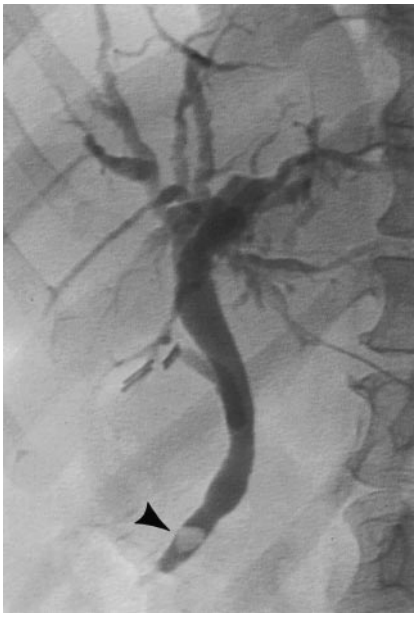


FIGURE 6 Oval filling defect (arrowhead) in the distal common bile duct represents a gallstone.

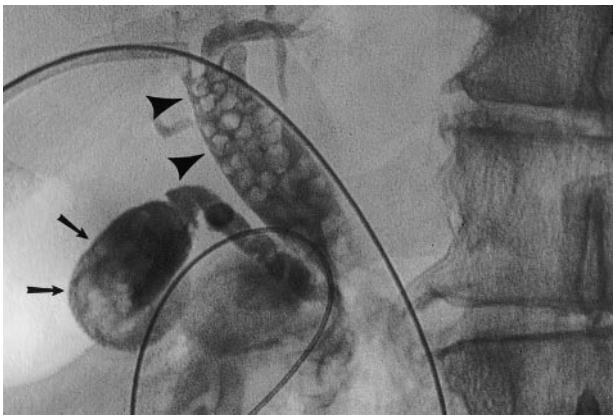


FIGURE 7 Multiple filling defects represent gallstones in the common hepatic duct (arrowheads) and gallbladder (arrows).

bile collections under CT or US. ERCP is usually successful at demonstrating a cystic duct leak or leakage from a partially or completely lacerated bile duct. PTC becomes necessary when ERCP is unrevealing or technically unsuccessful. PTC is difficult in this patient group because the decompressed bile ducts are difficult to access with a needle and subsequently to cannulate.

Aberrant extrahepatic bile ducts are at risk for injury during operation because the surgeon may not anticipate their presence. PTC may demonstrate an aberrant right hepatic duct that was torn intraoperatively and no longer communicates with the biliary tree. This torn aberrant bile duct may go unappreciated on ERCP if a vascular clip is placed on the stump of the transected bile duct that arises from the main biliary tree. In this case, the radiologist passes a catheter percutaneously into the aberrant bile duct, allowing the surgeon to identify this transected duct intraoperatively for anastomosis of the duct to jejunum.

See Also the Following Articles

Bile Duct Injuries and Fistulas • Cholangiocarcinoma • Computed Tomography (CT) • Gallbladder Cancer • Gallstones, Pathophysiology of • Magnetic Resonance Imaging (MRI) • Pancreatic Ductal Adenocarcinoma • Radiology, Interventional • Ultrasonography

Further Reading

Clavien, P. A., and Baillie, J. (eds.) (2001). "Diseases of the Gallbladder and Bile Ducts." Blackwell Science, United Kingdom.

Meyers, W. C., and Jones, R. S. (eds.) (1990). "Textbook of Liver and Biliary Surgery." J. B. Lippincott, Philadelphia.

Pitt, H. A., Carr-Locke, D. L., and Ferrucci, J. T. (eds.) (1995). "Hepatobiliary and Pancreatic Disease: The Team Approach to Management." Little, Brown and Company, Boston.

Shlansky-Goldberg, R., and Weintraub, J. (1997). Cholangiography. *Semin. Roentgenol.* **XXXII**, 150–160.

Perforation

DAVID B. LAUTZ AND GLEN W. BARRISFORD
Brigham and Women's Hospital, Boston

angularis Angular "notch" in the stomach that demarcates the division between the body and antrum of the stomach.
diathermy Local elevation of temperature within the tissues, produced by high-frequency current, ultrasonic waves, or microwave radiation.
diverticulum Pouch or sac opening from a tubular structure or saccular organ such as the gut or bladder.
perforation Abnormal opening in a hollow organ or viscus.

Perforation is defined as an abnormal opening in a hollow organ or viscus. It is derived from the Latin *perforatus*, meaning "to bore through." Perforation is one of the most common etiologies of surgically treatable abdominal pain.

INTRODUCTION

Abdominal pain is a leading cause of hospital admission and physician visits. The severity ranges from benign self-limiting etiologies to life-threatening surgical emergencies. Subtle differences in the presentation cause this to be a very complex diagnostic dilemma. A thorough systematic evaluation is required for significant abdominal pain. Consequently, abdominal pain presents a universal challenge that transcends a wide spectrum of medical and surgical specialties.

Most abdominal pathology presents with associated abdominal pain. A complete understanding of the pathophysiology coupled with this abdominal pain is essential in the recognition of disease. Visceral, parietal, and referred pain are the commonly acknowledged divisions. The onset, quality, duration, location, and associated symptoms classify the pain. Visceral pain is perceived via autonomic nerve fibers (sympathetic/parasympathetic), which symmetrically innervate abdominal organs. As a consequence of this symmetry, the sensory perception is along the midline. Visceral pain is characterized as intermittent, deep, dull, aching, crampy pain associated with nausea and diaphoresis. Stimuli on the visceral organs include stretching/traction, compression, torsion, and chemicals. Somatic (parietal) pain is perceived via spinal nerves ipsilaterally innervating the body wall. Greater localization is achieved for this severe, persistent, sharp, stabbing pain that results from local inflammation and irritation. Referred pain typically arises from visceral organs, travels via sympathetic visceral nerve fibers, and is perceived in a location remote from the site of pathology. This double innervation is a consequence of common embryonic origin. Common examples include back pain resulting from pancreatitis, right shoulder pain from hepatobiliary disease, and left shoulder pain secondary to splenic pathology.

The initial evaluation of the acute abdomen begins with a complete history and physical, and continues with inspection, auscultation, and palpation. Nausea, vomiting, fever, chills, constipation, diarrhea, and natural history/quality/distribution of symptoms are important historical facts. Distension on inspection and decreased bowel sounds on auscultation are consistent with acute abdominal pathology. On palpation, guarding/rigidity, diffuse or point tenderness, and rebound tenderness can be observed.

Abdominal pain can arise from an array of varied etiologies presenting in a similar fashion. Subtle differences in the history, presentation, evaluation, and diagnostic workup guide the clinician. Nonsurgical pathology can mimic surgical disease. The clinician must consider nonsurgical pathology (gastroenteritis, acute hepatitis/pancreatitis, sickle cell crisis, lead toxicity, acute porphyria, or pneumonia) when considering potential surgical emergencies. Surgical disease can be further subdivided into nonperforated (acute cholecystitis) and perforated (duodenal ulcer) pathology. The commonly encountered gastroenteral perforations are presented here.

PERFORATED ESOPHAGUS

Esophageal perforation is not typically considered a cause of abdominal pain. However, in order to complete the discussion of gastroenteral perforation, esophageal perforation is briefly considered. Esophageal perforation is a surgical emergency requiring rapid diagnosis. It most commonly occurs as a result of instrumentation, trauma, spontaneous perforation, and swallowing of a foreign body. It presents with pain in the cervical area,

 Copyright 2004, Elsevier (USA). All rights reserved.

dysphagia, and odynophagia. Symptoms are notably exacerbated with swallowing or movement. A recent history of instrumentation or vomiting is usually noted. Diagnostic workup includes posteroanterior and lateral chest films with findings of mediastinal emphysema/widening, cervical emphysema, or pneumothorax. Barium esophagram in the right lateral decubitis position is the gold standard for evaluation. Computed tomography and endoscopy have a limited role in the presence of an esophagram, but may be useful in complicated cases with equivocal studies. Treatment objectives include limitation of extravasation, prevention of infection, nutritional maintenance, and restoration of structural integrity. Nonoperative treatments offer a limited role in isolated cases. Operative exploration, wide drainage, and primary closure within 24 hours (if possible) remain the treatments of choice. Outcomes are variable and are contingent on duration to diagnosis and treatment, cause and location of injury, comorbidities, and existing esophageal disease.

PERFORATED GASTRIC ULCER

Gastric ulceration often presents with acute-onset epigastric pain that radiates to the back and is exacerbated by the ingestion of food. Common risk factors include use of nonsteroidal antiinflammatory drugs (NSAIDs), cigarette smoking, and chemotherapeutic agents (5-fluorouracil, cisplatin, doxorubicin, or mitomycin C). Gastric ulcer disease presents in approximately 100,000 new cases annually and is most commonly seen in men and the elderly. Complications of gastric ulceration result in several thousand deaths and a 10% rate of gastric malignancy annually. Normal or decreased levels of acid secretion are observed with a breakdown in the gastric mucosal barrier. Ulceration is most common along the lesser curvature of the stomach at the junction between the antral and fundic mucosa. Of ulcers, 70% occur superior to the incisura angularis and 20% are distal. Gastric ulcers are classified into five categories defined by location and secretory status.

The gold standard for the evaluation of gastric ulceration is endoscopy. Treatment options include medical, endoscopic, and surgical management. Initial medical management includes antimicrobials directed at *Helicobacter pylori* (isolated in 85–90% of gastric ulcers) and antisecretory medications. Typically, a 12- to 24-week trial is undertaken. Endoscopic diathermy or vasoconstrictive injection can treat or temporize active bleeding. However, surgical management is indicated in cases complicated by nonhealing/recurrent ulcers, suspected malignancy, obstruction, hemorrhage (most common), and perforation.

Several surgical treatment options exist. Typical treatment includes excision of the lesion, a vagotomy, and a drainage procedure. The lesion should be excised if possible. Truncal or proximal gastric (highly selective) vagotomy is performed to limit acid secretion. Truncal vagotomy is undertaken at the level of the diaphragm as the vagal nerve fibers enter the abdomen. It is associated with a greater reduction in acid secretion and a lower recurrence rate compared to proximal gastric vagotomy. However, the gastric drainage mechanism and vagally supplied viscera are denervated and impaired. Thus, a gastric drainage procedure is mandatory with truncal vagotomy. However, antral nerve fibers (and gastric drainage) are preserved in cases of proximal gastric vagotomy. Unfortunately, this comes at the expense of a higher recurrence rate and greater acid secretion.

Drainage procedures include pyloroplasty and antrectomy. Several different types of pyloroplasty procedures are performed to improve drainage. However, each utilizes a longitudinal incision at the pylorus that is horizontally closed. Antrectomy is the resection of the distal stomach and removal of the pylorus. The defect can be repaired with a gastroduodenostomy (Billroth I) or gastrojejunostomy (Billroth II). Complicated proximal gastric ulceration may require a subtotal gastrectomy, whereas widespread gastric disease may require a near-total gastrectomy or gastric devascularization. These latter procedures are often associated with a very high morbidity and mortality and represent less commonly employed techniques. Following gastric drainage procedures or gastric resection, postgastrectomy syndromes can emerge. Early and late dumping, afferent and blind loop obstruction, alkaline reflux, gastric atony, and nutritional disturbances are all well-described complications of gastric surgery.

PERFORATED DUODENAL ULCER

Duodenal and gastric ulcer diseases are jointly referred to as peptic ulcer disease. Although the pathophysiology and treatments are similar, duodenal ulcer disease warrants a brief independent discussion. Duodenal ulcer disease is three times as common as gastric ulcer disease (300,000 patients yearly). Malignancy is far less commonly associated with duodenal ulceration. Typically, the presentation is similar to that of gastric ulceration. However, the symptoms of duodenal ulceration are often alleviated with consumption of food. As in gastric ulceration, a better understanding of the pathophysiology has led to a decrease in operative indications. However, the operative indications/interventions are

similar. Evaluation utilizes endoscopy as the diagnostic standard.

As with gastric ulceration, treatment options may include medical, endoscopic, and surgical approaches. Surgical options are based on the same operative indications (bleeding, perforation, obstruction, and intractability). In the case of perforation, a duodenal ulcer is oversewn and a well-vascularized Graham patch is used to cover the defect. Highly selective vagotomy is preferred, but truncal vagotomy and drainage procedures are within the standard of care. Duration of symptoms, previous treatments, medication usage (NSAIDs), and comorbidity assist in operative planning. Postgastrectomy syndromes are, again, a well-documented complication.

PERFORATED APPENDICITIS

Appendicitis is the most common emergent surgical procedure performed in Western countries. Reginald Herber Fitz, a Harvard Medical School pathologist, first described the pathologic process in 1889. He illustrated an event sequence of luminal obstruction, inflammation, perforation, abscess formation, and peritonitis. Presentation involves acute or gradual onset of midline pain (usually periumbilical, but occasionally epigastric) that begins as dull cramping (visceral pain). As the inflammatory process progresses, sharp stabbing pain is localized in the right lower quadrant (somatic pain). This is often associated with nausea, anorexia, and vomiting. In the predominant number of cases, fever will be absent.

Evaluation begins with a physical examination, laboratory studies, and plain films of the chest and abdomen. On examination, peritoneal signs (rebound/guarding) and pain with any movement (Dunphy's sign), with internal rotation of the hip (obturator sign), during extension of the hip (ileopsoas sign), and initiating in the right lower quadrant during palpation of the left lower quadrant (Rosving's sign) support the diagnosis without confirmation. Laboratory studies are often unremarkable. Leukocytosis of greater than 10,000/μl is seen in approximately two-thirds of patients. In cases in which the leukocytosis exceeds 20,000/μl, suspicion for perforation greatly increases. Hematuria, proteinuria, and pyuria are commonly encountered in the absence of urologic disease. Fecaliths and free intraperitoneal air are occasionally seen on plain films. Ultrasound findings of a target lesion or a thickened incompressible appendiceal wall can be 80% sensitive and 90% specific. Computed tomography is increasingly used in establishing diagnosis. With improvement in technology, computed tomography is often more accurate than ultrasound, and more capable of identifying an abnormally positioned appendix. Additionally, a negative study rules out a variety of abdominal pathologies and diminishes the need for costly hospitalization/observation. Barium contrast studies are safe and available, but are infrequently implemented in the presence of ultrasound or computed tomography.

In general, patients who present with a clinical picture suggesting appendicitis require exploration. This can be performed laparoscopically or via laparotomy. Laparoscopic appendectomy is diagnostic and therapeutic, often resulting in a reduction of postoperative pain and length of hospitalization when compared with laparotomy. Unfortunately, laparoscopy is associated with increased operative time (expense) and postoperative emesis. However, it is often beneficial in female patients when there exists complicated differentiation from gynecologic pathology. Appendectomy via laparotomy utilizing a muscle-splitting incision remains the standard of care. In cases of perforation, the fascia is closed while the skin and subcutaneous tissues are left to close by secondary intention.

The factor associated with the greatest morbidity and mortality in cases of appendicitis is delay in diagnosis and treatment. Approximately 64% of patients present with a retrocecal appendix, which can delay the diagnosis. Diagnostic delay results in increased rates of perforation. Perforation is associated with less than 5% mortality, compared to 0.6% in nonperforated appendicitis. The perforation rate is approximately 4% in young patients and is greatly increased in the elderly. A negative appendectomy rate of approximately 25% is considered acceptable when compared with the complications associated with delay in treatment.

PERFORATED DIVERTICULITIS

The word "diverticulum" is derived from the Latin *deverticulum*, a "by-road" or "diversion" (French, *de-verto;* "to turn aside"). The diverticula arise as the colonic mucosal tissue herniates through the muscularis. Generally, two types of diverticula are seen. Right colon diverticula are usually congenital and are more commonly seen in Asian countries. Conversely, sigmoid diverticula usually occur as a consequence of a low-fiber diet (high intraluminal pressures) and are more prevalent in Western societies. Diverticulitis was initially described in the early twentieth century and has significantly increased in prevalence. It is incidentally noted in one-third of the population over 45 years of age (two-thirds over 85 years).

Diverticulitis typically presents with acute or subacute onset of left lower quadrant abdominal pain,

constipation, diarrhea, fever, and, infrequently, a palpable mass. The expression "left-sided appendicitis" has described this clinical presentation. The diagnosis is largely based on clinical findings and can be made in the absence of further imaging. However, computed tomography (CT) with intravenous contrast has been used as the initial diagnostic exam to provide objective data, assist in treatment planning, and elucidate difficult clinical scenarios. Characteristic findings on CT include thickening of the colonic wall, pericolonic fat infiltration, and pericolonic/distant abscess formation. The role of endoscopy and barium enema is limited during the acute episode due to the risk of perforation.

Diverticular disease can be categorized in two groups, asymptomatic diverticulosis and acute uncomplicated or complicated diverticulitis. Asymptomatic diverticular disease is found incidentally and can benefit from a high-fiber diet. No additional treatment may be required. Acute uncomplicated diverticulitis presents with left lower quadrant pain, fever, and leukocytosis (occasionally constipation/diarrhea, nausea/vomiting, and dysuria). It is defined as local perforation without abscess, bleeding, or free perforation. Therapy includes a high-fiber diet and broad-spectrum antibiotics. Coverage for gram-negative and anaerobic bacteria is implemented. Stable patients without diffuse abdominal findings on examination can be treated as outpatients, with close followup as judged by an experienced surgeon. In the event that these patients worsen clinically (pain/fever), admission and further evaluation/intervention are required. If no findings are noted on CT, but clinical improvement does not occur within 72–96 hours, surgical intervention can be considered. For immunocompromised patients or those unable to tolerate oral medications, admission for administration of intravenous antibiotics is preferred.

The presence of an abscess, fistula, intestinal obstruction, or free perforation classifies diverticulitis as complicated (complicated diverticulitis). The incidence of free perforation is rare, but the mortality can be as high as one-third. Traditional practice has been to resect with the initial episode. However, more conservative methods have emerged with great success. Small pericolic abscesses can be managed conservatively, whereas larger and more distant fluid collections can be successfully drained under CT guidance. In cases when drainage is inadequate, urgent surgery is required. Surgical treatment generally includes resection of the diseased sigmoid, proximal diversion, and the creation of a Hartman's pouch. Primary bowel anastomosis can be considered only under the best possible conditions and in experienced hands. Recent advent of laparoscopic resection has offered a reduction in pain and hospitalization when performed under elective circumstances.

See Also the Following Articles

Appendicitis • Diverticulosis • Duodenal Ulcer • Endoscopy, Complications of • Gastric Ulcer

Further Reading

Author Unknown. (1995). *In* "Steadman's Medical Dictionary," 26th Ed., pp. 477, 513, 1325. Williams & Wilkins, Philadelphia.

Cameron, J. L. (2001). *In* "Current Surgical Therapy," 7th Ed., pp. 7–12, 77–81, 267–272. Mosby, Philadelphia.

Greenfield, L. J., *et al.* (1997). *In* "Surgery: Scientific Principles and Practice," 2nd Ed., pp. 1246–1261. Lippincott-Raven, Philadelphia.

Greenfield, L. J., *et al.* (2001). *In* "Surgery: Scientific Principles and Practice," 3rd Ed., pp. 750–766, 1224–1235. Lippincott Williams & Wilkins, Philadelphia.

Lawrence, P. F. (1988). *In* "Essentials of General Surgery," pp. 187–200. Williams & Wilkins, Baltimore.

Sabiston, D. C. (1991). *In* "Textbook of Surgery: The Biologic Basis of Modern Surgical Practice," 14th Ed., pp. 736–755. W. B. Saunders, Philadelphia.

Sabiston, D. C. (1997). *In* "Textbook of Surgery: The Biologic Basis of Modern Surgical Practice," 15th Ed., pp. 825–846, 964–970. W. B. Saunders, Philadelphia.

Silen, W. (2000). *In* "Copes Early Diagnosis of the Abdomen," pp. 83–103, 105–119, 268–286. Oxford University Press, Cambridge.

Peristalsis

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

monosynaptic spinal reflex Neural reflex circuit in the spinal cord consisting of a sensory neuron and a motor neuron with a single synaptic connection between the two.

reflexes Motor responses to sensory stimulation; each time a "hardwired" neural circuit is activated, the responses are repeated in the same way.

Peristalsis is a term used in reference to the organized propulsion of material over variable distances within the lumen of the esophagus, the small and large intestine, and sometimes the distal stomach. The muscle layers of the intestine and esophagus behave in a stereotypical pattern to achieve peristaltic propulsion. Integrated neural circuits of the enteric nervous system control the behavior of the intestinal musculature. Peristaltic propulsion in the esophagus is controlled by signals transmitted from the brain stem to the esophagus by the vagus nerves and by the enteric nervous system.

PERISTALTIC PROPULSION

During peristalsis (Fig. 1), the longitudinally oriented muscle in the segment ahead of the advancing intraluminal contents contracts while the circumferentially oriented muscle layer relaxes in the same segment. The esophagus and intestine are tubes that behave physically like a cylinder with constant surface area. Shortening of the longitudinal axis of the cylinder is accompanied by widening of the cross-sectional diameter. Simultaneous shortening of the longitudinal axis and relaxation of the circular muscle result in expansion of the lumen. This prepares a receiving segment for the forward-moving intraluminal contents during peristalsis.

The second component of stereotypic peristaltic behavior is contraction of the circumferentially oriented muscle layer in the segment behind the advancing intraluminal contents. The longitudinally oriented muscle layer in this segment relaxes simultaneously with contraction of the circular muscle, resulting in conversion of this region to a propulsive segment that propels the luminal contents ahead, into the receiving segment. Intestinal segments ahead of the advancing front become receiving segments and then propulsive segments in succession as the complex of propulsive and receiving segments travels along the intestine.

PERISTALTIC REFLEX

Peristaltic reflex is a term sometimes applied inappropriately; it is not synonymous with the multiple patterns of coordinated propulsive motility that occur in the

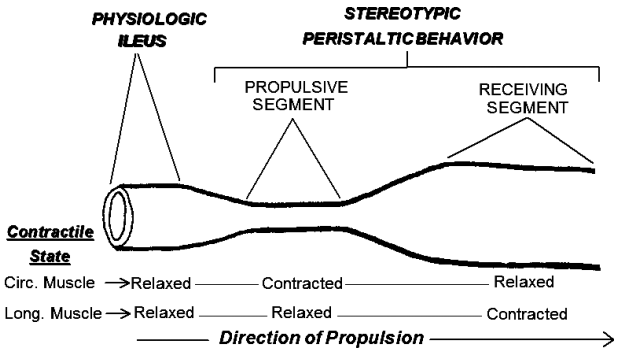


FIGURE 1 The circumferential and longitudinal muscle layers of the intestines behave in a stereotypical pattern during peristaltic propulsion. A "hardwired" reflex circuit in the enteric nervous system determines the pattern of behavior of the two muscle layers. During peristaltic propulsion, the longitudinal muscle layer in the segment ahead of the advancing intraluminal contents contracts while the circumferential muscle layer relaxes simultaneously. Simultaneous shortening of the longitudinal intestinal axis and relaxation of the circumferential muscle in the same segment result in expansion of the lumen, which becomes a receiving segment for the forward-moving contents. The second component of the reflex is contraction of the circular muscle in the segment behind the advancing intraluminal contents. The longitudinal muscle layer in the same segment relaxes simultaneously with contraction of the circular muscle, which results in conversion of this region to a propulsive segment that propels the luminal contents ahead into the receiving segment. The reflex circuits are coupled in series along the intestine, such that receiving segments convert to propulsive segments as the next segment in line becomes a receiving segment. Propulsive segments then return to their previous state of physiologic ileus. The distance over which the peristaltic reflex circuit for the formation of propulsive and receiving segments is activated in sequence down the bowel determines the length of bowel over which propulsion occurs in one or the other of the intestinal motility patterns.

 Copyright 2004, Elsevier (USA). All rights reserved.

small and large intestine during different digestive states. The peristaltic reflex is, rather, the intestinal analogue of spinal motor reflexes, such as the monosynaptic patellar and Achilles tendon reflexes. Monosynaptic spinal reflexes are investigator-evoked artifacts arising from connections of stretch receptors in the muscle to alpha spinal motor neurons that innervate the same muscle. They reflect the effects of sudden activation of stretch receptors (i.e., muscle spindles) in the muscle and have little relevance for understanding the complexity of neural control of movement. The peristaltic reflex is much the same in that it is a fixed response evoked by investigational stretching of the intestinal wall or stroking of the mucosa. It is like a spinal reflex in that it is a motor response to sensory stimulation, and is repeated the same way each time the "hardwired" reflex circuit is activated. The peristaltic reflex circuit is "wired" such that it evokes relaxation of the circumferentially oriented muscle layer and contraction of the longitudinal muscle below the point of stimulation, and contraction of the circumferentially oriented muscle layer above the point of stimulation.

Like the spinal reflexes, the peristaltic reflex is positioned at the lowest level of the hierarchical organization of neural control of intestinal motility, and undoubtedly underlies each of the various patterns of propulsive motility that impart functionality to the intestine during daily life. As with a spinal motor reflex, the sequencing of the pattern of behavior of the various muscle groups is hardwired into the circuitry, whereas the repetition rate of the pattern and strength of each motor component of the pattern are adjusted by sensory feedback or other commands to compensate automatically for local loads and higher functional demands on the intestine as a whole. Added factors requiring a higher order of neural control are the distance and direction in which propulsion occurs for the patterns of motility that characterize the various digestive states. Short-distance propulsion in the postprandial digestive state, propulsion over intermediate distances during interdigestive motility (i.e., migrating motor complex), long-distance power propulsion all in the orthograde direction, and retropulsion during emesis are neural control requirements unique to the enteric nervous system. Better understanding of the neural basis for intestinal motility will require moving forward from the overworked concept of the peristaltic reflex and on to investigation of microcircuits in positions at levels of organization beyond the reflex hardwiring that faithfully reproduces the muscle behavior each time the investigator stretches the intestinal wall or strokes the mucosa.

See Also the Following Articles

Emesis • Enteric Nervous System • Migrating Motor Complex • Postprandial Motility • Power Propulsion

Further Reading

Kosterlitz, H. (1967). Intrinsic intestinal reflexes. *Am. J. Dig. Dis.* **12**, 245–254.

Kosterlitz, H., and Robinson, J. (1959). Reflex contractions of the longitudinal muscle coat of the isolated guinea-pig ileum. *J. Physiol. (Lond.)* **146**, 369–379.

Sifrim, D., Janssens, J., and Vantrappen, G. (1992). A wave of inhibition precedes primary peristaltic contractions in the human esophagus. *Gastroenterology* **103**, 876–882.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Wood, J., Alpers, D., and Andrews, P. (1999). Fundamentals of neurogastroenterology. *Gut* **45**, 1–44.

Peritoneal Disorders

BUCKMINSTER FARROW AND B. MARK EVERS
University of Texas, Galveston

peritonitis Inflammation of the peritoneal linings of the abdominal cavity; may be of infectious, chemical, or unknown origin.
peritoneal mesothelioma A rare malignant tumor arising from the peritoneal lining of the abdominal cavity.
pseudomyxoma peritonei A rare condition manifested by diffuse, gelatinous implants of the peritoneal cavity and omentum arising from mucinous neoplasms of the ovary or appendix.

Peritonitis, an inflammation of the peritoneal lining of the abdominal cavity, is the most common peritoneal disease and may be of infectious, chemical, or unknown origin. Other peritoneal disorders include peritoneal mesothelioma, a rare malignant tumor that arises from the peritoneal lining of the abdominal cavity, and pseudomyxoma peritonei, a rare condition manifested by diffuse, gelatinous implants of the peritoneal cavity and omentum that arises from mucinous neoplasms of the ovary or appendix.

PERITONITIS

The most common disorder of the peritoneum is inflammation, or peritonitis, which is usually of infectious origin. Peritonitis can be classified into two types based on etiology.

Primary Peritonitis

Primary peritonitis occurs without an identifiable source of infection, usually in patients with preexisting ascites (most commonly cirrhosis in adults or nephrotic syndrome in children). Diagnosis is made by cell count and culture of the ascitic fluid. Treatment requires antibiotic coverage appropriate for both gram-positive and gram-negative bacteria, because *Escherichia coli* and *Klebsiella pneumoniae* are increasingly common in ascitic fluid cultures.

Secondary Peritonitis

Secondary peritonitis, much more common than primary peritonitis, is caused by an intra-abdominal infection, such as perforation of the bowel, and can be further divided into acute suppurative, granulomatous, and aseptic (chemical) forms. Acute suppurative peritonitis is usually caused by spillage of intestinal contents into the peritoneal cavity as a result of primary intra-abdominal disease (e.g., perforated peptic ulcer, appendicitis, diverticulitis, perforated carcinoma), penetrating trauma, or iatrogenic perforation after instrumentation or radiologic procedures. Early signs of peritoneal inflammation include nausea, vomiting, anorexia, and vague poorly localized abdominal pain. As the infection progresses, these symptoms may worsen and the pain becomes more focal as the inflammation of the visceral peritoneum extends to the parietal peritoneal. Fever and signs of hypovolemia (tachycardia, dry mucous membranes, low urinary output) may also be present. Treatment consists of proper fluid resuscitation, antibiotic therapy, and prompt treatment of the underlying pathology usually through surgical intervention.

Granulomatous peritonitis may be from fungal, amebic, or parasitic sources, but tubercular infection is by far the most common etiology. There has been an unfortunate resurgence of tuberculous peritonitis (Fig. 1) due to an increase in the prevalence of acquired immune deficiency syndrome, although it still

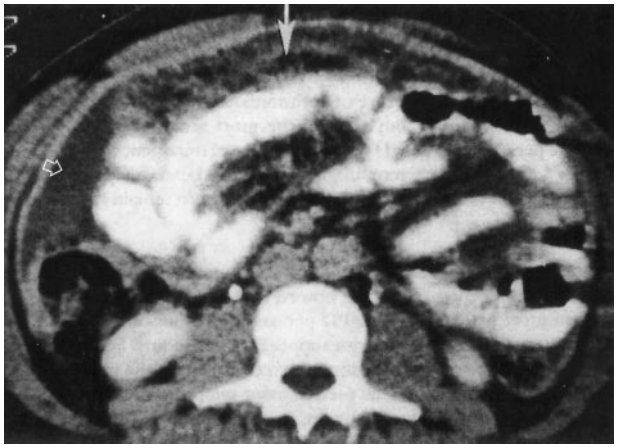


FIGURE 1 Computed tomography scan demonstrates ascites (open arrow) and small bowel thickening (filled arrow) in a patient with tuberculous peritonitis. Courtesy of Charles J. Fagan, M.D. (Galveston, TX).

 Copyright 2004, Elsevier (USA). All rights reserved.

complicates less than 1% of all *Mycobacterium tuberculosis* infections. The infection usually originates outside the peritoneum from diseased bowel, from salpingitis, or through hematogenous spread of a primary pulmonary infection. Unlike suppurative peritonitis, onset of symptoms is quite insidious with 70% of patients displaying fever, malaise, anorexia, weakness, or weight loss for more than 4 months prior to diagnosis. Ascites and diffuse tenderness are usually present and should suggest the illness in high-risk or immunocompromised patients with unexplained fever and malaise. The disease usually responds rapidly to anti-tuberculosis therapy with the exception of cases involving newly emergent drug-resistant mycobacteria.

Chemical (aseptic) peritonitis results from spillage of irritant materials that are initially sterile, but with time become secondarily infected and present as suppurative peritonitis. Bile, urine, and chyle are potential endogenous causes, whereas iatrogenic etiologies include barium from radiologic studies (with a concurrent perforation) and starch powder from surgical gloves. Patients who undergo continuous ambulatory peritoneal dialysis may acquire infections of the normally sterile catheter and peritoneum, leading to bacterial peritonitis. Treatment is similar to suppurative peritonitis—fluid resuscitation, appropriate antibiotic coverage, and surgery to control the source of peritoneal contamination.

PRIMARY MESOTHELIOMA

Similar to pleural forms, primary peritoneal mesothelioma is linked to asbestos exposure, although only 20–40% of mesotheliomas occur in the peritoneum (Fig. 2). Clinical presentation includes nonspecific abdominal pain, nausea, vomiting, weight loss, and diarrhea; ascites is found in 90% of patients. Most cases are not identified until laparoscopy or laparotomy is performed. Curative treatment is rare because the disease is usually quite advanced at the time of diagnosis. Surgery serves only to confirm the diagnosis and provide palliative procedures to relieve obstruction. Chemotherapy and radiotherapy have shown minimal success although newer intraperitoneal chemotherapy and immunotherapy may improve the prognosis.

PSEUDOMYXOMA PERITONEI

Pseudomyxoma peritonei is a rare condition manifested by diffuse, gelatinous implants of the peritoneal cavity and omentum arising from mucinous neoplasms of the ovary or appendix (Fig. 3). Most cases occur in women

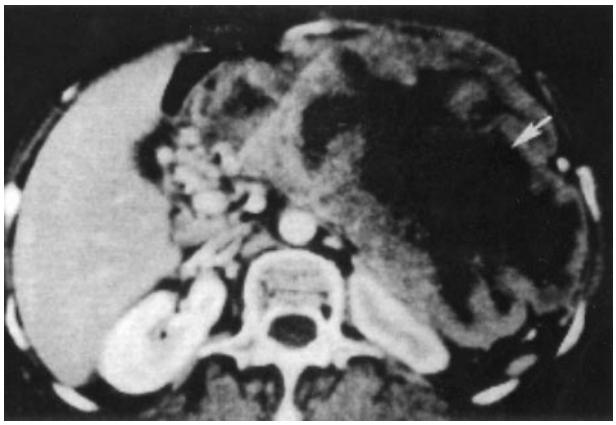


FIGURE 2 Computed tomography scan demonstrates diffuse mesenteric and peritoneal involvement of a soft tissue mass (arrow) causing displacement of intra-abdominal organs in a patient with peritoneal mesothelioma. Courtesy of Charles J. Fagan, M.D. (Galveston, TX).

who have disease in both organs; thus, the tissue of origin is not established. Patients may present with abdominal pain or increasing girth due to mucinous ascites, but most diagnoses are made following surgical exploration for appendicitis or ovarian tumors. Computed tomography may reveal characteristic "scalloping" of the hepatic and bowel margins. Treatment is primarily surgical, to include aggressive debulking, appendectomy, bilateral oophorectomy, and omentectomy. Recurrence rates are high (approximately 75% of patients); however, due to the low grade of this malignancy, repeat debulking procedures are indicated and may improve outcome. Chemotherapy

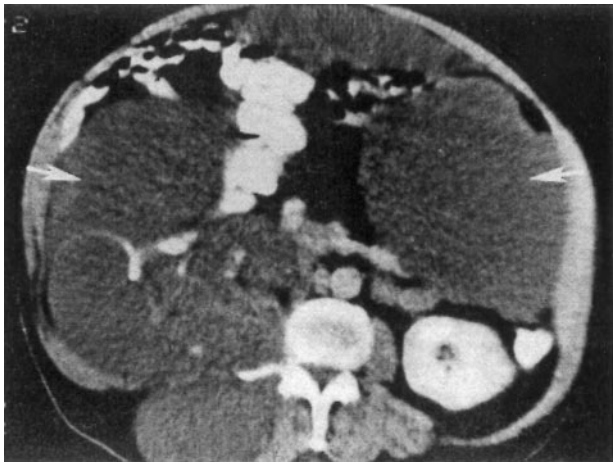


FIGURE 3 Computed tomography scan demonstrates large gelatinous masses (arrows) in a patient with pseudomyxoma peritonei. Courtesy of Charles J. Fagan, M.D. (Galveston, TX).

may produce a modest survival advantage over surgery alone; radiotherapy appears to have no role in the treatment of pseudomyxoma peritonei.

See Also the Following Articles

Ascites • Peritoneum, Anatomy and Development

Further Reading

Bosscha, K., van Vroonhoven, T. J., and van der Werken, C. (1999). Surgical management of severe secondary peritonitis. *Br. J. Surg.* **86**, 1371–1377.

Marshall, J. B. (1993). Tuberculosis of the gastrointestinal tract and peritoneum. *Am. J. Gastroenterol.* **88**, 989–999.

Piazza, D., Caruso, F., Scaringi, S., Ferrara, M., Latteri, F., and Dell'Erba, D. (2000). Primary diffuse malignant peritoneal mesothelioma: Case report and update of therapy. *J. Surg. Oncol.* **75**, 55–58.

Rimola, A., Garcia-Tsao, G., Navasa, M., Piddock, L. J., Planas, R., Bernard, B., and Inadomi, J. M. (2000). Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: A consensus document. *J. Hepatol.* **32**, 142–153.

Sugarbaker, P. H., Shmookler, B., Ronnett, B. M., Chang, D., and Ambrose, N. S. (1999). Pseudomyxoma peritonei. *Br. J. Surg.* **86**, 842.

Peritoneum, Anatomy and Development

MICHELE I. SLOGOFF AND B. MARK EVERS
University of Texas Medical Branch, Galveston

coelom Body cavity of the embryo.
mesothelium Layer of flat cells derived from the mesoderm that lines the coelom.
peritoneum Serous membrane lining the abdominopelvic walls and investing the viscera.

The peritoneum is the mesothelial lining of the peritoneal cavity and its contained viscera; it functions as a bidirectional dialysis membrane and thus plays a major role in defending against inflammatory processes of the abdomen. The peritoneum is the key organ involved in sensation of abdominal pain and is thus integral to diagnosing abdominal pathology.

EMBRYOLOGY AND DEVELOPMENT

The peritoneum and other body cavities begin to develop from the intraembryonic coelom near the end of the third week of gestation. The coelom has a parietal wall and a visceral wall, both lined by mesothelium; the parietal mesothelium is derived from somatic mesoderm whereas the visceral mesothelium is derived from splanchnic mesoderm. By the fourth week of gestation, the coelom appears as a horseshoe-shaped cavity in the cardiogenic and lateral mesoderm. The curve of the "horseshoe" represents the future pericardial cavity, and the lateral and caudal extensions represent the eventual pleural and peritoneal cavities (Fig. 1). These lateral areas communicate with the extraembryonic coelom. The development of the midgut involves a herniation through this communication into the umbilical cord, where the midgut develops into the small intestine and part of the large intestine. At this point, the intraembryonic coelom is divided into right and left halves that are divided by the ventral and dorsal mesenteries.

Toward the end of the fourth week of gestation, the lateral parts of the intraembryonic coelom move onto the ventral aspect of the embryo and the ventral mesentery degenerates, creating a large peritoneal cavity with a dorsal mesentery (Fig. 2). Until the seventh week of gestation, this peritoneal cavity communicates with the pericardial and pleural cavities through pericardioperitoneal canals; during the fifth and sixth weeks of gestation, folds form near the cranial and caudal ends of these canals. Fusion of these membranous folds with mesoderm ventral to the esophagus separates the pericardial and pleural cavities while fusion of the caudal pleuroperitoneal membranes forms the diaphragm and separates the pleural and peritoneal cavities. The peritoneal cavity then loses its connection with the extraembryonic coelom during the tenth week of gestation, when the intestines return to the abdomen from the

Copyright 2004, Elsevier (USA). All rights reserved.

may produce a modest survival advantage over surgery alone; radiotherapy appears to have no role in the treatment of pseudomyxoma peritonei.

See Also the Following Articles

Ascites • Peritoneum, Anatomy and Development

Further Reading

Bosscha, K., van Vroonhoven, T. J., and van der Werken, C. (1999). Surgical management of severe secondary peritonitis. *Br. J. Surg.* **86**, 1371–1377.

Marshall, J. B. (1993). Tuberculosis of the gastrointestinal tract and peritoneum. *Am. J. Gastroenterol.* **88**, 989–999.

Piazza, D., Caruso, F., Scaringi, S., Ferrara, M., Latteri, F., and Dell'Erba, D. (2000). Primary diffuse malignant peritoneal mesothelioma: Case report and update of therapy. *J. Surg. Oncol.* **75**, 55–58.

Rimola, A., Garcia-Tsao, G., Navasa, M., Piddock, L. J., Planas, R., Bernard, B., and Inadomi, J. M. (2000). Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: A consensus document. *J. Hepatol.* **32**, 142–153.

Sugarbaker, P. H., Shmookler, B., Ronnett, B. M., Chang, D., and Ambrose, N. S. (1999). Pseudomyxoma peritonei. *Br. J. Surg.* **86**, 842.

Peritoneum, Anatomy and Development

MICHELE I. SLOGOFF AND B. MARK EVERS
University of Texas Medical Branch, Galveston

coelom Body cavity of the embryo.
mesothelium Layer of flat cells derived from the mesoderm that lines the coelom.
peritoneum Serous membrane lining the abdominopelvic walls and investing the viscera.

The peritoneum is the mesothelial lining of the peritoneal cavity and its contained viscera; it functions as a bidirectional dialysis membrane and thus plays a major role in defending against inflammatory processes of the abdomen. The peritoneum is the key organ involved in sensation of abdominal pain and is thus integral to diagnosing abdominal pathology.

EMBRYOLOGY AND DEVELOPMENT

The peritoneum and other body cavities begin to develop from the intraembryonic coelom near the end of the third week of gestation. The coelom has a parietal wall and a visceral wall, both lined by mesothelium; the parietal mesothelium is derived from somatic mesoderm whereas the visceral mesothelium is derived from splanchnic mesoderm. By the fourth week of gestation, the coelom appears as a horseshoe-shaped cavity in the cardiogenic and lateral mesoderm. The curve of the "horseshoe" represents the future pericardial cavity, and the lateral and caudal extensions represent the eventual pleural and peritoneal cavities (Fig. 1). These lateral areas communicate with the extraembryonic coelom. The development of the midgut involves a herniation through this communication into the umbilical cord, where the midgut develops into the small intestine and part of the large intestine. At this point, the intraembryonic coelom is divided into right and left halves that are divided by the ventral and dorsal mesenteries.

Toward the end of the fourth week of gestation, the lateral parts of the intraembryonic coelom move onto the ventral aspect of the embryo and the ventral mesentery degenerates, creating a large peritoneal cavity with a dorsal mesentery (Fig. 2). Until the seventh week of gestation, this peritoneal cavity communicates with the pericardial and pleural cavities through pericardioperitoneal canals; during the fifth and sixth weeks of gestation, folds form near the cranial and caudal ends of these canals. Fusion of these membranous folds with mesoderm ventral to the esophagus separates the pericardial and pleural cavities while fusion of the caudal pleuroperitoneal membranes forms the diaphragm and separates the pleural and peritoneal cavities. The peritoneal cavity then loses its connection with the extraembryonic coelom during the tenth week of gestation, when the intestines return to the abdomen from the

Copyright 2004, Elsevier (USA). All rights reserved.

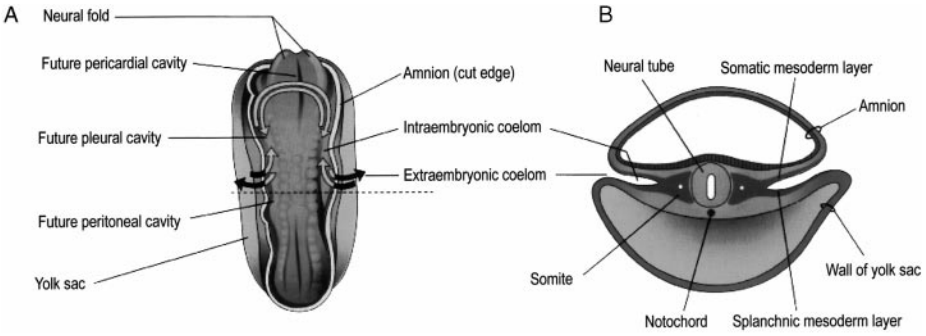


FIGURE 1 (A) Drawing of a dorsal view of a 22-day-old embryo, showing outline of the horseshoe-shaped intraembryonic coelom. (B) Transverse section through embryo at the level shown in A. Reproduced with permission from Moore and Persaud (1998).

umbilical cord. At this point, the peritoneum lines the abdominal wall and invests the abdominal viscera.

ANATOMY AND PHYSIOLOGY

The peritoneum is made up of the visceral and parietal layers; it has a total surface area of approximately 2 m^2. The visceral peritoneum covers the intraperitoneal organs and forms the mesenteries by which they are suspended. The peritoneum and its mesentery are supplied mainly by the splanchnic blood vessels, and, to a lesser extent, by branches of the visceral and parietal peritoneum.

Differences arise in the innervations of the visceral and parietal peritoneum, leading to differing patterns of sensation of painful stimuli. The visceral peritoneum receives its innervation from the autonomic nervous system and responds primarily to traction and pressure or distension; painful stimuli are perceived as a poorly localized, dull pain. In general, pain in the epigastric region is referred from a foregut structure (e.g., stomach, duodenum, pancreas, or biliary tract); pain in the periumbilical area is the result of stimulus to a midgut structure (appendix, jejunum, or ileum) and pain in the hypogastric or suprapubic region is from a hindgut source (distal colon or rectum). In contrast, both somatic and visceral afferent nerves innervate the parietal peritoneum. Therefore, noxious stimuli are perceived as a localized, sharp pain with rebound tenderness and are referred to as "peritonitis."

Functionally, the peritoneum serves as a bidirectional dialysis membrane through which both large- and small-molecular-weight solutes pass by simple passive diffusion. Absorption can be altered by changes in intraabdominal pressure, temperature, pH, and portal pressure in addition to lymphatic blockade and peritoneal scarring. The peritoneum contains a complex defense system that protects against inflammatory processes. This system consists mainly of mechanical clearance of the peritoneal cavity, the bactericidal

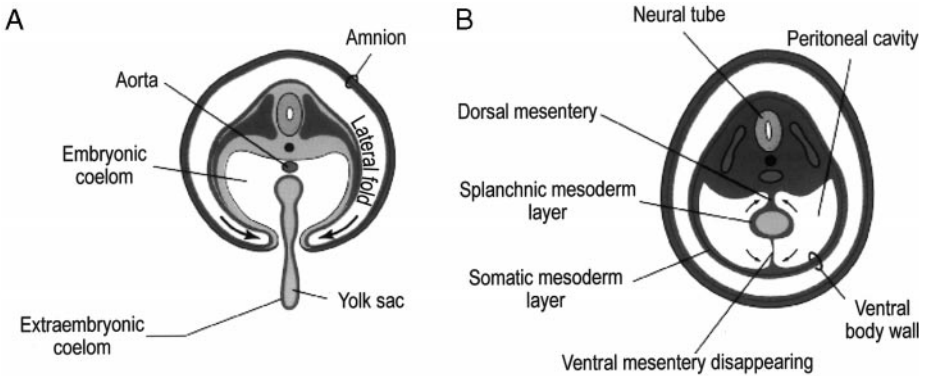


FIGURE 2 (A) Transverse section illustrating embryonic folding and its effects on the intraembryonic coelom and other structures. (B) Transverse section illustrating formation of the ventral body wall and disappearance of the ventral mesentery. Arrows indicate the junction of the somatic and splanchnic mesoderm. Reproduced with permission from Moore and Persaud (1998).

mechanisms of polymorphonuclear leukocytes, and sequestration mechanisms such as fibrin trapping of bacterial activation of complement. The peritoneum is the key organ in sensation of abdominal pain and in the physical diagnosis of ongoing intraabdominal pathology.

See Also the Following Articles

Development, Overview • Gastrointestinal Tract Anatomy, Overview • Peritoneal Disorders

Further Reading

Aune, S. (1970). The effect of transperitoneal fluid transport on the transfer of solutes. *Scan. J. Gastroenterol.* **5**, 241–252.

Berndt, W. O., and Gosselin, R. E. (1962). Differential changes in permeability of mesentery to rubidium and phosphate. *Am. J. Physiol.* **202**, 761–767.

Heel, K. A., and Hall, J. C. (1996). Peritoneal defenses and peritoneum-associated lymphoid tissue. *Br. J. Surg.* **83**, 1031–1036.

Moore, K. L., and Persaud, T. V. N. (1998). "The Developing Human," 6th Ed. W. B. Saunders, Philadelphia.

Shear, L., Castellot, J. J., Shinaberger, J. H., Poole, L., and Barry, K. G. (1966). Enhancement of peritoneal fluid absorption by dehydration, mercaptomerin and vasopressin. *J. Pharmacol. Exp. Ther.* **154**, 289–297.

Shear, L., Swartz, C., Shinaberger, J. A., and Barry, K. G. (1965). Kinetics of peritoneal fluid absorption in adult man. *N. Engl. J. Med.* **272**, 123–127.

Topley, N., and Williams, J. D. (1994). Role of the peritoneal membrane in the control of inflammation in the peritoneal cavity. *Kidney Int. Suppl.* **48**, S71–S78.

Pernicious Anemia

RALPH CARMEL

New York Methodist Hospital, Brooklyn and Weill Medical College of Cornell University

intrinsic factor A glycoprotein that binds cobalamin specifically, attaches to a receptor (cubilin) on the apical membrane of the ileal epithelial cell, and is internalized together with its cobalamin by that cell.

Pernicious anemia (PA) is defined as cobalamin (vitamin B_{12}) deficiency caused by that vitamin's malabsorption because of the patient's failure to secrete gastric intrinsic factor. This anachronistic hematological name for what is really a gastroenterological disorder was coined originally to reflect the megaloblastic anemia that cobalamin deficiency often causes. However, many patients diagnosed with PA have little or no anemia; moreover, folate deficiency and rare metabolic disorders can cause the same anemia. PA does not refer to cobalamin deficiency produced by any other mechanisms, such as intestinal disease.

PATHOGENESIS AND DIAGNOSIS

The parietal cell synthesizes intrinsic factor (IF) and secretes it, as well as acid, into the gastric lumen. The pathogenetic basis of pernicious anemia (PA) is the inability of the parietal cell to do so. Without IF, cobalamin cannot be absorbed efficiently.

In the more common, acquired form of PA, the lack of IF results from severe atrophic gastritis with loss of parietal cells. The gastritis tends to involve the fundus while sparing the antrum and the disorder has many features of an autoimmune process. Antibody directed at the H^+,K^+-ATPase pump of the parietal cell circulates in the blood in 70–90% of patients, but it is characteristic for atrophic gastritis in general and is not specific for PA. Approximately 60–70% of patients with PA have circulating antibodies to IF, whose presence is virtually diagnostic for pernicious anemia. Serum gastrin levels are elevated in >80% of cases and often massively so. The gastritis in approximately 10% of patients with acquired PA is diffuse, does not spare the antrum, and is accompanied by fewer immune phenomena. A much rarer form of PA is an unrelated inborn error of IF synthesis in which the stomach and parietal cell function are otherwise normal.

Copyright 2004, Elsevier (USA). All rights reserved.

mechanisms of polymorphonuclear leukocytes, and sequestration mechanisms such as fibrin trapping of bacterial activation of complement. The peritoneum is the key organ in sensation of abdominal pain and in the physical diagnosis of ongoing intraabdominal pathology.

See Also the Following Articles

Development, Overview • Gastrointestinal Tract Anatomy, Overview • Peritoneal Disorders

Further Reading

Aune, S. (1970). The effect of transperitoneal fluid transport on the transfer of solutes. *Scan. J. Gastroenterol.* **5**, 241–252.

Berndt, W. O., and Gosselin, R. E. (1962). Differential changes in permeability of mesentery to rubidium and phosphate. *Am. J. Physiol.* **202**, 761–767.

Heel, K. A., and Hall, J. C. (1996). Peritoneal defenses and peritoneum-associated lymphoid tissue. *Br. J. Surg.* **83**, 1031–1036.

Moore, K. L., and Persaud, T. V. N. (1998). "The Developing Human," 6th Ed. W. B. Saunders, Philadelphia.

Shear, L., Castellot, J. J., Shinaberger, J. H., Poole, L., and Barry, K. G. (1966). Enhancement of peritoneal fluid absorption by dehydration, mercaptomerin and vasopressin. *J. Pharmacol. Exp. Ther.* **154**, 289–297.

Shear, L., Swartz, C., Shinaberger, J. A., and Barry, K. G. (1965). Kinetics of peritoneal fluid absorption in adult man. *N. Engl. J. Med.* **272**, 123–127.

Topley, N., and Williams, J. D. (1994). Role of the peritoneal membrane in the control of inflammation in the peritoneal cavity. *Kidney Int. Suppl.* **48**, S71–S78.

Pernicious Anemia

RALPH CARMEL

New York Methodist Hospital, Brooklyn and Weill Medical College of Cornell University

intrinsic factor A glycoprotein that binds cobalamin specifically, attaches to a receptor (cubilin) on the apical membrane of the ileal epithelial cell, and is internalized together with its cobalamin by that cell.

Pernicious anemia (PA) is defined as cobalamin (vitamin B_{12}) deficiency caused by that vitamin's malabsorption because of the patient's failure to secrete gastric intrinsic factor. This anachronistic hematological name for what is really a gastroenterological disorder was coined originally to reflect the megaloblastic anemia that cobalamin deficiency often causes. However, many patients diagnosed with PA have little or no anemia; moreover, folate deficiency and rare metabolic disorders can cause the same anemia. PA does not refer to cobalamin deficiency produced by any other mechanisms, such as intestinal disease.

PATHOGENESIS AND DIAGNOSIS

The parietal cell synthesizes intrinsic factor (IF) and secretes it, as well as acid, into the gastric lumen. The pathogenetic basis of pernicious anemia (PA) is the inability of the parietal cell to do so. Without IF, cobalamin cannot be absorbed efficiently.

In the more common, acquired form of PA, the lack of IF results from severe atrophic gastritis with loss of parietal cells. The gastritis tends to involve the fundus while sparing the antrum and the disorder has many features of an autoimmune process. Antibody directed at the H^+,K^+-ATPase pump of the parietal cell circulates in the blood in 70–90% of patients, but it is characteristic for atrophic gastritis in general and is not specific for PA. Approximately 60–70% of patients with PA have circulating antibodies to IF, whose presence is virtually diagnostic for pernicious anemia. Serum gastrin levels are elevated in >80% of cases and often massively so. The gastritis in approximately 10% of patients with acquired PA is diffuse, does not spare the antrum, and is accompanied by fewer immune phenomena. A much rarer form of PA is an unrelated inborn error of IF synthesis in which the stomach and parietal cell function are otherwise normal.

Copyright 2004, Elsevier (USA). All rights reserved.

Whereas atrophic gastritis itself is quite common, PA is relatively uncommon. Although PA predominates in the elderly, approximately 2% of whom have unsuspected PA, it affects younger individuals too, especially among black women.

The diagnosis of PA is usually established by proving IF absence directly by assay of gastric juice or indirectly by showing abnormal absorption of radioactive cobalamin with correction when IF is given exogenously (Schilling test). A presumptive diagnosis can also be made by demonstrating antibody to IF in the blood.

DEFICIENCY OF COBALAMIN

When IF secretion fails, cobalamin absorption virtually ceases and a negative cobalamin balance gradually develops. The progression is slow because the daily turnover of cobalamin is only approximately 0.1% of the body stores. Several years usually elapse before the clinical signs of cobalamin deficiency begin to appear. Although all cells require cobalamin, the most common manifestations of cobalamin deficiency, whether due to PA or other diseases, are hematological and neurological ones. Patients can have both or just one of these two types of manifestations—or, if caught early, neither one of them.

Megaloblastic anemia is characterized by large red blood cells (a high mean corpuscular volume is often the earliest expression of cobalamin deficiency), abnormal nuclei, and eventually pancytopenia as ineffective hematopoiesis progresses. The "methylfolate trap" produced by cobalamin deficiency causes the impaired DNA synthesis of this anemia, which is indistinguishable from that in folate deficiency. The neurological problems typically involve the posterior, and occasionally lateral, columns of the spinal cord and peripheral nerves. However, all parts of the nervous system can be affected, including cerebral function. The biochemical basis for the neurological dysfunction is unknown.

COMPLICATIONS

Two types of disorders complicate PA more often than expected. One is organ-specific immune disorders such as endocrinopathy, especially hypothyroidism, which can affect 5% of patients. The other complication is gastric neoplasia. The risk of gastric carcinoma is increased several-fold in PA and gastric carcinoids may be even more common than cancer.

THERAPY AND MANAGEMENT

The cobalamin deficiency is easily treated: megaloblastic anemia is readily reversed and, if cobalamin therapy is begun early enough, so are the neurologic abnormalities. However, the gastric defect persists, making lifelong cobalamin replacement mandatory. Because cobalamin is absorbed poorly in PA, treatment is usually by injection (oral cobalamin can be used, but very large daily doses must be given). Most experts advise endoscopy for neoplasia in every patient at the time of diagnosis and thyroid function should be monitored periodically.

See Also the Following Articles

Cobalamin Deficiency • Gastrin • Intrinsic Factor • Parietal Cells • Vitamin B12: Absorption, Metabolism, and Deficiency

Further Reading

Carmel, R. (1996). Prevalence of undiagnosed pernicious anemia in the elderly. *Arch. Intern. Med.* **156**, 1097–1100.

Carmel, R., and Rosenblatt, D. S. (2003). Disorders of cobalamin and folic acid metabolism. *In* "Blood: Principles and Practice of Hematology" (R. I. Handin, S. E. Lux, and T. P. Stossel, eds.), 2nd Ed., pp. 1361–1397. Lippincott Williams & Wilkins, New York, in press.

Chanarin, I. (1979). "The Megaloblastic Anaemias," 2nd Ed. Blackwell Scientific, Oxford, UK.

Kokkola, A., Sjöblom, S.-M., Haapiainen, R., Sipponen, P., Puolakkainen, P., and Järvinen, H. (1998). The risk of gastric carcinoma and carcinoid tumours in patients with pernicious anaemia: A prospective follow-up study. *Scand. J. Gastroenterol.* **33**, 88–92.

Nicolas, J.-P., and Guéant, J.-L. (1995). Gastric intrinsic factor and its receptor. *Bailliere Clin. Hematol.* **8**, 515–531.

Toh, B.-H., van Driel, I. R., and Gleeson, P. A. (1997). Pernicious anemia. *N. Engl. J. Med.* **337**, 1441–1448.

Pharmacology, Overview

BEVERLEY GREENWOOD-VAN MEERVELD AND KALINA VENKOVA
Oklahoma Foundation for Digestive Research and Veteran's Affairs Medical Center, Oklahoma City

agonist Drug or substance that binds to a receptor on a cell to induce a response.
antagonist Drug or substance that binds to a receptor on a cell to prevent or reverse the response of an agonist.
inflammatory bowel disease A severe inflammatory disorder of the gastrointestinal tract characterized by abdominal pain, rectal bleeding, diarrhea, fever, and severe weight loss.
irritable bowel syndrome A functional disorder of the gastrointestinal (GI) tract characterized by abdominal pain and discomfort associated with altered bowel habits, occurring in the absence of structural or biochemical abnormalities within the GI tract.
neurotransmitter A chemical released by neurons in the brain or peripheral nervous system to communicate with other neurons.
peptic ulcer disease A disorder of the upper gastrointestinal tract (esophagus, stomach, and duodenum), characterized by inflammation and ulceration.
prebiotic Agents that stimulate selectively the growth of bifidobacteria and lactobacilli in the gut.
probiotic A live microbial food supplement that affects the host by improving its intestinal microbial balance.
prokinetics Agents that stimulate the movement of lumenal contents along the gastrointestinal tract.
receptor A protein structure to which neurotransmitters, hormones, and pharmaceutical bind selectively to produce a functional effect.

Disorders of the gastrointestinal (GI) tract are common, unpleasant, and complex, affecting the mucosa, musculature, and innervation from the esophagus to the colon. These disorders are manifested as ulceration, inflammation, obstruction, diarrhea, constipation, and abdominal pain. Over the past few years, although the pharmacological treatment of many GI disorders has improved significantly, the management of GI disorders still continues to pose a significant challenge to clinicians. However, as understanding of the biology of GI disorders expands, the potential for the future development of novel pharmacological agents increases. This article provides a brief pharmacological overview of the approaches to treat a series of important GI disorders.

PEPTIC ULCER AND GASTROESOPHAGEAL REFLUX DISEASE THERAPY

Under physiological conditions, cytoprotective mechanisms exist to prevent intraluminal contents from damaging the mucosal lining of the gastrointestinal (GI) tract. However, under pathophysiological conditions, there is an imbalance between defensive factors, on the one hand, and offensive factors, on the other, leading to mucosal inflammation and injury. Nonsteroidal anti-inflammatory drugs (NSAIDs) are the mainstay for the treatment of musculoskeletal inflammatory diseases, yet they represent one of the major causes of serious GI injury. Traditional NSAIDs are nonselective inhibitors of cyclooxygenase (COX) isoforms COX-1 and COX-2 and destroy the mucosal defense system in part through decreasing the synthesis of cytoprotective COX-1-derived anti-inflammatory prostaglandins in the upper GI tract. GI toxicity associated with NSAIDs includes dyspepsia, gastroduodenal ulcers, and severe gastric bleeding. Newer NSAIDs that act selectively to inhibit COX-2-derived pro-inflammatory prostaglandins, but not COX-1, have significantly reduced GI toxicity. Another major cause of peptic ulcer disease is *Helicobacter pylori* infection. Gastric damage also occurs in critically ill patients due to stress-related mucosal disease, thought to be caused by decreased blood flow, mucosal ischemia, and hypoperfusion–reperfusion injury.

The goals of treatment for peptic ulcer disease include relieving symptoms, increasing healing rate, preventing complications such as bleeding, perforations, and obstruction, and finally, avoiding recurrence. Pharmacological treatment for peptic ulcer disease includes the use of over-the-counter (OTC) antacids to neutralize the pH of gastric contents. Subsequently, histamine type 2 (H_2)-receptor antagonists, such as cimetidine (Tagamet), ranitidine (Zantac), nizatidine (Axid), and famotidine (Pepcid), are used to inhibit acid secretion. Proton pump inhibitors, which include omeprazole (Prilosec), lansoperazole (Prevacid),

 Copyright 2004, Elsevier (USA). All rights reserved.

rabeprazole (Aciphex), pantoperazole (Protonix), and esomeprazole (Nexium), are able to effectively suppress basal and stimulated gastric acid secretion by noncompetitive inhibition of the H^+,K^+-ATPase pump on the parietal cell. Other agents that can be employed to treat peptic ulcer disease include coating agents, such as sucralfate (Carafate), an aluminum salt of a sulfated disaccaride that coats the mucosal lining and allows healing to occur. For those patients who present with NSAID-induced peptic ulcer disease, the use of the synthetic prostaglandin analogue misoprostil (Cytotec) in combination with the NSAID has reduced the incidence of peptic ulcer disease. For patients who test positive for *H. pylori* infection, eradication of the microorganism results in ulcer healing and reduces the risk of ulcer recurrence and complications. Eradication of *H. pylori* is complex, requiring a 10- to 14-day multidrug regimen of antibiotics and acid suppression. However, newer approaches to improve the eradication of *H. pylori* will likely improve patient compliance, be devoid of the potential for antimicrobial resistance, and have a lower cost.

Another common disorder of the upper GI tract is gastroesophageal reflux disease (GERD), which is defined as the excessive backflow of gastric contents into the esophagus. Symptoms of GERD include frequent or daily heartburn, which can progress to esophageal injury. Extraesophageal injury can also occur in association with GERD, resulting in hoarseness, vocal cord granulomas, dental erosions, chronic cough, bronchitis, asthma, and unexplained chest pain. Those at risk for GERD include infants younger than 12 months, though the condition is often self-limiting. Older adults are also at risk for GERD due to prolonged acid exposure over many years. Additional risk factors for GERD include increased acid secretion, use of drugs that reduce lower esophageal sphincter (LES) pressure, abnormalities in esophageal clearance mechanisms, gastroparesis (delayed gastric emptying), hiatal hernia, obesity, spinal cord injury, thyroid disease, Zollinger-Ellison syndrome, diabetes, and connective tissue disorders. The management of GERD requires a stepwise approach, and if lifestyle modifications fail to control symptoms, pharmacological therapy can be added. Drugs used for the management of GERD include OTC antacids to neutralize the pH of gastric contents. However, acid suppression remains the cornerstone of treatment for GERD since it provides adequate symptom relief, restores the quality of life, and prevents many of the potential complications associated with the disorder. H2-receptor antagonists are able to inhibit acid secretion, provide heartburn relief, and can be purchased OTC. Proton pump inhibitors suppress acid secretion and thus provide symptomatic relief for patients with GERD and represent the preferred therapy for the long-term management of patients with erosive esophagitis. Prokinetics (see next section) can be used in patients with GERD to increase LES pressure, promote gastric emptying, and enhance gastroduodenal coordination; however, the side effects of currently available agents often limit their usefulness.

PROKINETIC AGENTS

Disorders such as gastroparesis, postoperative ileus, and pseudo-obstruction are recognized motility disorders that can affect propulsion and cause a delay in intestinal transit. Agents that accelerate GI transit are known as prokinetics and examples include cholinomimetics, such as bethanechol (Urecholine), which causes contraction of the GI musculature through stimulation of the cholinergic receptor (M_2) present on smooth muscle cells. However, the use of bethanechol is not popular since smooth muscle contractions induced by bethanechol are simultaneous rather than peristaltic. Additionally, cholinergic agonists have many side effects, including abdominal pain, diarrhea, salivation, and gastric acid secretion. Newer prokinetics came from a chemical class known as substituted benzamides, which include metoclopramide (Reglan), a dopamine D_2 receptor antagonist, and cisapride (Propulsid), which works through a serotonin receptor mechanism. Both agents enhance gastric emptying and increase intestinal transit through these different mechanisms. However, recent clinical experience has linked cisapride to life-threatening cardiac arrhythmias due to slowing of cardiac repolarization (QT prolongation), resulting in withdrawal of cisapride. This has created a gap in the prokinetic arena and appears to have resulted in the increased use of metoclopramide. Unfortunately, metoclopramide is associated with its own set of serious side effects, including gynecomastia, galactorrhea, fatigue, tremor or rigidity, and tardive dyskinesia. Motilides represent another class of prokinetic agents, so named because they stimulate the motilin receptor present on smooth muscle cells and enteric neurons. Examples include erythromycin, a macrolide antibiotic that can accelerate gastric emptying, enhance coordination between the stomach and duodenum, and stimulate the migrating motor complex. Although recent advances in the area of prokinetics have been limited, a new serotonin receptor agonist, tegaserod (Zelnorm), has been shown to stimulate intestinal transit and relieve the symptoms of constipation.

PHARMACOLOGICAL TREATMENT OF INFLAMMATORY BOWEL DISEASE

Ulcerative colitis (UC) and Crohn's disease (CD), collectively known as inflammatory bowel diseases (IBDs), are immunoregulatory disorders of the GI tract, characterized by an abnormal, intense, and sustained inflammatory response to antigen stimulation. The current hypothesis is that hyperimmune reactivity toward the antigens of enteric bacteria, in a genetically susceptible host, is the basis for disease pathogenesis. Since IBD is characterized by spontaneous "flares" and remission, therapies are directed at treating the active inflammation and maintaining remission. 5-Aminosalicylates are the accepted therapy in the treatment of acute inflammation and maintaining remission in UC. Sulfasalazine (Azulfidine) was the first agent introduced for the treatment of UC and its pharmacologically active moiety is 5-aminosalicylic acid (5-ASA). Since 5-ASA is absorbed in the small intestine, delivery systems enabling 5-ASA to reach the colon before absorption occurs have been established. Specifically, such systems include enteric coatings that dissolve at the alkaline pH present in the terminal ileum or via a prodrug, such as balsalazide (Colozol), which is metabolized by bacterial azoreductases in the colon to release the therapeutically active moiety (5-ASA) and an inert carrier molecule. Topical therapy by enemas, suppositories, or foams has been an advance in the management of distal IBD. For severe active disease, glucocorticoids have clinical importance based on their immunosuppressive capacity against humoral and cell-mediated immune responses. Glucocorticosteroids bind to specific receptors that, on activation, translocate to the nucleus and either increase or decrease the expression of responsive genes. The effects of natural (cortisol) and synthetic [prednisolone (Prelone, Pedoapred, Orapred), dexamethasone (Decadron, Dexone, Hexadrol), budesonide] glucocorticoids involve regulation of the transcription and production of pro-inflammatory cytokines and tumor necrosis factor (TNF). A major limitation in the use of glucocorticoids for chronic treatment is that therapeutic doses are much greater than the daily glucocorticoid production and thus cause significant side effects. A partial solution to this problem is the development of topical-use glucocorticoids, which have low systemic bioavailability due to poor absorption or extensive local or first-pass hepatic metabolism. Topical instillation of budesonide, beclomethasone dipropionate, or prednisolone metasulfobenzoate into the colorectal area has been shown to be effective in patients with active distal IBD. New formulations of budesonide for controlled ileal release (EntocortEC) show high topical efficacy with minimal side effects in the treatment of mild to moderate, active CD. Immunosuppressive drugs with steroid-sparing properties may be beneficial in patients who failed to respond or who develop severe side effects to glucocorticoids. Cytotoxic agents, such as azathioprine (Imuran) and 6-mercaptopurine, have been found to be helpful in patients with CD. Cyclosporine, which suppresses cellular immunity, has been shown to be effective in patients with UC. Methotrexate, a folic acid antimetabolite that is structurally related to interleukin-1 (IL-1) and can interfere with its function, is currently used in patients with refractory IBD. However, the benefits of immunosuppressants are limited by the risk of high toxicity and serious side effects. Recent therapeutic strategies for the treatment of IBD are based on the ability of monoclonal antibodies to neutralize the effect of immune molecules. Infliximab (Remicade), a chimeric monoclonal immunoglobulin G1 antibody that neutralizes TNFα, has been found to have efficacy for refractory or fistulizing CD. Additionally, a therapy based on antibodies to CD4 is under investigation. Specific inhibitors of intercellular adhesion molecules have been designed to eliminate mucosal tissue damage by reducing leukocyte infiltration. Most recently, synthetic compounds (CNI-1493) or small protein molecules (granulocyte/macrophage colony-stimulating factor, bacterial/permeability-increasing protein $rBPI_{21}$) that are normally produced by the immune system have been manufactured by recombinant gene technology and are undergoing clinical trials. Recombinant anti-inflammatory cytokines IL-10 and IL-11, as well as peptide growth factors, are also in clinical development. Despite some promising results, the clinical significance of these new mediator-targeted strategies for IBD treatment is still limited by the inconvenience and high cost imposed by the intravenous or subcutaneous routes of drug administration. Future advances in the pharmacologic approach toward IBD and optimizing the clinical management of patients with IBD will lie in identifying factors predictive of response.

TREATMENT OF DIARRHEA

Diarrhea, the frequent passage of watery/semiformed stool, is not only distressing but may be debilitating or even life-threatening. Acute uncomplicated diarrhea is often self-limiting; however, when diarrhea is chronic, it represents a major health problem and leads to significant dehydration and electrolyte imbalances. Diarrhea can be caused by infectious organisms or malabsorption and can also be drug-induced by agents such as tetracycline, specific antacids, antihypertensive

drugs such as reserpine or guanethidine, and chemotherapy. Diarrhea is also a prominent symptom of IBD. The management and pharmacological therapy of diarrhea depend on its severity. Acute uncomplicated diarrhea is often treated by OTC medications such as loperamide (Imodium), a peripherally acting opiate agent that decreases motility and limits epithelial secretion. Loperamide is efficacious and free of opiate-like and unwanted central nervous system side effects. In patients with chemotherapy-induced diarrhea, loperamide is moderately effective but for those patients refractory to loperamide, evidence suggests that octreotide, a somatostatin analogue, may be helpful in promoting intestinal absorption and relieving diarrhea. A common health problem among travelers to developing countries is diarrhea. The primary objective with pharmacotherapy for traveler's diarrhea is to reduce the symptoms and their duration with a combination of an antimicrobial agent and loperamide. Fluoroquinolones are the first drugs of choice for moderate to severe traveler's diarrhea; however, with the emergence of resistance, especially in *Campylobacter jejuni* ententis, other agents such as azithromycin and rifaximin are being investigated. For more severe forms of diarrhea, such as that seen with cholera infection, oral rehydration therapy can successfully treat the diarrhea. Future therapeutic approaches to prevent episodes of diarrhea may involve the use of prebiotics and probiotics to enhance the presence of bifidobacteria and lactobacilli in the gut to increase the body's natural defense against invading pathogens. However, future studies are needed to determine the efficacy and usefulness of these types of agents for the treatment of diarrhea.

TREATMENT OF CONSTIPATION

Constipation is one of the most common medical complaints in Western countries and is defined as difficulty in passing or straining to pass stool, with less than three bowel movements per week. Constipation often results from disordered motility, leading to a delay in intestinal propulsion and slow transit. Other patients with constipation have a disorder in defecation with normal colonic transit; however, in the majority of cases, constipation results from a diet low in fiber. Constipation can also result from diabetes, pregnancy, and structural diseases of the colon and rectum. Constipation is a very common problem affecting the elderly and often results from polypharmacy since many drugs used by the elderly are known to be associated with constipation. Usually patients with constipation have mild to moderate symptoms and can be treated with increased fluid and fiber intake accompanied by regular exercise. Laxatives are commonly used to treat constipation and most are available OTC. Laxatives exert their effects by increasing stool frequency and stool weight. They are divided into (1) dietary fiber and bulk-forming laxatives, (2) saline and osmotic laxatives, and (3) stimulant laxatives. The dietary fiber and bulk-forming laxatives, such as bran, psyllium, and methylcellulose, increase the water content and mass of the stools, thereby stimulating intestinal peristalsis and decreasing colonic transit time. Saline and osmotic laxatives cause the retention of water within the GI tract because they are poorly absorbed. Examples include magnesium salts, which are charged and thus do not readily cross cell membranes and so remain in the lumen of the GI tract. Magnesium salts have a rapid onset of action and are used for cleansing the bowel prior to radiological procedures. Nonabsorbable sugars, such as lactulose, are also examples of osmotic laxatives, as are sorbitol, glycerin, and polyethylene glycol-based solutions. Stimulant laxatives, such as bisacodyl, senna, cascara, castor oil, and colace, increase colonic peristalsis and act as powerful cathartics. Taken together, however, there is a lack of published evidence for the long-term effectiveness of laxatives for the treatment of constipation. Probiotics have been suggested to improve motility and reduce fecal enzyme activity in patients with constipation. However, large carefully controlled clinical trials are needed to determine the usefulness of probiotic therapeutic approaches for relieving constipation.

THERAPIES FOR IRRITABLE BOWEL SYNDROME

Irritable bowel syndrome (IBS) is a disorder of the GI tract that often begins in early adult life and may affect up to 15–20% of the population; thus, it is associated with major health and economic effects. Currently, IBS is characterized according to symptoms of abdominal pain and discomfort associated with altered bowel function occurring in the absence of structural and biochemical abnormalities. The pathophysiology of IBS is due at least in part to heightened visceral sensitivity, altered intestinal motility, and psychosocial factors, which likely are mediated via alterations in the bidirectional cross talk between the neuronal networks in the brain (central nervous system) and the gut (enteric nervous system). Because functional bowel disorders are multifaceted, drug treatment of patients with IBS is complex; however, the management of IBS in North America was recently reviewed by the American College of Gastroenterology Functional Gastrointestinal Disorders Task Force and evidence-based recommendations were made.

Traditional IBS therapeutic approaches have included the use of antispasmodic agents, such as hyoscyamine and dicyclomine, which are thought to reduce high-amplitude contractions in the GI tract. Anticholinergic side effects, such as dry mouth, blurred vision, and constipation, limit their usefulness. Moreover, there are insufficient data from clinical trials to determine the effectiveness of antispasmodic agents in patients with IBS. Antidiarrheal agents, such as loperamide, have been shown to decrease stool frequency and improve stool consistency in IBS patients; however, there is no demonstrable effect of loperamide on the relief of abdominal pain and bloating. Laxatives, such as fiber, are effective at relieving constipation but there is no demonstrable effect of bulking agents on the relief of abdominal pain and bloating in IBS. There is a limited amount of data suggesting that antidepressant therapy may provide pain relief in IBS patients. However, tricylic antidepressants, such as amitriptyline, desipramine, trimipramine, and doxepin, have not been shown to be more effective than placebo in clinical trials. The effectiveness of serotonin reuptake inhibitors for the treatment of IBS has not been carefully documented, though preliminary evidence is promising.

Recent advances in the understanding of gut physiology have led to the introduction of two novel pharmacological approaches for the treatment of IBS, both targeted toward specific receptor systems, which are likely to improve the treatment of patients with IBS. The 5-hydroxytryptamine type 3 (5-HT_3) receptor antagonist alosetron (Lotronex) has been shown to reduce the symptoms of IBS in female patients with diarrhea. Alosetron has received Food and Drug Administration (FDA) approval for the treatment of women with severe diarrhea-predominant IBS who have failed to respond to conventional IBS therapy. This cautious approach by the FDA is based on incidents of ischemic colitis that have been identified in patients taking alosetron. The efficacy of alosetron is thought to result from a delay in colonic transit and a decrease in abdominal discomfort during colonic distension. The mechanism(s) underlying the effect of alosteron remains incompletely understood; however, evidence suggests that 5-HT_3 receptors are present on afferent nerves. Thus, a reduction in symptoms may occur through a decrease in the activity of central autonomic networks. Another recent therapy designed to treat IBS is tegaserod (Zelnorm), a 5-hydroxytryptamine type 4 (5-HT_4) receptor partial agonist, which possesses promotility effects through the stimulation of intestinal and colonic transit. In preclinical studies, tegaserod reduced afferent firing rates in cats and inhibited visceral sensitivity in rodents. Tegaserod recently received FDA approval for the short-term treatment of women with IBS who have constipation. In clinical trials, tegaserod was more effective than placebo at relieving global IBS symptoms in female patients with constipation. Tegaserod also significantly improved abdominal discomfort, bloating, and constipation. Unlike cisapride, cardiac repolarization is unaffected by tegaserod and thus is not associated with potential life-threatening cardiac arrhythmias.

To improve pharmacological treatment of patients with IBS, a strong scientific rationale exists to examine the effectiveness of other agents targeted toward specific neurotransmitter systems involved in the brain–gut axis. Newer generation anticholinergics that selectively antagonize the muscarinic M_3 receptor, as well as selective antagonists of neurokinin receptors (NK_1, NK_2, and NK_3), corticotropin-releasing factor receptor type 1, somatostatin, and cholecystokinin receptor, may reduce the symptoms of IBS. Finally, as knowledge of the role of intestinal microflora grows, future approaches for the treatment of IBS to reduce pain and abdominal bloating may include dietary supplementation with prebiotics or probiotic bacteria to enhance the levels of specific, beneficial intestinal bacteria, such as *Lactobacillus plantarium 299* and *casei GG* or *Bifidobacterium infantis*. However, to date, the evidence in support of probiotic bacteria for the treatment of IBS remains controversial and requires further investigation.

FINAL COMMENTS

Substantial progress has been made in understanding the effective pharmacological approaches to the treatment of GI disorders. Future research needs to be directed at better defining GI physiology and pathophysiology so that current treatment approaches can be modified and new therapies that possess enhanced efficacy and fewer side effects can be developed.

Acknowledgment

Thanks are extended to the Presbyterian Health Foundation and the Veterans' Affairs Research Foundation for their continued support.

See Also the Following Articles

Antacids • Anti-Diarrheal Drugs • H2-Receptor Antagonists • Laxatives • Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) • Over-the-Counter Drugs • Probiotics • Proton Pump Inhibitors

Further Reading

Berman, S. M., Chang, L., Suyenobu, B., Derbyshire, S. W., Stains, J., Fitzgerald, L., Mandelkern, M., Hamm, L., Vogt, B., Naliboff, B. D., and Mayer, E. A. (2002). Condition-specific deactivation of

brain regions by 5-HT3 receptor antagonist Alosteron. *Gastroenterology* **123**, 969–977.
Bombardier, C. (2002). An evidence-based evaluation of the gastrointestinal safety of coxibs. *Am. J. Cardiol.* **89**, 3D–9D.
Brandt, L. J., Locke, R. G., Olden, K., Quigley, E., Schoenfeld, P., Schuster, M., and Talley, N. (2002). An evidence-based approach to the management of irritable bowel syndrome. *Am. J. Gastroenterol.* **97**, S1–S26.
Campieri, M. (2002). New steroids and new salicylates in inflammatory bowel disease: A critical appraisal. *Gut* **50**, III43–III6.
De Vos, M. (2002). Standard and innovative therapy of inflammatory bowel diseases. *Clin. Exp. Rheumatol.* **20**, S95–S100.
Elson, C. O. (2002). Genes, microbes, and T cells: New therapeutic targets in Crohn's disease. *N. Engl. J. Med.* **346**, 614–616.
Floch, M. H., and Narayan, R. (2002). Diet in the irritable bowel syndrome. *J. Clin. Gastroenterol.* **35**, S45–S52.
Madden, J. A. J., and Hunter, J. O. (2002). A review of the role of gut microflora in irritable bowel syndrome and the effect of probiotics. *Br. J. Nutr.* **88**, S67–S72.
Wingate, D., Phillips, S. F., Lewis, S. J., Malagelada, J. R., Speelman, P., Steffen, R., and Tytgat, G. N. (2001). Guidelines for adults on self-mediation for the treatment of acute diarrhea. *Aliment. Pharmacol. Ther.* **15**, 773–782.

Physiologic Ileus

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

interdigestive state Physiologic conditions in the digestive tract in the time between completion of digestion and absorption of a meal and the ingestion of the next meal.
migrating motor complex A specific pattern of small intestinal motility that begins when digestion of a meal is complete and ends with intake of the next meal. Also called interdigestive motility.
paralytic ileus/adynamic ileus A condition in which the prolonged absence of intestinal motility is pathologic.

Physiologic ileus is the term used to describe the normal absence of motility along a length of intestine. It reflects the operation of one of the neural programs in the spectrum of programs present in the enteric nervous system. Physiologic ileus is a fundamental behavioral state of the intestine in which quiescence of motor function is neurally programmed. For example, in the interdigestive state, the musculature of the small intestine is quiescent except in a limited region where the activity front of the migrating motor complex is present.

NEURAL MECHANISM

The state of physiologic ileus disappears following ablation of the enteric nervous system. In situations in which enteric neural functions are blocked by anesthetics or in patients in which pathological factors have destroyed the enteric nervous system, contractile behavior that is disorganized and nonpropulsive occurs continuously.

Physiologic ileus requires the continuous activity of subpopulations of inhibitory motor neurons that are present in the enteric nervous system. Ongoing discharge of impulses by these neurons releases inhibitory neurotransmitters at neuromuscular junctions and this maintains a state of inhibition in the muscles, which are autogenically active in the absence of neural inhibition. In the absence of ongoing neural inhibition, myogenic pacemakers (i.e., electrical slow waves) evoke contractions with each slow cycle and this results in continuous and uncontrolled contractions. Physiologic ileus refers to the normal state of motor quiescence, whereas the terms paralytic ileus and adynamic ileus refer to conditions in which the prolonged absence of intestinal motility is pathological. Pathological ileus is well known as a consequence of manipulation of the bowel and irritation of the peritoneal membranes during abdominal surgery.

See Also the Following Articles

Basic Electrical Rhythm • Enteric Nervous System • Migrating Motor Complex • Toxic Megacolon

Copyright 2004, Elsevier (USA). All rights reserved.

brain regions by 5-HT3 receptor antagonist Alosteron. *Gastroenterology* **123**, 969–977.
Bombardier, C. (2002). An evidence-based evaluation of the gastrointestinal safety of coxibs. *Am. J. Cardiol.* **89**, 3D–9D.
Brandt, L. J., Locke, R. G., Olden, K., Quigley, E., Schoenfeld, P., Schuster, M., and Talley, N. (2002). An evidence-based approach to the management of irritable bowel syndrome. *Am. J. Gastroenterol.* **97**, S1–S26.
Campieri, M. (2002). New steroids and new salicylates in inflammatory bowel disease: A critical appraisal. *Gut* **50**, III43–III6.
De Vos, M. (2002). Standard and innovative therapy of inflammatory bowel diseases. *Clin. Exp. Rheumatol.* **20**, S95–S100.
Elson, C. O. (2002). Genes, microbes, and T cells: New therapeutic targets in Crohn's disease. *N. Engl. J. Med.* **346**, 614–616.
Floch, M. H., and Narayan, R. (2002). Diet in the irritable bowel syndrome. *J. Clin. Gastroenterol.* **35**, S45–S52.
Madden, J. A. J., and Hunter, J. O. (2002). A review of the role of gut microflora in irritable bowel syndrome and the effect of probiotics. *Br. J. Nutr.* **88**, S67–S72.
Wingate, D., Phillips, S. F., Lewis, S. J., Malagelada, J. R., Speelman, P., Steffen, R., and Tytgat, G. N. (2001). Guidelines for adults on self-mediation for the treatment of acute diarrhea. *Aliment. Pharmacol. Ther.* **15**, 773–782.

Physiologic Ileus

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

interdigestive state Physiologic conditions in the digestive tract in the time between completion of digestion and absorption of a meal and the ingestion of the next meal.
migrating motor complex A specific pattern of small intestinal motility that begins when digestion of a meal is complete and ends with intake of the next meal. Also called interdigestive motility.
paralytic ileus/adynamic ileus A condition in which the prolonged absence of intestinal motility is pathologic.

Physiologic ileus is the term used to describe the normal absence of motility along a length of intestine. It reflects the operation of one of the neural programs in the spectrum of programs present in the enteric nervous system. Physiologic ileus is a fundamental behavioral state of the intestine in which quiescence of motor function is neurally programmed. For example, in the interdigestive state, the musculature of the small intestine is quiescent except in a limited region where the activity front of the migrating motor complex is present.

NEURAL MECHANISM

The state of physiologic ileus disappears following ablation of the enteric nervous system. In situations in which enteric neural functions are blocked by anesthetics or in patients in which pathological factors have destroyed the enteric nervous system, contractile behavior that is disorganized and nonpropulsive occurs continuously.

Physiologic ileus requires the continuous activity of subpopulations of inhibitory motor neurons that are present in the enteric nervous system. Ongoing discharge of impulses by these neurons releases inhibitory neurotransmitters at neuromuscular junctions and this maintains a state of inhibition in the muscles, which are autogenically active in the absence of neural inhibition. In the absence of ongoing neural inhibition, myogenic pacemakers (i.e., electrical slow waves) evoke contractions with each slow cycle and this results in continuous and uncontrolled contractions. Physiologic ileus refers to the normal state of motor quiescence, whereas the terms paralytic ileus and adynamic ileus refer to conditions in which the prolonged absence of intestinal motility is pathological. Pathological ileus is well known as a consequence of manipulation of the bowel and irritation of the peritoneal membranes during abdominal surgery.

See Also the Following Articles

Basic Electrical Rhythm • Enteric Nervous System • Migrating Motor Complex • Toxic Megacolon

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Cullen, J., Caropreso, D., Hemann, L., Hinkhouse, M., Conklin, J., and Ephgrave, K. (1997). Pathophysiology of adynamic ileus. *Digest. Dis. Sci.* **42**, 731–737.

Kalff, J., Schraut, W., Simmons, R., and Bauer, A. (1998). Surgical manipulation of the gut elicits an intestinal muscularis inflammatory response resulting in postsurgical ileus. *Ann. Surg.* **228**, 652–663.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Pica

EDWARD A. ROSE AND ANNE VICTORIA NEALE
Wayne State University, Detroit

geophagia Clay eating.
pagophagia Ice eating.
pica Persistent eating of nonnutritive substances for a period of at least 1 month, without an associated aversion to food.

Pica, the persistent eating of nonnutritive substances, without an associated food aversion, occurs in people of all ages and both sexes, particularly in young children and pregnant women. The term "pica" comes from the Latin word meaning magpie, presumably named after this bird's peculiar eating behaviors involving an indiscriminate preference for both foods and nonfoods. The causes of pica are uncertain, thus treatment is difficult. Although pica is not always dangerous, the underlying possible etiologies should be assessed.

HISTORY

The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) defines pica as the persistent eating of nonnutritive substances for a period of at least 1 month, without an associated aversion to food. The behavior must be developmentally inappropriate and not part of a culturally sanctioned practice, and severe enough to warrant clinical attention. Some clinicians argue that a diagnosis of pica can include the compulsive consumption of certain foods, blurring the distinction between pica and food cravings. Pica is most frequently reported in pregnant women, patients of lower socioeconomic status, and children. It is also found in some cases of iron-deficiency anemia as well as in deficiencies of other nutrients, such as zinc. In some cultures, pica is considered therapeutic and is used in treating maladies such as anemia and anxiety. Interestingly, the range of reported items of consumption has not changed much during the past four centuries. Pica of dirt and clay was known to the Greeks and the Romans and was recorded in a thirteenth century Latin work. Pica was first addressed in a medical book in 1563, in which geophagia was described in pregnant women and in children.

EVALUATION

The cause of pica behavior has eluded researchers for centuries. Researchers have described several theoretical approaches that attempt to explain the etiology from nutritional, sensory, physiologic, neuropsychiatric, cultural, or psychosocial perspectives. There have been few epidemiologic studies detailing the prevalence of pica. Estimates have varied widely within a particular population, depending on the criteria used. Studies of pregnant, otherwise healthy women have found pica in approximately 8% of respondents. Many of these patients had low serum ferritin levels, suggesting a link between pica and iron deficiency.

In the absence of complications that might signal abnormal eating patterns, diagnosis depends on self-reporting. Patients are likely to underreport pica behavior because of embarrassment or because they are not aware that such behavior might be worth reporting.

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Cullen, J., Caropreso, D., Hemann, L., Hinkhouse, M., Conklin, J., and Ephgrave, K. (1997). Pathophysiology of adynamic ileus. *Digest. Dis. Sci.* **42**, 731–737.

Kalff, J., Schraut, W., Simmons, R., and Bauer, A. (1998). Surgical manipulation of the gut elicits an intestinal muscularis inflammatory response resulting in postsurgical ileus. *Ann. Surg.* **228**, 652–663.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Pica

EDWARD A. ROSE AND ANNE VICTORIA NEALE
Wayne State University, Detroit

geophagia Clay eating.
pagophagia Ice eating.
pica Persistent eating of nonnutritive substances for a period of at least 1 month, without an associated aversion to food.

Pica, the persistent eating of nonnutritive substances, without an associated food aversion, occurs in people of all ages and both sexes, particularly in young children and pregnant women. The term "pica" comes from the Latin word meaning magpie, presumably named after this bird's peculiar eating behaviors involving an indiscriminate preference for both foods and nonfoods. The causes of pica are uncertain, thus treatment is difficult. Although pica is not always dangerous, the underlying possible etiologies should be assessed.

HISTORY

The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) defines pica as the persistent eating of nonnutritive substances for a period of at least 1 month, without an associated aversion to food. The behavior must be developmentally inappropriate and not part of a culturally sanctioned practice, and severe enough to warrant clinical attention. Some clinicians argue that a diagnosis of pica can include the compulsive consumption of certain foods, blurring the distinction between pica and food cravings. Pica is most frequently reported in pregnant women, patients of lower socioeconomic status, and children. It is also found in some cases of iron-deficiency anemia as well as in deficiencies of other nutrients, such as zinc. In some cultures, pica is considered therapeutic and is used in treating maladies such as anemia and anxiety. Interestingly, the range of reported items of consumption has not changed much during the past four centuries. Pica of dirt and clay was known to the Greeks and the Romans and was recorded in a thirteenth century Latin work. Pica was first addressed in a medical book in 1563, in which geophagia was described in pregnant women and in children.

EVALUATION

The cause of pica behavior has eluded researchers for centuries. Researchers have described several theoretical approaches that attempt to explain the etiology from nutritional, sensory, physiologic, neuropsychiatric, cultural, or psychosocial perspectives. There have been few epidemiologic studies detailing the prevalence of pica. Estimates have varied widely within a particular population, depending on the criteria used. Studies of pregnant, otherwise healthy women have found pica in approximately 8% of respondents. Many of these patients had low serum ferritin levels, suggesting a link between pica and iron deficiency.

In the absence of complications that might signal abnormal eating patterns, diagnosis depends on self-reporting. Patients are likely to underreport pica behavior because of embarrassment or because they are not aware that such behavior might be worth reporting.

Copyright 2004, Elsevier (USA). All rights reserved.

Clinical suspicion is therefore required to diagnose pica in the ambulatory setting. Iron or zinc nutrient deficiencies may be suggestive; likewise, lead or mercury poisoning may lead to the diagnosis of pica. Gastrointestinal complications such as bezoars or obstructions may also suggest pica.

TREATMENT

Given the difficulty inherent in diagnosing pica and the multitude of possible etiologies, treatment is difficult. Many studies have described diminished pica behavior in patients following iron or zinc replacement to treat low iron or zinc levels, although the empiric evidence implicating zinc deficiency in pica is less convincing than is the evidence for iron. In any case, if iron deficiency leads to pica, then pica behavior should cease once iron is replaced. Cessation of pica behavior with iron replacement may not happen, however, suggesting that continued pica behavior may constitute an addiction or a learned pattern of behavior.

Not all forms of pica are dangerous, and some cases might not require intervention. However, physicians must be prepared for cases of pica in their daily practice. Education about nutrition, along with iron therapy or transfusions, might be the first wave of intervention. Psychological counseling or behavior therapy can also be useful adjuncts. Recently, there has been some evidence that pica is a part of the obsessive–compulsive disorder (OCD) spectrum of diseases. In support of this theory, selective serotonin reuptake inhibitors (SSRIs) have shown some promise. Severe or recalcitrant cases could require referral to a mental health specialist.

Review of the literature on pica confirms just how little is known about this common but commonly missed condition. Its cause is related to many factors, and there are questions about whether pica is a cause of or an effect of metabolic or behavioral states. Accurate diagnosis is hindered by the need for self-reporting on the part of the patient and by a low index of suspicion on the part of the clinician. No specific screening tests for pica exist, but an accurate and timely diagnosis can help to avoid some of the many nutritional, gastrointestinal, and psychological complications. Finally, when pica is diagnosed, there are no proven treatments. Although selective serotonin reuptake inhibitors can be helpful in some cases, diagnosis and treatment must be individualized, and the practicing physician will most likely need to rely on help from mental health professionals.

Acknowledgment

Adapted with permission from the publisher from Rose, E. A., Porcerelli, J., and Neale, A. V. (2000). Pica: Common but commonly missed. *JABFP* **13**, 353–358.

See Also the Following Articles

Bezoars • Foreign Bodies • Psychosociology of Irritable Bowel Syndrome

Further Reading

American Psychiatric Association (1994). "Diagnostic and Statistical Manual of Mental Disorders: DSM-IV." American Psychiatric Press, Washington, D.C.

Parry-Jones, B., and Parry-Jones, W. L. (1992). Pica: Symptom or eating disorder? A historical assessment. *Br. J. Psychiatry* **160**, 341–354.

Rose, E. A., Porcerelli, J. H., and Neale, A. V. (2000). Pica: Common but commonly missed. *J. Am. Board Fam. Pract.* **13**, 353–358.

Picture Archiving and Communication Systems (PACS)

MARC J. GOLLUB
Memorial Sloan-Kettering Cancer Center, New York

analog imaging system Plain radiograph (e.g., chest X ray, mammogram) scan devices; images are not obtained using digital computer technology.
archive Storage of data.
bit Binary unit; smallest unit of information used by a computer; 8 bits equals 1 byte.
computed radiography System using computers to capture and digitize an image rather than print it on film.
digital display protocol Preselected method of displaying a patient's new and old radiographic images.
digital fluoroscopy Capturing and transmitting fluoroscopy information to a computer rather than intensifying it and using it to expose film.
digital images Binary-coded representations that can be interpreted by a computer and displayed on a monitor.
ergonomics Design of environment or equipment related to natural body position and physical comfort.
film digitizer Machine containing a laser that converts analog information on conventional film into digital data, which can be stored in a computer and displayed on a monitor.
film file room Place where X rays are stored.
film screen Combination of silver-coated photographic film layers, sensitive to X rays only, enclosed in a light-proof cassette.
gigabyte One billion (1×10^9) bytes.
graphical user interface Setup that allows humans to interface with software, using words, pictures, and computer tools (keyboard and mouse); a workstation monitor, for example, on which two-dimensional and three-dimensional images are manipulated.
hardware components Computers, monitors, printers, and any other peripheral computer-connected devices.
hospital information system Computerized system containing patient medical and personal identity information.
imaging device Machinery that scans and captures body images (as in computer tomography, magnetic resonance imaging, radiography, and fluoroscopy).
image fusion Digital overlap of data from different types of images.
image noise Information that interferes with interpretation of utilizable data.
imaging plate System analogous to silver nitrate film, but using phosphorescent technology to capture images digitally, rather than converting silver to form an image.
image resolution Degree to which very small objects located in close proximity can be distinguished in an image.
lossless compression Method to compress digital data without losing critical information.
lossy compression Method to compress voluminous data with some loss of information.
network Interconnected computer systems (for example, in a hospital or community health care system).
open architecture Network design that allows interconnected components to communicate using the same standards and language.
operating system Programs (software) that determine the functions and abilities of a computer; disk operating system (DOS) (examples include Windows 95 and Apple DOS).
pixel Picture element.
processing speed Time it takes a computer to accomplish a given task.
radiology information system Computerized system used for scheduling and storing pertinent patient data.
region of interest Area indicated by a circle or square on a cross-sectional image in which the computer is programmed to display a reading that may indicate a solid or fluid component.
storage archiving units Disks that are recorded and read using laser light (examples include read-only memory compact and digital video disks).
telecommunications Transmission of sounds, images, and digital data using electrical and computer technology.
three-dimensional computed tomography angiography Three-dimensional reconstruction of blood vessels, extracting background data from conventional computed tomography images to allow images to appear as if they were simply an angiogram.
three-dimensional rendering Computer reconstruction of two-dimensional data to appear as though three-dimensional.
virtual colonoscopy Three-dimensional images of the colon, simulating an endoscopic view, created using information obtained from a computer tomography scan.

A picture archiving and communication system is a network of computers and imaging devices, databases, storage devices, and display monitors, all of which communicate with one another throughout a hospital, health

 Copyright 2004, Elsevier (USA). All rights reserved.

network, or community to store, transmit, and display digital images and patient information in a rapid and convenient manner. The system eliminates conventional hardcopy film and replaces the conventional film file room and associated personnel, avoiding the inadvertent duplication and loss of information associated with an analog imaging system. A fully networked hospital-wide system facilitates simultaneous image viewing and consultation of radiographic studies by, for example, a surgeon in the operating room, a radiologist in the reading room, and an oncologist in the clinic. The system incorporates the hospital information system and radiology information system for seamless data transmission, billing, and patient scheduling.

BRIEF HISTORY

Great technological advances in the past century that set the stage for the creation of picture archiving and communication system (PACS) have included the development of digital fluoroscopy, digital image systems, computed radiography, high-resolution (2048×2048 pixels $\times$ 12 bits) liquid crystal display flat-screen monitors, personal computers with high-capacity memory, self-contained units using dry film and laser technology, and gigabyte-speed Ethernet wide area networks has enabled systematic linkages of picture archives and communication.

Dr. Paul Capp introduced the concept of digital radiology in the early 1970s. Soon after, the invention of computed tomography (CT) by Godfrey Hounsfield represented a landmark advance in imaging science. CT introduced cross-sectional imaging, but also demonstrated the value of computers in image production. Plain radiography (film screen) and its associated shortcomings, including inability to change the image once the film is exposed, inability to view the film at multiple locations, and inefficiency of manual filing and retrieval of films, are well-known to radiologists. The development of imaging plate (IP) systems (based on photostimulable phosphor technology) in the late 1970s and early 1980s allowed for replacement of film-screen imaging and creation of computed radiography (CR) in 1987. Concurrent with these advances, computer technology evolved rapidly, with increases in processing speed, more powerful operating systems and graphical user interfaces, increased storage capacity, and decreased cost. Finally, the development of the first film digitizer and large-capacity optical storage disk by Kodak further provided essential components for PACS development. Eventually, storage devices, essentially redundant arrays of inexpensive disks (RAIDs, see later), were created.

The first international conference and workshop on PACS for medical applications took place in January 1982, sponsored by the International Society of Optical Engineering (SPIE). At the time, however, much of medical imaging technology had yet to be developed. During the 1980s, Ethernet, a local area network (LAN) communications system via coaxial cable radiofrequencies, became a telecommunications standard, with data transmission rates quickly maturing from 1 megabyte per second to the current 100 megabytes per second (gigabyte Ethernet). Concurrently, data compression was created as an alternative way to speed network delivery. Lossless compression rates of 3 : 1 have evolved toward lossy compression techniques promising rates on the order of 20 : 1 or 30 : 1. The American College of Radiology and National Electrical Manufacturer's Association (ACR-NEMA) group proposed standards for communication of medical images over networks, which became official in 1985. These limited standards were revised in the 1990s into the Digital Imaging and Communications in Medicine (DICOM) standard to embrace traditional computer networks. The saga of PACS development continues as technology improves and various companies have begun to offer fiber-distributed data interfaces (FDDIs), T3 lines (carrying large-bandwidth signals on fiber-optic cable), and asynchronous transfer modes (ATMs), an international standard adopted by the telecommunications industry, promising memory-to-memory transfer rates of 2.4 gigabytes per second.

GENERAL PRINCIPLES

PACS consists of image acquisition devices [e.g., computed radiography/tomography and magnetic resonance imaging (MRI) scanners], storage archiving units (e.g., RAIDs, optical disks, read-only memory digital video disks), display workstations, computer processors, and databases. A communications network and a database management system integrate these components. The hardware components are integrated by standardized, flexible software systems for communications, error handling, database and storage management, job scheduling, interprocessor communication, and network monitoring (Fig. 1).

PACS must include system standardization, open architecture, connectivity, reliability, and security. System standards might include UNIX and Windows NT operating systems, transmission control protocol/Internet protocol (TCP/IP), DICOM communication protocols, structured query language, ACR-NEMA and DICOM image formats, C and C++ programming languages, X Window user interface, American

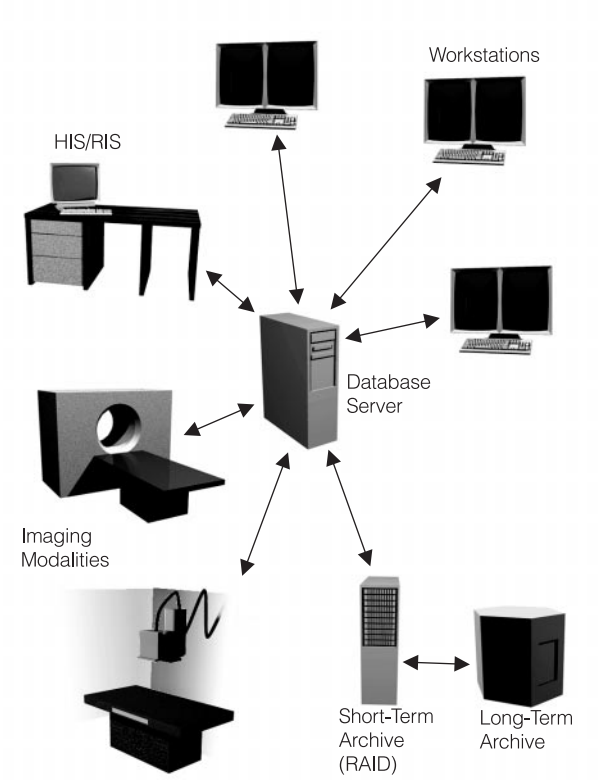


FIGURE 1 An example of a basic PACS configuration. HIS, Hospital information system; RIS, radiology information system; RAID, redundant array of inexpensive disks. © ECRI (2000). Reprinted with permission from *Health Devices* 2000 Nov; 29(11): 387.

Standard Code for Information Interchange (ASCII) text representation for message passing, and health level-7 (HL-7) for health care database information exchange.

Open architecture and connectivity describe a system in which components that are to be connected are able to "speak" to one another in the same "language," so that they are not isolated. They can also be simultaneously upgraded. Reliability (thus high "up time") is critical to patient care. Extended periods of down time cannot be tolerated. Systems must use fault-tolerant measures, including error detection and logging software, external auditing programs, and hardware redundancy. Finally, security is important in the realm of medicolegal and patient confidentiality issues. Violations can include physical intrusion and misuse of privilege.

Image compression is used in PACS to decrease data storage and transfer requirements and to increase data transfer speed. Compression can involve either no loss of information (lossless) or reduction of data with preservation of important information (lossy). Typical acceptable data compression ratios used today are 3 : 1 using the Joint Photographic Experts Group (JPEG) file format, which is DICOM compatible. Wavelet transform can allow higher rates of compression, but is not currently DICOM compatible. Image compression has become a more important issue with the emergence of thinner and more voluminous multislice CT data sets. Megibow *et al.* recently evaluated the ability to diagnose appendicitis, depending on the amount of lossy wavelet compression from 8 : 1 to 24 : 1. Beyond 8 : 1, sensitivity and eventually accuracy (at 24 : 1) decreased, suggesting that finite levels of wavelet transform may be applied to CT images without compromising diagnostic performance.

Using effective networking, many components can be interconnected in PACS. This capability is a key feature of PACS and is facilitated by the DICOM communication standard developed by ACR-NEMA to allow the sharing of information using products from different suppliers. Networks vary according to data transmission methods, architecture of the network hardware, image distribution methods, and image identification and retrieval methods. Typical setups include (1) a regular Ethernet LAN between the image modality device and the image acquisition gateway (low speed; 10 megabits/sec, because the CT scanner is slow to generate images), (2) an Ethernet or FDDI (medium speed; 100 megabits/sec) or ATM (high speed; 155–622 megabits/sec) between the gateway and the PACS controller, and (3) an ATM or FDDI between the PACS controller and the display workstations, because clinicians must access images at a rapid rate.

Network hardware architecture may be centralized (a central location for all images, with direct high-speed links to any workstation at any time), local (tailored to a specific modality, e.g., ultrasound miniPACS), or distributed (in which local modality archives—CT, MRI, and ultrasound (US)—are connected by a high-speed standardized network). Each of these network organizations has advantages and disadvantages regarding cost and demands on the server.

Image distribution methods include on-demand and routed (cached). An on-demand system allows a user to log on and access any image from any workstation (usually in centralized systems). In routed systems, studies are stored locally at a workstation or at specified workstations, but not at all of them. An example of a centralized, on-demand type system (Fig. 2) is the General Electric Pathspeed PACS. Images are sent from an imaging modality to a central database server and then to a central archive (e.g., a RAID) for short-term memory (images are typically copied from the RAID to long-term storage as well). When a user selects a study at a workstation, it is sent from the RAID to the workstation's

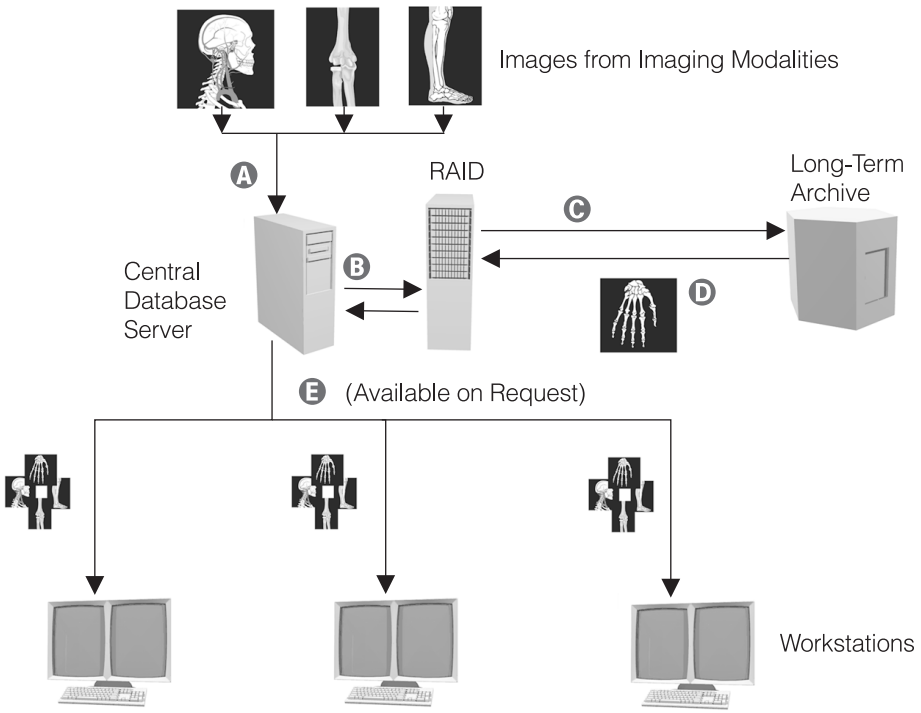


FIGURE 2 Simplified representation of the workflow in a PACS that uses an on-demand image distribution method with a centralized architecture. (A) Images are sent from an imaging modality to a central database server, which in turn (B) sends the images to a redundant array of inexpensive disks (RAID). (C) At this point, images may also be copied to the long-term archive. (D) Images requested from the long-term archive, either those retrieved before the procedure or those requested by the user at the workstation, are also moved to the RAID. (E) Users at any workstation can query the database server to receive any image in the system. © ECRI (2000). Reprinted with permission from *Health Devices* 2000 Nov; **29**(11), 395.

memory. When a user updates a study, the updates are sent back to the RAID (and also copied to long-term storage). Similarly, when a user requests a study from the long-term archive, it is sent from there directly to the RAID and then routed to the user's workstation. This same process occurs in "prefetching," whereby PACS automatically retrieves relevant studies from the long-term archive when a patient is admitted or scheduled at the radiology information system (see later). One drawback of centralized versus distributed systems is that a system failure of the central database disables the whole PACS. However, PACS today have significant built-in redundancy to prevent such system-wide failures.

ANATOMY OF PACS

Input

The database server and archive system together comprise the PACS controller. The controller organizes data from multiple sources into a coherent package. It acquires images, archives them, and distributes them to display workstations and processes image retrieval requests. PACS can also be interrogated in order to retrieve images, review patient and study information, study practice statistics, and perform outcomes analysis.

Images from the acquisition devices (e.g., CT, MRI, and CR scanners) and from devices that convert analog images to digital images (digitizers), as well as pertinent patient information, arrive at the acquisition gateway computer (AGC). The job of the AGC is to receive images from the respective modality components, extract text information describing the received study, update the database management system, determine the destination workstation, automatically retrieve necessary comparison images, determine optimal contrast and brightness parameters for display from lookup tables for cross-sectional modalities (e.g., window and level), perform image data compression, archive new studies onto the optical disk library, and service archive retrieval requests from workstations and other PACS

controllers. The AGC ensures the integrity of image data at various points along the path. At the imaging device, an image will not be deleted from the local storage device until verified by a technologist that it has been archived via the PACS connections. At the AGC, images remain on the local magnetic disk until the archive subsystem acknowledges that a successful archive has occurred, at which time space is made on the disk and these are cleared. At the PACS central node, images arriving in the archive server from various nodes are not deleted until permanently archived.

Other sources of information arriving at the database server of the PACS controller are the hospital information system (HIS) and radiology information system (RIS). The HIS serves three functions: (1) it supports clinical and medical patient care activities in the hospital, (2) it facilitates administration of the hospital's daily business transactions, including financial, personal, payroll, and bed census, and (3) it assists the evaluation of hospital performance, costs, and long-term forecast projections. RIS manages patient demographics, billing, procedure descriptions, scheduling, diagnostic reports, patient arrival scheduling, film location, film movement, and exam room scheduling. It consists of a computer system with peripheral devices such as alphanumeric terminals, printers, and bar code readers. Using the HL-7 standard, scanning a simple bar-code entry can download the relevant data from the RIS to the acquisition unit, which would then incorporate the data into the DICOM packet with the image.

RIS/HIS interfacing occurs using a standard Ethernet. Information and messages concerning patient admission, discharge, and transfer are sent to PACS only when a patient is scheduled for an examination in radiology. An HL-7 standard data format is used running on TCP/IP communication protocols. Prefetching is also initiated as soon as HIS/RIS gets an admission, discharge, or transfer message. Prefetched data are sent to display workstations prior to current exam completion. The prefetched algorithm is based on predefined parameters such as examination type, disease category, radiologist, referring physician, and location of display workstation. Integration of these systems into PACS reduces redundant data entry and storage and avoids the expense and clutter of separate office computers, terminals, and workstations.

Archive

An important advantage of PACS has been the elimination of film. Nonetheless, a busy department can generate many gigabytes (10^9) of data per day and several terabytes (10^{12}) per year. This presents a technological and economic challenge. Digitized images created over several years can be accessed on a variety of storage media, including magnetic and optical disks and tapes, on "jukebox" systems for storage of multiple disks and tapes.

A tiered approach between short- and long-term methods offers two advantages—cost control and maintaining useful retrieval times based on likelihood and immediacy of need; less expensive storage media can be used for long-term storage because the level of functionality required is less than that needed for short-term storage. Short-term storage, usually accomplished using RAIDs, can involve storing studies for only a few weeks; users can specify any length of time. The demand for rapid display of images at the display workstation is well served by RAIDs, whereby multiple disks are accessed to aggregate data fragments. As the cost of RAIDs decreases there is a greater tendency to use this method for intermediate and even long-term storage, allowing images to be retrieved quickly regardless of when they were archived.

To provide both temporary and permanent storage, long-term archive employs a variety of media, including magneto-optical disks and read-only memory compact and digital videodisks. Typically, PACS copies images to both short-term and long-term storage immediately to protect from loss of short-term memory in a system failure. Images in short-term storage are removed once they have been read.

Output

With PACS, film is no longer the medium of the radiologist. The image output is to a display workstation or to an inexpensive paper printer. This reduces costs associated with film, chemicals, film storage, and film management. Different display workstations offer various levels of functionality, resolution, and other display characteristics and can meet the needs of different users. From prior surveys, radiologists have deemed most important the ability to compare new with old studies, and rapid access to images. Certain workstation display features are considered standard, such as zooming, changing window and level, rotating an image, and obtaining density measurements. Each display workstation consists of a host computer and an image display board, a display video monitor, and a local storage device.

Workstations will differ in their physical (spatial resolution, focus, luminance, computation speed, local storage space, bus speed, and network access speed) and functional (software capabilities and human−computer interface) requirements. Important features include ergonomics, lighting conditions,

glare, and acoustic noise from hardware. A key issue determining workstation quality is image resolution. Resolution increases with an increased number of pixels. Bit depth (number of bits/pixel) is another important feature in grayscale depiction. Although 12 bits/pixel is standard for CT/MR and CR for all tissues, US requires only 8 bits/pixel. The development of monitor technology to overcome limitations and inadequacies for the accurate display of mammograms and skeletal radiographs has been a challenge.

Six types of workstations can be described: diagnostic, review, analysis, digitizing and printing, interactive teaching, and desktop. At a diagnostic workstation, radiologists make primary diagnoses using the best resolution possible. This includes 2K ($K = 1024$ linear pixels; 2048×2048 pixel imaging array) monitors for plain radiography (necessary for analysis of fine detail) and 1K (1024×1024 pixels) monitors for CT and MR. At the diagnostic monitors, a digital Dictaphone may be available. A worklist may also be available to include all studies that meet the user's selection criteria, e.g., body-MR cases. The user also can choose how to arrange the images on the monitor, including positioning of old studies, using various digital display protocols (DDPs). The resolution of review workstations in clinics and hospital wards need not be as high, because cases are primarily being reviewed after they have been interpreted, and not every minute detail need be observed.

Analysis workstations are packaged with specialized image analysis software; they can perform, for example, three-dimensional rendering, image fusion, and color enhancement. Models include the GE Navigator software for virtual colonoscopy and Vital Images Vitrea software for three-dimensional CT angiography. A digitizing and printing workstation is used by the film file room personnel to digitize historical films or films acquired outside the hospital. Elimination of film is the ideal objective of PACS, but sometimes hardcopy images will be necessary. This is accomplished with a multiformat camera and laser film, a laser scanner, and a laser or paper printer. Interactive teaching workstations emulate the role of teaching files in the film library, but with more interactive features. Finally, a desktop workstation is used to generate lecture slides and teaching and research material from images on PACS. This is done at a personal computer and requires only a 512×512-pixel monitor.

WORKFLOW SCHEMA

In a film-based department: a procedure is scheduled and patient data are entered in the RIS. Before the patient arrives, the film file room personnel retrieves prior studies for that patient and sends them to the radiologist's reading room. When the patient arrives for the study, the identifying data are reentered into the imaging modality. The patient is scanned and films are exposed at the modality. These are developed and brought to the reading room for posting on a film alternator. These are read while also reviewing prior images. If additional prior imaging studies are needed, the radiologist requests these from the file room. Once they arrive, the radiologist finishes the case and dictates a report. A transcriptionist types the dictated report and sends a printed report to the radiologist, who edits it as necessary. Once revised by the transcriptionist, the report is sent to the referring physician. This whole process may take 3–4 days.

After full PACS implementation, the workflow would proceed as follows: scheduling information and patient data are entered in the RIS and sent to the relevant imaging modality automatically. Before the patient arrives, PACS automatically retrieves prior studies from its long-term archive based on the information in the RIS and the rules established. The images are made available in local storage for ready access. At the scheduled time, the patient is imaged and digital images are transmitted immediately to PACS. These are read along with relevant prior images at a PACS workstation. Images can be easily adjusted for improved viewing or analysis. Radiologists at other locations can be consulted on the same study if necessary. If older studies are needed, they can be accessed from long-term storage within a few minutes. When review is complete, a report is dictated using speech-recognition tools. The radiologist edits the electronic report directly and sends it to the referring physician immediately through RIS or e-mail. This can take several hours or less from start to finish.

ADVANTAGES

General

As a new and evolving costly technology, PACS will need to undergo formal critical evaluation by physicians, the health care system, and society. The methodology for doing this has been set forth by Fryback and Thornbury. Categories of evaluation include safety, technical efficacy, diagnostic accuracy efficacy, diagnostic thinking, therapeutic efficacy, patient outcome efficacy, and societal efficacy.

Widely publicized studies from the hospital of the University of Pennsylvania in the early 1990s demonstrated a statistically significant reduction in time spent

TABLE I Summary of Benefits of the Picture Archiving and Communications Systems[a]

Category	Benefit
1	**Benefits to the diagnostician**
	Improved access to current patient records
	Improved access to patient history records
	File integrity and speed of retrieval
	Better diagnosis
	Quicker diagnosis/improved productivity
2	**Benefits to the referring physician**
	Better patient management/earlier intervention
	Better patient outcome
	Reduced length of stay
	Reduced legal costs arising from maladministration claims, based on loss of films, lack of patient history, etc.
3	**Benefits to the patient**
	Reduced radiation exposure from X-ray equipment
	Shorter examination times
	Reduced radiation exposure as a result of less need for retakes of images
	Reduced patient inconvenience in attending hospitals for examination and reexaminations
	Reduced chance of adverse reaction to contrast agents
4	**Benefits to the hospital**
	Better communication with physicians
	Better hospital administration
	Better training of radiology and other students through access to on-line image files and to digital teaching files
	Greater staff retention due to improved morale

[a] Data reprinted with permission from Crowe (1992).

by clinicians performing clinical activities when images were available on digital display, compared with traditional film retrieval protocols. Imagine the advantages in gastrointestinal radiology if more immediate clinical action could be taken in cases of bowel obstruction, perforation, gastrointestinal bleeding, ischemia, postoperative leaks, and the like. Other studies looking at time savings have shown that nonphysician employee staff can be reduced 26% and that significantly increased time is available to technologists when film handling is eliminated. The overall benefits to the diagnostician, referring physician, patient, and hospital are summarized in Table I. In a practice of gastrointestinal radiology, PACS can facilitate daily practice in both fluoroscopy and cross-sectional modalities.

Gastrointestinal Fluoroscopy

Fluoroscopy is a unique modality that involves the hands-on, potentially time-intensive radiological examination of a patient with repeated and continuous radiation exposure (at the radiologist's discretion). A constantly moving tube (because of peristalsis) is being examined within the patient. Characterization of this is useful in addition to information gained from static images. Intermittent distension and collapse of the bowel and variable progression of administered contrast through the gastrointestinal tract limit the examiner. The window of opportunity to capture time-dependent changes can be missed or capitalized on in a manner quite unlike the radiological examination of any other organ system.

Examining the patient using digital technology and PACS, compared to film, allows more rapid and well-timed exposure and capture of pertinent information because a cassette is not mechanically moving into place. Acquisition of rapid-sequence digital exposures (8/sec) also becomes possible (versus mechanically derived exposures of 100-mm film, limited to 3/sec). Digital exposures can be immediately reviewed on the monitor before fluoroscopy resumes. The adequacy of each film can be immediately assessed, and assessed sooner en masse, compared to cassette film. Immediate supervision of residents can be done for quality control and can minimize the number of errors caused by technically suboptimal exams. The immediate or rapid availability of old studies and correlative modalities enables very quick answers to any queries and allows reexamination while the patient is still in the department; this is a very problem-oriented and tailored approach. Waiting for development of spot radiographs is avoided, eliminating the long delays that may prevent reexamining a patient, only to find suboptimal coating and distension. Finally, compared to conventional fluoroscopic equipment, digital technology allows lower radiation exposure, depending on the number of exposures, the use of pulsed fluoroscopy, and screen capture.

Immediate image review has practical benefits. In cases in which overheads are not required, patients need no longer wait in the department while a series of cassettes is developed in a film processor. They can leave immediately after review of spot films on the workstation. This also frees up the room and the technologist more quickly, further expediting overall throughput. Significant advantages are also available during final film readout sessions at the PACS monitor. Old studies are readily available for comparison. Even old studies, performed at an outside hospital but pertinent to the current fluoroscopic studies, can be digitized in advance for later review. In addition, immediate side-by-side correlative modality comparison (to CT, MRI, PET, etc.) is a useful educational tool in gastroenterology, because the study is often being performed as followup to questionable findings on cross-sectional studies.

This process may also allow demonstration of findings more rapidly.

The software tools available at the workstation monitor allow control of brightness, contrast, magnification, and edge enhancement so that a granular or reticular mucosal pattern or small polyps or ulcers may be more easily seen. Small postoperative leaks and extraluminal gas collections may be made more apparent with these adjustments as well. The scrolling function allows scrolling through captured rapid-sequence studies for a dynamic effect and avoids obtaining more films.

PACS has improved the efficiency of communication and reporting of results to referring clinicians. Placing arrows and annotations to mark significant images allows a quick review of the pertinent findings by clinicians at remote display monitors. PACS makes a positive contribution to teaching and learning. More images can be displayed more quickly without the burden of handling many films. Residents and fellows can scroll through these images to arrive at a better gestalt for the abnormality. Instead of displaying suboptimal copy films on a light box, digital images (which can be manipulated) are automatically accessible and can be viewed and demonstrated in a conference room from a PACS monitor connected to audiovisual equipment. This facilitates better visualization for all residents/fellows in the conference room, rather than the few sitting up front or actually "taking the case."

PACS can also assist in research and publication efforts. It allows more efficient collection of similar cases, by saving them in a teaching folder in the PACS module/worklist. In the academic setting, using a personal computer web browser or workstation, production of slides quickly and without loss of image quality is an added benefit for both teaching files and publication uses. Finally, teleradiology has great potential for aiding consultations and extending expertise around the world.

Cross-Sectional

The contribution of PACS to the capture, transmission, and display of images from cross-sectional modalities has been nothing short of revolutionary. If "a picture is worth a thousand words," then one can barely do justice trying to express in any amount of words the incremental advantages of displaying a 500-"picture" MRI study on soft copy, versus on innumerable pages of laser-printed film. For example, viewing images in a rapid sequential manner on PACS offers a teleological three-dimensional gestalt that cannot be achieved with film reading. In PACS, the imaging specialist has the ability to scroll or cineloop through images and mentally reconstruct anatomy and pathology, based on training and experience. Stated in another way, many structures are anatomically aligned more longitudinally in the body, such that interpretation of axial images gives an incomplete picture of the overall organ or process. Radiologists have all had the experience of deciding if a dot is a lung nodule or lymph node based on whether it is visible on several slices above and below (is longitudinally continuous and therefore a vessel or other tubular structure), or is isolated to that slice (real pathology). Scrolling allows more rapid assessment of longitudinality. When dealing with the alimentary tract, not only is tracing the course of the colon facilitated, but an orderly tracing of the entire small bowel can be nearly achieved. Localizing obstructions and masses, intussusceptions and even accurately identifying the course of bowel and its relationship to adjacent structures become more reliable. The teleologically gifted can often draw a coronal "road map" for the surgeon after viewing the intestines in this manner. The use of specialized display workstations (e.g., GE Advantage Windows 3.1) to reformat images into coronal or sagittal views or create angled multiplanar reconstructions may assist with these tasks as well.

In the solid gastrointestinal tract (liver, gallbladder, pancreas, and intestinal mesentery), PACS offers advantages in image manipulation, archive, volume handling, and image viewing. Manipulation of images may include distance measurements, volumes (tumor burden changes with treatment, liver transplant volume estimation, lung reduction surgery evaluation, etc.), region of interest density measurements, and magnification/minification. In addition, the window and level can be easily adjusted to the observer's liking. Annotations using arrows and text can be added, and significant images can be marked for later reference or remote conferencing with other interested clinicians.

Image archive is now digital. It is not necessary to look through an entire film folder for old/misplaced/jumbled CT examinations. All prior studies are available either in short-term (prefetched) or long-term archive, which may take 1−2 minutes to retrieve, saving time and manual labor. Comparison with other types of studies is also facilitated. This efficient, invaluable process has the potential to help us learn how to better interpret various studies—in a sense, providing personal quality control.

The volume of images in a PACS folder may challenge memory capacity, depending on image compression types and ratios, but still represents a major advantage over film. Filming extra sequences (delayed images, thinner slice reconstructions, coronal or other reformats, etc.) no longer requires using more

film and personnel time, both valuable and expensive resources, but rather merely requires sending the images to PACS.

Viewing images is facilitated with PACS. Any two images from any study and any date can be placed directly next to one another. This can be accomplished with preset digital display protocols. These DDPs can be personalized and saved and shared among different users. The scrolling function, in addition to being useful in distinguishing longitudinally oriented structures from lymph nodes or lung nodules, can also be used to simulate motion in vascular perfusion studies. Finally, the timely arrival of PACS has facilitated the viewing of large data sets associated with virtual colonoscopy.

Some difficulties, disadvantages, and pitfalls have been found with PACS. Several institutions have initially found that cost savings for film attributable to PACS can be largely offset by PACS equipment maintenance costs. On the other hand, cost effectiveness also varies with intangible benefits, such as referring physicians' and support staff productivity gains. Pitfalls in PACS are usually a result of human error, whereas bottlenecks are due to imperfect design. Three errors involving the acquisition device are entering the wrong input parameters, stopping an image transmission process improperly, and incorrect patient positioning (in CR, the most common error occurs when the technologist places the imaging plate under the patient in the wrong direction).

A recent United States Department of Defense survey of PACS deficiencies at 14 sites found that the components most commonly cited for deficiencies were the radiologist's workstation (25%), the HIS/RIS/PACS broker interface (16%), the RIS (14%), monitor displays (10%), and the web-based image distribution systems (6%). Larger systems had more failures than smaller ones.

FUTURE DEVELOPMENTS/SUMMARY

PACS originated as an image management system for improving the efficiency of radiology practice. It has evolved into a hospital-integrated system dealing with information media in many forms, including voice, text, medical records, images, and video recordings. To integrate these various types of information requires the technology of multimedia: hardware platforms, information systems and databases, communication protocols, display technology and system interfacing and integration.

Future developments may include enterprise- and community-wide image distribution systems to balance workload and maximize scarce professional resources, improvements in the quality of service to referring physicians with web-based information distribution systems, wavelet compression techniques, flat-panel monitors, reduced cost of RAIDs, analysis workstations on every monitor, seamless incorporation of HIS/RIS, voice-recognition and computer-automated detection software, and temporal image database management systems to perform outcomes analysis (for example, in lung cancer screening studies by extracting volumes of lung nodules in longitudinal scans).

See Also the Following Articles

Computed Tomography (CT) • Magnetic Resonance Imaging (MRI) • Ultrasonography

Further Reading

Arenson, R. L., Sehadri, S. B., Kundel, H. L., *et al.* (1988). Clinical evaluation of a medical image management system for chest images. *Am. J. Roentgenol.* **150**, 55–59.

Beall, D. P., Horton, M. C., Lewis, T. E., Monroe, B. S., Baker, M. S., Sweet, C. F., *et al.* (2001). Department of Defense picture archiving and communication system acceptance testing: Review of results and identification of problem components. *Radiology* (Suppl.) **221**, 275.

Becker, S. H., and Arenson, R. L. (1994). Costs and benefits of picture archiving and communication systems. *J. Am. Med. Informatics Assoc.* **1**, 361–371.

Blaine, G. J., Cox, J. R., and Jost, R. G. (1996). Networks for electronic radiology. *Radiol. Clin. North Am.* **34**(3), 505–523.

Crowe, B. L. (1992). Overview of some methodological problems in assessment of PACS. *J. Biomed. Comput.* **30**, 181–186.

Dwyer, S. J. III. (1996). Imaging system architectures for picture archiving and communications systems. *Radiol. Clin. North Am.* **34**(3), 495–505.

ECRI. Emergency Care Research Institute. (2000). Picture archiving and communications systems. *Health Devices* **29**(11), 385–427.

Fryback, D. G., and Thornbury, J. R. (1991). The efficacy of diagnostic imaging. *Med. Decis. Making* **11**, 88–94.

Honeyman, J. C., Frost, M. M., Huda, W., Loeffler, W., Ott, M., and Staab, E. V. (1994). Picture archiving and communications systems (PACS). *Curr. Prob. Diagnost. Radiol.* **23**, 102–158.

Huang, H. K. (1999). "PACS. Basic Principles and Applications." Wiley-Liss, New York.

Krohmer, J. S. (1989). Radiography and fluoroscopy 1920 to the present. *Radiographics* **9**, 1129–1153.

Levine, M. S., and Laufer, I. (1998). The gastrointestinal tract: Dos and don'ts of digital imaging. *Radiology* **207**, 311–316.

Lou, S. L., Huang, H. K., and Arenson, R. L. (1996). Workstation design: Image manipulation, image set handling, and display issues. *Radiol. Clin. North Am.* **34**(3), 525–545.

Megibow, A. J., Rusinek, H., Macari, M. J., Bennett, G. L., and Israel, G. M. (2001). CT interpretation utilizing compressed image data: Acute appendicitis as a model for assessing diagnostic quality. *Radiology* (Suppl.) **221**, 377.

Saarinen, A. O., Wilson, M. C., Iverson, S. C., and Loop, J. W. (1990). Potential time savings to radiology department personnel in a PACS-based environment. Proceedings of SPIE Medical Imaging IV: PACS System Design and Evaluation, Vol. 1234, pp. 823–831.

Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)

JOSEPH R. PISEGNA

VA Greater Los Angeles Healthcare System, and David Geffen School of Medicine, University of California

hip (SV1) Incorporates a 20-amino-acid insertion in the third intracellular loop of PAC1.
hip-hop (SV3) Incorporates a 40-amino-acid insertion composed of both the hip and hop cassettes.
hop (SV2) Incorporates a 20-amino-acid insertion in the third intracellular loop of PAC1.
PAC1 Pituitary adenylate cyclase-activating polypeptide type 1 receptor.

Pituitary adenylate cyclase-activating peptide (PACAP) is the most recently discovered neuropeptide in the vasoactive intestinal polypeptide, secretin, glucagon, and related family of peptide hormones. It is so named because, following its discovery in the rat anterior pituitary, this 38-amino-acid peptide hormone was shown to be a potent stimulator of adenylyl cyclase. This hormone was discovered in 1993 and all of its physiological functions are still not completely understood. Given the distribution of this hormone in the peripheral systems, including the gastrointestinal tract, it would be expected to have a major role in physiological regulation.

PACAP HORMONE

Discovery and Gene Structure

Pituitary adenylate cyclase-activating peptide (PACAP) was discovered by Arimura and co-workers using isolated fractions of ovine hypothalamic extracts that stimulated adenylyl cyclase activity in anterior pituitary cells. PACAP circulates as two biologically active forms, one containing 38 amino acids (PACAP-38) and the other containing 27 amino acids (PACAP-27). There is significant interspecies conservation of the amino acid structure of the PACAP hormone.

The primary sequence of the PACAP hormone has a 68% homology with vasoactive intestinal polypeptide (VIP), therefore PACAP is a member of a broader category of hormones that includes VIP, secretin, glucagon, and growth hormone-releasing factor (GRF). As demonstrated in Table I, there is significant homology amongst these peptides. The three-dimensional structure for hormones in this family also demonstrates similarities. The primary amino acid sequence similarity reflects the close similarity in gene structure among the relatives of PACAP, suggesting a similar ancestral gene and perhaps also gene duplication. The PACAP hormone gene has been cloned in several species, including mice and humans. The human gene is composed of five exons and has a structure that is similar to that of other members of this family of peptides. This close similarity suggests that all of the members of this family of peptides may have originated from a similar ancestral gene through duplication.

TABLE I Alignment of Amino Acid Sequences for PACAP and Related Hormones

Hormone[a]	Sequence[b]
PACAP-38	**HSDGIFTDSYSRYRKQMAVKKLAAVLGKRYKQRVKNK-NH2**
PACAP-27	**HSDGIFTDSYSRYRKQMAVKKLAAVLG-NH2**
VIP	**HSD**AVF**T**D**NYTRLRKQMAVKKL**NSILN**K-NH2**
Secretin	**HSDGTFTSELSRLR**EGARLQRLLQGLV**G-NH2**
Helodermin	**HSDAIFTYS**KLLARLALQ**KYLAS**ILGS**R**TSPPP-NH2
Glucagon	**HSQGTFTSDYSKYLDSRRAQDFVQWLMNT-NH2**
GRF	YA**DAIFTNSYSKVLGQLSARKLLQDIMSRQQGESNQERGARARL-NH2**

[a] Abbreviations: PACAP, pituitary adenylate cyclase-activating peptide; VIP, vasoactive intestinal peptide; GRF, growth hormone-releasing factor.
[b] Boldface type designates amino acids with complete identity at those positions of the PACAP-38 hormone.

 Copyright 2003, Elsevier (USA). All rights reserved.

PACAP Hormone Distribution

In mammals, the highest concentration of PACAP is in the central nervous system (CNS). The regions of the brain that appear to have the highest concentration (in the rat) are the paraventricular and supraoptic nuclei within the hypothalamus. It is presumed that PACAP is transported from the hypothalamus to the anterior pituitary, where it can exert its effects on the endocrine system. PACAP-38 immunoreactivity has also been shown to be present in extrahypothalamic sites, such as the substantia nigra, cerebellum, pons, and the paraventricular nuclei of the thalamus. In addition, the spinal cord also contains PACAP that is mainly localized in the dorsal root ganglia and dorsal horn. Outside of the CNS, PACAP is present in the adrenal medulla, where it appears to be a potent stimulator of catecholamine release. PACAP is also present in the enteric neural plexus, where it is an important mediator of gastric acid secretion and intestinal motility.

THE PACAP TYPE 1 RECEPTOR (PAC1)

Cloning, Pharmacology, and Signaling

Through cloning studies, PACAP has been determined to have high affinity for three receptors. The first of these receptors to be cloned was the classical VIP receptor (VPAC1); subsequently cloned were the type 1 PACAP receptor (PAC1) and the VIP2 receptor (VPAC2). Although PACAP has high affinity for all three receptors, each can be distinguished by their affinities for the ligands VIP, PACAP, and helodermin, as shown in Table II (see also Table III for nomenclature of PACAP and receptors). Interestingly, PAC1 has affinity for only PACAP-38 and PACAP-27.

TABLE II Classification of Receptors in the PACAP Superfamily Based on Relative Affinities to Related Peptides[a]

IUP[b] nomenclature	Relative affinities
PAC_1	PACAP-27 = PACAP-38 ≫ VIP > Helodermin
$VPAC_1$	PACAP-27 = PACAP-38 = VIP ≫ Helodermin
$VPAC_2$	Helodermin > PACAP-27 = PACAP-38 = VIP

[a] Adapted from Harmer *et al.* (1998).
[b] IUP, International Union of Pharmacology.

Cloning of rat PAC1 cDNA shows that it is homologous to the VIP and secretin receptors. The PAC1 cDNA encodes a 495-amino-acid protein with a molecular mass of approximately 50 kDa. Cloning of the rat PAC1 gene shows that the receptor can exist as one of four major splice variants (Fig. 1). The major difference between these four potential splice variants is the structure of the third intracellular domain. There is a "hip" splice variant (SV1) and a "hop" splice variant (SV2), each with a 20-amino-acid third-loop insertion; a "hip-hop" splice variant (SV3) incorporates a combined

TABLE III Nomenclature for the PACAP-Related Receptor Superfamily in Comparison to Previous Nomenclature and Hormone Affinity[a]

Receptor type					
IUP[b] nomenclature	Previous nomenclature	Selective agonists	Selective antagonists	Fluorescent agonists	Selective antagonists
PAC_1	PACAP type 1 PVR1	PACAP-38 PACAP-27 Maxadilan?	PACAP 6–38 PACAP 6–27	Fluor-PACAP	PACAP(6–38)
$VPAC_1$	VIP	[Arg^{16}]chicken secretin		Fluo-VIP	[Ac-His^1, D-Phe^2, Lys^{15}, Arg^{16}]
	VIP_1	[$K^{15}R^{16}L^{27}$]VIP (1–7) GRF(8–27)-NH_2			VIP(3–7) GRF(8–27)-NH_2
	PACAP type 2 PVR2 $VIP/PACAP_1$				
$VPAC_2$	VIP_2 PACAP-3 PVR3 $VIP/PACAP_2$	Helodermin Ro 25-1553 Ro 25-1392			

[a] Adapted from Harmar *et al.* (1998).
[b] IUP, International Union of Pharmacology.

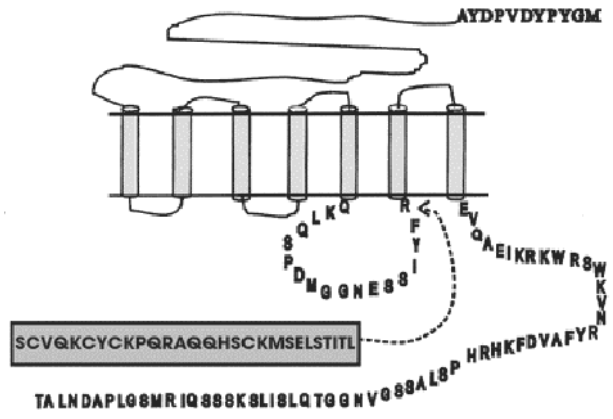


FIGURE 1 Structure of the heptahelical receptor for PACAP, PAC1. The receptor has seven transmembrane domains indicated by the cylinders and a long third intracellular loop. Splice variants are indicated in the shaded rectangle.

40-amino-acid third-loop insertion. These differences in the structure of the third intracellular loop may account for variations in signal transduction coupling to phospholipase C and differences in the tissue distribution of the splice variant. Additional splice variants, affecting either the N-terminus or the second and fourth exons, have also been identified and all of the human PAC1 receptor cDNA splice variants have been cloned. However, unlike the rat splice variants, differences in signal transduction coupling are not observed and the human gene is localized to chromosome 7, whereas in the rat it is localized to chromosome 4. What was observed was a higher efficacy for the hop variant, an intermediate coupling for the hip-hop splice variant, and a lower efficacy of the hip splice variant for coupling to phospholipase C.

The PAC1 receptor, unlike the VPAC1 and VPAC2 receptors, is coupled to a dual signal transduction pathway. A fourth transmembrane splice variant is not coupled to either adenylyl cyclase or phospholipase C yet couples to an L-type Ca^{2+} channel. The region of the native PAC1 shown to be responsible for signal transduction coupling is the COOH terminus, and two amino acids, Ser and Arg, appear to be coupled to signaling.

Localization of PAC1 Receptors

The PAC1 receptor appears to be expressed in both the CNS and peripheral tissues. In the CNS, the greatest density of receptors occurs in the hypothalamus (i.e., the supraoptic nucleus, periventricular nucleus, and lateral hypothalamus), with the predominant splice variant being the null variant without a splice variant cassette. In the retina, PAC1 has been detected and distributed in the inner plexiform layer. In peripheral tissues, the greatest density of PAC1 is in the adrenal medulla, where the predominant splice variant is the hop type; a similar predominance is found in the anterior pituitary. The human prostate gland contains PAC1 receptors that appear to be up-regulated in conditions such as benign hyperplastic prostate. PAC1 receptors are also present on germ cells of the testis as well as on spermatogonia and Sertoli and Leydig cells. PAC1 and VPAC1 receptors are expressed within the gastrointestinal (GI) tract; the PAC1 receptor is expressed on the gastric enterochromaffin like (ECL) cells and the VPAC1 receptor is expressed on the somatostatin-containing D cells and chief cells of the stomach. The liver appears to contain predominantly VPAC1 receptors. The smooth muscles of the GI tract contain VPAC1 and PAC1 receptors, and the PAC1 receptor has been described in the rat tanei coli. The PAC1 receptor has also been found in macrophages, and other PACAP receptor types have been described in a number of tumor cell lines. PAC1 receptor expression has been reported in human lung and breast cancer cell lines. Differences in splice variant expression of PAC1 receptors in pituitary tumors have also been demonstrated.

PHYSIOLOGY OF PACAP

Central Nervous System

Based on the high concentration of PACAP and its receptor, it would be expected that PACAP exerts significant neurophysiological effects. PACAP increases the activity level as well as amount of vasopressin release, and intracisternal administration of PACAP results in the release of gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), prolactin, somatostatin, and the dopamine analogue 3,4-dihydroxyphenylacetic acid (DOPAC). In the pineal gland, there is a high concentration of PAC1, and PACAP can stimulate melatonin secretion. Given its expression in the hypothalamus, another potential action of PACAP is in the control of appetite. In cultured cell systems, PACAP has been shown to activate the c-fos, c-Jun, and mitogen-activated protein (MAP) kinase signaling system, indicating its role in regulating proliferation of cells. In cerebellar granule cells, PACAP appears to reduce apoptosis and have a protective effect against gp120 cultured neuroblasts, again supporting a role for this hormone as a neuroprotective factor.

Endocrine Organs

PACAP regulates the anterior pituitary to release growth hormone (GH), LH, follicle-stimulating hormone (FSH), prolactin, and adrenocorticotropic

hormone (ACTH). PACAP has been shown to stimulate somatotropic cells that release growth hormone in a way that is additive to the effects of growth hormone-releasing factor (GRF). PACAP acts synergistically with GnRH to stimulate the release of LH and FSH. No direct effect of PACAP has been shown on thyrotropic cells of the pituitary. The second major endocrine site of physiological activity is in the male and female reproductive tracts. PACAP has been localized to the smooth muscles of the female reproductive tract, where it plays a role in muscle relaxation. VPAC2 receptors have been identified in placental tissue, which was the site of initial cloning of the VPAC2 receptor. The ovary contains PACAP in the granulosa zone, where PACAP stimulates progesterone production in the preovulatory phase. PACAP and PAC1 receptors are expressed in large numbers in the male gonadal germ cells, and PACAP stimulates testosterone release in the epididymis and is involved in sperm release. A physiological reduction in PACAP may therefore play a clinical role in male impotence. Outside of the CNS, the highest concentration of PACAP is found in the adrenal gland, where PACAP is the most potent stimulator of catecholamine release.

Respiratory Organs

The major effect of PACAP in the respiratory tract is bronchodilation, an effect that is mediated primarily through the VPAC1 receptor. Expression of PAC1 in this system is unlikely.

Gastrointestinal Tract

PACAP-containing enteric nerve fibers are present and colocalized with PAC1 in the stomach and intestine. PACAP is the major neural regulator of gastric acid secretion and may account for the observed nocturnal increase in secretion. VPAC1 receptors are expressed on the surface of the D cell, where PACAP, along with galanin, inhibits acid secretion through the release of somatostatin from the gastric D cell (Fig. 2). Another important effect of PACAP is in the regulation of intestinal motility. PACAP acts mainly to promote VPAC1 receptor-mediated relaxation. In the rat colon, PACAP stimulates apamin-sensitive K^+ channels, leading to relaxation. The major colonic peristaltic reflex is mediated via VIP whereas the descending relaxation phase of intestinal peristalsis appears to be regulated by PACAP through the VPAC1 receptor. Another novel effect that has been discovered is related to nitric oxide synthesis and the interplay between the hormones VIP and PACAP, although the exact mechanism has not been discovered.

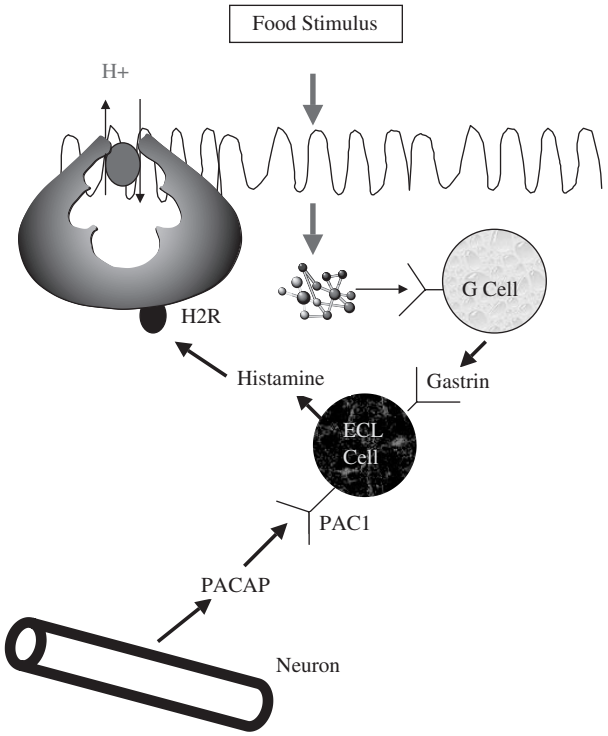


FIGURE 2 Model for the regulation of gastric acid secretion by the hormone PACAP. Released by neurons of the gastrointestinal enteric neural plexus, PACAP binds to the PAC1 receptors of nearby enterochromaffin-like (ECL) cells. PACAP stimulation of ECL cells triggers the release of histamine. Liberated histamine then binds to histamine-2 receptors (H2R) located on the surface of stomach parietal cells, thereby regulating gastric acid secretion.

Cardiovascular System

Similar to the respiratory system, the predominant effects of PACAP in the cardiovascular system occur through smooth muscle relaxation that is mediated through the VPAC1 receptor. In the cardiovascular system, PACAP relaxes smooth muscle through cAMP and protein kinase A, resulting in hypotension. In animals, PACAP administration results in a biphasic effect, with initial vasodilation and a later catecholamine release reflex causing an increase in blood pressure and heart rate. In cultured cardiac myocytes, PACAP exerts positive inotropic and chronotropic effects).

Immune System

The effects of PACAP on cell-mediated immunity have not been thoroughly investigated and only recently has the expression of PAC1 been discovered on lymphocytes. In mice, PACAP stimulates murine macrophages, which in turn stimulate T cell proliferation through VPAC1 receptors, thereby releasing

interleukin-10 (IL-10) and inhibiting IL-6 and IL-12 production. VIP and PACAP inhibit IL-2 transcription in T cells by inhibiting c-Jun. VIP and PACAP have been shown to inhibit nuclear factor κB (NF-κB) by inhibiting p65 nuclear translocation and NF-κB DNA binding. In macrophages, VIP and PACAP inhibit interferon γ (IFN γ)-induced activation of the Jak1/Jak2/STAT/IRF-1 signaling cascade.

Tumor Biology

The majority of the early work on pharmacology and signal transduction relied on tumor cells consisting of the AR42 J rat pancreatic cancer, NB-OK1, human astrocytoma, and PC-12 rat pheochromocytoma cell lines. In these cell systems, PACAP stimulates c-*fos*, c-*myc*, and c-*jun* and is a potent stimulator of cell proliferation. In human lung cancer cell lines, PACAP stimulates growth. Radioligand binding studies demonstrate expression of receptor in a large percentage of human tumors, including those of breast, prostate, pancreas, lung, colon, stomach, and liver as well as lymphomas and meningiomas.

CONCLUSIONS

PACAP is one of the most recently described neuropeptides and therefore very little is known about its function. Its physiological relevance in health and disease is now only beginning to be understood. Recent studies of gene deletion in mice may elucidate further the functional significance of the PACAP hormone and its receptor, PAC1.

See Also the Following Articles

Enteroglucagon • Gastric Acid Secretion • Secretin • Vasoactive Intestinal Peptide (VIP)

Further Reading

Arimura, A. (1992). Pituitary adenylate cyclase-activating polypeptide (PACAP): Discovery and status of current research. *Regul. Pept.* **37**, 287–303.

Delgado, M., and Ganea, D. (2000). Inhibition of IFN-gamma-induced Janus kinase-1–STAT1 activation in macrophages by VIP and PACAP. *J. Immunol.* **165**, 3051–3057.

Delgado, M., and Ganea, D. (2001). Vasoactive intestinal peptide and pituitary adenylate cyclase activating polypeptide inhibit nuclear factor kappa B-dependent gene activation at multiple levels in the human monocytic cell line THP-1. *J. Biol. Chem.* **276**, 369–380.

Ganea, D. (1996). Regulator effects of vasoactive intestinal peptide on cytokine production in central and peripheral lymphoid organs. *Adv. Neuroimmunol.* **6**, 61–74.

Grider, J. R., and Makhlouf, G. M. (1986). Colonic peristaltic reflex: Identification of VIP as a mediator of descending relaxation. *Am. J. Physiol.* **251**, G40–G45.

Harmar, A. J., Arimura, A., Gozes, I., *et al.* (1998). International Union of Pharmacology. XVIII. Nomenclature of receptors for vasoactive intestinal polypeptide and pituitary adenylate cyclase activating polypeptide. *Pharmacol. Rev.* **50**, 265–270.

Koves, K., Arimura, A., Gorcs, T. G., and Somogyvari-Vigh, A. (1991). Comparative distribution of immunoreactive pituitary adenylate cyclase-activating polypeptide and vasoactive intestinal polypeptide in rat forebrain. *Neuroendocrinology* **54**, 159–169.

Leyton, J., Gozes, Y., Pisegna, J. R., *et al.* (1999). PACAP (6–38) is a PACAP receptor antagonist for breast cancer cells. *Breast Cancer Res. Treat.* **56**, 177–186.

Lutz, E. M., Sheward, W. J., West, K. M., *et al.* (1993). The VIP2 receptor: Molecular characterization of a cDNA encoding a novel receptor for vasoactive intestinal peptide. *FEBS Lett.* **324**, 3–8.

Lyu, R.-M., Germano, P. M., Choi, J. K., Le, S., and Pisegna, J. R. (2000). The crucial amino acids within the carboxyl terminus of the PACAP receptor that are involved with signal transduction and receptor internalization: Identification of a conserved amino acid motif for members of the Class II receptor family. *J. Biol. Chem.* **275**, 36134–36142.

Miampamba, M., Germano, P. M., Arli, S., Wong, H. H., Scott, D., Tache, Y., and Pisegna, J. R. (2002). Expression of pituitary adenylate cyclase-activating polypeptide and PACAP type 1 receptor in the rat gastric and colonic myenteric neurons. *Regul. Pept.* **105**(3), 145–154.

Miyata, A., Arimura, A., Dahl, R. R., *et al.* (1989). Isolation of a novel 38-residue hypothalamic polypeptide which stimulates adenylate cyclase in pituitary cells. *Biochem. Biophys. Res. Commun.* **164**, 567–574.

Pisegna, J. R, Leyton, J., Coelho, T., *et al.* (1997). Differential activation of immediate-early gene expression by four splice variants of the human pituitary activating polypeptide receptor: Evidence for an activation by PACAP hybrid and the phospholipase C inhibitor U73122. *Life Sci.* **61**, 631–639.

Pisegna, J. R., and Wank, S. A. (1993). Molecular cloning and functional expression of the pituitary adenylate cyclase-activating polypeptide type I receptor. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6345–6349.

Pisegna, J. R., and Wank, S. A. (1996). Cloning and characterization of the signal transduction of four splice variants of the human pituitary adenylate cyclase activating polypeptide receptor: Evidence for dual coupling to adenylate cyclase and phospholipase C. *J. Biol. Chem.* **271**, 17267–17271.

Rawlings, S. R., and Hezareh, M. (1996). Pituitary adenylate cyclase-activating polypeptide (PACAP) and PACAP/VIP receptors: Actions on the anterior pituitary gland. *Endocrinol. Rev.* **17**, 4–29.

Reubi, J. C., Laderach, U., Waser, B., Gebber, J. O., Robberecht, P., and Laissue, J. A. (2000). Vasoactive intestinal peptide/PACAP receptor subtypes in human tumors and their tissues of origin *Cancer Res.* **60**, 3105–3112.

Schworer, H., Katsoulis, S., Creutzfeldt, W., and Schmidt, W. E. (1992). Pituitary adenylate cyclase-activating polypeptide, a novel VIP-like gut-brain peptide, relaxes the guinea pig taenia caeci via apamin-sensitive potassium channels. *Naunyn Schmiedebergs Arch. Pharmacol.* **346**, 511–514.

Seki, T., Shioda, S., Izumi, S., *et al.* (2000). Electron microscopic observation of PACAP-containing neurons in the rat retina *Regul. Pept.* **21**, 109–113.

Shivers, B. D., Gorcs, T. J., Gottschall, P. E., and Arimura, A. (1991). Two high affinitiy binding sites for pituitary adenylate cyclase-activating polypeptide have different tissue distributions. *Endocrinology* **128**, 3055–3065.

Spengler, D., Waeber, C., Pantaloni, C., *et al.* (1993). Differential signal transduction by five splice variants of the PACAP receptor. *Nature* **365**, 170–175.

Vaudry, D., Gonzalez, B. J., Basille, M., *et al.* (2000). Pituitary adenylate cyclase-activating polypeptide and its receptors: From structure to functions. *Pharmacol. Rev.* **52**, 269–324.

Wang, H. Y., Jiang, X. M., and Ganea, D. J. (2000). The neuropeptides VIP and PACAP inhibit IL-2 transcription by decreasing c-Jun and increasing Jun B expression in T cells. *Neuroimmunology* **104**, 68–78.

Zeng, N., Kang, T., Wong, H., *et al.* (1999). PACAP type I receptor activation regulates ECL cells and gastric acid secretion. *J. Clin. Invest.* **104**, 1383–1391.

Zia, F., Fagarasan, M., Bitar, K., Coy, D. H., Pisegna, J. R., Wank, S. A., and Moody, T. W. (1996). PACAP receptors regulate the growth of non-small cell lung cancer cells. *Cancer Res.* **55**, 4886–4891.

Pneumatosis, Benign and Serious

PUNITA GUPTA AND ROY MCCAULEY
Tufts–New England Medical Center

hematochezia Passage of bloody stools.

lymphangioma Well-circumscribed nodule of lymphatic vessels; the vessels vary in size, are usually greatly dilated, and are lined with normal endothelial cells. They are most frequently found in the neck, axilla, arm, mesentery, and retroperitoneum.

necrotizing enterocolitis Extensive ulceration and necrosis of the ileum and colon in premature infants in the neonatal period; possibly due to perinatal intestinal ischemia and bacterial invasion.

pneumoperitoneum Presence of air or gas in the peritoneal cavity as a result of disease, but can also be artificially produced (i.e., postsurgical intervention).

pseudomembranous colitis Inflammation of the small and large bowels, with the formation and passage of pseudomembranous material in the stools usually secondary to *Clostridium difficile* toxin.

tenesmus Painful spasm of the anal sphincter, causing an urgent desire to evacuate the bowel or bladder and involuntary straining, with the passage of little fecal matter or urine.

volvulus Twisting of the intestine, causing obstruction.

Pneumatosis intestinalis is defined as multiple submucosal, subserosal, or muscularis gas-filled cysts in the wall of the gastrointestinal tract from the stomach to the rectum. Originally, these intramural cysts were referred to as pneumatosis cystoides intestinalis, but because a linear and curvilinear distribution of the cysts also exists, the term "pneumatosis intestinalis" is now used to include all intramural gas collections. The most common locations for these cysts are, in order of prevalence, the jejunum, ileocecal region, and colon. Other structures that can be involved include the mesentery, peritoneum, and falciform ligament.

INTRODUCTION

Pneumatosis intestinalis (PI) is an uncommon condition that has the highest incidence in the fourth through seventh decades of life. PI is usually an unexpected finding on plain abdominal and barium radiographs and is generally a benign condition. In fact, it is the most common cause of benign pneumoperitoneum. However, when associated with fulminant conditions such as hemorrhage or intestinal necrosis, PI has serious implications. In neonates, pneumatosis is frequently associated with necrotizing enterocolitis (NEC). As a result, it is important to recognize that the prognosis of patients with PI is a function of their underlying condition and management should be based on clinical grounds.

PATHOGENESIS

The cause of pneumatosis may be primary or secondary. The primary, or idiopathic, form is not associated with any other lesions and usually does not require

Copyright 2004, Elsevier (USA). All rights reserved.

Shivers, B. D., Gorcs, T. J., Gottschall, P. E., and Arimura, A. (1991). Two high affinitiy binding sites for pituitary adenylate cyclase-activating polypeptide have different tissue distributions. *Endocrinology* **128**, 3055–3065.

Spengler, D., Waeber, C., Pantaloni, C., *et al.* (1993). Differential signal transduction by five splice variants of the PACAP receptor. *Nature* **365**, 170–175.

Vaudry, D., Gonzalez, B. J., Basille, M., *et al.* (2000). Pituitary adenylate cyclase-activating polypeptide and its receptors: From structure to functions. *Pharmacol. Rev.* **52**, 269–324.

Wang, H. Y., Jiang, X. M., and Ganea, D. J. (2000). The neuropeptides VIP and PACAP inhibit IL-2 transcription by decreasing c-Jun and increasing Jun B expression in T cells. *Neuroimmunology* **104**, 68–78.

Zeng, N., Kang, T., Wong, H., *et al.* (1999). PACAP type I receptor activation regulates ECL cells and gastric acid secretion. *J. Clin. Invest.* **104**, 1383–1391.

Zia, F., Fagarasan, M., Bitar, K., Coy, D. H., Pisegna, J. R., Wank, S. A., and Moody, T. W. (1996). PACAP receptors regulate the growth of non-small cell lung cancer cells. *Cancer Res.* **55**, 4886–4891.

Pneumatosis, Benign and Serious

PUNITA GUPTA AND ROY MCCAULEY
Tufts–New England Medical Center

hematochezia Passage of bloody stools.

lymphangioma Well-circumscribed nodule of lymphatic vessels; the vessels vary in size, are usually greatly dilated, and are lined with normal endothelial cells. They are most frequently found in the neck, axilla, arm, mesentery, and retroperitoneum.

necrotizing enterocolitis Extensive ulceration and necrosis of the ileum and colon in premature infants in the neonatal period; possibly due to perinatal intestinal ischemia and bacterial invasion.

pneumoperitoneum Presence of air or gas in the peritoneal cavity as a result of disease, but can also be artificially produced (i.e., postsurgical intervention).

pseudomembranous colitis Inflammation of the small and large bowels, with the formation and passage of pseudomembranous material in the stools usually secondary to *Clostridium difficile* toxin.

tenesmus Painful spasm of the anal sphincter, causing an urgent desire to evacuate the bowel or bladder and involuntary straining, with the passage of little fecal matter or urine.

volvulus Twisting of the intestine, causing obstruction.

Pneumatosis intestinalis is defined as multiple submucosal, subserosal, or muscularis gas-filled cysts in the wall of the gastrointestinal tract from the stomach to the rectum. Originally, these intramural cysts were referred to as pneumatosis cystoides intestinalis, but because a linear and curvilinear distribution of the cysts also exists, the term "pneumatosis intestinalis" is now used to include all intramural gas collections. The most common locations for these cysts are, in order of prevalence, the jejunum, ileocecal region, and colon. Other structures that can be involved include the mesentery, peritoneum, and falciform ligament.

INTRODUCTION

Pneumatosis intestinalis (PI) is an uncommon condition that has the highest incidence in the fourth through seventh decades of life. PI is usually an unexpected finding on plain abdominal and barium radiographs and is generally a benign condition. In fact, it is the most common cause of benign pneumoperitoneum. However, when associated with fulminant conditions such as hemorrhage or intestinal necrosis, PI has serious implications. In neonates, pneumatosis is frequently associated with necrotizing enterocolitis (NEC). As a result, it is important to recognize that the prognosis of patients with PI is a function of their underlying condition and management should be based on clinical grounds.

PATHOGENESIS

The cause of pneumatosis may be primary or secondary. The primary, or idiopathic, form is not associated with any other lesions and usually does not require

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Pulmonary Conditions Associated with Pneumatosis Intestinalis

Chronic obstructive pulmonary disease
Asthma
Cystic fibrosis (with pancreatic involvement)
High-pressure pulmonary ventilation
Chest trauma

treatment. The secondary form has a variety of causes. The majority of cases, which are benign, are associated with chronic obstructive pulmonary disease (e.g., cystic fibrosis) or the immunocompromised state, (i.e., AIDS, high-dose steroid treatment, or chemotherapy). The more serious cases are classically related to conditions that cause bowel ischemia. A variety of related pulmonary and gastrointestinal (GI) disorders are listed in Table I and Table II. In the appropriate clinical setting, these conditions should be considered in the differential diagnosis of pneumatosis intestinalis.

The defining pathogenesis of pneumatosis is unknown, but as shown in Table III, several theories have been proposed. The proposed mechanisms can be summarized as follows:

- Primary: idiopathic cysts tend to resolve spontaneously, but can sometimes recur.
- Secondary:

A. The mucosal disruption hypothesis suggests that gas from the bowel lumen adjacent to the damaged mucosa dissects under pressure along the tissue spaces into the intestinal wall. Distant spread is thought to be through the mesentery. The mechanism for PI in collagen vascular disorders is unclear, but is likely related to vasculitis that eventually results in mucosal disruption.

B. Another mechanism suggests that the intestinal wall is penetrated by gas-forming organisms. For example, in bowel ischemia, there is microscopic infiltration of polymorphonuclear leukocytes with debris containing bacteria. Although this theory is based on strong clinical and experimental evidence, most cases of PI are not related to infection, and most cyst ruptures do not produce peritonitis.

C. Bacterial fermentation of carbohydrates may result in high intraluminal hydrogen tension. This hydrogen diffuses into the intestinal wall and attracts nitrogen, oxygen, and carbon dioxide from the blood. As a result, an intramural bubble, or cyst, is created, consisting of nitrogen as the other gases get absorbed.

D. A theory that explains the pulmonary causes of PI is the dissection of air from alveolar rupture along the bronchopulmonary bundles into the mediastinum, tracking along the major blood vessels into the retroperitoneum, and then along the mesenteric vessels to the visceral surface of the bowel.

E. Increased bowel wall permeability due to loss of structural integrity from the shrinkage of submucosal lymphoid tissue is a mechanism proposed to explain PI in immunocompromised patients.

PATHOLOGY

Other names for PI, cystic lymphomatosis and enteromesenteric emphysema, are based on pathologic appearance. Grossly, the cysts resemble cystic lymphangiomas, hydatid cysts, or sessile polyps. A thin layer of endothelial or simple cuboidal epithelial cell lines the cysts. The diameter of the cysts can vary, if the cysts grow, from a few millimeters to a few centimeters. On cross-section, the cysts can have a

TABLE II GI Conditions Associated with Pneumatosis Intestinalis

Peptic ulcer disease	Whipple's disease
Intestinal obstruction	Volvulus of stomach and sigmoid colon
Mesenteric vascular occlusion	Intestinal lymphosarcoma, leukemia, Hodgkin's disease
Acute necrotizing enterocolitis	Enteric feeding via needle catheter jejunostomy
Chronic inflammatory bowel disease (Crohn's disease, ulcerative colitis)	Jejunoileal bypass for obesity
Hirschsprung's disease	Postsurgical bowel anastomosis
Perforated diverticulum	Abdominal trauma
Appendicitis	Mucosa damage from caustic agent ingestion
Collagen vascular disorders (scleroderma, dermatomyositis, lupus)	After sigmoidoscopy, colonoscopy, or endoscopy
Immunocompromised (intestinal graft-vs.-host disease, organ transplantation, bone marrow transplantation, AIDS, steroids, chemotherapy)	Biliary stent perforation
	Sclerotherapy
Cytomegalovirus infection of the intestinal mucosa	Intestinal dysmotility
Intestinal parasites and tuberculosis	N_2O anesthesia
Diabetic enteropathy	Treatment with lactulose
	Systemic amyloiosis

TABLE III Mechanisms of Pneumatosis Intestinalis

I. Primary
Idiopathic
II. Secondary
A. Mucosal disruption
1. Increased intraluminal gas pressure
2. Vasculitis
B. Intestinal wall penetration by organisms
C. Carbohydrate fermentation
D. Retroperitoneal air dissection from lungs
E. Decreased submucosal lymphoid tissue

"honeycomb" appearance without communication to the bowel lumen. This can be difficult to distinguish from dilated lymphatic channels, which have a similar appearance. The connective tissue surrounding the cysts usually consists of inflammatory changes. A fibrotic reaction can occur if the cysts do not spontaneously resolve, leading to their obliteration. Ultimately, this can lead to a rigid or dysfunctional intestinal wall. Puncture of these cysts during biopsy results in their collapse and sometimes a "popping" sound. Although variable, the pattern of the intramural cysts can aid in distinguishing the cause. A bubbly pattern is most common in necrotic bowel, whereas, in non-life-threatening causes the cysts are usually well localized, larger, and more spherical. They also tend to assume a linear or clusterlike pattern.

SIGNS AND SYMPTOMS

Symptoms are nonspecific and often depend on the location of the cysts and the extent of bowel involvement, especially in the primary form. Colonic cysts are common in the primary form and can result in crampy abdominal pain, hematochezia, mucus in stools, weight loss, diarrhea, tenesmus, and flatus. Steatorrhea has been reported in idiopathic small bowel cysts, but abdominal distension and vomiting are more common when the small bowel is involved. In the secondary form, symptoms are usually related to the associated disease. Cyst rupture, usually from the small bowel, is the most common cause of nonsurgical pneumoperitoneum that is not procedure related. In healthy patients with vague abdominal complaints and free air under the diaphragm, without evidence of a perforated viscous, PI should be suspected.

In 3% of the cases, indications require immediate intervention. These include volvulus, obstruction, hemorrhage, and intestinal perforation. Peritonitis can also occur, but is unusual. When PI complicates intestinal ischemia or pseudomembranous colitis, it should be regarded as an ominous sign. In neonates, one of the serious causes of PI is NEC. These babies often present with emesis, feeding intolerance, abdominal distension, increased gastric residuals, and bloody stools. Nonetheless, the physical exam can be misleading. The presence of abdominal free air can cause the abdomen to be tympanitic. Free air between the liver and diaphragm may obscure normal liver dullness on percussion. If the cysts are large, they may present as a palpable mobile mass. Cysts in the rectal wall can even be felt as firm nodules during a rectal exam.

DIAGNOSIS

Adults

Upright or decubitus plain abdominal radiograph or barium studies and computed tomography (CT) usually initially suggest the diagnosis. On plain radiographs, predominantly linear or segmentally clustered intramural air is seen (see Fig. 1). The cystic pattern does not consistently correlate with the severity of PI, but the location may suggest an underlying condition or predisposing factor. In the gastric wall, linear pneumatosis should raise the question of gastric outlet obstruction or mucosal disruption from prior intubation. A mottled cystic pattern has been associated with phlegmonous gastritis. Small bowel cysts have been associated with collagen vascular diseases. Colonic cysts are common in the idiopathic form, but are also present in the setting of colitis and ischemia. The latter becomes a grave sign when associated with portal venous gas (see Fig. 2).

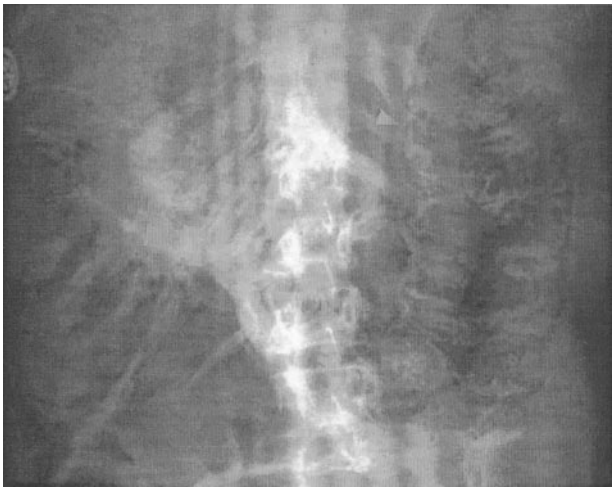


FIGURE 1 Plain frontal radiograph of the abdomen. The arrowheads point to a linear pattern of PI secondary to inflammatory bowel disease in this 22-year-old patient complaining of abdominal pain. After steroid treatment, the patient's pain resolved and she was discharged home. Radiograph courtesy of Dr. Stephen Bloom.

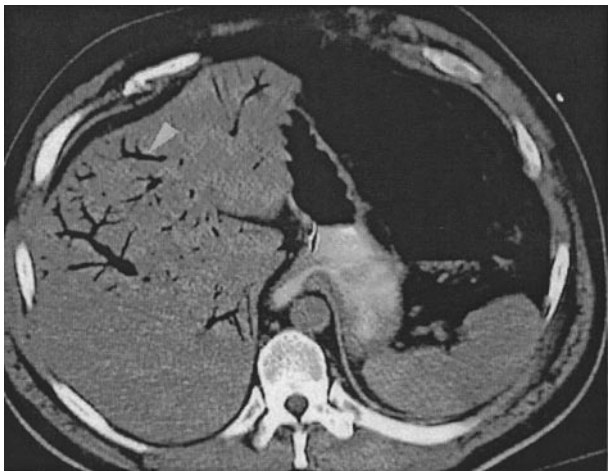


FIGURE 2 Axial abdominal CT image without contrast. Portal venous gas is present throughout the hepatic veins in this 54-year-old female with sepsis (arrowhead). Shortly, after obtaining the CT, the patient continued to decompensate, developed shock, and then expired.

Procedures such as colonoscopy or sigmoidoscopy can result in linear colonic cysts. Air between the bowel loops may represent free air from a ruptured cyst, but a large mesenteric cyst can have a similar appearance. In the presence of isolated PI and pneumoperitoneum, idiopathic cyst rupture or ruptured pulmonary alveoli from prior general anesthesia administration should be considered. Pulmonary causes of PI also tend to be transient. On barium studies, radiolucencies indent the barium column as they line the intestinal wall, resulting in a smooth and undulating marginal pattern. CT with lung windows is probably the most sensitive tool in evaluating PI (see Fig. 3). Ultrasound (US) is another modality that can be used to make the diagnosis. On sonography, the intramural bubbles are seen as hyperechoic structures with shadowing, sometimes referred to as the "circle sign." Magnetic resonance imaging (MRI) can also aid in the diagnosis.

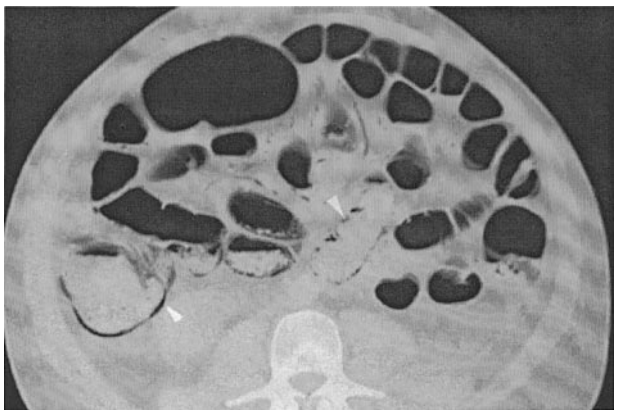


FIGURE 3 Axial abdominal CT image with contrast. Pneumatosis is present throughout the dilated large and small bowel (arrowheads) in this 45-year-old male with ischemic bowel secondary to infection. The patient was admitted and treated with antibiotics. Clinically, the patient improved and was eventually discharged home.

Pediatrics

NEC is a common GI emergency in neonates. The most common risk factor is prematurity. PI is present in 75% of patients with NEC and should be considered a diagnostic sign in the appropriate clinical setting. Although rare, patients with NEC can also have air in the gastric wall, known as pneumatosis gastralis, or emphysematous gastritis. No individual laboratory features are diagnostic or specific for NEC, which is why abdominal radiographs remain essential in the diagnosis and management of NEC (see Fig. 4). The presence of PI is considered a relatively early stage of NEC and can precede the clinical findings by several hours. The bubbly intramural gas collections can be confused with stool or meconium in a normal colon. Prone views or followup radiographs can help to make the distinction. In NEC, PI most commonly involves the terminal ileum and colon. It is usually associated with dilated bowel loops and an asymmetric bowel gas pattern. The cystic intramural

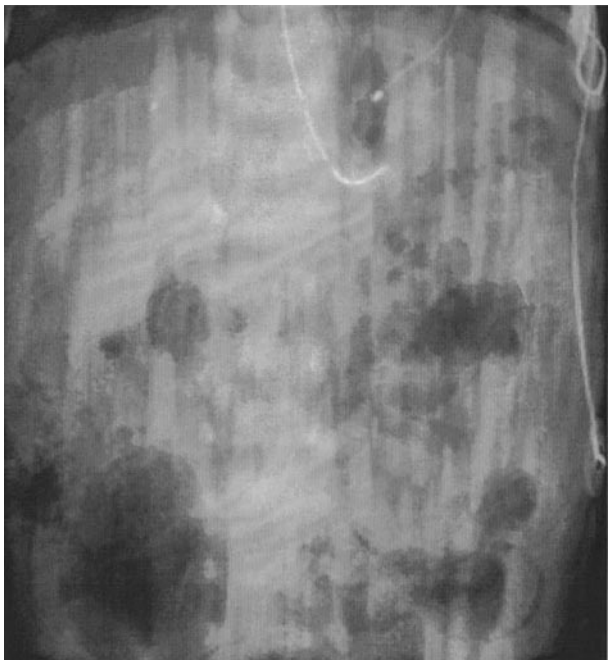


FIGURE 4 Plain frontal radiograph of the abdomen. This premature infant presented with abdominal distension. Note the diffuse pneumatosis throughout the bowel wall. The baby was treated for NEC.

pattern may present as linear stripes, a ring, or a localized cluster. If it occurs in a more diffuse pattern, it should be a marker for a more serious condition. Ileal involvement of PI can also simulate meconium ileus. Portal venous gas is usually associated with the more severe cases of NEC and heralds surgical intervention. Sonography of the portal vein during NEC will often show bubbles of gas traveling into the liver, where they collect anteriorly, giving a characteristic echogenic loss of definition of the normal liver structure.

Colonic PI has also been associated with Hirschsprung's disease. The presence of perinatal PI should also raise the question of congenital cardiac anomalies such as hypoplastic left heart. Patients with PI who are more likely to have a poor outcome are those with underlying congenital heart disease and tissue transplants. Graft-versus-host disease, colitis, and bowel ischemia are also ominous preceding events. Contrast enemas, which are not routinely performed because of the risk of perforation, especially in NEC, show mucosal irregularity and edema. Additional causes of pediatric PI include complications of chemotherapy for leukemia and lymphoma and cardiac surgery (as listed in Table IV). The disappearance of PI or portal venous gas does not always correlate with clinical improvement. The only universally agreed indication for surgery when clinical suspicion of NEC is high is pneumoperitoneum, which is not always present in all babies with a perforation.

Several conditions may be included in the differential diagnosis of PI. However, the presence of specific secondary findings helps to differentiate PI. For example, unlike PI, enterogenous cysts are usually single. Nodules of lymphosarcoma may clinically resemble PI, but on imaging studies they are not radiolucent. The reversible thumbprint-like deformity of the mesenteric border in bowel ischemia can sometimes be misinterpreted as PI. Infectious etiologies or pseudomembranous colitis should be correlated with the clinical presentation. Mucosal polyps on barium studies result in intraluminal filling defects as opposed to the smooth, scalloped margins of PI. Crohn's colitis and ulcerative colitis also result in barium filling defects, but without cystic lucencies. It is important to realize that any of the related conditions mentioned here can be included in the differential assessment of PI, but it is the clinical setting that should guide the correct diagnosis.

TABLE IV Pediatric Conditions Associated with Pneumatosis Instestinalis

Necrotizing enterocolitis (usually involves the terminal ileum, colon)
Hirschsprung's disease
Congenital cardiac anomalies (i.e., hypoplastic left heart)
Graft-versus-host disease
Ischemic bowel
Postchemotherapy for leukemia and lymphoma
Postcardiac surgery

MANAGEMENT AND CONCLUSION

PI is generally an asymptomatic benign entity, but on rare occasions can also be serious. In the absence of intestinal inflammation or necrosis or a high white blood count, and with a benign physical exam, conservative treatment is safe. Hyperbaric oxygen to decompress the cysts and antibiotics have demonstrated some efficacy, but watchful management can be sufficient. When a life-threatening complication such as bowel perforation, obstruction, hemorrhage, or intestinal ischemia occurs in the setting of PI in the adult or pediatric population, immediate surgical intervention may be crucial to the patient's outcome.

See Also the Following Articles

Cystic Fibrosis • Necrotizing Enterocolitis

Further Reading

Buonomo, C. (1999). Neonatal imaging. *Radiol. Clin. North Am.* **37**(6), 1187–1198.

Hoer, J., Truong, S., Virnich, N., Fuzesi, L., and Schumpelick, V. (1998). Pneumatosis cystoides intestinalis: Confirmation of diagnosis by endoscopic puncture a review of pathogenesis, associated disease and therapy and a new theory of cyst formation. *Endoscopy* **30**(9), 793–799.

Longo, W., Ballantyne, G., and Graham, A. (1987). Pneumatosis cystoides intestinalis. *J. Clin. Gastroenterol.* **9**, 571–573.

Meyers, M., Ghahremani, G., Clements, J., and Goodman, K. (1977). Pneumatosis intestinalis. *Gastrointest. Radiol.* **2**, 91–105.

Mularski, R., Sippel, J., and Osborne, M. (2000). Pneumoperitoneum: A review of nonsurgical causes. *Crit. Care Med.* **28**(7), 2638–2643.

Pear, B. L. (1998). Pneumatosis intestinalis: A review. *Radiology* **207**, 13–19.

Davila, A. D. and Willenbucher, R. F. (1998). Pneumatosis cystoides intestinalis. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., pp. 1979–1982. W. B. Saunders, Philadelphia.

Evers, B. M. (2001). Pneumatosis intestinalis. *In* "Sabiston Textbook of Surgery" (C. M. Townsend, ed.), 16th Ed., pp. 910–911. W. B. Saunders, Philadelphia.

Shimada, M., *et al.* (2001). Pneumatosis cystoides intestinalis treated with hyperbaric oxygen therapy: Usefulness of an endoscopic ultrasonic catheter probe for diagnosis. *Int. Med.* **40**(9), 896–900.

Virginia Commonwealth University Health Systems. (2003). Department of Radiology web site: The spectrum of pneumatosis intestinalis in children. Retrieved at www.radiology.vcu.edu/pneumo/pneumatosis.htm.

Polyamines

SHIRLEY A. MCCORMACK AND LEONARD R. JOHNSON
University of Tennessee Health Science Center

antizyme An enzyme homologue of ornithine decarboxylase capable of binding and inhibiting it.
apoptosis An endogenous program of cellular reactions that results in cell death without allowing toxic degradation products to be released into the cellular milieu.
confluent Cells in culture at high density and in contact with one another so that there are no empty spaces.
cytoskeleton The inner framework of the cell. It maintains and adapts cell shape, makes directed migration possible, and provides strength and organization for cellular functions.
dimer A molecule made up of two identical units.
down-regulating Reducing the level of a product by inhibiting its synthesis, degrading it, or transporting it out of the cell.
filopodia Narrow spike-like extensions of cell borders in cells that are not necessarily migrating.
focal adhesions Concentrated patches of stress fibers, associated proteins, and integrin receptors on the plasma membrane by which integrins attach to the extracellular matrix.
lamellipodia Wide wave-like extensions of the cell border in migrating cells.
lamina propria A layer of mostly fibroblastic cells that forms the extracellular matrix for epithelial cell attachment. It lies between the epithelium and the muscularis.
neoplastic Rapidly growing cells that are not necessarily transformed.
organelles Small intracellular "organs" that perform specialized tasks.
overexpression Production of excessive amounts of a product by cells because extra copies of the gene encoding the product have been introduced into the cell chromosomes.
26S-proteosome A specialized lysosome in cytoplasm that degrades discarded molecules.
subconfluent Cells at low density that have not yet formed a continuous layer.
transfected Cells with artificially introduced genes in their chromosomes.
transformed Cells with altered DNA that have lost normal regulatory control.
villi Finger-like projections of mucosa covered with differentiating epithelial cells that increase the absorptive area of the intestine.

Although polyamines are required in a large, but still unknown, number of the most basic functions in all forms of eukaryotic and prokaryotic life, an understanding of their participation in the life of the cell has been many years developing. With every discovery of another biological function that requires them, an entirely new interdependent set of cellular reactions previously unsuspected is found. This article attempts to describe the current knowledge of polyamines in the cellular physiology of the gastrointestinal mucosal epithelium.

INTRODUCTION

A Brief History of Polyamine Discovery

During the period of the Scientific Revolution, Antonie van Leeuwenhoek, using an early microscope, began to explore the previously invisible world of various plant and animal samples. In 1678, he wrote a now famous letter to the Royal Society of London in which he reported observing sperm as well as a slowly crystallizing substance in human semen. Two hundred forty-six years later, the "slowly crystallizing" substance was identified as spermine phosphate and the letter became the earliest documented record of a polyamine.

As chemistry evolved from alchemy, the gases oxygen, carbon dioxide, and nitrogen were fractionated from air. Methods were developed for their analyses and studies of oxidized and reduced nitrogen (ammonia) followed. In 1838, F. Wohler converted ammonium cyanate to urea. His studies led to the detection of nitrogen in more complex compounds such as uric acid, allantoin, and asparagine and established organic chemistry as a new field of chemistry. Organic chemistry, in turn, revealed the structures of the amines and polyamines and resulted in the development of other branches of chemistry, i.e., analytical chemistry, pharmacology, biochemistry, microbiology, and immunology. The beer and wine industries, searching for new methods to improve fermentation and avoid spoilage, isolated and identified the basic amino acids arginine, lysine, and ornithine and other amino acids soon followed. Other key discoveries were the structural role of nitrogen, an ornithine cycle in mammalian liver, and specific amino acid decarboxylases and oxidases. Spermine, cadaverine, putrescine and spermidine were

 Copyright 2004, Elsevier (USA). All rights reserved.

identified in 1881, 1886, 1889, and 1927, respectively. After that, further progress toward understanding their biological functions stalled until the 1960s when biochemists began to investigate the structure of polyamines in relation to their effects on organ function, animal metabolism, and nutrition. Research on polyamines has greatly accelerated in recent years. Over the period 2001 to 2002 alone, the National Library of Medicine lists more than 1000 reports dealing with polyamines.

The authors have borrowed much of this short history of polyamines from their first observation to the period ending with World War II from a detailed timeline published as an appendix in the 1998 book *Polyamines* by Seymour S. Cohen (see Further Reading).

The Distribution of Polyamines in Nature

Originally, biologists thought that spermine was found only in higher eukaryotes and that prokaryotes had only putrescine and spermidine. As an increasing variety of microbes, plants, and animals were investigated, many exceptions appeared. For example, in prokaryotic thermophilic microbes, nucleic acids are distorted by hydrogen bonding at the temperatures at which they live (65–75°C) and proteins can be synthesized only when the nucleic acids are bound to spermine. Some halophiles living in very-high-salt environments lack polyamines entirely. Eukaryotic photosynthetic algae, yeast, and slime molds contain putrescine and spermidine but often do not contain spermine. Spermine is absent in protozoa and shrimp but present in mollusks, arthropods, echinoderms, and tunicates. Unusual polyamines, i.e., hydroxypolyamines, branched tertiary polyamines, and quaternary polyamines, are found in prokaryotic thermophilic microbes and in some eukaryotic animals. Even more unusual are two eukaryotes, the plant *Arabidopsis thaliana* and the protozoan *Trypanosoma cruzi*, neither of which has ornithine decarboxylase (ODC), the rate-limiting enzyme that controls the biosynthesis of polyamines. Putrescine, spermidine, and spermine are present in all cells of all vertebrates.

BIOCHEMISTRY

Polyamine Structure

The four basic polyamines in mammals and in human are putrescine, cadaverine, spermidine, and spermine. Putrescine and cadaverine are primary diamines, 1,4-diaminobutane, and 1,5-diaminopentane, respectively. Cadaverine is usually a product of bacteria in the gut. Strictly speaking, the diamines are not polyamines, but they are often included with the polyamines for the sake of convenience. Spermidine is a triamine and spermine is a tetramine. Both contain primary and secondary amines (R_1R_2NH). Other natural polyamines that include tertiary (R_3) and quaternary (R_4N^+OH) amines have been found as have moieties with one carbon less than the usual 1,4-diaminobutane, called norspermidine or norspermine.

Biosynthesis of the Polyamines

Three key enzymes, all with short half-lives, are responsible for the biosynthesis of the natural polyamines, putrescine, spermidine, and spermine (see Fig. 1). The three enzymes are ODC, *S*-adenosylmethionine decarboxylase (AdoMetDC), and spermidine/spermine acetyltransferase. ODC is the rate-limiting enzyme because, by decarboxylating ornithine to synthesize putrescine, it provides both an active polyamine and a precursor of other polyamines. AdoMetDC decarboxylates *S*-adenosyl methionine to make spermidine by transferring its aminopropyl moiety to putrescine and, by another aminopropyl transfer to spermidine, to make spermine. Spermidine/spermine acetyltransferase regulates polyamine interconversion and allows cells to adjust the levels of the three polyamines. The higher eukaryotes also have another pathway by which, with acetylation and oxidation, spermine can be converted to spermidine and spermidine to putrescine. Spermidine synthase, spermine synthase, and polyamine oxidase, an enzyme that oxidatively splits the monoacetyl derivatives of spermidine and spermine, are usually present in excess and are not rate-limiting under physiological conditions. In nontumorigenic rapidly dividing cells such as embryonal cells and gut mucosal cells, ODC synthesizes transient high levels of putrescine. In nonproliferating cells, AdoMetDC provides a basal level of putrescine by conversion from spermidine. Research on the structure of ODC has shown that it is a dimer of 52 to 55 kDa subunits and is well conserved in eukaryotes from fungi to humans. In addition, it has been crystallized and found to be a group IV pyridoxal phosphate-dependent enzyme.

POLYAMINE REGULATION

How Is ODC Regulated?

The response of ODC to hormones or growth factors in tissues and cultured cells is typically a rapid 10- to 100-fold increase in ODC activity followed by an equally rapid decrease even before the cells begin DNA synthesis. The speed and extent of these responses were unexpected and investigators have wondered

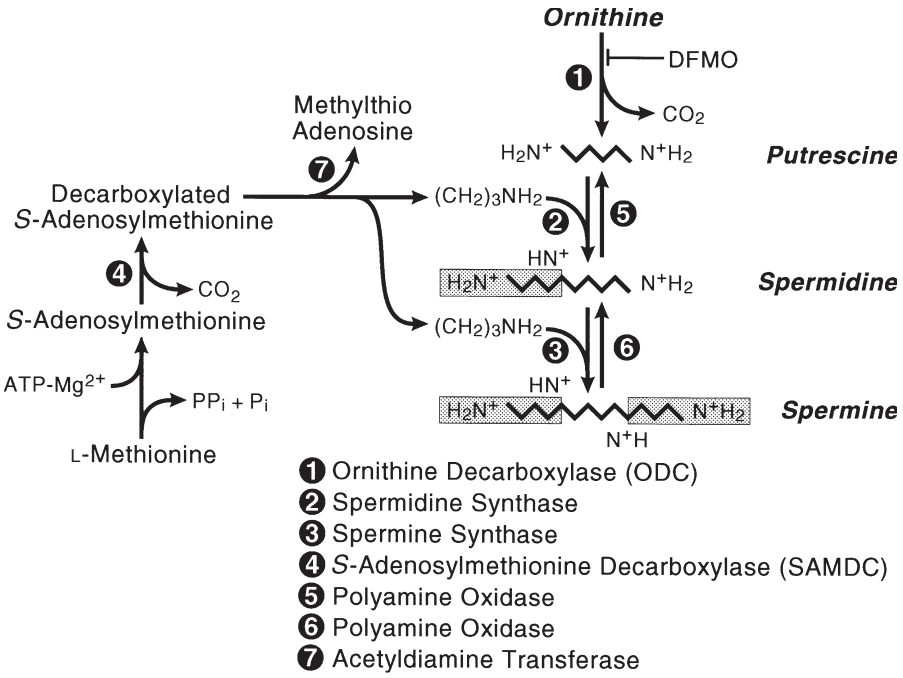


FIGURE 1 The biosynthesis of the major physiological polyamines and their precursor, putrescine. The polyamines are shown in shorthand representation, i.e., putrescine is $H_2NCH_2CH_2CH_2CH_2NH_2$. Thus, putrescine has 4 CH groups, spermidine has 7, and spermine has 10. They carry two, three, and four positive charges that reflect the strength of their ability to bind to anions. The enzymes involved in the reactions are numbered. Reprinted from McCormack, S. A., Ray, R. M., and Johnson, L. R. (2002). Polyamines in intestinal epithelial restitution. *In* "Gastrointestinal Mucosal Repair and Experimental Therapeutics" (C.-H. Cho and J.-L. Wang, eds.), 3rd Ed., Vol. 1, pp. 43–56. Karger, Basel, Switzerland, with permission.

how the sudden decreases in ODC and polyamines are controlled especially since prolonged high levels of polyamines are toxic to cells. Several regulatory processes for the polyamines have been proposed, i.e., negative feedback and repression, antizymes, and segregation by macromolecules. Negative feedback and repression is a process common to many enzymes. In the case of ODC, negative feedback is mediated through translational expression at the 3′-untranslated region of ODC mRNA. Antizymes, analogues and suicide inhibitors of ODC, are novel endogenous ODC inhibitors. These inhibitors are unique in two respects. They target the enzyme ODC for degradation (most enzymes are targeted by ubiquitin) and they are induced by rising levels of ODC's product, spermidine. Also, in addition to down-regulating ODC, antizymes inhibit polyamine uptake and stimulate its outward transport. Finally, some investigators have suggested that negative feedback may not function in the low levels of free polyamines. Most of the polyamines produced at any one time are quickly segregated because they are strongly basic due to their amine groups and bind to proteins, DNA, RNA, and other negatively charged macromolecules. The free polyamines remaining can be regulated by antizymes. Some investigators have suggested that ODC may be regulated for the control of another as yet unknown reaction. If this should be demonstrated, it would open new and exciting possibilities for polyamine research. At present, the most obvious and only proven role of ODC is to provide a biosynthetic route for putrescine and cadaverine.

Synthetic Inhibitors of Polyamines

The rapid rise and fall of the polyamine biosynthetic enzymes ODC and AdoMetDC intrigued biochemists when they observed it during the 1960s. When polyamines were found in large quantities in the tumors, blood, and urine of cancer patients, biochemists began to synthesize inhibitors in the hope that they could lower the levels of polyamines and stop the growth of the tumors. α-Difluoromethylornithine (DFMO) and methylglyoxal-bis-guanylhydrazone (MGBG) emerged as promising inhibitors.

DFMO is one of the most useful of the synthesized inhibitors of ODC because it irreversibly and

specifically inhibits ODC, blocks the production of putrescine from ornithine, inhibits or reduces any cell functions that require putrescine, is relatively nontoxic, and is easily dissolved in aqueous medium. Experiments with animals and cells in culture have shown that the specificity of polyamine deficiency can be demonstrated by supplying a polyamine at the same time as the DFMO because the combination can maintain any inhibited function at near normal levels. DFMO is widely used experimentally to study the effects of polyamine depletion on cell growth, migration, attachment, the actin cytoskeleton, signaling, etc., in culture and *in vivo*.

MGBG is an effective inhibitor of spermidine synthesis by AdoMetDC. However, it is more toxic than DFMO, may reduce the availability of methionine for protein synthesis, and, unexpectedly, has been shown to increase polyamines to very high levels in the brain. Both MGBG and DFMO have been used in cancer therapy and, in some cases, have caused improvement. Researchers continue to synthesize new inhibitors in the hope of improving their effectiveness. Unfortunately, reducing polyamines significantly in the gastrointestinal tract is difficult even when polyamines are severely restricted in the diet because they are contributed by food, bacteria, and sloughed cells in the lumen.

POLYAMINE ROLES IN ESSENTIAL PHYSIOLOGICAL PROCESSES

Cytoskeletal Maintenance

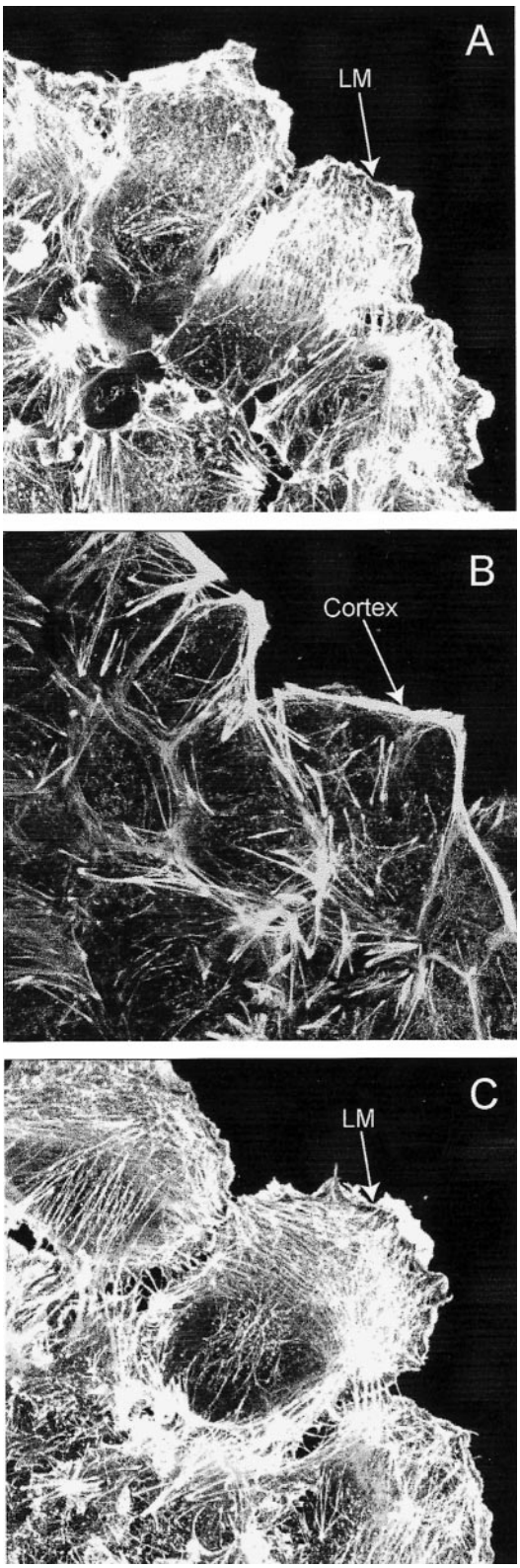
The cytoskeleton is the framework of cellular structure and organization. Actin is its major protein and is present in two forms. The filamentous form of actin, F-actin, polymerizes and depolymerizes continuously to meet the demands of the moment with the help of an array of associated proteins. In the process, F-actin provides shape, rigidity, and strength to the cell. The monomeric form, G-actin, is the result of F-actin depolymerization as well as a source of new monomers for polymerization. Nonmuscle myosin II, a second protein intimately involved with the cytoskeleton, is the motor for cytoskeletal movement. It also adjusts cell shape, responds to calcium levels, and transports and distributes organelles and proteins. Polyamine deficiency decreases nonmuscle myosin II protein by 75% and changes its distribution so that it no longer binds to actin filaments but aggregates in apparently nonfunctional clumps. Other actin-binding proteins important for cytoskeletal function are also adversely affected by polyamine deficiency.

Under the microscope, the cytoskeleton presents a beautiful, intricate pattern when stained with fluorescently linked phalloidin, a toxin of *Amanita phalloides* that binds and stabilizes polymeric actin. The effects of polyamine deficiency on the cytoskeleton can be easily recognized because the distribution of F-actin in the cell is radically altered (see Fig. 2). Actin filaments and stress fibers disappear from the cell interior and concentrate at the cortex rather than maintain their usual position throughout the cell. Even when subconfluent, the cells are rounded and have few of the extensions (lamellipodia and filopodia) seen in normal subconfluent cells. This rearrangement shows that cytoskeletal functions necessary for cell migration, such as polarization, control of direction, cell shape, signaling, and ability to generate force, are disrupted, severely disabling attachment, spreading, and migration. These phenotypic changes were first noticed over 20 years ago in cells with mutant ODC genes.

Cell Growth

Polyamines are required for growth in almost all cells. An adequate supply of polyamines is especially important to the gastrointestinal epithelium where 72 h is the length of an average cell's life. Cell growth (as proliferation) is the basis of the continual renewal of the gastrointestinal epithelium. The epithelium lies in a single layer on the mucosa. It is not a level layer but an arrangement of pits (crypts) and their associated villi that expand the tract's absorptive surface enormously. The depth of the crypts and height of the villi vary in different parts of the tract. A single stem cell at the bottom of each crypt divides continuously; its progeny divide further and move up onto the villi, differentiating as they go into the various types of functional cells of the gastrointestinal epithelium. The oldest cells at the villous tips undergo apoptosis (a programmed death process that minimizes toxicity to surrounding cells) and are sloughed off into the lumen after approximately 72 h. The precise and timely progression of this cycle is vital to the health of the gastrointestinal tract and is dependent on polyamines.

Polyamines are supplied to the gastrointestinal epithelial mucosa from four sources, making its experimental elimination exceedingly difficult. The sources are as follows: (1) new synthesis via ODC in proliferating cells; (2) absorption of dietary polyamines, bacterial products, and intestinal secretions in the lumen; (3) absorption of polyamines from sloughed intestinal villous cells in the lumen: and (4) synthesis via ODC in villous cells transported to the crypt cells through mucosal circulation (see Fig. 3). Polyamines appear in the human gut after a meal and soon disappear from the duodenal and jejunal lumen as a consequence of



absorption and distribution to remote organs and tissues. Normal or neoplastic epithelial cells of the gut mucosa take up polyamines by an active transport process that can be stimulated by mitogens and peptide growth factors.

In culture, normal, transformed, or ODC gene-deleted mutants also depend on polyamines for growth. After the removal of supplemental polyamines from the medium of ODC gene-deleted cells or after treatment of normal cells with DFMO, proliferation is reduced nearly to zero by the time all cells have reached cell cycle stage G1. The cell cycle is arrested at G1 because of an increase in cell cycle inhibitors p21, p27, and p53. Without ODC, the pool of putrescine rapidly falls to very low levels and the source of putrescine from which AdoMetDC synthesizes spermidine is no longer available either. The spermine pool does not fall lower than 40% of normal, probably because it is largely bound to macromolecules. Conversely, cells transfected to overexpress the ODC gene do not require supplemental polyamines and DFMO can depress their growth only marginally.

Cell Migration

When damage to the gastrointestinal epithelial tract occurs, it must be repaired quickly in order to maintain a barrier to the spread of infection throughout the body. Repair occurs in two stages. The first, restitution, can last up to 12 h and consists of a response to cytokines, hormones, and factors secreted by surrounding cells. The restitution response in gastrointestinal epithelial cells consists of extending the cytoplasm in lamellipodia and filopodia into the wounded area. These specialized structures attach to the extracellular matrix through focal adhesions that form on their leading edge. Focal adhesions at the rear of the cell detach, allowing the trailing end of the cell to be drawn up. The process is

FIGURE 2 The migrating edge of rat intestinal cells (IEC-6) in culture. The cells are shown at 3 h of migration after 4 days of (A) no treatment (control), (B) DFMO, or (C) DFMO plus putrescine. F-actin is stained with Texas red phalloidin. The abundance of F-actin is obvious in the control and DFMO plus putrescine groups. Cell polarization toward the open area and lamellipodia (LM) are also plainly visible. The DFMO group shows scarce interior stress fibers, a thickened cell cortex (cortex), and the lack of lamellipodia characteristic of polyamine deficiency. Reprinted from McCormack, S. A., Ray, R. M., and Johnson, L. R. (2002). Polyamines in intestinal epithelial restitution. *In* "Gastrointestinal Mucosal Repair and Experimental Therapeutics" (C.-C. Cho and J.-L. Wang, eds.), 3rd Ed., Vol. 1, pp. 43–56. Karger, Basel, Switzerland, with permission.

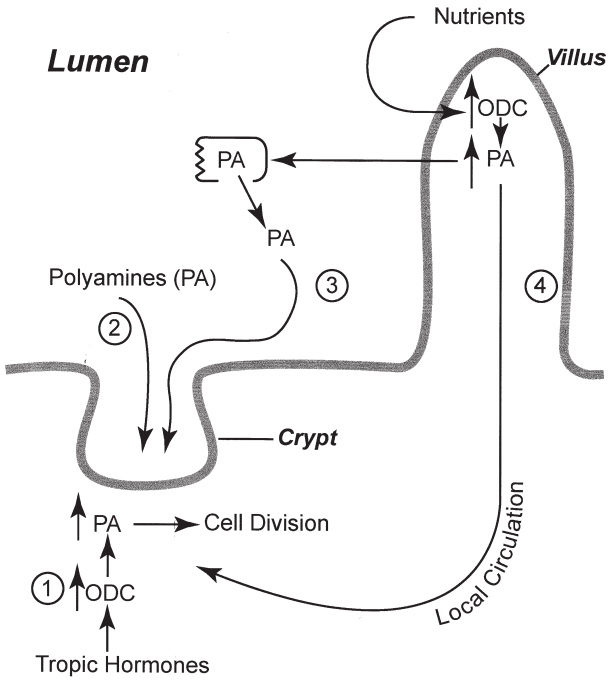


FIGURE 3 Model depicting how polyamines from various sources could reach the proliferating cells of the intestinal crypt and influence growth. The model reconciles the effects of various lumenal stimulators and hormones with known changes in mucosal growth. The numbers refer to four different mechanisms by which polyamines reach proliferating cells: (1) new synthesis via the ODC within the proliferating cell; (2) absorption from the lumen of polyamines supplied by the diet, intestinal secretions, or bacterial synthesis; (3) absorption from the lumen of polyamines contained within sloughed villous cells; and (4) synthesis via ODC in villous cells and transportation to the crypt cells in the mucosal circulation. Reprinted from Johnson, L. R., Tseng, C.-C., Wang, P., Tipnis, U. R., and Haddox, M. K. (1989). Mucosal ornithine decarboxylase in the small intestine: localization and stimulation. *Am. J. Physiol.* **256** (*Gastrointest. Liver Physiol.* **19**), G624–G630, with permission.

repeated to cover as much of the wound as possible. In other words, restitution is primarily a function of spreading and migration. After 12 h, the second stage of repair begins with the advent of cell proliferation. Hormones and growth factors play an important role during this stage as well. Growth continues until the wounded area is again covered with epithelium as before.

The severe inhibiting effects of polyamine deficiency on cell migration have been known since the early 1990s. Gastrointestinal mucosal wounds heal slowly in polyamine-deficient animals. Cultured cells from intestinal and gastric cells as well as cells from a variety of tissues show greatly reduced migration in quantitative migration studies. Investigators have found that this reduction is due to a number of specific effects that impinge on cell migration. These include rearrangement of the actin cytoskeleton, reduced focal adhesion signaling, impaired attachment, reduction of RhoA activity, reduced phosphorylation of specific transcription factors, and other cytoskeletal damaging effects. These observations open the door to undiscovered functions of the polyamines.

Since restitution primarily relies on the machinery of migration, it is heavily dependent on polyamines. Not only is it important to reinstate the barrier to microorganisms in the gastrointestinal tract quickly if it is breached, but the success of the second stage of repair depends to some extent on the success of the first.

The deceptively simple process of extending the cell's border into lamellipodia deserves some explanation in order to appreciate at least part of the elaborate mechanism involved. The lamellipodia have dense, branching actin filaments formed by the activity of the Arp2/3 complex at their outer leading edge. As more actin filaments form near the leading edge, the cell membrane is pushed outward. Focal adhesion sites form on the outer membrane through which integrin receptors can bind to ligands in the extracellular matrix. The key players coordinating this integrated system are the Rho family of small GTPases, specifically RhoA, Rac1, and Cdc42. RhoA regulates the assembly of actin stress fibers, focal adhesions, and contractility. Rac1 stimulates actin polymerization and the formation of ruffles and Cdc42 controls the formation of filopodia. New investigations show that the Rho family proteins have many other molecular links to the actin cytoskeleton through which they regulate actin polymerization, depolymerization, the activity of actin myosins, the speed of migration, and other processes.

ARE POLYAMINES INVOLVED IN GASTROINTESTINAL DISEASE?

Cancer

High levels of ODC activity and intracellular polyamines are common in cancer tissue. In premalignant tissue, they are a reliable sign of increased proliferation rates. Therefore, extracellular fluid polyamine levels that reflect intracellular events can be useful indicators of the effectiveness of therapy. For instance, in Barrett's-associated adenocarcinoma of the esophagus, an increase in tissue polyamine levels has been used to detect the disease still in an occult stage. In colon cancer patients on a low-polyamine diet, polyamine levels in urine may be useful for evaluating the effectiveness

of therapy. DFMO and MGBG have been effective treatments in some cases of colon cancer. Polyamine analogues that inhibit polyamine metabolism are also a possible adjunct to chemotherapy.

Gastrointestinal Immune System Diseases

The mucosal immune system is the first line of defense against microbial and dietary antigens. It connects closely regulated inductive (Peyer's patches) and effector (lamina propria) tissues for the induction of the immune (IgA) response sites that maintain immunological homeostasis in the gut. If homeostasis is lost, inflammatory bowel disease, Crohn's disease, colitis, gastric ulcer, cancer, celiac disease, and other gastrointestinal problems can result. Approximately 50% of the population is infected with *Helicobacter pylori*, bacteria responsible for most cases of peptic ulcer. These bacteria elicit a strong inflammatory response that becomes chronic and, instead of providing protection, eventually contributes to tissue damage and ulcer. Fortunately, the disease develops only in people with a specific combination of bacterial, environmental, and genetic factors. The polyamine content of the lumen is essential for normal healing in peptic and intestinal ulcer. Polyamines can protect cells against apoptosis in growth situations, but when normal cellular regulatory functions are lost, they can also encourage apoptosis in cells that are part of the immune defense system. Inhibitors of polyamine biosynthesis, polyamine analogues, and combinations of polyamine analogues with oligonucleotides may also be candidates for prevention and treatment in these diseases.

Aging

Polyamines are implicated in several changes that accompany aging in humans. One of these is increased susceptibility to gastrointestinal mucosal damage. The gastric mucosa is especially subject to mechanical and chemical damage, bacterial attack, ischemic damage, aberrant immunological responses, and stress. These types of damage must be repaired promptly by restitution and proliferation to avoid impairment of mucosal function. The gastrointestinal mucosa is dependent on epidermal growth factor and other hormones to maintain growth and cell differentiation as cells advance from the crypts onto the villi. With aging, polyamine levels, prostaglandins, cyclooxygenases, and epidermal growth factor receptors are decreased. The sum of these age-related changes retards restitution, cell migration, and proliferation and seriously impedes the repair of gastrointestinal mucosal damage.

Intestinal Parasites

Two billion people in developing countries are seriously debilitated by intestinal parasites. Yet, investigation of polyamines in these parasites was begun only in the past 10–15 years. More detailed knowledge on a variety of parasites is essential to prevent and cure these diseases.

All helminthic parasites that have been studied contain spermine and spermidine but lack significant amounts of putrescine. They also do not have ornithine, arginine, or *S*-adenosylmethionine decarboxylases and, therefore, their survival and growth are not affected by ODC or AdoMetDC inhibitors. Apparently they cannot biosynthesize polyamines and must depend on uptake from the lumen of their hosts. This fact offers a unique and convenient treatment modality. Depending on the parasite, a polyamine-free diet or the inhibition of polyamine transport may be an effective treatment. Other helminthic parasites have unusual polyamine homologues, a degradative pathway operating through *N*-acetyl polyamines, possibly a new type of ODC, etc. The fact that some parasites must obtain necessary polyamines in unusual ways may offer effective and specific treatments that could be utilized to a greater degree than at present.

SUMMARY

The history of the polyamines throughout the development of the biological sciences is a fascinating story. Biologists have been slow to appreciate the roles polyamines play in physiological processes, perhaps because there are so many roles and also for lack of necessary tools. The distribution of polyamines in nature is more varied and individual than was expected. All vertebrate mammals do have the same three natural polyamines, putrescine, spermine, and spermidine, and they synthesize, transport, interconvert, and degrade them by similar pathways. Although all of the facts about polyamine regulation are not yet at hand, it is clear that they can be toxic and that regulation is necessary. The association between rapid cell proliferation, polyamine levels, and cancer stimulated a search for synthetic polyamine inhibitors that is still ongoing in pharmaceutical laboratories. Inhibitors for both polyamine biosynthetic pathways are now available. DFMO and MGBG are especially useful in cell culture studies though less successful in treating human cancer. Recently, experiments using deletion and transfection of the ODC gene have corroborated the effects of DFMO. Polyamines are essential for cell proliferation, differentiation, migration, signaling, and attachment.

The cytoskeleton requires polyamines for virtually all of its normal functions. Repair of the gastrointestinal mucosa requires polyamines for the coordination of restitution and proliferation. Finally, polyamines are important in human gastrointestinal diseases, both from the standpoint of excess and of deficiency. Cancer, immune diseases, and diseases of aging can all result from failure to maintain polyamines at normal levels. Finally, polyamines may underlie other essential processes that have not yet been discovered. In short, continued vigorous investigation of polyamines promises to benefit both medical and basic science.

Acknowledgment

This work was supported by Grants DK16505 and DK52874 from the National Institutes of Health.

See Also the Following Articles

Epithelium, Proliferation of • Epithelium, Repair of

Further Reading

Brabender, J., Lord, R., Danenberg, K., Metzger, R., Schneider, P., Uetake, H., Kawakami, K., Park, J., Salonga, D., Peters, J., DeMeester, T., Holscher, A., and Danenberg, P. (2001). Upregulation of ornithine decarboxylase mRNA expression in Barrett's esophagus and Barrett's-associated adenocarcinoma. *J. Gastrointest. Surg.* **5**, 174–182.

Brzozowski, T., Konturek, P. C., Konturek, S. J., Schuppan, D., Drozdowicz, D., Kwiecien, S., Majka, J., Nakamura, T., and Hahn, E. (2001). Effect of local application of growth factors on gastric ulcer healing and mucosal expression of cyclooxygenase-1 and -2. *Digestion* **64**, 15–29.

Coffino, P. (2001). Antizyme, a mediator of ubiquitin-independent proteasomal degradation. *Biochimie* **83**, 319–323.

Cohen, S. S. (1998). "A Guide to the Polyamines." Oxford University Press, New York.

Davis, R. H., Morris, D. R., and Coffino, P. (1992). Sequestered end products and enzyme regulation: The case of ornithine decarboxylase. *Microbiol. Rev.* **56**, 280–290.

Deloyer, P., Peulen, O., and Dandrifosse, G. (2001). Dietary polyamines and non-neoplastic growth and disease. *Eur. J. Gastroenterol. Hepatol.* **13**, 1027–1032.

Ernst, P. B., Takaishi, H., and Crowe, S. E. (2001). *Helicobacter pylori* infection as a model for gastrointestinal immunity and chronic inflammatory diseases. *Digest. Dis.* **19**, 104–111.

Halter, F., Tarnawski, A. S., Schmassmann, A., and Peskar, B. M. (2001). Cyclooxygenase 2—Implications on maintenance of gastric mucosal integrity and ulcer healing: Controversial issues and perspectives. *Gut* **49**, 443–453.

Johnson, R. L., and McCormack, S. A. (1994). Regulation of gastrointestinal mucosal growth. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 1, pp. 611–641. Raven Press, Washington, DC.

Kiyono, H., Kweon, M. N., Hiroi, T., and Takahashi, I. (2001). The mucosal immune system: From specialized immune defense to inflammation and allergy. *Acta Odontol. Scand.* **59**, 145–153.

McCormack, S., and Johnson, L. (2001). Polyamines and cell migration. *J. Physiol. Pharmacol.* **52**, 327–349.

McCormack, S., Ray, R., and Johnson, L. (2002). Polyamines in intestinal epithelial restitution. *In* "Gastrointestinal Mucosal Repair and Experimental Therapeutics" (C.-H. Cho, J.-Y. Wang, eds.), 3rd Ed., Vol. 1, pp. 43–56. Karger, Basel.

Pollard, T. D., Blanchoin, L., and Mullins, R. D. (2000). Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576.

Ridley, A. J. (2001). Rho GTPases and cell migration. *J. Cell Sci.* **114**, 2713–2722.

Robertson, D. J., and Grimm, I. S. (2001). Inflammatory bowel disease in the elderly. *Gastroenterol. Clin. North Am.* **30**, 409–426.

Silen, W. (1987). Gastric mucosal defense and repair. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 1, pp. 1055–1069. Raven, New York.

Thomas, T., and Thomas, T. J. (2001). Polyamines in cell growth and cell death: Molecular mechanisms and therapeutic applications. *Cell. Mol. Life Sci.* **58**, 244–258.

Wallace, H. M., and Caslake, R. (2001). Polyamines and colon cancer. *Eur. J. Gastroenterol. Hepatol.* **13**, 1033–1039.

Porcelain Gallbladder

SHIH-TING PATRICIA TSAI
University of Southern California and LAC/USC Medical Center

carcinoma Any cancer that arises from the epithelium.
cholecystectomy Surgical removal of the gallbladder.
cholelithiasis The formation of stones in the gallbladder.
hyperechoic Ultrasound description of lesions that reflect sound waves.

Porcelain gallbladder is a rare condition and is defined as intramural calcification of the gallbladder. Patients are generally asymptomatic and diagnosis is usually incidental. The pathogenesis remains unknown but is highly associated with cholelithiasis. Ultrasound and computed tomography are used to assist in diagnosis. There is a high incidence of gallbladder carcinoma in patients with this condition and therefore cholecystectomy is recommended. However, some recent studies show controversial results and definitive management remains to be determined.

INTRODUCTION

Porcelain gallbladder is a rare disease; Grandchamps first described it in 1797 and Florcken first used the term in 1929. Many terms have been used to describe the appearance of this gallbladder condition: "calcifying cholecystitis," "china gallbladder," "cholecystopathy chronic calcanea," "calcified gallbladder," and "porcelain gallbladder." The term porcelain gallbladder describes the brittle consistency and bluish discoloration of the gallbladder wall.

INCIDENCE

The incidence is low, ranging between 0.06 and 0.8% of cholecystectomy specimens. The disease is five times more common in women between the ages of 38 and 70 years (with a mean of 54 years) than in the general population. Rarely, pediatric cases are reported.

PATHOGENESIS

The pathogenesis of the calcification remains controversial. It occurs either as a broad, continuous band in the muscularis or as multiple punctate areas in both the glandular spaces and sinuses. Three potential mechanisms are proposed. One mechanism may be cystic duct obstruction that leads to mucosal deposition of calcium carbonate salts, resulting in bile stagnation within the gallbladder. Another mechanism of injury may be related to chronic irritation of the wall by stones or another foreign body. This theory is supported by the observation that more than 95% of patients with porcelain gallbladder also have stones. The third mechanism may be related to a dystrophic process due to chronic low-grade infection and compromised circulation from cystic duct obstruction. This process results in hemorrhage, scarring, and hyalinization of the wall, which in turn provides a matrix for the deposition of lime salts.

CLINICAL PRESENTATION

Most patients are asymptomatic and the lesion is generally detected incidentally.

DIAGNOSIS

Plain film usually shows a large solitary calcified mass in the right upper quadrant. However, computed tomography and ultrasound are the usual imaging modalities that provide more definitive diagnosis of porcelain gallbladder. In 1984, Kane reported three distinct sonographic patterns in nine patients. The type I pattern was identified by a hyperechoic semilunar structure with a posterior shadow and no gallstones. Type II had a biconvex, curvilinear echogenic structure with acoustic shadowing and stones, and type III had irregular echoes with posterior shadowing. In 1989, Shimizu and co-workers reviewed 30 cases in the world's literature in which ultrasound features of porcelain gallbladder were described. They recommended a simple classification: complete (complete replacement of the mucosa with dense connective tissue and calcification, correlated to type I) and incomplete (some mucosa remaining, correlated to types II and III). They found no gallstone and no cancer in the complete type. However, in the incomplete type, gallstones are detected and there is a 41% incidence of cancer. They

 Copyright 2004, Elsevier (USA). All rights reserved.

hypothesized that malignancies arise only from the mucosal epithelium and therefore gallbladder malignancy is improbable in the type I complete group.

MANAGEMENT AND RELATIONSHIP TO GALLBLADDER CARCINOMA

Prophylactic cholecyctectomy is generally recommended due to the risk of malignancy. The incidence reported in the older literature varies from 12.5 to 61%. However, a recent study by Hines that reviewed 10,741 cholecystectomy specimens showed no carcinoma among 15 patients with porcelain gallbladder. Another study by Berger and colleagues that reviewed 25,900 gallbladder specimens confirmed the association between gallbladder carcinoma and calcified gallbladder but at a lower rate than previously estimated. The incidence of cancer depends on the pattern of calcification; selective mucosal calcification poses a significant risk of cancer (7%), whereas diffuse intramural calcification does not. Therefore, management of asymptomatic patients with debilitating comorbidities remains controversial. The natural history of porcelain gallbladder remains to be elucidated.

See Also the Following Articles

Cholecystectomy • Computed Tomography (CT) • Gallbladder Cancer • Gallstones, Pathophysiology of • Ultrasonography

Further Reading

Ashur, H. (1978). Calcified gallbladder. *Arch. Surg.* **113**, 594–596.
Berger, D. (2001). Carcinoma in the porcelain gallbladder: A relationship revisited. *Surgery* **129**, 699–703.
Hines, O. (2001). Porcelain gallbladder is not associated with gallbladder carcinoma. *Am. Surgeon* **67**, 7–10.
Horton, J. (1993). Gallstone disease and its complications. *In* "Gastrointestinal and Liver Disease" (M. Sleisenger, and B. Scharschmidt, eds.), 6th Ed., Vol. 1, pp. 948–970.
Kang, R. (1984). Porcelain gallbladder: Ultrasound and CT appearance. *Radiology* **152**, 137–141.
Levin, B. (1999). Gallbladder carcinoma. *Ann. Oncol.* **10**, 129–130.
Shimizu, M. (1989). Porcelain gallbladder: Relation between its type by ultrasound and incidence of cancer. *J. Clin. Gastroenterol.* **11**, 471–476.

Porphyria

JOSEPH R. BLOOMER
University of Alabama at Birmingham

genetic heterogeneity Several different mutations in the same gene are found in a genetic disorder.
porphyrin Compound with a chemical structure consisting of four pyrrole groups linked by methene bridges; the compound is pigmented and exhibits red fluorescence when exposed to ultraviolet light around 400 nm (Soret band).
porphyrinogen Reduced form of the porphyrin; not pigmented and does not exhibit fluorescence.
porphyrin precursors Early intermediates of the heme biosynthetic pathway (δ-aminolevulinic acid and porphobilinogen) from which pyrrole groups are formed.

The porphyrias are genetic/metabolic disorders that are characterized biochemically by the excessive accumulation and excretion of porphyrins and porphyrin precursors. These compounds are intermediates of the heme biosynthetic pathway, and each of the porphyrias is associated with deficient activity of a specific enzyme in the pathway. The major clinical manifestations—photocutaneous lesions, neuropsychiatric dysfunction, and structural liver disease—are linked to the biochemical abnormalities. Therapy is directed to ameliorate the biochemical abnormalities, which improves the clinical status of patients.

INTRODUCTION

The first descriptions of the porphyrias appeared in the latter part of the nineteenth century. During the twentieth century, the individual porphyrias were

Copyright 2004, Elsevier (USA). All rights reserved.

hypothesized that malignancies arise only from the mucosal epithelium and therefore gallbladder malignancy is improbable in the type I complete group.

MANAGEMENT AND RELATIONSHIP TO GALLBLADDER CARCINOMA

Prophylactic cholecyctectomy is generally recommended due to the risk of malignancy. The incidence reported in the older literature varies from 12.5 to 61%. However, a recent study by Hines that reviewed 10,741 cholecystectomy specimens showed no carcinoma among 15 patients with porcelain gallbladder. Another study by Berger and colleagues that reviewed 25,900 gallbladder specimens confirmed the association between gallbladder carcinoma and calcified gallbladder but at a lower rate than previously estimated. The incidence of cancer depends on the pattern of calcification; selective mucosal calcification poses a significant risk of cancer (7%), whereas diffuse intramural calcification does not. Therefore, management of asymptomatic patients with debilitating comorbidities remains controversial. The natural history of porcelain gallbladder remains to be elucidated.

See Also the Following Articles

Cholecystectomy • Computed Tomography (CT) • Gallbladder Cancer • Gallstones, Pathophysiology of • Ultrasonography

Further Reading

Ashur, H. (1978). Calcified gallbladder. *Arch. Surg.* **113**, 594–596.
Berger, D. (2001). Carcinoma in the porcelain gallbladder: A relationship revisited. *Surgery* **129**, 699–703.
Hines, O. (2001). Porcelain gallbladder is not associated with gallbladder carcinoma. *Am. Surgeon* **67**, 7–10.
Horton, J. (1993). Gallstone disease and its complications. *In* "Gastrointestinal and Liver Disease" (M. Sleisenger, and B. Scharschmidt, eds.), 6th Ed., Vol. 1, pp. 948–970.
Kang, R. (1984). Porcelain gallbladder: Ultrasound and CT appearance. *Radiology* **152**, 137–141.
Levin, B. (1999). Gallbladder carcinoma. *Ann. Oncol.* **10**, 129–130.
Shimizu, M. (1989). Porcelain gallbladder: Relation between its type by ultrasound and incidence of cancer. *J. Clin. Gastroenterol.* **11**, 471–476.

Porphyria

JOSEPH R. BLOOMER
University of Alabama at Birmingham

genetic heterogeneity Several different mutations in the same gene are found in a genetic disorder.
porphyrin Compound with a chemical structure consisting of four pyrrole groups linked by methene bridges; the compound is pigmented and exhibits red fluorescence when exposed to ultraviolet light around 400 nm (Soret band).
porphyrinogen Reduced form of the porphyrin; not pigmented and does not exhibit fluorescence.
porphyrin precursors Early intermediates of the heme biosynthetic pathway (δ-aminolevulinic acid and porphobilinogen) from which pyrrole groups are formed.

The porphyrias are genetic/metabolic disorders that are characterized biochemically by the excessive accumulation and excretion of porphyrins and porphyrin precursors. These compounds are intermediates of the heme biosynthetic pathway, and each of the porphyrias is associated with deficient activity of a specific enzyme in the pathway. The major clinical manifestations—photocutaneous lesions, neuropsychiatric dysfunction, and structural liver disease—are linked to the biochemical abnormalities. Therapy is directed to ameliorate the biochemical abnormalities, which improves the clinical status of patients.

INTRODUCTION

The first descriptions of the porphyrias appeared in the latter part of the nineteenth century. During the twentieth century, the individual porphyrias were

Copyright 2004, Elsevier (USA). All rights reserved.

identified and their biochemical and clinical features were defined. The porphyrias have been classified as hepatic or erythropoietic, depending on which tissue is the major site of expression of the biochemical abnormalities. Some are also classified as acute or inducible because they are associated with episodic attacks of neuropsychiatric dysfunction. In this article, the biochemical and clinical features of the eight types of porphyria are outlined. Two conditions that may be confused with the porphyrias, secondary porphyrinuria and pseudoporphyria, are also described.

BIOCHEMICAL ABNORMALITIES IN PORPHYRIAS

Porphyrins/porphyrinogens and porphyrin precursors are intermediates of the heme biosynthetic pathway, a critical metabolic process that involves eight enzymes. The first step in the pathway, which is the condensation of succinyl coenzyme A and glycine to form δ-aminolevulinic acid (ALA), is catalyzed by the mitochondrial enzyme ALA synthase and is rate limiting in the liver. The last step, in which ferrous iron is inserted into protoporphyrin to produce heme, also takes place in the mitochondria. Intermediate steps to form porphobilinogen (PBG) and porphyrinogens occur in the cytoplasm.

Deficient activity of an enzyme in the pathway causes a specific pattern of accumulation and excretion of porphyrins and porphyrin precursors (Table I). In the acute porphyrias, this is exacerbated during an attack due to a marked increase in hepatic ALA synthase activity. The pattern of biochemical abnormalities is used to diagnose porphyria in a patient with compatible clinical features. Demonstration of deficient enzyme activity in cells/tissue is also used in diagnosis, particularly in acute intermittent porphyria and the familial form of porphyria cutanea tarda.

The cloning and sequencing of cDNA and genomic DNA for enzymes of the pathway have made it possible to identify gene mutations that underlie the enzyme defects in the porphyrias. Genetic heterogeneity has been found in each. Thus, molecular analysis has not yet found widespread use in diagnosis, but it is helpful in identifying asymptomatic carriers of the gene defect in families in whom a mutation has been found and for evaluating individuals in geographic areas where a specific mutation has a high prevalence. There is not usually a clear relationship between specific gene mutations and the severity of clinical and biochemical manifestations, and expression of the disease is variable even among members of a

TABLE I Biochemical Abnormalities in the Porphyrias[a]

Type of porphyria	Enzyme defect	Location/ biosynthetic step	Major site of expression	Principal biochemical features
ALA dehydrase deficiency	ALA dehydrase	Cytoplasm/2	Liver	↑ ALA in urine
Acute intermittent porphyria	PBG deaminase	Cytoplasm/3	Liver	↑ ALA and PBG in urine
Hereditary coproporphyria	Coproporphyrinogen oxidase	Mitochondria/6	Liver	↑ ALA, PBG, coproporphyrin in urine ↑ Coproporphyrin in feces
Variegate porphyria	Protoporphyrinogen oxidase	Mitochondria/7	Liver	↑ ALA, PBG, coproporphyrin in urine ↑ Protoporphyrin in feces
Porphyria cutanea tarda	Uroporphyrinogen decarboxylase	Cytoplasm/5	Liver and bone marrow	↑ Uroporphyrin in urine Isocoproporphyrin in feces
Hepatoerythropoietic porphyria	Uroporphyrinogen decarboxylase	Cytoplasm/5	Liver	↑ Zn-protoporphyrin in red cells ↑ Uroporphyrin in urine Isocoproporphyrin in feces
Erythropoietic porphyria	Uroporphyrinogen III synthase	Cytoplasm/4	Bone marrow	↑ Uroporphyrin in red cells ↑ Uroporphyrin in urine
Erythropoietic protoporphyria	Ferrochelatase	Mitochondria/8	Bone marrow (liver variable)	↑ Protoporphyrin in red cells ↑ Protoporphyrin in feces

[a] Abbreviations: ALA, δ-aminolevulinic acid; PBG, porphobilinogen.

family. Thus, other genetic and/or acquired factors are often critical to the phenotypic expression of the disorder.

HEPATIC PORPHYRIAS

Several porphyrias are classified as hepatic because the liver is the major site of expression of the biochemical abnormalities (Table I). Four of the hepatic porphyrias are also termed acute because there occur episodes of severe neuropsychiatric dysfunction that are separated by asymptomatic periods (Table II). The acute attacks are precipitated by ingestion of drugs, fasting, alcoholism, infection, and hormonal effects. The most common symptom is abdominal pain, which may be accompanied by hypertension and tachycardia as manifestations of autonomic nerve dysfunction. Peripheral neuropathy causes paralysis and respiratory compromise if the attack is severe. Psychiatric manifestations include hysteria, psychosis, and depression, which sometimes persist after the attack has subsided. Seizures also occur and present a difficult problem because most antiepileptic drugs can exacerbate the attack.

During an acute attack, urinary excretion of the porphyrin precursors ALA and PBG increases markedly. Clinical and basic studies indicate that ALA may cause the neurological dysfunction. An alternate possibility is that heme deficiency in nerve tissue is the cause. Finally, both (and other) factors may underlie the attack.

Therapy of the acute attack consists of stopping the precipitating factor, carefully managing fluid and electrolyte status, and providing adequate caloric intake that is high in carbohydrates. Intravenous administration of hematin (ferriheme hydroxide) has become standard, because this may promptly ameliorate biochemical and clinical manifestations. During asymptomatic periods, patients should not take drugs that precipitate attacks, should avoid fasting and excess intake of alcohol, and should have infections treated promptly.

In four of the hepatic porphyrias, photocutaneous lesions occur (Table II). Skin fragility develops as a consequence of the photoactive properties of porphyrins deposited in skin tissue and/or circulating in dermal blood vessels. This causes blisters to form after minor trauma to sun-exposed areas. Erosions, scarring, pigment changes, and small white papules called milia subsequently develop. Sclerodermoid changes of the skin may occur in long-standing untreated disease.

In porphyria cutanea tarda, hepatic iron overload is frequent, and there is an increased prevalence of mutations in the *HFE* gene, which is associated with hereditary hemochromatosis. Patients also have a higher rate of chronic hepatitis C. In long-standing untreated porphyria cutanea tarda, there is an increased incidence of hepatocellular carcinoma. This is also found in the acute porphyrias, particularly acute intermittent porphyria, the reason for which is unclear.

The photocutaneous lesions in porphyria cutanea tarda are managed by phlebotomy to deplete excess hepatic iron, as uroporphyrin formation decreases concomitantly. Removal of 4–8 liters of blood usually resolves the clinical and biochemical abnormalities. Chloroquine or related compounds are used if phlebotomy is not tolerated. Once in remission, most patients with porphyria cutanea tarda remain free of photocutaneous lesions provided that they avoid taking iron-containing compounds and remain abstinent from alcohol. Photocutaneous lesions in variegate porphyria and hereditary coproporphyria do not respond to phlebotomy, however.

TABLE II Clinical Features in the Porphyrias

Type of porphyria	Usual inheritance[a]	Usual onset of disease	Photocutaneous lesions	Neuropsychiatric symptoms	Chronic liver disease	Hepatoma
ALA dehydrase deficiency	AR	Childhood	−	+	−	−
Acute intermittent porphyria	AD	Early adulthood	−	+	−	+
Hereditary coproporphyria	AD	Early adulthood	+	+	−	+
Variegate porphyria	AD	Early adulthood	+	+	−	+
Porphyria cutanea tarda	AD (familial type)	Adulthood	+	−	+	+
Hepatoerythropoietic porphyria	AR	Childhood	+	−	+	−
Erythropoietic porphyria	AR	Infancy	+	−	−	−
Erythropoietic protoporphyria	Triallelic	Childhood	+	+ (cholestatic crisis)	+	−

[a] Abbreviations: AR, autosomal recessive; AD, autosomal dominant.

ERYTHROPOIETIC PORPHYRIAS

In two porphyrias, erythropoietic protoporphyria (EPP) and erythropoietic porphyria, the bone marrow is the major site of expression of the biochemical abnormalities. The major clinical manifestation in EPP is lifelong photosensitivity. In contrast to the other porphyrias, photosensitivity occurs acutely during sun exposure. Erythema and edema of the skin develop but blisters and erosions are rare. Chronic skin changes involve thickening and lichenification of the skin on the nose and dorsal aspects of the hands. Oral administration of β-carotene reduces photosensitivity in many patients. In some cases, the only effective management is use of opaque sunscreens or avoidance of sun exposure, even through window glass.

Hepatobiliary disease is another feature of EPP, and in approximately 5% of individuals, the occurrence of structural damage to the liver may cause liver failure and necessitate liver transplantation. Liver damage is due to the toxic effect of protoporphyrin on liver function and structure, particularly when there is progressive accumulation of protoporphyrin in the liver due to impaired excretion of protoporphyrin in bile. Therapies for this condition involve interruption of the enterohepatic circulation of protoporphyrin by using cholestyramine or activated charcoal and decreasing the excess production of protoporphyrin and improving liver function through the intravenous administration of hematin. However, when liver damage is advanced, liver transplantation is the only effective treatment. Unfortunately, disease frequently recurs in the graft because excessive production of protoporphyrin in the bone marrow is not significantly changed by liver transplantation.

Erythropoietic porphyria is a recessive disorder that usually has onset in infancy. A few cases of adult onset have been reported. Skin lesions are similar to those in porphyria cutanea tarda. As patients age, there may be progressive destruction of the fingertips, ears, and nose. Hemolytic anemia and splenomegaly are common. Therapy is generally supportive, consisting of protection from sun exposure and prompt treatment of skin infections. Red blood cell transfusion and intravenous administration of hematin are used to decrease the production of porphyrin. Splenectomy effects remission of disease in some patients.

SECONDARY PORPHYRINURIA AND PSEUDOPORPHYRIA

Several diseases are associated with an increase in the urinary excretion of porphyrin, particularly coproporphyrin excretion, which is termed secondary porphyrinuria. These diseases include various types of anemia and malignancy, hepatobiliary diseases, diabetes, and infections. Some patients have abdominal pain and other symptoms of acute porphyria. However, with the exception of lead poisoning and hereditary tyrosinemia, urinary excretion of ALA and PBG is normal. The patients also do not develop photocutaneous lesions like those with the porphyrias. Thus, the secondary porphyrinurias can usually be distinguished from the porphyrias.

Pseudoporphyria is a condition in which there are skin lesions similar to those in porphyria cutanea tarda, but serum and urine porphyrin levels are normal or only minimally elevated. This occurs in renal failure, from use of medications such as nonsteroidal antiinflammatory drugs and tetracycline, and from ultraviolet A exposure. Treatment consists of discontinuing the causative factor and protection from the sun.

See Also the Following Articles

Hepatitis C • Hepatocellular Carcinoma (HCC)

Further Reading

Anderson, K. E. (2003). The porphyrias. *In* "Hepatology" (D. Zakim and T. Boyer, eds.), 4th Ed., pp. 291–346. W. B. Saunders, Philadelphia.

Bloomer, J., and Brenner, D. (2003). The porphyrias. *In* "Schiff's Diseases of the Liver" (E. Schiff, M. Sorrell, and W. Maddrey, eds.), 9th Ed., pp. 1231–1260. Lippincott Williams & Wilkins, Philadelphia.

Portal Hypertension and Esophageal Varices

KEVIN M. COMAR AND ARUN J. SANYAL
Virginia Commonwealth University

ascites Excessive accumulation of fluid in the peritoneal cavity.
cirrhosis Advanced liver disease characterized by distorted architecture secondary to hepatic fibrosis and regenerative nodules.
esophageal varices Dilated blood vessels around the esophagus.
portal hypertension A portal pressure >5 mmHg in the portal circulation.

Portal hypertension is directly responsible for the development of variceal hemorrhage and ascites, the two major complications of cirrhosis. Of these complications, hemorrhage from esophageal varices is an immediately life-threatening event that is associated with a high mortality. The prevention and treatment of esophageal varices is therefore a cornerstone in the management of cirrhosis.

PATHOGENESIS

Development of Portal Hypertension

The portal vein is formed by the confluence of the superior mesenteric vein and the splenic vein. It drains blood from the splanchnic bed to the liver, which is drained by hepatic veins in the inferior vena cava. Following the principles of Ohm's law, the pressure in the portal vein is determined by the portal venous inflow and the resistance to outflow from the portal vein:

$$\text{Portal Pressure} = \text{Portal venous inflow} \times \text{Resistance to outflow from the portal vein.} \quad [1]$$

Portal hypertension is initiated by an increase in resistance to portal venous outflow. Depending on the site of the increase in resistance, portal hypertension may also be classified as presinusoidal, sinusoidal, or postsinusoidal (Table I). The most common cause of portal hypertension is cirrhosis. The increased resistance found in cirrhosis is intrahepatic and primarily sinusoidal in origin. Increased sinusoidal resistance results from both a fixed component due to the architectural distortion associated with cirrhosis and a dynamic component due to altered regulation of sinusoidal vascular resistance. The latter is related to decreased nitric oxide (NO) production by the sinusoidal endothelium in cirrhosis. Cirrhosis is also associated with increased NO production in the systemic circulation, which leads to systemic arterial vasodilation and a hyperdynamic circulatory state. Mesenteric arterial dilation increases portal venous inflow and further compounds the severity of the portal hypertension.

TABLE I Causes of Portal Hypertension

Presinusoidal
Prehepatic
Splenic vein thrombosis
Portal vein thrombosis
Cavernous transformation of vein
Extrinsic compression of the portal vein
Intrahepatic
Primary biliary cirrhosis (precirrhotic stages)
Primary sclerosing cholangitis
Sarcoidosis
Schistosomiasis
Sinusoidal
Cirrhosis
Alcoholic hepatitis
Nodular regenerative hyperplasia
Vitamin A toxicity
Posthepatic
Budd-Chiari syndrome
Veno-occlusive disease
Constrictive pericarditis
Tricuspid valve disease
Severe congestive cardiomyopathy

Measurement of Portal Hemodynamics

Portal venous hemodynamics can be measured by hepatic venous catheterization. During this procedure, a balloon catheter is passed into the hepatic vein and the following parameters are measured:

Free hepatic venous pressure (FHVP)
= pressure in the hepatic vein measured with the balloon deflated.

 Copyright 2004, Elsevier (USA). All rights reserved.

Wedged hepatic venous pressure (WHVP)
= pressure in the hepatic vein measured with the balloon inflated to occlude the hepatic vein. [2]
Hepatic venous pressure gradient (HVPG)
= WHVP − FHVP.

The HVPG represents the pressure in the hepatic sinusoids and portal vein and is a measure of portal pressure.

Development of Variceal Hemorrhage

Once portal hypertension occurs, nature decompresses the portal vein by the opening of a collateral circulation that diverts blood from the portal venous bed directly to the systemic circulation. The most common site for development of such collaterals is the gastroesophageal junction, where the collaterals form thin-walled varicose veins.

As noted above, the pressure in the varix is determined by the product of variceal blood flow and resistance. Esophageal varices do not develop at HPVG values less than 12 mm Hg. On the other hand, HVPG values greater than 12 mmHg do not directly correlate with risk of bleeding. This suggests that local factors at the level of the varices also determine an individual subject's probability of bleeding (Table II).

The risk of rupture of esophageal varices is determined by the wall tension, which is dictated by Laplace's law:

Wall tension = Transmural pressure gradient × radius / thickness of the variceal wall. [3]

Based on Laplace's law (Eq. [3] above), large varices with thin walls and a high intramural pressure are most likely to bleed. Varices in the distal esophagus have the least amount of tissue covering the esophageal veins and therefore are the most prone to bleed. Thinning of the mucosa over a varix also contributes to the risk of hemorrhage and is noted clinically by the presence of "red signs" on endoscopy (Table III). These include red wale marks, cherry red spots, hematocystic spots, and diffuse erythema. Clinical features, such as the degree of liver dysfunction and history of previous variceal bleeds, are also significant predictors of the risk of developing variceal hemorrhages.

TABLE II Risk Factors That Affect the Likelihood of Variceal Hemorrhage

Portal pressure
Variceal pressure
Variceal location
Variceal size
Variceal appearance (red signs)
Liver function
Previous variceal hemorrhage

TABLE III Types of Endoscopic "Red Signs" on Varices

Red signs
Red wale marks
Cherry red spots
Hematocystic spots
Diffuse erythema

NATURAL HISTORY

Approximately 30% of patients with compensated cirrhosis and 60% of patients with decompensated cirrhosis have esophageal varices. In those without varices, the risk of *de novo* varix development is approximately 5% annually. One-third of all patients with varices experience variceal hemorrhage. The risk of bleeding is greatest in the first year after diagnosis and can be estimated by assessment of variceal size, liver function, and presence of red signs (Table IV).

Once varices start bleeding, spontaneous hemostasis occurs in only 50% of cases. The risk factors for continued bleeding include advanced liver failure and large spurting varices. Once hemostasis occurs, there is a high risk of recurrent bleeding over the next 2−3 days, which subsides to baseline levels by 6 weeks. The risk factors for such "early rebleeding" include age >60 years, hemoglobin less than 8 g/dl on admission, renal failure, ascites, active bleeding, red signs or platelet clots on varices, and overly aggressive volume replacement. Overall, each episode of variceal hemorrhage continues to be associated with a 20−30% mortality rate.

In the long term, over 70% of survivors of an index bleed will experience recurrent hemorrhage if left untreated and a similar number will die within 1 year. The risk of late rebleeding >6 weeks is linked to severity of liver failure, ascites, presence of hepatoma, active alcoholism, and red signs.

PRIMARY PREVENTION

Selection of Patients for Primary Prophylaxis

The risks of variceal hemorrhage varies greatly from one patient to the next. Given the potential toxicity of

TABLE IV Estimated 1-Year Percentage Probability of Bleeding as a Function of All Possible Combinations of the Endoscopic Variables

Red wale markings	Child's class								
	A			B			C		
	*F1	F2	F3	F1	F2	F3	F1	F2	F3
−	6	10	15	10	16	26	20	30	42
+	8	12	19	15	23	33	28	38	54
++	12	16	24	20	30	42	36	48	64
+++	16	23	34	28	40	52	44	60	76

Based on data from Defranchis, R. (1998). Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices. *N. Engl. J. Med.* **319**, 983.

* F1, F2, and F3 are progressively larger varices.

the available treatment approaches, it is imperative to identify those at greatest risk of bleeding and thus most likely to benefit from primary prophylaxis. This is achieved by clinical assessment and by endoscopy of all patients with cirrhosis (Table IV). Those with medium to large varices are generally considered for primary prophylaxis. If no varices are seen, follow-up endoscopy is generally recommended at 2-year intervals although the cost-effectiveness of this approach remains to be established.

Beta-Blockers

Nonselective beta-blockers are the first-line treatment for primary prophylaxis of variceal bleeding in cirrhosis. They decrease portal pressure by causing beta-blockade, which allows unopposed α-adrenergic-mediated mesenteric arteriolar vasoconstriction. This decreases portal venous inflow and thus portal pressure. At high doses, bradycardia and decreased cardiac output also contribute to this effect.

The efficacy and safety of propranolol and nadolol have been evaluated in a large number of clinical trials. Meta-analyses have shown that beta-blockers reduce the index bleed by approximately 45% and also decrease mortality due to bleeding by 50%. However, this is not accompanied by an overall improvement in survival. The best predictor of successful primary prophylaxis is the ability to produce a sustained drop in HVPG by 20% or to values less than 12 mm Hg. It has therefore been recommended that hepatic venous catheterization be performed prior to and 3 months after initiation of treatment with beta-blockers. The cost-effectiveness of such an approach remains to be validated and the facilities to perform HVPG measurements are not universally available. In their absence, the dose of beta-blockers is titrated to achieve a reduction in heart rate (HR) to 25% of baseline, or to 55−60 beats per minute, and is limited by the development of side effects. Regrettably there is a poor correlation between HR and reduction in portal pressure.

Approximately 20% of patients do not respond to beta-blockers and an additional 20% are unable to tolerate beta-blockers. Risk factors associated with high failure rates include a smaller decrement in resting heart rate, younger age, advanced liver failure, large variceal size, and lower doses of beta-blockers. Over time, many patients also develop tachyphylaxis due to increased veno-collateral resistance within the liver. The side effects of beta-blockers include precipitation or worsening of congestive heart failure, sinus bradycardia, increased airway resistance, exacerbation of peripheral vascular disease, facilitation of hypoglycemia, depression, fatigue, and sexual dysfunction.

Nitrates

Nitrates act as venodilators, in turn decreasing venous return and therefore decreasing cardiac output. Systemic venodilation also decreases postsinusoidal resistance and consequently decreases portal hypertension. Initial clinical trials comparing isosorbride mononitrate (ISMN) alone versus propranolol alone found ISMN to be as effective as propranolol for the prevention of bleeding; however, those over 50 years of age had a poorer survival with ISMN. Another recent study also showed that ISMN did not reduce the incidence of variceal bleeding or survival in patients with varices unable to tolerate beta-blockers. Thus, nitrates are not recommended as monotherapy for primary prophylaxis of variceal hemorrhage.

Nitrates Plus Beta-Blockers

A combination of ISMN with a nonselective beta-blocker has a synergistic effect on portal pressures. Clinical trials have shown combination therapy to be superior to beta-blockers alone. Such benefits are most pronounced in those with relatively preserved liver function. The use of combination therapy is usually restricted to such individuals or those who do not have a sustained improvement in HVPG after starting beta-blockers.

Endoscopic Therapy

Endoscopic sclerotherapy (EST) is performed by injection of a sclerosant into varices that produce variceal thrombosis and obliteration. Endoscopic variceal ligation (EVL) is performed by placing an elastic O ring around the neck of a varix and subsequently causing strangulation of the vein. Thrombosis and

ischemic necrosis of the necrotic mucosa follow, leading to the obliteration of the varix. The use of EST for primary prophylaxis has been discontinued since a large multicenter trial showed increased mortality in those undergoing EST. On the other hand, compared to no treatment, EVL decreases the risk of index bleeding by 64% and decreases overall mortality by 45%. Compared to beta-blockers, EVL reduces the risk of bleeding by approximately 52% but this does not translate into a survival advantage. EVL is currently recommended for those with large varices who are unable to tolerate beta-blockers. There are currently no published data comparing a combination of beta-blockers with band ligation to either beta-blockers or EVL alone.

Summary

All patients with cirrhosis should undergo endoscopy to assess the risk of variceal hemorrhage. Those at intermediate or high risk should undergo treatment with a nonselective beta-blocker, e.g., propranolol or nadolol. The dose should be titrated to achieve a resting heart rate of 55–60 beats/min. Ideally, the HVPG should be measured before and 1–3 months after starting therapy to identify poor responders to therapy. Such individuals may be treated by addition of nitrates or by EVL. EVL should also be considered in those who cannot tolerate pharmacologic treatment (Fig 1).

MANAGEMENT OF ACTIVE VARICEAL BLEEDING

There are three goals of management of active variceal hemorrhage: (1) hemodynamic resuscitation; (2) prevention of complications; and (3) achievement of hemostasis. Therapy must be directed at all three goals simultaneously in order to optimize outcomes. Hemodynamic resuscitation must be aggressively pursued and the hemoglobin maintained between 9 and 10 g/dl (Table V). The urine output should be carefully monitored and fluid infusions titrated to maintain a urine output greater than 50 cc/h. The airway must be protected and the patient intubated if he or she is unable to clear secretions in the airway. Blood and ascites fluid should be obtained for culture, and prophylactic broad-spectrum antibiotics, e.g., cefotaxime, should be administered intravenously until culture results are determined. In the absence of positive cultures or spontaneous bacterial peritonitis (SBP), treatment may be switched to an oral quinolone, e.g., norfloxacin, for up to 2 weeks for primary prophylaxis of SBP.

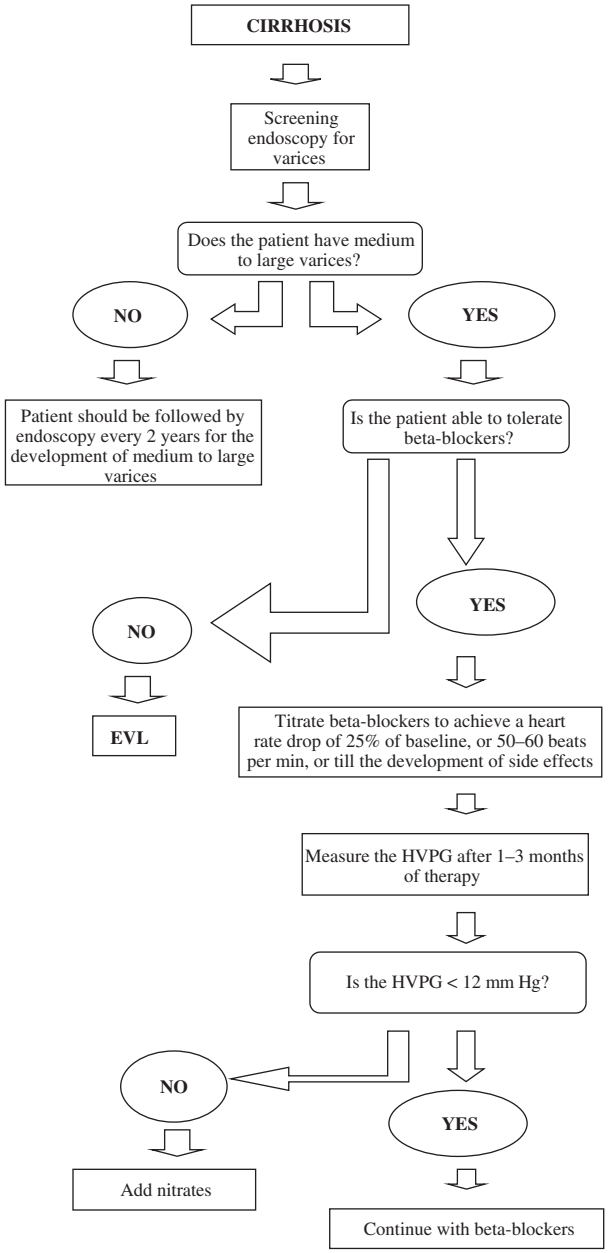


FIGURE 1 Primary prevention. These are only general guidelines and appropriate therapy should be based on the patient's individual circumstances and the expertise available.

First-Line Therapy for Achievement of Hemostasis

Pharmacologic Treatment

Vasopressin and its analogues Vasopressin interacts with arterial smooth muscle receptors to induce splanchnic arteriolar vasoconstriction, decreasing blood flow to all splanchnic organs and resulting in a decline in portal pressure. The use of vasopressin has

TABLE V Management of Patients with Active Variceal Hemorrhage

Hemodynamic resuscitation
Maintain hemoglobin between 9 and 10 g/dl
Replace platelets in actively bleeding patients when levels fall below 50,000/ml
Airway protection
Intubate for aspiration pneumonia prevention
Renal support
Preserve urine output above 50 ml/h
Avoid nephrotoxins
Sepsis prevention
Obtain blood and ascites fluid for culture and prophylactic broad-spectrum antibiotics
Neurologic support
Monitor mental status and avoid sedation
Metabolic support
Monitor and treat electrolyte abnormalities
Monitor and treat endocrine abnormalities
Treat alcohol withdrawal when indicated

declined due to its side effects, which include myocardial and intestinal ischemia, hypertension, bradycardia, hyponatremia, and fluid retention. Some of these effects can be minimized while enhancing therapeutic efficacy by concomitant administration of nitroglycerin.

Terilipressin is an inert analogue of vasopressin; due to its slow activation, terlipressin has less toxicity than vasopressin alone or in combination with nitroglycerin. Terlipressin has also been shown to be better than placebo or vasopressin and similar to somatostatin or balloon tamponade in acute variceal bleeding management. Currently, Terlipressin is not available for clinical use in the United States.

Somatostatin and its analogues Somatostatin inhibits the release of splanchnic vasodilators hormones, such as glucagon and vasoactive intestinal peptide, thereby increasing mesenteric arteriolar tone and decreasing portal venous inflow. Somatostatin is as effective as vasopressin in the control of the bleeding and has a lower risk of adverse effects. The beneficial effects of somatostain over vasopressin have made somatostatin the pharmacological treatment of choice for active bleeding even though no decrease in mortality has been clearly proven. Circulating somatostatin has a half-life of only a few minutes and this has led to use of its longer acting analogue, octreotide. Although there is some controversy over the relative efficacy of octreotide versus somatostatin, the use of octreotide with EVL or EST has been found to be superior to either EVL or EST alone in terms of achieving hemostasis or prevention of early rebleeding. However, no change in overall mortality has been demonstrated with either somatostatin or octreotide.

Endoscopic therapy EST is highly effective in achieving hemostasis (70–90%) and decreasing the risk of early rebleeding (20–30%) compared to vasopressin or balloon tamponade. EST is as effective as somatostatin or octreotide but less effective than combination therapy. Unfortunately, endoscopic sclerotherapy has a 10 to 30% complication rate (Table VI) and a 0.5 to 2% mortality rate. EST and EVL are comparable in terms of efficacy when used for active variceal hemorrhage. EVL has a lower rebleeding rate, lower mortality rate, and lower incidence of complication than EST; however, in cases of severe bleeding, limitation of the field of vision makes EVL more challenging.

FAILURE OF FIRST-LINE THERAPY (SECOND-LINE TREATMENT)

Approximately 10–20% of patients either continue to bleed despite first-line therapy or experience early rebleeding. Such patients are at great risk of dying from exsanguination and from the complications of variceal hemorrhage, such as aspiration, sepsis, and hepatorenal syndrome. It is therefore imperative to quickly stabilize the patient and perform a salvage procedure in such cases.

Balloon Tamponade

This technique utilizes a balloon catheter that mechanically tamponades the bleeding varix. There are several types of balloon catheters that are used for this purpose and all can quickly achieve hemostasis. Unfortunately, there is a very high risk of bleeding when the

TABLE VI Complications of Sclerotherapy

Local
Stricture formation
Ulceration
Perforation
Bleeding
Esophageal dysmotility
Pain/odynophagia
Regional
Pleural effusion
Mediastinitis
Acute gastric dilation
Systemic
Pulmonary edema
Aspiration pneumonia
Sepsis
Spontaneous bacterial peritonitis
Adult respiratory distress syndrome
Portal vein thrombosis

balloon is deflated. Moreover, prolonged insufflation of the balloon can lead to mucosal necrosis and ulceration. Finally, balloon tamponades occlude the gastroesophageal junction and prevent clearance of swallowed saliva and secretions, thereby creating a high-risk situation for pulmonary aspiration. These risks can be minimized by protecting the airway by intubation and by the use of an esophageal aspiration port. Balloon tamponade is used only as a temporary measure while preparations are made for a definitive procedure.

Surgery

The surgical options available for acute variceal bleeding consist of shunt or nonshunt operations. Shunt operations decompress the portal system and are total, partial, or selective. Total shunts redirect all portal blood away from the liver and are highly effective in terminating active bleeding and preventing future bleeds. Unfortunately, approximately 40 to 50% patients who undergo total shunts will have chronic or recurrent encephalopathy and accelerated progression of underlying liver failure. Such operations are not performed as first-line therapy because the outcomes with EST have been shown to be comparable to those for urgent portacaval shunt surgery.

Partial shunts preserve some hepatic portal perfusion and in turn reduce the rates of encephalopathy and liver failure. Nonetheless, the incidence of early shunt thrombosis is relatively high. Selective shunts divide the portal system into a decompressed variceal compartment and a hypertensive superior mesenteric–portal vein compartment. The most prominent selective shunt is distal splenorenal shunt (DSRS), which maintains a certain degree of liver perfusion, avoids dissection of the hilus, and therefore does not interfere with subsequent liver transplantation. The drawback to DSRS is that sinusoidal hypertension persists, and ascites, which does not resolve, can be difficult to manage. Nonshunt operations include esophageal transection and devascularization. Esophageal transection is highly effective in arresting bleeding from esophageal varices unresponsive to medical therapy, but rebleeding from gastric varices and the transection line is a problem in 50% of patients. The complete devascularization (Suguira) procedure is highly effective in controlling active hemorrhage. Such operations carry a rebleeding rate of 2 to 37% and have a reported mortality rate of 5% or higher.

Transjugular Intrahepatic Portosystemic Shunts

Transjugular intrahepatic portosystemic shunts (TIPS) is an angiographic procedure in which a fenestrated metal stent is deployed between the intrahepatic portion of the portal vein and the hepatic vein, thereby creating a "side to side" porta-systemic anastamosis. The ability to decompress the portal vein without the need for general anesthesia or major surgery has led to a resurgence in interest in portal decompression in patients with active and refractory variceal hemorrhage. TIPS can be performed successfully in over 90% of cases and achieves hemostasis in over 90% of cases. It also has been shown to improve survival in this group of very sick individuals. Thus, TIPS is the procedure of choice for salvage therapy in high-risk individuals who continue to bleed or have severe early rebleeding after first-line therapy.

TIPS is associated with numerous complications (Table VII) including encephalopathy, hemolysis, and thrombosis. In the long term, virtually all patients develop recurrent portal hypertension due to the ingrowth of tissue from the surrounding liver, which forms a pseudo-intimal lining in the shunt. This requires patients to undergo periodic sonographic screening for shunt patency after TIPS placement. Unfortunately, sonography is specific but relatively insensitive for the diagnosis of shunt stenosis.

Summary

Management of active variceal bleeding includes hemodynamic resuscitation, prevention and treatment of complications, and halting the bleed. Somatostatin, octreotide, or terlipressin may be started in the emergency room and given in conjunction with EST or EVL. Those who fail first-line therapy should be stabilized and a definitive procedure (TIPS for high-risk cases and TIPS or surgery for average-risk cases) performed quickly (Fig. 2).

TABLE VII Complications of Transjugular Intrahepatic Portosystemic Shunts (TIPS)

Procedure-linked complications
Neck hematoma
Perihepatic hematoma
Cardiac arrhythmias
Rupture of liver capsule
Puncture of portal vein
Complications related to portosystemic shunting
Hepatic encephalopathy
Increased susceptibility to bacteremia
Liver failure
Stent-related complications
TIPS-associated hemolysis
Infection of the stents
Stent stenosis or malfunction

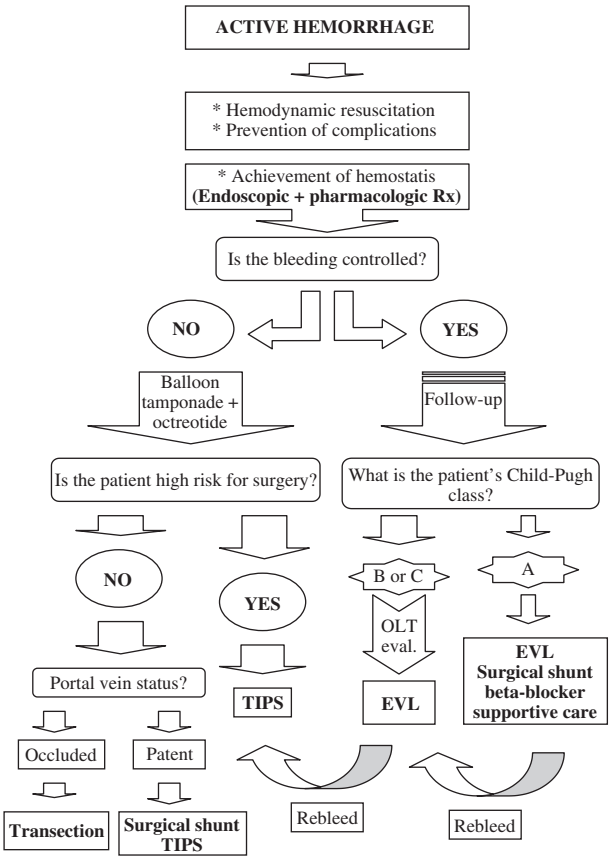


FIGURE 2 Management of active hemorrhage. These are only general guidelines and appropriate therapy should be based on the patient's individual circumstances and the expertise available.

SECONDARY PREVENTION

Given the high recurrence rate of rebleeding, attention must be given to providing secondary prophylaxis to those who survive an index bleed. The treatment modalities available are identical to those used for primary prophylaxis and acute management.

Endoscopic Therapy

EVL is currently the procedure of choice for the secondary prevention of esophageal variceal hemorrhage. Compared to sclerotherapy, EVL showed earlier variceal obliteration, fewer complications, less rebleeding, and a trend toward decreased mortality. The combination of EST and EVL does not provide any advantage over EVL alone and is actually associated with higher mortality.

Beta-Blockers

Studies on the effect of beta-blockers as secondary prevention suggest that beta-blockers reduce the risk of bleeding by approximately 40% and risk of death by 20%. A crucial predictive factor in the effectiveness of beta-blockers is its ability to decrease HPVG by 20% or more. Good responders to beta-blockers are considered patients with 20% or more reduction in HPVG and are much less likely (2 of 25 versus 23 of 44) to have episodes of rebleeding within the first 28 months of therapy. Patients with well-preserved hepatic synthetic function seem to benefit the most from beta-blocker therapy. Overall, EST is associated with a lower rebleed rate than beta-blockers but the effects of these therapies on mortality are similar. Recent studies indicate that EVL is at least as effective as beta-blockers for the prevention of recurrent variceal bleeding.

Beta-Blocker Plus Oral Nitrates

Combination pharmacologic therapy has been shown to be more effective than EST or EVL in two studies of patients with well-preserved liver function from the same center. In two other studies, combination therapy was found to be equally as effective as EVL. The potential role of combination therapy is evolving and it may become an alternative to EVL for the prevention of recurrent variceal hemorrhage.

Transjugular Intrahepatic Portosystemic Shunts

TIPS decreases the risk of recurrent bleeding more effectively than endoscopic therapy. However, this is not associated with a survival advantage. There is also an increased risk of encephalopathy following TIPS. In some cases, TIPS is followed by progressive liver failure. Finally, patients require multiple follow-up sonograms and angiograms for the detection and treatment of shunt stenosis. These considerations have relegated the role of TIPS to a salvage treatment in those who rebleed despite adequate endoscopic and pharmacologic therapy.

Surgery

The efficiency of surgical treatment for rebleeding prevention is offset by studies showing that the survival outcome after sclerotherapy is identical to or better than that achieved by surgery. In addition, postoperative mortality and complications are more considerable with surgery. Regardless of the limitations of surgery, surgery remains a valuable form of treatment in selected patients who are refractory to endoscopic intervention. The outcomes after surgery are best in those with well-preserved liver functions.

Liver Transplant

Orthotopic liver transplant (OLT) is the only remedy that provides long-term treatment for prevention of

rebleeding, hepatic decompensation, and death. OLT provides an 80–90% 1-year survival rate and a 60% 5-year survival rate. OLT may be considered in those with recurrent variceal bleeds or those with severe liver failure. Unfortunately, this option is not available to many subjects and, when available, is limited by organ availability.

See Also the Following Articles

Ascites • Budd–Chiari Syndrome • Cholangitis, Sclerosing • Cholestatic Diseases, Chronic • Cirrhosis • Hepatic Circulation • Portal Vein Thrombosis • Sinusoidal Obstruction Syndrome (Hepatic Venoocclusive Disease) • Somatostatin • Upper Gastrointestinal Bleeding • Variceal Bleeding

Further Reading

Banares, R., Albillos, A., Rincon, D., *et al*. (2002). Endoscopic treatment versus endoscopic plus pharmacologic treatment for acute variceal bleeding: A meta-analysis. *Hepatology* **35**, 609–615.

Beppu, K., Inokuchi, K., Koyanagi, N., *et al*. (1981). Prediction of variceal hemorrhage by esophageal endoscopy. *Gastrointest. Endosc.* **27**, 213–218.

Bomzon, A., and Blendis, L. M. (1994). The nitric oxide hypothesis and the hyperdynamic circulation in cirrhosis. *Hepatology* **20**, 1343–1350.

Cales, P., Desmorat, H., Vinel, J. P., *et al*. (1990). Incidence of large oesophageal varices in patients with cirrhosis: Application to prophylaxis of first bleeding. *Gut* **31**, 1298–1302.

Cello, J. P., Grendell, J. H., Crass, R. A., Weber, T. E., and Trunkey, D. D. (1987). Endoscopic sclerotherapy versus portacaval shunt in patient with severe cirrhosis and acute variceal hemorrhage: Long-term follow-up. *N. Engl. J. Med.* **316**, 11–15.

D'amico, G., and Luca, A. (1997). Natural history. Clinical–haemodynamic correlations: Prediction of the risk of bleeding. *Bailliere Clin. Gastroenterol.* **11**, 243–256.

D'amico, G., Morabito, A., and Pagliaro, L. (1986). Six week prognostic indicators in upper gastrointestinal hemorrhage in cirrhotics. *Front. Gastrointestinal. Res.* **9**, 247.

D'amico, G., Pagliaro, L., and Bosch, J. (1995). The treatment of portal hypertension: A meta-analytic review. *Hepatology* **22**, 332–354.

Garcia-Tsao, G., Groszmann, R. J., Fisher, R. L., Conn, H. O., Atterbury, C. E., and Glickman M. (1985). Portal pressure, presence of gastroesophageal varices and variceal bleeding. *Hepatology* **5**, 419–424.

Graham, D. Y., and Smith, J. L. (1981). The course of patients after variceal hemorrhage. *Gastroenterology* **80**, 800–809.

Imperiale, T. F., and Chalasani, N. (2001). A meta-analysis of endoscopic variceal ligation for primary prophylaxis of esophageal variceal bleeding. *Hepatology* **33**, 802–807.

Imperiale, T. F., Teran, J. C., and McCullough, A. J. (1995). A meta-analysis of somatostatin versus vasopressin in the management of acute esophageal variceal hemorrhage. *Gastroenterology* **109**, 1289–1294.

Laine, L., and Cook, D. (1995). Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding: A meta-analysis 3. *Ann. Intern. Med.* **123**, 280–287.

Poynard, T., Cales, P., Pasta, L., *et al*. (1991). Beta-adrenergic-antagonist drugs in the prevention of gastrointestinal bleeding in patients with cirrhosis and esophageal varices: An analysis of data and prognostic factors in 589 patients from four randomized clinical trials. Franco-Italian Multicenter Study Group. *N. Engl. J. Med.* **324**, 1532–1538.

Sanyal, A. J., Freedman, A. M., Luketic, V. A., *et al*. (1997). The natural history of portal hypertension after transjugular intrahepatic portosystemic shunts. *Gastroenterology* **112**, 889–898.

Sanyal, A. J., Freedman, A. M., Luketic, V. A., *et al*. (1996). Transjugular intrahepatic portosystemic shunts for patients with active variceal hemorrhage unresponsive to sclerotherapy. *Gastroenterology* **111**, 138–146.

Portal Vein Thrombosis

PATRICK S. KAMATH
Mayo Clinic and Foundation

ascites Presence of fluid in the peritoneal cavity.
hepatic encephalopathy Neuropsychiatric manifestations associated with liver disease.
hypersplenism Reduction in two or more of the formed elements of blood as a manifestation of splenomegaly; associated with a normal bone marrow.
portal hypertension Elevation in pressure in the portal venous system. Bleeding from esophageal varices, ascites, and hepatic encephalopathy are complications of portal hypertension.
portal venous system Vessel pathway beginning and ending in the capillaries.

Portal vein thrombosis is a common cause of portal hypertension in the absence of chronic liver disease, especially in children. The clinical manifestations of portal vein thrombosis depend both on the extent and the duration of the thrombosis. Thus, patients may present with abdominal pain, diarrhea, or gastrointestinal bleeding.

NORMAL PORTAL CIRCULATION

The portal vein serves to drain almost the entire gastrointestinal tract, spleen, pancreas, and gallbladder. The portal vein is approximately 7 cm in length and courses in the hepatoduodenal ligament, usually dorsal to the bile duct and hepatic artery, and divides into two lobar veins, the left and the right, before entering the portal fissure. The superior pancreaticoduodenal and left gastric veins drain into the portal vein near its origin. The upper 5 cm of the portal vein usually receives no venous branches. The umbilical vein and the paraumbilical veins may drain into the left branch of the portal vein, whereas the cystic vein drains into the right portal vein.

The portal vein constitutes the major oxygen supply to the liver. Portal venous flow is approximately 1150 ml/min, as opposed to hepatic arterial flow, which is only 350 ml/min, resulting in total liver blood flow of 1500 ml/min. The portal vein supplies between 50 and 70% of the oxygen to the liver, especially during the fasting state.

PATHOGENESIS OF PORTAL VEIN THROMBOSIS

Portal vein thrombosis in children is usually a result of umbilical cord sepsis or follows catheterization of the umbilical vein. In adults, portal vein thrombosis is secondary to diseases that result in increased coagulability of the blood. These include heritable or acquired disorders of coagulation, cancer (especially hepatocellular cancer), intraabdominal sepsis, inflammatory bowel disease, and postoperative states. Portal vein thrombosis may result from extension of thrombosis in the vein following splenectomy. In young women, oral contraceptive use is an additional risk factor. In patients with cirrhosis of the liver, portal vein thrombosis occurs in less than 1% of patients. However, among patients with cirrhosis and with portal vein thrombosis, hepatocellular carcinoma occurs in up to 25% of patients.

CLINICAL PRESENTATION

Portal vein thrombosis is classified as acute, subacute, or chronic. Acute portal vein thrombosis manifests as abdominal pain that may be associated with diarrhea. The pain is nonspecific. Radiological imaging in these patients demonstrates thrombosis of the portal vein without significant collateral circulation. Symptoms are usually present for days. In subacute portal vein thrombosis, symptoms are present for days to weeks, and imaging studies show a collateral circulation in addition to the portal vein thrombosis. In chronic portal vein thrombosis, the portal vein may not be visualized and is replaced by an extensive network of collateral veins, the so-called cavernous transformation of the portal vein. When portal vein thrombosis occurs in patients with cirrhosis of the liver, additional clinical features include further deterioration in liver function manifesting as ascites, and worsening encephalopathy.

Physical examination of patients with portal vein thrombosis is nonspecific. Patients in whom thrombosis has extended into the smaller veins of the portal

 Copyright 2004, Elsevier (USA). All rights reserved.

circulation, especially those closely applied to the bowel, may have abdominal tenderness. When the small vessels are thrombosed, there is an increased risk of developing ischemia of the bowel that may progress to infarction and peritonitis. Such patients are more likely to have hemodynamic instability.

DIAGNOSIS

Routine blood tests are not helpful in making a diagnosis of portal vein thrombosis. Diagnosis is usually made with radiological investigations, such as Doppler ultrasonography or, preferably, computer tomography (CT) scans of the abdomen. Magnetic resonance imaging is helpful but not often used in the initial investigation of these patients. Mesenteric angiogram is the gold standard in demonstrating thrombosis in the portal vein, but is usually required only if the plan involves surgery to decompress the portal system. Investigations are also carried out to exclude inherited and acquired disorders of coagulation.

TREATMENT

Treatment of portal vein thrombosis is challenging. Long-term anticoagulation is recommended in all patients with portal vein thrombosis with an underlying thrombophilia. Anticoagulation is also recommended in patients with an acute or subacute presentation, provided these are no contraindications such as active bleeding. In the rare patient who presents within 24–48 hours of the onset of portal vein thrombosis, thrombolytic therapy in an attempt to dissolve the clot may be tried. Patients with chronic portal vein thrombosis present predominantly as portal hypertension manifesting with variceal bleeding and hypersplenism. In such patients with large esophageal varices, anticoagulation is not recommended. Patients who present with gastrointestinal bleeding can be treated either by injection of a sclerosant into the varices or by band ligation of the varices. Surgical shunts are used in patients in whom bleeding cannot be controlled by conservative measures; this involves connecting a branch of the portal venous system, either the splenic vein or the superior mesenteric vein, to a systemic vein, such as the renal vein or inferior vena cava.

See Also the Following Articles

Ascites • Hepatic Encephalopathy • Portal Hypertension and Esophageal Varices

Further Reading

Condat, B., Pessionem, F., Hillarre, S., Denninger, M. H., Guillin, M. C., Polliquin, M., Hadengue, A., Erlinger, S., and Valla, D. (2001). Current outcome of portal vein thrombosis in adults: Risk and benefit of anticoagulant therapy. *Gastroenterology* **120**, 490–497.

Kumar, S., Sarr, M. G., and Kamath, P. S. (2001). Mesenteric venous thrombosis. *N. Engl. J. Med.* **345**, 1683–1688.

Okuda, K., Ohnishi, K., Kimura, K., Matsutani, S., Sumida, M., Goto, N., Musha, H., Takashi, M., Suzuki, N., and Shinagava, T. (1985). Incidence of portal vein thrombosis in liver cirrhosis: An angiographic study in 708 patients. *Gastroenterology* **89**, 279–286.

Postprandial Motility

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

gastrocolic reflex Changes in the motility of the large intestine following ingestion of a meal.
migrating motor complex Motility pattern of the interdigestive state in the small intestine.
power propulsion Motility pattern that propels the intraluminal contents rapidly over long distances in the small and large intestine.

Ingestion of a nutrient meal initiates specialized patterns of postprandial motility in the stomach and small and large intestine. Postprandial motility in the reservoir of the upper stomach relaxes the gastric wall to accommodate the increasing volume during active ingestion of a meal. Postprandial motility in the antral region of the stomach triturates the contents to particles of sufficiently small size for effective emptying into the small intestine. Mixing movements are the characteristic motility pattern in the small intestine, and the increased incidence of mass movements and a generalized increase in mixing-like movements occur in the large intestine following ingestion of a meal.

INTRODUCTION

Following ingestion of a meal, when the gastric wall is not relaxing in response to the swallowing and filling process that initiates postprandial motility in the upper stomach, the muscles in the wall of the gastric reservoir contract, exerting controlled compressive forces on the contents of the upper stomach. In the antral stomach region, the postprandial motility process effects pulverization of food particles. Particulates in the stomach are not emptied until they are reduced to sizes less than about 5 mm. The reduced size of the particles increases the surface area for action by digestive enzymes in the small intestine. In the small intestine, mixing movements, the characteristic motility pattern of the fed state, are initiated. (The terms "digestive state" and "fed state" are used alternatively to describe the state of the gut after a meal.) The mixing movements start with the first ingestion of a nutrient meal and are coincident with termination of the motility pattern that characterizes the interdigestive state (i.e., the migrating motor complex). Mixing movements function to blend pancreatic, biliary, and intestinal secretions with nutrients in the small intestine and bring products of digestion into contact with the absorptive surfaces of the mucosa.

Power propulsion, a programmed motor event in the transverse and descending colon, is also associated with intake of a meal. This form of motor behavior fits the general pattern of neurally coordinated peristaltic propulsion and results in the mass movement of feces over extended distances toward the anus. Mass movements may be triggered by increased delivery of ileal contents into the ascending colon following a meal. Gastrocolic reflex is a term used to refer to the increased incidence of mass movements and a generalized increase in mixing-like movements that occur in the large intestine following ingestion of a meal.

MIXING MOVEMENTS

Mixing movements of the small intestine are also called segmenting movements or segmentation due to their appearance on X-ray films of the small intestine. The mixing pattern of motility is programmed by the enteric nervous system. In the mixing pattern, the behavior of the musculature is organized to propel luminal contents in both directions over short distances. This is in contrast to other forms of propulsive motility that move the luminal contents in one direction over extended lengths of intestine. Mixing movements consist of circumferential muscle contraction in segments separated on either end by relaxed receiving segments. The receiving segments appear as sacculations with increased cross-sectional diameter on X-ray images of the small intestine (Fig. 1). Mixing occurs as the contents are forcefully propelled into the receiving segments from both directions. Each segmental contraction and relaxed receiving segment reflects the occurrence of stereotypic peristalsis that does not propagate beyond a single segment. Mixing movements are, in effect, the occurrence of ultrashort peristalsis that is repeated at multiple sites along the intestine. As is the case for propulsion over greater distances in other patterns of motility, the basic peristaltic neural reflex (i.e., contraction above

 Copyright 2004, Elsevier (USA). All rights reserved.

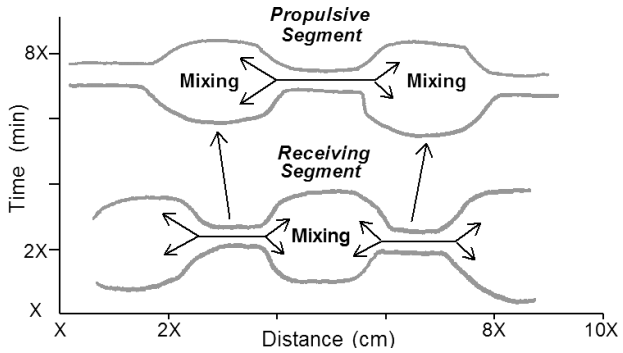


FIGURE 1 The postprandial pattern of small intestinal motility is characterized by mixing movements. Mixing movements consist of propulsive and receiving segments as described for the stereotypic pattern of behavior of the circumferential and longitudinal muscle layers during peristalsis. They can be viewed as short-distance peristaltic propulsion. The postprandial pattern of motility consists of propulsive segments separated by receiving segments, occurring with apparent randomness at many sites along the intestine. Mixing of the luminal contents occurs in the receiving segments. Over time, cyclic conversion of receiving segments to propulsive segments occurs coincident with the conversion of propulsive segments to receiving segments.

and relaxation of the intestinal wall below) underlies the mixing motility pattern. Postprandial segmentation is a mechanism for the mixing and stirring of luminal contents in the receiving segments. The enteric neural program for mixing is cyclic, with receiving segments converting to contracting segments and contracting segments becoming receiving segments. Segmenting movements occur with the same intervals as electrical slow waves or at multiples of the shortest slow-wave interval in the particular region of intestine.

Command signals transmitted from the brain to the small intestine by the vagus nerves are important for the conversion from the interdigestive motility pattern to the digestive pattern. After interruption of transmission in the vagus nerves, a larger quantity of ingested food is necessary for termination of the interdigestive motor pattern, and interruption of the migrating motor complex is often incomplete.

Evidence of vagal commands for the digestive motor pattern has been obtained in animals with cooling cuffs placed surgically around each vagus nerve. During the fed pattern, cooling and blockade of impulse transmission in the nerves results in interruption of the fed pattern of mixing movements. When the vagus nerves are blocked during the fed pattern, migrating motor complexes reappear in the intestine, but not in the stomach. With warming of the nerves and release of neural blockade, the fed pattern returns.

See Also the Following Articles

Barostat • Basic Electrical Rhythm • Colonic Motility • Duodenal Motility • Gastric Emptying • Gastric Motility • Gastro-colic Reflex • Ileal Brake • Migrating Motor Complex • Power Propulsion • Pylorus • Small Intestinal Motility

Further Reading

Hall, K., el-Sharkawy, T., and Diamant, N. (1986). Vagal control of canine postprandial upper gastrointestinal motility. *Am. J. Physiol.* **250**, G501–G510.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480. Lippincott Williams & Wilkins, Baltimore.

Pouchitis

DAVID B. SACHAR
Mount Sinai School of Medicine, New York

diversion colitis A characterisitic mucosal inflammation that typically occurs whenever the colon is excluded from the intestinal stream and that subsides when intestinal continuity is restored.

kock pouch (continent ileostomy) A reservoir constructed, following colectomy, from approximately 40 cm of ileum much like a pelvic pouch (q.v.), but with the modification of a one-way nipple valve accessed by intubation directly through the anterior abdominal wall.

probiotics Live microorganisms that confer health benefits through any one of a number of mechanisms, such as therapeutic modification of the enteric flora.

Pouchitis denotes mucosal inflammation occurring in an ileal reservoir following total colectomy, usually for ulcerative colitis or familial adenomatous polyposis. The incidence of this disorder is much higher in patients with underlying chronic idiopathic inflammatory bowel disease than in those with genetic neoplastic syndromes, but it occurs in both Kock pouches (continent ileostomies) and pelvic pouches. The condition is manifested clinically by increased stool frequency, urgency, diarrhea, bleeding, and abdominal pain and often by fever or other constitutional complications.

DIAGNOSIS

Since other conditions may produce the same clinical syndrome as pouchitis, definitive diagnosis requires endoscopic and histologic confirmation. Characteristic endoscopic features include diffuse mucosal erythema, edema, friability, hemorrhage, and ulceration. The histologic picture is typically one of acute and/or chronic inflammatory infiltrate, villous atrophy, and crypt hyperplasia.

DIFFERENTIAL DIAGNOSIS

Conditions that mimic pouchitis generally carry very different prognostic and therapeutic implications. It is therefore essential to distinguish pouchitis from other complications such as mechanical outflow obstruction, local sepsis, Crohn's disease, or intestinal dysmotility. A defunctionalized pouch may also manifest the mucosal changes associated with "diversion colitis," but these should regress once intestinal continuity is reestablished.

NATURAL HISTORY

Nearly half of all ulcerative colitis patients with ileal pouches will experience at least one acute episode of pouchitis. For almost any purpose of discussion, however, it is useful to consider pouchitis as occurring in several distinct phenotypic categories:

- single acute episode;
- one or two acute episodes per year;
- more than two acute episodes per year, each responding to medical therapy;
- frequent episodes requiring chronic maintenance therapy to prevent recurrence;
- chronic refractory pouchitis, not responding to medical therapy.

Although approximately 50% of ulcerative colitis patients with ileal pouches will not suffer even one bout of pouchitis during a 5- to 10-year follow-up, the remaining 50% will be distributed approximately as follows: single episode or only one or two episodes per year, 10%; more than two responsive episodes per year, 25%; chronic pouchitis requiring maintenance therapy, 10%; and chronic refractory pouchitis, 5%. These specific frequencies may of course vary from series to series, but it is generally agreed that most pouchitis cases are responsive to acute medical therapy (see below), with only approximately 10–15% of cases (i.e., approximately 5–7% of all ulcerative colitis patients with pouches) proving to be chronic and unremitting.

PATHOPHYSIOLOGY

The current understanding of the mechanisms underlying pouchitis is limited to theory and conjecture. Nonetheless, most current concepts are based on two undisputed observations. The first is that the overwhelming majority of cases, over 90%, are at least

 Copyright 2004, Elsevier (USA). All rights reserved.

initially responsive to antibiotics. Hence, there must be a microbiologic component to the disease. The second universal finding is that although pouchitis is experienced by nearly half of ulcerative colitis patients with ileal pouches, this complication is extremely rare (albeit not totally unreported) in familial adenomatous polyposis. Therefore, there must also be an element of underlying host susceptibility, presumably immunogenetic in nature. For this reason, studies have been exploring cytokine profiles, inflammatory mediators, volatile fatty acid levels, bile acid composition, and other permeability and vascular factors as putative contributors to the pathologic process.

RISK FACTORS

Ideally, it would be helpful to know in advance which patients were at particular risk for pouchitis and which were not. The only absolute certainty in this regard is that it is patients with underlying chronic idiopathic inflammatory bowel disease who are most vulnerable; patients with familial adenomatous polyposis only rarely develop this complication.

Within the inflammatory bowel disease population, however, it is less clear what the risk factors are. Several putative markers of an immunogenetic disposition to pouchitis might include primary sclerosing cholangitis (PSC), high preoperative perinuclear anti-neutrophil cytoplasmic antibody titers, extensive backwash ileitis, and a strong preoperative history of extraintestinal complications. As with the original ulcerative colitis, smoking may exert some type of "protective" effect against pouchitis. Crohn's disease-like complications of an ileal pouch might also be anticipated in patients who had unmistakable Crohn's disease preoperatively, especially if the small bowel had been involved; but not every case of chronic refractory pouchitis, even when accompanied by granulomas or transmural inflammation, is *prima facie* evidence of preexisting Crohn's disease.

TREATMENT

Initially, virtually all cases of pouchitis will respond favorably to treatment with metronidazole or other broad-spectrum antibiotics, including ciprofloxacin. In fact, some investigators require such a therapeutic response to be manifested before they will even accept a definitive diagnosis of pouchitis. As noted above under Natural History, all but approximately 10–15% of cases with pouchitis can be managed exclusively with antibiotic therapy, whether intermittent or chronic; the worst dilemma in managing this complication is how to treat the small but not insignificant minority of cases that are refractory to antibiotics.

In these truly refractory cases, medical treatment reflects despair more than reliable evidence. Oral and topical aminosalicylates and steroids are often used with anecdotal impressions of success; some patients seem to respond to oral anti-metabolites, but even less evidence supports the occasional resort to oral cholestyramine or topical cyclosporin, glutamine, butyrate, or Kaopectate. In any event, all such medical approaches must be disappointing to the extent that they are reminiscent of the regimens that the colectomy was designed to eliminate. A potentially more promising approach to prophylaxis against recurrent pouchitis has been suggested by Italian investigators using probiotic therapy.

As many as half the cases of chronic refractory pouchitis may ultimately come to further surgery—either pouch revision or "salvage" or complete excision with conversion to the standard Brooke ileostomy, an alternative that may have been initially less appealing to the patient, but that may ultimately restore a more acceptable quality of life.

See Also the Following Articles

Colectomy • Colitis, Ulcerative • Crohn's Disease • Familial Adenomatous Polyposis (FAP) • Ileoanal Pouch

Further Reading

Gionchetti, P., Rizzello, F., Venturi, A., Brigidi, P., Matteuzzi, D., Bazzocchi, G., Poggioli, G., Miglioli, M., and Campieri, M. (2000). Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: A double-blind, placebo-controlled trial. *Gastroenterology* **119**, 305–309.

Lohmuller, J. L., Pemberton, J. H., Dozois, R. R., Ilstrup, D., and van Heerden, J. (1990). Pouchitis and extraintestinal manifestations of inflammatory bowel disease after ileal pouch–anal anastomosis. *Ann. Surg.* **211**, 622–627.

Madden, M. V., McIntyre, A. S., and Nicholls, R. J. (1994). Double-blind crossover trial of metronidazole versus placebo in chronic unremitting pouchitis. *Digest. Dis. Sci.* **39**, 1193–1196.

Nicholls, R. J., and Banerjee, A. K. (1998). Pouchitis: Risk factors, etiology, and treatment. *World J. Surg.* **22**, 347–351.

Phillips, S. F. (1998). Pouchitis. Up To Date, Inc. Available at http://www.uptodate.com.

Sandborn, W., McLeod, R., and Jewell, D. (2000). Pharmacotherapy for inducing and maintaining remission in pouchitis. *Cochrane Database Syst. Rev.*

Sartor, R. B. (2000). Probiotics in chronic pouchitis: Restoring luminal microbial balance. *Gastroenterology* **119**, 584–587.

Shepherd, N. A., Hulten, L., Tytgat, G. N., Nicholls, R. J., Nasmyth, D. G., Hill, M. J., Fernandez, F., Gertner, D. J., Rampton, D. S., Hill, M. J., *et al.* (1989). Pouchitis. *Int. J. Colorectal Dis.* **4**, 205–229.

Stahlberg, D., Gullberg, K., Liljeqvist, L., Hellers, G., and Lofberg, R. (1996). Pouchitis following pelvic pouch operation for ulcerative colitis: Incidence, cumulative risk, and risk factors. *Dis. Colon Rectum* **39**, 1012–1018.

Power Propulsion

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

manometric recording Graphic measure of changes in pressure in localized regions of the digestive tract, indicating contractile behavior of the musculature.

migrating motor complex Specific pattern of small intestinal motility that begins when digestion of a meal is complete and ends with intake of the next meal; also called interdigestive motility.

mixing movements Specific pattern of small intestinal motility that starts with the ingestion of a meal and accomplishes mixing of the contents in the lumen; also called digestive motility.

Power propulsion is a motility pattern that is specialized for the rapid transport of contents over long distances in the small and large intestine. It appears on mechanical records of intestinal motility (e.g., manometric recordings) as strong, long-lasting circular muscle contractions that propagate for extended distances along the bowel. These "giant" migrating contractions are considerably stronger than the circular muscle contractions that occur during the migrating motor complex or the mixing movements in the small intestine.

INTRODUCTION

Power propulsion, the motility pattern of giant migrating circular muscle contractions that last 18–20 seconds and span several cycles of the electrical slow waves, is a component of a highly efficient propulsive mechanism that rapidly strips the lumen clean. The propulsive movement travels at about 1 cm/sec over long lengths of intestine.

Intestinal power propulsion differs from peristaltic propulsion during the migrating motor complex and the contractions of mixing movements in that the circular contractions in the propulsive segments are stronger and propagation occurs over longer reaches of intestine. The circular contractions are not time locked to the electrical slow waves and probably reflect strong activation of the muscle by excitatory motor neurons. Power propulsion represents another of the motility programs stored in the program library of the enteric nervous system.

PHYSIOLOGIC SIGNIFICANCE

Noxious stimulation of the mucosa starts the neural program for power propulsion in the small and large intestine. Power propulsion starts in the midjejunum and travels toward the stomach during vomiting; otherwise, the direction of travel in both the small and large intestine is in the anal direction. Abdominal cramping sensations and diarrhea are associated with power propulsion when it occurs in response to luminal events that threaten whole-body integrity. Application of irritants to the mucosa, the introduction of luminal parasites, release of enterotoxins from pathogenic bacteria, allergic reactions, and exposure to ionizing radiation all trigger the response. This suggests that power propulsion is a defense mechanism in both the upper and lower regions of the intestinal tract. It is a protective adaptation for rapid clearance of undesirable elements from the intestinal lumen. Oral clearance from the upper small intestine is achieved by vomiting of the material. Clearance from the lower intestinal tract is by way of watery feces.

Aside from its involvement in intestinal defense in potentially pathologic circumstances, operation of the power propulsion motility program also accomplishes mass movement of intraluminal material in normal states, especially in the large intestine. Mass movements in the colon following ingestion of a meal and during defecation are produced by the power propulsion program and reflect normal physiology.

See Also the Following Articles

Basic Electrical Rhythm • Borborygmus • Colonic Motility • Migrating Motor Complex • Postprandial Motility • Small Intestinal Motility

Further Reading

Mathias, J., and Clench, M. (1989). Alterations of small intestine motility by bacteria and their enterotoxins. *In* "Handbook of Physiology. The Gastrointestinal System I" (J. Wood, ed.), Sect. 6, Vol. 1, Pt. 2, pp. 1153–1177. American Physiological Society, Bethesda.

 Copyright 2004, Elsevier (USA). All rights reserved.

Sarna, S., and Lang, I. (1989). Motor and myoelectric activity associated with vomiting, regurgitation, and nausea. *In* "Handbook of Physiology, The Gastrointestinal System I" (J. Wood, ed.), Sect. 6, Vol. 1, Pt. 2, pp. 1179–1198. American Physiological Society, Bethesda.

Sarna, S., and Otterson, M. (1993). Myoelectric and contractile activities. *In* "Atlas of Gastrointestinal Motility in Health and Disease" (M. Schuster, ed.), pp. 3–42. Williams & Wilkins, Baltimore.

Wood, J. (1998). Enteric neuropathobiology. *In* "Functional Disorders of the Gut" (S. Phillips and D. Wingate, eds.), pp. 19–42. Churchill Livingstone, New York.

Wood, J. (2002). Neuro-pathophysiology of IBS. *J. Clin. Gastroenterol.* (in press).

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480. Lippincott Williams & Wilkins, Baltimore.

Prader–Willi Syndrome

WILLIAM B. ZIPF AND BELINDA PINYERD

Ohio State University and Central Ohio Pediatric Endocrinology and Diabetes Services, Columbus

centromere The site on a chromosome that pulls the chromosome toward one of the poles of the spindle during mitosis and meiosis; it is also the point of attachment of the sister chromatids.

imprinting A germ-line process that "presets" or predetermines the potential of a transmitted gene to be active or inactive without changing the actual sequence of the base pairs. It presumably reflects a modification of the DNA or proteins in such a way as to preset the activity of genes in the embryo.

methylation The process of adding a methyl group onto an existing molecule of a base pair that changes (often impedes) the ability of nuclear enzymes to "read" the genetic code.

provocative pituitary testing Analysis of the pituitary with specific medications previously shown to elicit a release of one or more pituitary hormones in a reproducible manner characteristic of normal or disease states.

uniparental disomy Two exact copies of the same chromosome or gene cluster from one parent.

Prader–Willi syndrome is a syndrome caused by a genetic abnormality located on the long arm of chromosome 15 near the centromere. The condition is characterized mainly by hyperphagia, severe obesity, mental retardation, short stature, hypogonadism, hypotonia, behavioral abnormalities, and shortened life span if untreated. The exact cause of the condition is unknown, although abnormal hypothalamic function is thought to be a primary result of the chromosome abnormality with resultant abnormalities in growth, metabolism, and development. No cure for the condition is available. However, with early diagnosis and aggressive management of the clinical and medical problems, most of the major and life-threatening medical problems associated with this condition can be either avoided or mitigated.

INTRODUCTION

Prader–Willi syndrome (PWS) is the single most common genetic cause of obesity. The obesity of PWS is a consequence of both severe hyperphagia and decreased overall metabolism. However, obesity is just one component of this extremely complex condition. Overall, the condition is characterized by a constellation of neurological, autonomic nervous system, hypothalamic, endocrine, and behavioral abnormalities. Scientific and medical interest in this condition is increasing because improved understanding may lead to elucidation of the genetic influences over weight control, obesity, and obesity co-morbidities. Descriptions of PWS date as far back as the 17th century. A famous painting of Eugenia Martinez Vallejo by Juan Carreno de Miranda (*The Monstrua*) depicts a girl with the condition prior to its discovery. The first report of PWS, published in 1956 by doctors A. Prader, H. Willi, and A. Labhart, described a syndrome characterized by obesity, short stature, a severe lack of muscle tone during infancy that persists into adulthood, delayed and incomplete puberty, amenorrhea in women, hypogonadism with cryptorchidism in men, and mental retardation.

Copyright 2004, Elsevier (USA). All rights reserved.

Sarna, S., and Lang, I. (1989). Motor and myoelectric activity associated with vomiting, regurgitation, and nausea. *In* "Handbook of Physiology, The Gastrointestinal System I" (J. Wood, ed.), Sect. 6, Vol. 1, Pt. 2, pp. 1179–1198. American Physiological Society, Bethesda.

Sarna, S., and Otterson, M. (1993). Myoelectric and contractile activities. *In* "Atlas of Gastrointestinal Motility in Health and Disease" (M. Schuster, ed.), pp. 3–42. Williams & Wilkins, Baltimore.

Wood, J. (1998). Enteric neuropathobiology. *In* "Functional Disorders of the Gut" (S. Phillips and D. Wingate, eds.), pp. 19–42. Churchill Livingstone, New York.

Wood, J. (2002). Neuro-pathophysiology of IBS. *J. Clin. Gastroenterol.* (in press).

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480. Lippincott Williams & Wilkins, Baltimore.

Prader–Willi Syndrome

WILLIAM B. ZIPF AND BELINDA PINYERD

Ohio State University and Central Ohio Pediatric Endocrinology and Diabetes Services, Columbus

centromere The site on a chromosome that pulls the chromosome toward one of the poles of the spindle during mitosis and meiosis; it is also the point of attachment of the sister chromatids.

imprinting A germ-line process that "presets" or predetermines the potential of a transmitted gene to be active or inactive without changing the actual sequence of the base pairs. It presumably reflects a modification of the DNA or proteins in such a way as to preset the activity of genes in the embryo.

methylation The process of adding a methyl group onto an existing molecule of a base pair that changes (often impedes) the ability of nuclear enzymes to "read" the genetic code.

provocative pituitary testing Analysis of the pituitary with specific medications previously shown to elicit a release of one or more pituitary hormones in a reproducible manner characteristic of normal or disease states.

uniparental disomy Two exact copies of the same chromosome or gene cluster from one parent.

Prader–Willi syndrome is a syndrome caused by a genetic abnormality located on the long arm of chromosome 15 near the centromere. The condition is characterized mainly by hyperphagia, severe obesity, mental retardation, short stature, hypogonadism, hypotonia, behavioral abnormalities, and shortened life span if untreated. The exact cause of the condition is unknown, although abnormal hypothalamic function is thought to be a primary result of the chromosome abnormality with resultant abnormalities in growth, metabolism, and development. No cure for the condition is available. However, with early diagnosis and aggressive management of the clinical and medical problems, most of the major and life-threatening medical problems associated with this condition can be either avoided or mitigated.

INTRODUCTION

Prader–Willi syndrome (PWS) is the single most common genetic cause of obesity. The obesity of PWS is a consequence of both severe hyperphagia and decreased overall metabolism. However, obesity is just one component of this extremely complex condition. Overall, the condition is characterized by a constellation of neurological, autonomic nervous system, hypothalamic, endocrine, and behavioral abnormalities. Scientific and medical interest in this condition is increasing because improved understanding may lead to elucidation of the genetic influences over weight control, obesity, and obesity co-morbidities. Descriptions of PWS date as far back as the 17th century. A famous painting of Eugenia Martinez Vallejo by Juan Carreno de Miranda (*The Monstrua*) depicts a girl with the condition prior to its discovery. The first report of PWS, published in 1956 by doctors A. Prader, H. Willi, and A. Labhart, described a syndrome characterized by obesity, short stature, a severe lack of muscle tone during infancy that persists into adulthood, delayed and incomplete puberty, amenorrhea in women, hypogonadism with cryptorchidism in men, and mental retardation.

Copyright 2004, Elsevier (USA). All rights reserved.

GENETICS

PWS is seen in all ethnic and socioeconomic groups and in all countries and is equally common in females and males, with an incidence between 1 : 10,000 and 1 : 15,000. PWS is an autosomal dominant disorder caused most often (70%) by a deletion of the paternal chromosome 15 at band q11–q13. In approximately 25% of individuals, PWS is a result of maternal uniparental disomy (UPD) of chromosome 15. In this situation, two copies of the maternal chromosome are inherited with no paternal contribution. Without the presence of a chromosome donated from the father, normal imprinting on the two maternally donated chromosomes leads to the absence of gene expression in this interval. The other 5% of patients may have abnormalities in the mechanism of imprinting (leading to the absence of gene expression from the paternally donated chromosome) or translocations affecting chromosome 15q11–q13.

CLINICAL PRESENTATION

The hallmark features of the syndrome include infant hypotonia, delayed development, childhood-onset hyperphagia and obesity, short stature, muscle hypotonia, characteristic facieses, delayed and incomplete puberty, small hands and/or feet, and mild to moderate mental retardation (IQ of 65–70). Many other important but less frequent findings have also been identified. Because there is variability among individuals with PWS, clinical diagnostic criteria (see Table I) have been established to raise diagnostic suspicion. The presence of these clinical findings is an indication for further genetic studies and genetic testing should accurately diagnosis 99% of these individuals. Since the clinical picture of an individual with PWS changes with age, the diagnostic criteria need to be considered in this context. These changes are not only important to take into account when making the diagnosis, but also in management as the co-morbidities and problems change.

In the newborn or infant, the classic presentation of PWS is profound hypotonia, lethargy, diminished deep tendon reflexes, poor feeding, below-average weight, and often a history of fetal inactivity. The infants are so “floppy” that they have little facial expression, an abnormal, weak, or absent cry, and gross motor development is delayed. The infants often experience failure to thrive because of severe feeding problems characterized by an extremely weak suck. This problem can remain severe for months and persist for years; often the severity can require special feeding interventions such as nasogastric food supplementation. Some time in the second year of life, motor function begins to improve. Gross motor skills are acquired late but by 2 years most children with PWS are walking.

In the toddler years (some time between 1 and 3 years of age), the appetite abnormalities begin to emerge. Initially, parents and health care providers are delighted to see children eat after struggling with feeding problems during the first year of life. At approximately 2 years of age, excessive weight gain is observed and later the tendency and risks for obesity become obvious as these children become more focused on food.

During the early and mid-childhood years, the physical growth abnormalities of PWS and the behavior characteristics of this syndrome begin to become apparent. There is increased body fat that is central in distribution and there is slowing of linear growth. Motor development continues to be delayed, the characteristic speech and language problems begin to emerge, and the learning disabilities are more easily identified. Growth of the PWS child begins to diverge from normal. Between 3 and 13 years of age, the 50th percentile for height corresponds to the 5th percentile in the normal population. At this point, PWS can be clearly distinguished from endogenous obesity in which the linear growth rate is accelerated and final height is normal or increased. Increased body fat and low muscle mass are also observed in PWS children. This unique pattern of growth is consistent with an abnormality in growth hormone (GH) secretion or action and studies of GH secretion have documented abnormally low GH responses to standard provocative pituitary GH stimulation testing in children and adults with PWS.

By adolescence, constant efforts to satisfy hunger result in aggressive and bizarre food-seeking activities. Hoarding food is common, as is stealing or sneaking to circumvent dietary restrictions. Complete control of the environment is needed with respect to access to food. Locks on refrigerator and cupboards are necessary and monitoring for nonfood and poisonous ingestions may be necessary. Accompanying the uncontrollable hunger is decreased calorie utilization, largely due to low muscle mass, low muscle tone, and corresponding inactivity. This contributes to the obese condition and increases the propensity for development of sleep apnea and Pickwickian syndrome.

Emotional liability and obsessive and compulsive behaviors along with intolerance to frustration are characteristic of the PWS adolescent. The obsessive–compulsive behaviors can manifest as persistent skin picking. Many adolescents with PWS develop personality problems ranging from being dull, lethargic, and indifferent to being clever, secretive, and manipulative. Transitioning from one activity to

TABLE I Published Diagnostic Criteria for Prader–Willi Syndrome

Major criteria

1. Neonatal and infantile central hypotonia with poor suck, gradually improving with age
2. Feeding problems in infancy with need for special feeding techniques and poor weight gain/failure to thrive
3. Excessive or rapid weight gain on weight-for-length chart (excessive is defined as crossing two centile channels) after 12 months but before 6 years of age; central obesity in the absence of intervention
4. Characteristic facial features with dolichocephaly in infancy, narrow face or bifrontal diameter, almond-shaped eyes, small-appearing mouth with thin upper lip, down-turned corners of the mouth (3 or more are required)
5. Hypogonadism—with any of the following, depending on age:
   a. Genital hypoplasia (male: scrotal hypoplasia, cryptorchidism, small penis and/or testes for age (<5th percentile); female: absence or severe hypoplasia or labia minora and/or clitoris)
   b. Delayed or incomplete gonadal maturation with delayed pubertal signs in the absence of intervention after 16 years of age (male: small gonads, decreased facial and body hair, lack of voice change; female: amenorrhea/oligomenorrhea after age 16)
6. Global developmental delay in a child <6 years of age; mild to moderate mental retardation or learning problems in older children
7. Hyperphagia/food foraging/obsession with food
8. Deletion 15q11–13 on high resolution (>650 bands) or other cytogenetic molecular abnormality of the Prader–Willi chromosome region, including maternal disomy

Minor criteria

1. Decreased fetal movement or infantile lethargy or weak cry in infancy, improving with age
2. Characteristic behavior problems—temper tantrums, violent outbursts, and obsessive–compulsive behavior; tendency to be argumentative, oppositional, rigid, manipulative possessive, and stubborn; perseverating, stealing, and lying (5 or more of these symptoms required)
3. Sleep disturbance and sleep apnea
4. Short stature for genetic background by age 15 (in the absence of growth hormone intervention)
5. Hypopigmentation—fair skin and hair compared with family
6. Small hands (<25th percentile) and/or feet (<10th percentile) for height and age
7. Narrow hands with straight ulnar borders
8. Eye abnormalities (esotropia, myopia)
9. Thick viscous saliva with crusting at the corners of the mouth
10. Speech articulation defects
11. Skin-picking

Supportive findings

1. High pain threshold
2. Decreased vomiting
3. Temperature instability in infancy or altered temperature sensitivity in older children and adults
4. Scoliosis and/or kyphosis
5. Early adrenarche
6. Osteoporosis
7. Unusual skill with jigsaw puzzles
8. Normal neuromuscular studies

To Score: Major criteria are weighted at 1 point each and minor criteria are weighted at $\frac{1}{2}$ point each. Supportive findings increase the certainty of diagnosis but are not scored. For children 3 years of age or younger, 5 points are required, 4 of which should come from the major group. For children >3 years of age and for adults, a total score of 8 is required and major criteria must comprise 5 or more points of the total score.

another is difficult and signs of depression and occasionally psychotic episodes may emerge.

At this age, the obesity will become more problematic unless dietary restrictions have been enforced. Left untreated, adolescents and adults with PWS can become morbidly obese and develop the co-morbidities of diabetes mellitus, hypertension, and cardiopulmonary insufficiency.

Physically, sexual development remains incomplete and pubertal growth is reduced. Without exogenous hormone intervention, males and females usually do not progress beyond Tanner stage II. Females may have occasional spotting but few have normal menstrual cycles. Reduced bone mineral density is typically present and this may lead to osteoporosis. These children are also at increased risk for scoliosis and/or kyphosis. In the absence of GH therapy, short stature (often severe) will be present.

The picture of PWS in the adult is dependent on the severity of the multiple problems and the intensity of

treatment during the childhood and adolescent years. Adults with PWS who have not had the benefit of early identification and treatment with specialized care (aggressive diet management, activity program, and sex steroid and growth hormone replacement) are characterized by severe obesity, decreased muscle mass, Pickwickian syndrome, sleep apnea, reduced bone mineral density, mental retardation, and behavior and psychiatric disorders. Left untreated, the life span of PWS individuals is two to three decades, with death accompanying the complications of severe weight gain including diabetes, respiratory insufficiency, and cardiac failure.

DIAGNOSTIC TESTING

Gunay-Aygun and colleagues reviewed the sensitivity of PWS diagnostic criteria and proposed revised criteria for DNA testing. From birth to 2 years, any infant with hypotonia and poor suck should have DNA testing for PWS. From age 2 to 6 years, any child with hypotonia and a history of poor suck and global developmental delay should have DNA testing. From age 6 to 12 years, any child with a history of hypotonia and poor suck, global developmental delay, and excessive eating with central obesity should be tested for PWS. In the adult, the clinical pattern can be more varied and important historical facts may be unavailable. For adults, the authors recommend methylation testing for any individual with mental retardation, severe obesity, a history of hyperphagia, low muscle tone, hypogonadism, and short stature or adult height less than would be predicted from genetic background. Definitive diagnosis of PWS is made via methylation analysis since it detects all three groups of molecular defects described above. If methylation analysis is not available, high-resolution chromosome analysis and/or FISH (fluorescence *in situ* hybridization) can be the initial test. However, methylation analysis and (or) analysis of genomic DNA are (is) necessary to delineate the parent-of-origin for the deleted region. If biparental inheritance is identified, then PWS is ruled out. If the methylation pattern is abnormal, FISH can be used to document a deletion and/or microsatellite probes can be used to confirm maternal UPD. Abnormal methylation and negative FISH and UPD studies indicate an imprinting defect.

TREATMENT

The treatment of PWS is a challenge given the extent of the functional and metabolic limitations associated with the condition. Since patients with PWS require a variety of interventions to optimize their health, a comprehensive plan approach is required. Although it is essential that a primary physician be identified to manage this complex problem, it is equally essential that the primary physician have available a team of experienced physicians, including those with specialties in neurology, psychiatry, urology, orthopedics, and gynecology, with other professionals from nursing, dietetics, genetics, social services, and psychology that are familiar with the PWS condition. Ancillary resources that are necessary will in part be age-dependent, but ultimately these will include social services, early intervention programs for children, physical therapy, occupational therapy, and age-appropriate educational/vocational services. A controlled and caring environment with regular routines along with pharmacological therapy may be initiated to stabilize mood and behavior and improve self-esteem. For adults, self-care has not been shown to be possible and group home environments with strict dietary controls are essential.

Since the severity of the associated health problems is directly related to the degree of obesity, nutritional intervention is the single key component of any treatment program. Appetite suppressants are ineffective in controlling overeating in individuals with PWS. Bariatric surgery has not been generally successful and can be associated with complications due to appetite abnormality and mental retardation. Weight management needs to include complete and absolute control over access to food and a calorie-restrictive diet, usually approximately 10 cal/cm of height. Dietary restrictions often need to be maintained for life in the range of 900 to 1200 kcal/day and rarely can exceed 1400 kcal/day. The low-calorie diet requires attention to the prevention of vitamin, essential fatty acid, and calcium deficiencies. Behavior modification plans that reward positive behaviors with respect to weight management and activity programs may be beneficial for some patients. Programs should also include family therapy and behavioral management components that emphasize environmental controls while simultaneously encouraging social integration and independence. Structured physical activity programs to increase energy expenditure and build muscle mass can be initiated and may sustain the individual with PWS if it is well-defined and easy to accomplish.

Endocrine abnormalities are a significant part of the medical picture of PWS. These abnormalities relate to the primary abnormalities of the hypothalamus and secondarily to the complications of the obesity and diet modifications. Pituitary function abnormalities include hypogonadotropic hypogonadism with absent, delayed, and incomplete puberty. On occasion, precocious but

incomplete puberty has been observed. GH secretion deficiency has also been documented to be part of this hypothalamic–pituitary function abnormality. Other endocrine abnormalities that have been identified include abnormalities of secretion of pancreatic secretion of insulin and pancreatic polypeptide with increased risk of diabetes. The abnormalities of glucose homeostasis and insulin secretion are not simply a consequence of the obese condition and associated insulin resistance, but have been hypothesized to be secondary to abnormalities in normal autonomic control of islet cell secretion. Thyroid deficiencies may occur in PWS individuals, although they have not been identified as common problems. PWS individuals are at high risk for osteoporosis. This can be a major problem as a consequence of the lack of spontaneous secretion of sex steroids during adolescence, growth hormone deficiency, and the dietary restrictions that limit milk and calcium intake.

Treatment of the endocrine abnormalities includes appropriately timed replacement of sex steroids. This may include human chorionic gonadotropin treatment of the infant with cryptorchidism/hypogonadism and use of estrogens and androgens at the time of adolescence. Early recognition of symptoms and signs of diabetes can allow for aggressive intervention with weight control and diet as primary treatment. Several controlled studies have now been published on the beneficial effect of GH in children with PWS and as a result GH is a Food and Drug Administration-approved therapy for growth failure in children with PWS. Daily subcutaneous administration of GH replacement (at a weekly dose of 0.24 mg/kg body weight) normalizes linear growth, promotes increases in lean body mass, decreases fat mass, and improves bone mineral density, all of which are beneficial to weight management. Therefore, growth rates of children with PWS should be monitored and referral for growth hormone evaluation is appropriate if the child's growth velocity decreases or if height is less than the third percentile. Further investigation is needed to determine whether adult GH therapy would be of benefit to the adult PWS patient.

Recent studies involving children and adults with PWS report fasting ghrelin plasma concentrations significantly higher in PWS subjects than in obese and lean control subjects. Ghrelin, a hormone predominantly secreted by the stomach, increases before every meal and decreases after nutrient intake, suggesting a role in meal initiation. Plasma ghrelin levels are also lower in obese individuals than in lean individuals. Future studies are needed regarding the role of ghrelin as an orexigenic factor driving the insatiable appetite of PWS patients and whether ghrelin antagonists could effectively reduce food intake.

As patients with PWS become older and more independent, behavioral disorders, including obsessive–compulsive disorders and psychoses, can become major issues. Serotonin-specific reuptake inhibitors may be appropriate to stabilize irritability and perseveration as initial therapies.

With appropriate intervention, the very poor clinical course can be changed for the individual with the PWS condition. Individuals who have had the benefit of early diagnosis and interventions will have more normal although generally still excessive weight, less severe shortness of stature, and persistent muscle hypotonia compared to normal, but significantly improved mobility and activity patterns than would otherwise be possible. With proper care, the behavioral problems, though significant, are manageable. The expected life spans of PWS individuals who have received anticipatory care and appropriate attention to medical problems have yet to be determined, but can be beyond 30 to 40 years and are associated with the absence of the major co-morbidities and a markedly improved quality of life.

See Also the Following Articles

Appetite • Growth Hormone • Obesity, Treatment of

Further Reading

Angulo, M., Castro-Magana, M., and Uy, J. (1991). Pituitary evaluation and growth hormone treatment in Prader–Willi syndrome. *J. Pediatr. Endocrinol.* **4**, 167–173.

Aughton, D. J., and Cassidy, S. B. (1990). Physical features of Prader–Willi syndrome in neonates. *Am. J. Dis. Child.* **144**, 1251–1254.

Butler, M. G., and Meaney, F. (1991). Standards for selected anthropometric measurements in Prader–Willi syndrome. *Pediatrics* **88**, 853–860.

Carrel, A. L., Myers, S. E., Whitman, B. Y., and Allen, D. B. (1999). Growth hormone improves body composition, fat utilization, physical strength and agility, and growth in Prader–Willi syndrome: A controlled study. *J. Pediatr.* **134**, 215–221.

Cassidy, S. B. (1997). Prader–Willi syndrome. *J. Med. Genet.* **34**, 917–923.

Cummings, D. E., Purnell, J. Q., Frayo, R. S., Schmidova, K., Wisse, B. E., and Weigle, D. S. (2001). A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. *Diabetes* **50**, 1714–1719.

DelParigi, A. (2002). High circulating ghrelin: A potential cause for hyperphagia and obesity in Prader–Willi syndrome. *J. Clin. Endocrinol. Metab.* **87**, 5461–5464.

Eiholzer, U., Nordmann, Y., l'Allemand, D., Schlumpf, M., Schmid, S., and Kromeyer-Hauschild, K. (2003). Improving body composition and physical activity in Prader–Willi syndrome. *J. Pediatr.* **142**, 73–78.

Epstein, L. H., McCurley, J., Valoski, A., and Wing, R. R. (1990). Growth in obese children treated for obesity. *Am. J. Dis. Child.* **144**, 1360–1364.

Gunay-Aygun, M., Schwartz, S., Heeger, S., O'Riordan, M. A., and Cassidy, S. B. (2001). The changing purpose of Prader–Willi syndrome clinical diagnostic criteria and proposed revised criteria. *Pediatrics* **108**, E92.

Haqq, A. M., Farooqi, I. S., O'Rahilly, S., Stadler, D. D., Rosenfeld, R. G., Pratt, K. L., LaFranchi, S. H., and Purnell, J. Q. (2003). Serum ghrelin levels are inversely correlated with body mass index, age, and insulin concentrations in normal children and are markedly increased in Prader–Willi syndrome. *J. Clin. Endocrinol. Metab.* **88**, 174–178.

Holm, V. A., Cassidy, S. B., Butler, M. G., *et al.* (1993). Prader–Willi syndrome: Consensus diagnostic criteria. *Pediatrics* **91**, 398–402.

Ritzen, E. M., Lindgren, A. C., Hagenas, L., Marcus, C., Muller, J., and Blichfeldt, S. (1999). Growth hormone treatment of patients with Prader–Willi syndrome. Swedish Growth Hormone Advisory Group. *J. Pediatr. Endocrinol. Metab.* **12**, 345–349.

Schuster, D., Oswei, K., and Zipf, W. B. (1996). Characterization of alterations in glucose and insulin metabolism in Prader–Willi subjects. *Metabolism* **45**, 1514–1520.

Stadler, D. D. (1995). Nutritional management. *In* "Management of Prader–Willi Syndrome" (L. R. Greenswag and R. Alexander, eds.), pp. 88–114. Springer-Verlag, New York.

Whitman, B., Greenswag, L., and Boyt, M. (1998). The use and impact of psychotropic medications for managing behavior in persons with Prader–Willi syndrome. Proceedings of the 13th Annual Prader–Willi Scientific Conference, July 22, 1998, Columbus, OH.

Wren, A. M., Seal, L. J., Cohen, M. A., Brynes, A. E., Frost, G. S., Murphy, K. G., Dhillo, W. S., Ghatei, M. A., and Bloom, S. R. (2001). Ghrelin enhances appetite and increases food intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 5992.

Pregnancy and Gastrointestinal Disease

RENE DAVILA AND CLAUDIO R. TOMBAZZI
University of Tennessee Health Science Center

hyperemesis gravidarum Pernicious condition in which the pregnant patient develops severe and intractable nausea and vomiting associated with nutritional and fluid/electrolyte deficiencies.

trimester Portion of a pregnancy in time; first, second, and third trimesters roughly correspond to months 1–3, 4–6, and 6–9.

The development of common gastroenterological conditions during pregnancy may present a challenge to physicians who are not familiar with the altered physiology of the gravid state. Knowledge of the most common gastrointestinal diseases in pregnancy is critical to the health of both mother and child.

INFLAMMATORY BOWEL DISEASE

The rates of fertility, congenital malformation, stillbirths, and spontaneous abortions in women with ulcerative colitis (UC) are comparable to those of unaffected women. Ulcerative colitis may have its onset in the first two trimesters of gestation. Women with preexisting UC and whose UC is asymptomatic in the prenatal period usually do not experience exacerbation of disease during pregnancy. On the other hand, women who have active disease before pregnancy may experience worsening of symptoms during the first trimester and significant improvement during the second trimester. Most pregnancies of women with ulcerative colitis result in delivery of normal full-term babies.

Crohn's disease is rarely diagnosed initially during pregnancy. The fertility rate in women with Crohn's disease is affected by the multiple clinical complications of the disease, such as perineal scarring, dyspareunia, and decreased libido. Women with active Crohn's disease have an increased risk for premature birth. There are no increases, however, in congenital malformations, stillbirths, or spontaneous abortions. Treatment of Crohn's disease during pregnancy with sulfasalazine, 5-aminosalicylic acid, or corticosteroids has been undertaken without risk for fetal abnormalities.

The use of immunosuppressive therapy with azathioprine or 6-mercaptopurine has generally been avoided during pregnancy, although published data have not suggested an increased risk of complications.

 Copyright 2004, Elsevier (USA). All rights reserved.

Epstein, L. H., McCurley, J., Valoski, A., and Wing, R. R. (1990). Growth in obese children treated for obesity. *Am. J. Dis. Child.* **144**, 1360–1364.

Gunay-Aygun, M., Schwartz, S., Heeger, S., O'Riordan, M. A., and Cassidy, S. B. (2001). The changing purpose of Prader–Willi syndrome clinical diagnostic criteria and proposed revised criteria. *Pediatrics* **108**, E92.

Haqq, A. M., Farooqi, I. S., O'Rahilly, S., Stadler, D. D., Rosenfeld, R. G., Pratt, K. L., LaFranchi, S. H., and Purnell, J. Q. (2003). Serum ghrelin levels are inversely correlated with body mass index, age, and insulin concentrations in normal children and are markedly increased in Prader–Willi syndrome. *J. Clin. Endocrinol. Metab.* **88**, 174–178.

Holm, V. A., Cassidy, S. B., Butler, M. G., *et al.* (1993). Prader–Willi syndrome: Consensus diagnostic criteria. *Pediatrics* **91**, 398–402.

Ritzen, E. M., Lindgren, A. C., Hagenas, L., Marcus, C., Muller, J., and Blichfeldt, S. (1999). Growth hormone treatment of patients with Prader–Willi syndrome. Swedish Growth Hormone Advisory Group. *J. Pediatr. Endocrinol. Metab.* **12**, 345–349.

Schuster, D., Oswei, K., and Zipf, W. B. (1996). Characterization of alterations in glucose and insulin metabolism in Prader–Willi subjects. *Metabolism* **45**, 1514–1520.

Stadler, D. D. (1995). Nutritional management. *In* "Management of Prader–Willi Syndrome" (L. R. Greenswag and R. Alexander, eds.), pp. 88–114. Springer-Verlag, New York.

Whitman, B., Greenswag, L., and Boyt, M. (1998). The use and impact of psychotropic medications for managing behavior in persons with Prader–Willi syndrome. Proceedings of the 13th Annual Prader–Willi Scientific Conference, July 22, 1998, Columbus, OH.

Wren, A. M., Seal, L. J., Cohen, M. A., Brynes, A. E., Frost, G. S., Murphy, K. G., Dhillo, W. S., Ghatei, M. A., and Bloom, S. R. (2001). Ghrelin enhances appetite and increases food intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 5992.

Pregnancy and Gastrointestinal Disease

RENE DAVILA AND CLAUDIO R. TOMBAZZI
University of Tennessee Health Science Center

hyperemesis gravidarum Pernicious condition in which the pregnant patient develops severe and intractable nausea and vomiting associated with nutritional and fluid/electrolyte deficiencies.

trimester Portion of a pregnancy in time; first, second, and third trimesters roughly correspond to months 1–3, 4–6, and 6–9.

The development of common gastroenterological conditions during pregnancy may present a challenge to physicians who are not familiar with the altered physiology of the gravid state. Knowledge of the most common gastrointestinal diseases in pregnancy is critical to the health of both mother and child.

INFLAMMATORY BOWEL DISEASE

The rates of fertility, congenital malformation, stillbirths, and spontaneous abortions in women with ulcerative colitis (UC) are comparable to those of unaffected women. Ulcerative colitis may have its onset in the first two trimesters of gestation. Women with preexisting UC and whose UC is asymptomatic in the prenatal period usually do not experience exacerbation of disease during pregnancy. On the other hand, women who have active disease before pregnancy may experience worsening of symptoms during the first trimester and significant improvement during the second trimester. Most pregnancies of women with ulcerative colitis result in delivery of normal full-term babies.

Crohn's disease is rarely diagnosed initially during pregnancy. The fertility rate in women with Crohn's disease is affected by the multiple clinical complications of the disease, such as perineal scarring, dyspareunia, and decreased libido. Women with active Crohn's disease have an increased risk for premature birth. There are no increases, however, in congenital malformations, stillbirths, or spontaneous abortions. Treatment of Crohn's disease during pregnancy with sulfasalazine, 5-aminosalicylic acid, or corticosteroids has been undertaken without risk for fetal abnormalities.

The use of immunosuppressive therapy with azathioprine or 6-mercaptopurine has generally been avoided during pregnancy, although published data have not suggested an increased risk of complications.

 Copyright 2004, Elsevier (USA). All rights reserved.

Immunosuppressive therapy does not represent an absolute contraindication for pregnancy or an absolute indication for termination of pregnancy.

PANCREATITIS

The incidence of pancreatitis during pregnancy is 0.009% (less than 1 in 10,000). Approximately 90% of cases of pancreatitis during pregnancy are due to gallstones, with clinical manifestations and diagnosis similar to those in nonpregnant patients. Management should be medical, but in a setting of high-level of care. The role of endoscopic retrograde cholangiopancreatography (ERCP) in the management of gallstone pancreatitis during pregnancy is unclear and is potentially dangerous due to radiation exposure to the fetus. Surgery should be reserved for patients with life-threatening conditions.

GALLSTONE-RELATED DISEASE

The incidence of gallstones increases in pregnancy. The prevalence in asymptomatic pregnant women is 2.5−12%. The etiology for this increased incidence is related to abnormalities in gallbladder motility, hormonal changes, cholesterol supersaturation, and changes in bile acid composition of bile.

The clinical manifestations are similar than those in nonpregnant women. Diagnosis is usually made by the use of abdominal ultrasound. Radionuclide hepatobiliary iminodiacetic acid (HIDA) scan studies are contraindicated due to radiation exposure. In patients with cholecystitis during pregnancy, conservative management with antibiotics and intravenous fluid replacement is the treatment of choice. ERCP with stone removal has been performed in a limited number of cases with extreme lead shielding and minimal use of fluoroscopy. Cholecystectomy is indicated for patients who do not respond to conservative therapy. Laparoscopic cholecystectomy is not recommended because of the high risk of damage to the gravid uterus. Open cholecystectomy during the first trimester may contribute to a risk of abortion, and surgery during the third trimester may induce labor.

HEMORRHOIDAL DISEASE

Pregnant women commonly experience anorectal discomfort, bleeding, or anal pruritus during the third trimester of gestation. Factors commonly involved are mechanical compression of veins by the enlarging uterus and worsening constipation with increased straining during defecation. Treatment includes conservative measures such as sitz baths and suppositories. Surgical hemorrhoidectomy should be reserved for intractable cases.

CONSTIPATION

The frequency of constipation during pregnancy is 11−40%. The pathogenesis is unclear, but may be related to increased serum progesterone levels, extrinsic compression of the colon or rectum by the gravid uterus, oral iron supplementation, and increased absorption of water and electrolytes. The management should include increased intake of fluid and dietary fiber, as well as the use of bulk-forming agents. Laxatives containing anthraquinone or cascara derivatives should not be used during pregnancy because of the potential to cause congenital malformations. Castor oil should also be avoided because it increases premature uterine contractions. The use of phenolphthalein laxatives is contraindicated during breast-feeding.

GASTROESOPHAGEAL REFLUX DISEASE

The incidence of heartburn during pregnancy is 30−50%, and it often occurs daily. The pathogenesis of gastroesophageal reflux in pregnancy is related to the enlarging gravid uterus that causes increased intraabdominal and intragastric pressures, a hormone-mediated decrease in lower esophageal sphincter pressure, and decreased esophageal clearance. The clinical manifestations are similar to those in nonpregnant women. Esophagogastroduodenoscopy is safe during pregnancy, but is seldom needed to establish a diagnosis. The role of 24-hour pH monitoring remains to be determined. Complications of reflux such as severe erosive or ulcerated esophagitis are rarely seen.

Management includes dietary changes such as eating a diet low in fatty foods, as well as avoidance of foods with a high acidic content or substances that decrease the lower esophageal sphincter pressure, such as caffeine. Lifestyle changes such as elevation of the head of the bed and cessation of smoking are also recommended. The use of non-calcium-based antacids or sucralfate appears to be safe and effective.

Intractable gastroesophageal reflux may require the use of drugs that reduce or suppress gastric acid secretion. Cimetidine should be avoided because of antiandrogenic effects that may affect normal male fetal

development. Ranitidine use for short periods appears to be safe during pregnancy. Although certain proton pump inhibitors such as omeprazole and pantoprazole may be toxic to the fetus, lansoprazole does not appear to be toxic, at least in animals.

NAUSEA AND HYPEREMESIS GRAVIDARUM

The incidence of nausea and emesis during pregnancy is 60–70%, and is more common in the first trimester. Hyperemesis gravidarum occurs in 0.5 to 10 of 1000 pregnancies. It presents as severe, persistent emesis that usually requires hospitalization for fluid replacement and nutritional supplementation. Hospitalization is indicated when pregnant women present with tachycardia and hypotension, ketosis, or weight loss. Associated factors include nulliparity, multiple gestations, obesity, and hydatidiform mole. Other associated factors sometimes include abnormal gastric emptying, autonomic dysfunction, hyperthyroidism, and psychological factors. Conservative management is directed to fluid and electrolyte replacement. The use of antiemetics such as metoclopramide and prochlorperazine is believed to be safe during pregnancy.

See Also the Following Articles

Colitis, Ulcerative • Constipation • Crohn's Disease • Gallstones, Pathophysiology of • Gastroesophageal Reflux Disease (GERD) • Hemorrhoids • Hyperemesis Gravidarum • Liver Disease, Pregnancy and • Nausea

Further Reading

Cappell, M. (1998). The safety and efficacy of gastrointestinal endoscopy during pregnancy. *Gastroenterol. Clin. North Am.* **27**, 37–71.

Dixon, N. P., Faddis, D. M., and Silberman, H. (1987). Aggressive management of cholecystitis during pregnancy. *Am. J. Surg.* **154**, 292–294.

Riely, C. A., and Davila, R. (2002). Pregnancy-related hepatic and gastrointestinal disorders. *In* "Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., Vol. II, pp. 1448–1461. W. B. Saunders, Philadelphia.

Saleeby, R. G., Rosen, L., Stasik, J. J., Riether, R. D., Sheets, J., and Khubchandani, I. T. (1991). Hemorrhoidectomy during pregnancy: Risk or relief? *Dis. Colon Rectum* **34**, 260–261.

Woolfson, K., Cohen, Z., and McLeod, R. S. (1990). Crohn's disease and pregnancy. *Dis. Colon Rectum* **33**, 869–873.

Probiotics

FERGUS SHANAHAN

Alimentary Pharmabiotic Centre, University College Cork, National University of Ireland, and Cork University Hospital, Cork, Ireland

prebiotic A nondigestible food ingredient that can beneficially influence the health of the host by selectively altering the enteric flora.

probiotic A live microorganism that, when consumed in adequate amounts, can confer a health effect on the host.

synbiotic A mixture of prebiotic and probiotic elements.

Probiotics are live microorganisms that, when consumed in sufficient amounts, can favorably influence the microbial ecology of the host and thus confer a health benefit on the host. Probiotics may exert their beneficial effects by different mechanisms, including the production of antimicrobial factors, competition for binding sites or nutrients, and modulation of the immune system.

TERMINOLOGY

Probiotics are biological agents, usually consumed as food supplements, that can favorably influence the microbial ecology of the host. Precise definition continues to evolve and they have recently been described as "live microorganisms which, when consumed in adequate amounts, confer a health effect on the host." The emphasis on live microbes is in contrast to prebiotics, which are nondigestible food ingredients, often of an oligo- or a polysaccharide nature, that beneficially affect the health of the host by selectively stimulating the growth or activity of certain bacterial species already established within the colon of the host. Mixtures of probiotics and prebiotics are referred to as synbiotics. Each is an example of a growing list of functional foods or neutraceuticals that confer a health benefit beyond their nutritional content. Although the term probiotic ("for life") is relatively new, the concept of consuming selected bacteria for health promotion was recognized almost a century ago. The Russian-born Nobel laureate Eli Metchnikoff regarded enteric lactobacilli and the consumption of yogurt-like foods as important for health and longevity and the French pediatrician Tissier studied enteric "bifid" bacteria in relation to diarrheal illness and suggested that they could have a therapeutic role. Probiotics are generally gram-positive bacteria and members of the genera *Lactobacillus* and *Bifidobacterium*, although other bacteria including *Escherichia coli* and nonbacterial organisms such as *Saccharomyces boulardii* have been selected as potential probiotics. Criteria for selection of microorganisms as candidate probiotics include proliferative capacity and capability of transit and survival within the gastrointestinal tract. This requires relative resistance to acid and bile. Most important are safety criteria. Lactobacilli and bifidobacteria have a long history of usage without hazard, which is the greatest testimony to their safety. In rare or exceptional circumstances, lactobacilli have been linked with systemic translocation but there appears to be no increased frequency of bacteremia with increased usage of probiotics.

MECHANISM OF ACTION

Probiotics may exert their beneficial effects in various settings by different mechanisms. These include the production of antimicrobial factors such as bacteriocins, competitive exclusion of pathogen binding to the mucosal epithelium, competition for nutrients, conditioning of the mucosal epithelium and subepithelial structures, and modulation of the immune system. Evidence from studies *in vitro* supports each of these mechanisms. Persuasive evidence also indicates molecular signaling from commensal and probiotic organisms to the host mucosal cells. The mechanism by which the host immune system distinguishes pathogenic organisms from commensal and probiotic organisms is mediated in part by pattern recognition molecules (or toll-like receptors) on the surface of immune and epithelial mucosal cells. Ligands for these receptors include bacterial lipopolysaccharide, cell wall components, and bacterial nucleic acid.

THERAPEUTIC APPLICATIONS

In general, the promise and claims for probiotics have outstripped the level of supporting evidence. This initially shrouded the field with a measure of skepticism that is beginning to be replaced with more rigorous

 Copyright 2004, Elsevier (USA). All rights reserved.

scientific scrutiny of the role of the enteric microflora in gastrointestinal development, physiology, and disease. The potential for therapeutic or prophylactic manipulation of the gut flora with probiotics seems most obvious in the context of gastrointestinal infections and overgrowth syndromes including antibiotic-associated overgrowth with *Clostridium difficile*. The likely participation of the commensal flora in generating cocarcinogens from dietary substrates suggests a role for probiotics in prevention of colorectal cancer. There is also persuasive evidence for a role for probiotics in food allergy and other atopic disorders. Several lines of evidence have indicated that a subset of the resident commensal flora drive the inflammatory process in patients with Crohn's disease and ulcerative colitis. This has led to the use of probiotic preparations in these conditions. Pilot studies in animal models of inflammatory bowel disease have been very encouraging and the best evidence to date for efficacy in human inflammatory bowel disease has been in patients with pouchitis following colectomy for ulcerative colitis with the creation of an ileoanal pouch anastomosis. Finally, the scope and potential of probiotics may, in the future, be redefined with the use of genetically modified organisms that are engineered for delivery of biologically relevant molecules such as regulatory cytokines, enzymes, and vaccines.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Microflora, Overview • Pharmacology, Overview

Further Reading

Dunne, C., and Shanahan, F. (2002). Role of probiotics in the treatment of intestinal infections and inflammation. *Curr. Opin. Gastroenterol.* **18**, 40–45.

FAO/World Health Organization (2001). Report of a Joint FAO/WHO Expert Consultation: Health and nutritional properties of probiotics in food including powder milk and live lactic acid bacteria. Available at http://www.fao.org/es/ESN/Probio/report.

Shanahan, F. (2000). Probiotics in inflammatory bowel disease: Is there a scientific rationale? *Inflamm. Bowel Dis.* **6**, 107–115.

Proctitis and Proctopathy

JOSH LEVITSKY AND ELI D. EHRENPREIS
University of Chicago

diversion proctitis Inflammation and friability of the colonic mucosa after exclusion of a distal segment of colon from the fecal stream.

lymphoid follicular hyperplasia Increased growth and abundance of lymphoid follicles, often seen in pathologic specimens of patients with diversion proctitis and colitis.

mesalamine 5-Aminosalicylic acid.

proctitis Inflammation of the rectum.

proctopathy Any pathologic process involving the rectum without having a significant component of rectal inflammation.

short-chain fatty acids Chemicals derived from bacterial metabolism of nonabsorbed or poorly digested dietary carbohydrates; used in enema form for diversion proctitis.

Although the term proctitis has been applied to a number of disorders involving the rectum, its strict definition refers only to those conditions characterized by inflammation of the mucosa and/or deeper layers of the rectal wall. Other processes causing little to no rectal inflammation, such as chronic radiation injury, are served better by the term proctopathy. Some proctopathies are important clinical entities and receive a detailed discussion in this article. Rectal involvement may also be present in conditions limited to the colon, as in infectious colitis. Finally, rectal involvement may occur as a manifestation of a systemic disorder, such as amyloidosis.

SYMPTOMATOLOGY AND CLINICAL EVALUATION

The symptoms of proctitis and proctopathy occur due to damage of rectal mucosa and alteration of normal rectal function. Mucosal inflammation and disruption produce classic symptoms of rectal bleeding

 Copyright 2004, Elsevier (USA). All rights reserved.

scientific scrutiny of the role of the enteric microflora in gastrointestinal development, physiology, and disease. The potential for therapeutic or prophylactic manipulation of the gut flora with probiotics seems most obvious in the context of gastrointestinal infections and overgrowth syndromes including antibiotic-associated overgrowth with *Clostridium difficile*. The likely participation of the commensal flora in generating cocarcinogens from dietary substrates suggests a role for probiotics in prevention of colorectal cancer. There is also persuasive evidence for a role for probiotics in food allergy and other atopic disorders. Several lines of evidence have indicated that a subset of the resident commensal flora drive the inflammatory process in patients with Crohn's disease and ulcerative colitis. This has led to the use of probiotic preparations in these conditions. Pilot studies in animal models of inflammatory bowel disease have been very encouraging and the best evidence to date for efficacy in human inflammatory bowel disease has been in patients with pouchitis following colectomy for ulcerative colitis with the creation of an ileoanal pouch anastomosis. Finally, the scope and potential of probiotics may, in the future, be redefined with the use of genetically modified organisms that are engineered for delivery of biologically relevant molecules such as regulatory cytokines, enzymes, and vaccines.

See Also the Following Articles

Colitis, Ulcerative • Colorectal Adenocarcinoma • Crohn's Disease • Microflora, Overview • Pharmacology, Overview

Further Reading

Dunne, C., and Shanahan, F. (2002). Role of probiotics in the treatment of intestinal infections and inflammation. *Curr. Opin. Gastroenterol.* **18**, 40–45.

FAO/World Health Organization (2001). Report of a Joint FAO/WHO Expert Consultation: Health and nutritional properties of probiotics in food including powder milk and live lactic acid bacteria. Available at http://www.fao.org/es/ESN/Probio/report.

Shanahan, F. (2000). Probiotics in inflammatory bowel disease: Is there a scientific rationale? *Inflamm. Bowel Dis.* **6**, 107–115.

Proctitis and Proctopathy

JOSH LEVITSKY AND ELI D. EHRENPREIS
University of Chicago

diversion proctitis Inflammation and friability of the colonic mucosa after exclusion of a distal segment of colon from the fecal stream.

lymphoid follicular hyperplasia Increased growth and abundance of lymphoid follicles, often seen in pathologic specimens of patients with diversion proctitis and colitis.

mesalamine 5-Aminosalicylic acid.

proctitis Inflammation of the rectum.

proctopathy Any pathologic process involving the rectum without having a significant component of rectal inflammation.

short-chain fatty acids Chemicals derived from bacterial metabolism of nonabsorbed or poorly digested dietary carbohydrates; used in enema form for diversion proctitis.

Although the term proctitis has been applied to a number of disorders involving the rectum, its strict definition refers only to those conditions characterized by inflammation of the mucosa and/or deeper layers of the rectal wall. Other processes causing little to no rectal inflammation, such as chronic radiation injury, are served better by the term proctopathy. Some proctopathies are important clinical entities and receive a detailed discussion in this article. Rectal involvement may also be present in conditions limited to the colon, as in infectious colitis. Finally, rectal involvement may occur as a manifestation of a systemic disorder, such as amyloidosis.

SYMPTOMATOLOGY AND CLINICAL EVALUATION

The symptoms of proctitis and proctopathy occur due to damage of rectal mucosa and alteration of normal rectal function. Mucosal inflammation and disruption produce classic symptoms of rectal bleeding

 Copyright 2004, Elsevier (USA). All rights reserved.

(hematochezia), small-volume diarrhea, passage of mucus, and intermittent, crampy pain. Decreased rectal compliance results in the characteristic symptoms of increased stool frequency, urgency, painful anal spasm with limited evacuation (tenesmus), and at times, fecal incontinence. Symptoms may be exacerbated by anal involvement of the disease process. Because of the wide range of diagnoses possible with the development of the symptoms, a thorough, albeit organ-specific, history and physical examination is performed.

Pertinent information includes sexual practices, family history, prior pelvic irradiation or bowel surgeries, and a history of chronic constipation or anorectal disease. The physical examination should include an evaluation for abdominal tenderness, femoral and inguinal lymphadenopathy, and perianal ulceration, fistula, or tenderness. Rectal examination is useful for assessment of sphincter function, for the diagnosis of rectal masses, and to test for gross or occult blood. Stool cultures and tests for ova, parasites, and *Clostridium difficile* toxin are obtained when relevant. Flexible proctosigmoidoscopy and sometimes colonoscopy are used to determine the extent and severity of disease involvement, to sample colonic effluent for culture, and to obtain mucosal biopsies. Further management decisions are made based on the findings of the clinical evaluation. The etiologies of proctitis and proctopathy are shown in Table I.

ETIOLOGIES OF PROCTITIS

Infectious Proctitis

Sexually transmitted diseases account for most cases of infectious proctitis. Patients at highest risk include homosexual men, particularly those practicing anal receptive intercourse (ARI) and having multiple sexual partners. Patients with human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) are also at high risk.

TABLE I Etiologies of Proctitis and Proctopathy

Proctitis	Proctopathy
Infectious	Radiation
Bacterial	Ischemia
Neisseria gonorrhoeae	Trauma
Chlamydia trachomatis (LGV and non-LGV)	Solitary rectal ulcer syndrome
Shigella spp.	Hematologic diseases
Salmonella spp.	Lymphoma
Campylobacter spp.	Acute myelogenous leukemia
Streptococcus spp. (Group A)	Agranulocytosis
Clostridium difficile	Medullary aplasia
Pleisomonas shigelloides	Amyloidosis
Mycobacterium tuberculosis	
Parasitic	
Treponema pallidum	
Schistosoma mansoni	
Entamoeba histolytica	
Viral	
Cytomegalovirus	
Herpes simplex virus	
Ulcerative colitis	
Crohn's disease	
Medication-induced	
NSAIDs	
Gold	
Enema/suppository (bisacodyl, indomethacin, sodium phosphate, barium, hot water, acetaminophen, codeine, acetylsalicylic acid, hydrogen peroxide, vinegar, glutaraldehyde)	
Fecal diversion	
Allergy	
Collagen–vascular diseases	
Behçet's syndrome	
Systemic lupus erythematosus	
Rheumatoid arthritis	

Neisseria gonorrhoeae, a gram-negative intracellular diplococcus, infects the rectum 5 to 7 days after it is transmitted by ARI or inoculation by infected vaginal fluid. The prevalence of anororectal gonorrhea may be as high as 50% in women with gonococcal pelvic inflammatory disease. Homosexual men, particularly those visiting sexually transmitted disease clinics or having genital gonorrhea, are also at high risk. Most patients have minimal symptoms, although some develop arthritis, tenosynovitis, and skin rashes as well as symptoms of proctitis. Mucosal erythema, friability, and erosions are typically seen in the anorectum on sigmoidoscopy. The presence of a mucopurulent discharge in the anal canal in a high-risk patient is very suggestive of gonococcal proctitis. Diagnosis can be obtained by performing a Gram's stain and culture of the lower rectum by either swab or rectal biopsy; repeat cultures are often necessary given the difficulty in culturing gonococcus. The treatment of gonococcal proctitis is a single dose of ceftriaxone 125 mg intramuscularly in addition to treatment for presumed concomitant chlamydial infection (see below). Alternative regimens include the use of procaine penicillin 4.8 million units intramuscularly in two doses with 1 g of probenecid given at the time of injection, tetracycline 1.5 g orally followed by 500 mg four times a day for 4 days, or spectinomycin 2 g intramuscularly.

Chlamydia trachomatis is characterized by lymphogranuloma venereum (LGV) and non-LGV immunotypes, both of which cause proctitis approximately 10 days after inoculation. LGV strains typically cause more severe inflammation and symptoms than non-LGV strains. Untreated LGV may mimic Crohn's disease by causing ulcerations, fistulas, abscesses, and strictures. Findings on sigmoidoscopy include mucosal granularity, erythema, and ulceration. Biopsies may show a diffuse inflammatory infiltrate with crypt abscess, granulomas, and giant cells, similar to the pathology seen in Crohn's disease. Rectal culture, either by swab or by biopsy, microimmunofluorescence antibody staining, and complement fixation testing may be helpful but are often unsuccessful in obtaining diagnostic confirmation. A single 1 g oral dose of azithromycin, 7 days of oral doxycycline, 100 mg twice daily, or 21 days of trimethoprim-sulfamethoxaxole, double-strength twice daily, are all effective treatment regimens for chlamydia proctitis. In patients with chlamydia proctitis, the empiric treatment of gonococcal proctitis even before a confirmatory diagnosis is made is recommended.

Herpes simplex virus 2 is the most common anal herpetic infection. After an incubation period of approximately 3 weeks, herpes proctitis typically lasts for 7–10 days. Perianal vesicles and ulcerations on external examination in a patient with proctitis symptoms are highly suggestive of herpes infection. Severe herpes proctitis often causes intractable pain on defecation and anal manipulation, tenesmus, pruritis, and a mucopurulent discharge. Mucosal friability, ulcerations, vesicles, and pustules are usually limited to the distal 10 cm of rectum. Sigmoidoscopy, however, can rarely be performed without significant anesthesia because of severe discomfort. Diagnosis is made from viral culture of anorectal ulcer scrapings or biopsies. Giemsa staining of material scraped from the base of ulcers or vesicles reveals the characteristic multinucleated giant cells with intranuclear inclusion bodies. Treatment consists of 10 days of oral acyclovir, 400 mg five times daily, or in cases of resistance, foscarnet. Maintenance therapy with oral acyclovir may suppress further herpes outbreaks.

Anorectal syphilis infection with *Treponema pallidum* is commonly misdiagnosed as idiopathic anal ulcers or nonspecific proctitis. Infection occurs 2 to 8 weeks after sexual transmission. Anorectal syphilis is characterized by a variety of endoscopic appearances including anal ulceration, rectal ulceration, granulomatous proctitis, and tumor-like lesions. A combination of serologic testing, dark-field examination, and immunofluorescence staining usually leads to a diagnosis. Benzathine penicillin G, 2.4 million units intramuscularly, given initially and 7 days later, is the drug of choice for anorectal syphilis. Penicillin-allergic patients should be given either tetracycline, 500 mg orally four times daily for 15 days, or erythromycin, 500 mg orally four times daily for 15 days. Patients with significant penicillin allergies typically require desensitization to penicillin.

A number of other bacterial, parasitic, and viral etiologies can cause proctitis, although they typically cause a more proximal colitis. In homosexual men, particularly those with HIV infection, proctocolitis is often caused by *Campylobacter* spp., *Shigella* spp., non-typhi *Salmonella*, *Entamoeba histolytica*, and cytomegalovirus (CMV). CMV may also cause proctitis in recipients of bone marrow transplantation or chemotherapy. In Third World countries, tuberculosis is an often unrecognized cause of anorectal abscess and fistula and may occur in the absence of pulmonary infection. Cases of proctitis caused by *Pleisomonas shigelloides*, group A *Streptococcus*, and *Schistosoma mansoni* have been reported in the literature, mainly in homosexual men.

Ulcerative Proctitis

Ulcerative proctitis, a chronic inflammatory process limited to the rectum, affects approximately one-third to

one-half of patients diagnosed with ulcerative colitis. Although patients with ulcerative proctitis generally follow a more benign course with fewer symptoms than those with ulcerative colitis, 30–50% of them will develop more proximal disease distal to the splenic flexure within 10 years. The 10-year survival rate for ulcerative proctitis is high, almost 98% in one series. Annual incidence rates of ulcerative proctitis show that males are more likely to develop the disease and the disease is more common in urban than rural areas. The pathophysiology is similar to that of ulcerative colitis. Patients present with typical proctitis symptoms such as rectal bleeding, pain, and tenesmus. Endoscopic findings in the rectum include loss of vascular pattern, erythema, granularity, and friability of the mucosa. Crypt abscess formation, crypt architectural distortion, and a polymorphonuclear cell infiltrate in the lamina propria are the most common histopathologic findings.

Excellent treatments exist once the diagnosis of ulcerative proctitis is made. The initial goal of therapy is to induce remission. Mesalamine suppositories and enemas have been shown to be superior to oral mesalamine and steroid-based topical treatments in inducing remission after 2–4 weeks of therapy. Optimal doses for inducing remission range from 0.5 to 1.5 g of topical mesalamine per day. Once clinical remission is obtained, the treatments are focused on the maintenance of remission. Mesalamine suppositories and oral 5-aminosalicylic acid (5-ASA) maintain 1 year remission at similar rates, between 60 and 90%.

Crohn's Proctitis

Crohn's proctitis without colitis or perianal disease is extremely rare. 5-ASA suppositories or enemas, effective in ulcerative proctitis, have not been well studied in Crohn's proctitis but are often given due to lack of alternative therapies.

Medications

Many pharmacological agents can cause colitis, but only a few may affect the rectum without other gastrointestinal involvement (Table I). Nonsteroidal anti-inflammatory drugs (NSAIDs) are likely the most common cause of drug-induced proctitis, particularly if given in an enema form. Other enemas and suppositories can cause rectal inflammation from either local mucosal injury or pressure-induced injury. Treatments with gold for rheumatoid arthritis rarely lead to enterocolitis and/or ulcerative proctitis. Discontinuation of the offending agent is the most important intervention for these conditions.

Diversion Proctitis

Diversion colitis or proctitis is characterized by inflammation and friability of the colonic mucosa after exclusion of a distal segment of colon from the fecal stream. As early as 1 month after surgery, almost all excluded segments will have some degree of endoscopic or histologic abnormality. Histopathologic changes, including mucosal fissures, granulomas, crypt abscesses, architectural distortion, and transmural inflammation, can mimic changes seen in patients with inflammatory bowel disease. Lymphoid follicular hyperplasia, however, may be a distinctive feature of diversion colitis. The diagnosis is often confirmed when mucosal changes resolve after restoration of the fecal stream, although mucosal abnormalities persist in approximately 50% of cases. Short-chain fatty acids, important nutritional substrates for the colonic epithelium, are effective in the treatment of diversion proctitis when given locally as enemas.

Allergic Proctitis

Allergic proctitis is typically seen in infants who are sensitive to cow's milk protein or soy protein. Signs and symptoms vary from the presence of occult gastrointestinal blood loss to frank hematochezia and diarrhea. Other foods, such as egg white, peanuts, nuts, or fish, have also been implicated in patients with allergic gastroenteropathy and uncommonly in patients with allergic proctitis. Endoscopic findings of allergic proctitis are nonspecific, but rectal biopsy characteristically reveals eosinophilic infiltration within the lamina propria and intraepithelial eosinophils in surface and crypt epithelium. Treatment primarily involves elimination of the potential allergen and occasionally requires the use of corticosteroids.

Collagen Vascular Disease

A number of rheumatologic diseases can affect the gastrointestinal tract by autoimmune mechanisms, systemic inflammation, or vasculitis. Proctitis from these diseases is rare. Behçet's disease and systemic lupus erythematosus have been reported as causing proctitis. Rectal biopsies in patients with rheumatoid arthritis may reveal evidence of chronic inflammation and vasculitis. It is a generally accepted principle that anti-inflammatory and immune therapy directed at the systemic disease will improve accompanying gastrointestinal manifestations.

ETIOLOGIES OF PROCTOPATHY

Radiation Proctopathy

Radiation proctopathy commonly occurs after radiation therapy for the treatment of pelvic malignancies. Prostate, cervical, uterine, bladder, testicular, and rectal cancers are among the most common malignancies treated with radiation. A large discrepancy exists in the literature regarding the prevalence of radiation proctopathy in patients receiving pelvic radiation, with ranges between 5 and 80%. A dose of 65–70 Gy of external beam radiation is typically required to cause radiation proctopathy. Acute radiation proctopathy occurs in approximately one-third of patients and is characterized by diarrhea, urgency, and tenesmus, usually without hematochezia. Chronic radiation proctopathy occurring 6–12 months after radiation therapy is often characterized by rectal bleeding due to rupture of telengiectasias or oozing from friable, ischemic mucosa. Functional symptoms, such as difficulty with evacuation, fecal urgency and incontinence, small-volume diarrhea, and rectal pain, are not uncommon and can lead to significant morbidity.

Because of the lack of randomized controlled clinical trials, effective treatments for radiation proctopathy are often given empirically without significant supportive evidence. Generally, bleeding is best treated endoscopically, but it may respond to various pharmacotherapies. Noncontact methods of endoscopic therapy, such as laser and argon plasma coagulation, are typically more effective than contact methods, such as heater probe and bipolar electrocautery, in the obliteration of bleeding telengiectasias. Sucralfate, in oral and enema form, and topical 4% formalin applied endoscopically appear to be more effective than aminosalicylate or corticosteroid enemas in treating hemorrhagic radiation proctitis, although no direct comparisons have been performed in clinical trials. Less well-studied treatments include short-chain fatty acid enemas, misoprostol enemas, hyperbaric oxygen therapy, vitamins C and E, oral estrogen/progesterone, and oral sodium pentosan polysulfate. Vitamin A was recently found to be highly effective in one patient who developed a symptomatic radiation-induced anal ulcer. Unfortunately, because functional radiation proctopathy is poorly studied, it is unclear whether or not any of the treatments mentioned are effective.

Ischemic Proctopathy

Because of the extensive network of vascular collaterals supplying the rectum, ischemic proctopathy is rare. Ischemic proctopathy usually results from a sudden, major interruption in blood flow, such as that occurring after aortoiliac surgery. However, severe atheromatous disease combined with systemic hypotension may also compromise the otherwise abundant blood supply to the rectum. Diarrhea, abdominal pain, and unexplained sepsis in the first few days after an aortic operation are warning signs of intestinal ischemia. A sigmoidoscopy with biopsy confirms ischemic changes in the rectum, such as mucosal ulcerations and capillary dilation in overlying granulation tissue. Mild cases may be treated conservatively and are likely to resolve with hemodynamic support. The treatment of intractable rectal bleeding due to ischemic proctopathy usually requires surgical diversion, although endoscopic coagulation and local treatments such as formalin instillation have been reported.

Traumatic Proctopathy

Patients participating in ARI or, more commonly, in sexual practices involving the insertion of foreign bodies or various body parts (fists, forearms, etc.) into the anorectum may develop traumatic proctopathy. This etiology likely represents a pathophysiologic spectrum of rectal injury, from benign mucosal tearing to severe ischemia and/or perforation. If the sexual practice is found to be the cause of proctopathy, the physician must warn the patient about the risk of foreign body impaction and/or perforation and recommend that the behavior be discontinued. If an object is impacted in the rectum and cannot be retrieved at the bedside, it is removed in the operating room under general anesthesia.

Solitary Rectal Ulcer Syndrome

This poorly understood condition is seen in patients with chronic straining, rectal prolapse, and nonrelaxation of pelvic musculature with defecation. It is characterized by the presence of a large ulcer or multiple ulcers 4 to 15 cm from the anal verge. Solitary rectal ulcer syndrome may be mistaken for inflammatory bowel disease by inexperienced clinicians. Histologic features include fibromuscular proliferation in the lamina propria and glandular crypt distortion. Treatment with fiber, stool softeners, and biofeedback is generally preferable to surgical management.

Hematologic Diseases

Gastrointestinal involvement may occur in a variety of hematologic diseases. The anorectum alone is rarely involved. In a series of 514 patients hospitalized for miscellaneous hematologic diseases, anal lesions, such as infiltration and ulceration, were seen most

commonly in patients with agranulocytosis, medullary aplasia, or acute myeloid leukemia. Treatment included chemotherapy or surgical drainage when an abscess was present. Anorectal infiltration with lymphoma is uncommon but is an AIDS-defining illness in patients with HIV infection.

Amyloidosis

Amyloid protein often infiltrates the rectum and can be detected on submucosal rectal biopsies performed to diagnose systemic amyloidosis. Although gastrointestinal amyloidosis can cause chronic constipation, diarrhea, and malabsorption, rectal amyloid infiltration rarely leads to clinical manifestations associated with proctopathy.

See Also the Following Articles

Amyloidosis • Colitis, Ulcerative • Cow Milk Protein Allergy • Diarrhea, Infectious • Foreign Bodies • Sexually Transmitted Diseases • Solitary Rectal Ulcer Syndrome

Further Reading

Cohen, R. D., Woseth, D. M., Thisted, R. A., *et al.* (2000). A meta-analysis and overview of the literature on treatment options for left-sided ulcerative colitis and ulcerative proctitis. *Am. J. Gastroenterol.* **95**, 1263–1276.

Corman, M. L. (1998). "Colon and Rectal Surgery," 4th Ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Goldman, H., and Proujansky, R. (1986). Allergic proctitis and gastroenteritis in children: Clinical and mucosal biopsy features in 53 cases. *Am. J. Surg. Pathol.* **10**, 75–86.

Guillemot, F., Colombel, J. F., Neut, C., *et al.* (1991). Treatment of diversion colitis by short-chain-fatty acids. Prospective and double blind study. *Dis. Colon Rectum* **34**, 861–864.

Hong, J. J., Park, W., and Ehrenpreis, E. D. (2001). Review article: Current therapeutic options for radiation proctopathy. *Aliment. Pharm. Ther.* **15**, 1253–1262.

Ilnyckyj, A. (2001). Clinical evaluation and management of acute infectious diarrhea in adults. *Gastroenterol. Clin. North Am.* **30**, 599–609.

Meucci, G., Vecchi, M., Astegiano, M., *et al.* (2000). The natural history of ulcerative proctitis: A multicenter, retrospective study (GSMII). *Am. J. Gastroenterol.* **95**, 469–473.

Nelson, R. L., and Schuler, J. J. (1982). Ischemic proctitis. *Surg. Gynecol. Obstet.* **154**, 27–33.

Pare, A. A., and Gottesman, L. (1997). HIV infection and the gastrointestinal tract: Anorectal diseases. *Gastroenterol. Clin. North Am.* **26**, 367–377.

Parsons, D. P., and Isler, J. T. (1999). Office management of common anorectal problems: Approach to the patient. *Primary Care Clin. Office Pract.* **26**, 15–35.

Tjandra, J. J., Fasio, V. W., Church, J. M., *et al.* (1992). Clinical conundrum of solitary rectal ulcer. *Dis. Colon Rectum* **35**, 227–234.

Protein Digestion and Absorption of Amino Acids and Peptides

HANNELORE DANIEL
Technical University of Munich

oligopeptide An oligomer of amino acid units joined by peptide linkages. The term "oligopeptide" is commonly used to refer to structures containing 4 to 25 amino acid residues.

peptide bond The covalent chemical bond between two amino acid residues. It is formed by the subtraction of a water molecule from the amino group attached to the α-carbon of one amino acid and the carboxy group attached to the α-carbon of a second amino acid residue.

protein A macromolecular complex containing a large number of amino acid residues joined to one another via peptide bonds. Proteins may, in addition to the 20 different amino acids that they generally are made up of, contain various amino acid residues that have been modified posttranslationally by phosphorylation, hydroxylation, glycosylation, or attachment of fatty acid residues.

The dietary intake of proteins varies considerably with the nature of the diet and the main protein sources in the diet (plant versus animal origin). However, in almost all developed countries, the average protein intake is approximately 100 g/day. In addition, almost 60 g of endogenous proteins from gastrointestinal secretions and saliva enters the small intestine so that a total of approximately 160 g of protein per day undergoes hydrolysis by gastric, pancreatic, and brush border membrane-bound proteases and peptidases. The end products are absorbed by multiple transport systems in the apical membrane of the epithelial cells.

HYDROLYSIS OF PROTEINS AND DETERMINANTS OF THEIR DIGESTABILITY

The lumenal and brush border-bound digestive enzymes initially release a vast spectrum of medium-sized and short-chain peptides and eventually free amino acids. Although most proteins and oligopeptides are rapidly degraded, some structures are fairly resistant to hydrolysis. In particular, glycosylated peptides and those containing multiple prolyl residues are more stable against attack by proteases and peptidases. This intrinsic proteolytic stability is of particular importance for understanding the "survival" of the immunodominant epitopes from α-gliadin that account for the stimulatory activity of dietary gluten on intestinal and peripheral T lymphocytes in patients with celiac disease. The α-gliadin peptides are rich in proline and glutamine residues, which are exceptionally resistant to enzymatic processing. The very low activity of dipeptidyl peptidase IV and dipeptidyl carboxypeptidase I in the gastrointestinal tract determine as rate-limiting enzymes the digestive breakdown of these peptides. Various other biologically active peptides containing multiple prolyl groups have also been identified in digests, for example, of dietary proteins (mainly milk proteins), which led to the suggestion that these peptides that are released during digestion of dietary proteins in the gut may affect body functions by their opioid, immunomodulatory, or angiotensin-converting enzyme inhibitory activity. The extent to which and the speed with which a dietary protein is broken down to its constituent parts are therefore dependent on its composition (amino acid sequence) and on modifications that render it more resistant to hydrolysis. This also includes thermal effects of food processing that may cause the formation of Maillard products or that induce the conversion of free L-amino acids into their D-enantiomers. Studies with ^{15}N-labeled dietary proteins in humans have demonstrated a quite variable extent of digestion and absorption as indicated by different ileal losses of amino acids depending on the nature of the protein administered.

Even intact proteins or large oligopeptides can be absorbed in small quantities in intact form either via the paracellular route or by microfold cells, but the bulk of protein is taken up in the form of di- and tripeptides and as free amino acids. For these end products of digestion, specialized carriers are found in the apical membrane of enterocytes.

 Copyright 2004, Elsevier (USA). All rights reserved.

EPITHELIAL UPTAKE OF DI- AND TRIPEPTIDES

The multitude of intracellular peptidases with a strict specificity for the hydrolysis of di- and tripeptides suggested that such peptides may be absorbed in intact form followed by the intracellular release of free amino acids. The peptide transporter protein that carries di- and tripeptides into the cell was identified as a proton–peptide symporter that couples peptide uptake to the movement of protons down an electrochemical proton gradient. The PEPT1 protein contains 708 amino acids residues and 12 transmembrane domains with N- and C-terminal ends facing the cytosol. By coupling to proton flux, peptide transport occurs electrogenically regardless of the net charge of the substrate and causes intracellular acidification. The required proton gradient for peptide uptake is mainly, but not exclusively, provided by electroneutral proton /cation exchangers, such as the Na^+/H^+ antiporters that return protons to the lumen in exchange for Na^+ ions entering the cells. However, the main driving force for peptide transport is the negative membrane potential inside the cell. Normal dipeptides taken up into the cells are rapidly hydrolyzed by cytosolic peptidases and free amino acids are then delivered into the circulation. Although there is evidence for a basolateral efflux system for di- and tripeptides, the nature of this protein is not yet known.

Transport by PEPT1 shows a pronounced stereoselectivity, with peptides containing L-enantiomers of amino acid residues possessing higher affinity for transport than peptides containing D-enantiomers. Peptides consisting solely of D-amino acids do not show any relevant affinity for transport. One of the most striking features of PEPT1 is its capability of sequence-independent transport of all possible di- and tripeptides. This means that the 20 proteinogenic L-α-amino acids alone provide 400 different dipeptides and 8000 different tripeptides that can be transported by PEPT1. The ability of PEPT1 to also accept a variety of peptidomimetics, such as antibiotics of the aminocephalosporin and aminopenicillin classes, or selected angiotensin-converting enzyme inhibitors, such as captopril, explains the excellent oral availability of these drugs. The clinical importance of PEPT1 has been demonstrated in a variety of studies in different organisms including humans. Here, dipeptide mixtures have been shown to be superior to free amino acids for fast intestinal absorption and they are also more useful for enteral nutrition as they provide a lower osmolarity of the nutrition solution. Moreover, in a variety of gastrointestinal diseases, the peptide transporter has been found to be less affected by the pathophysiology than the amino acid transporters. Figure 1 displays the peptide transporter in the apical membrane of epithelial cells in the context of the various amino acid carriers that in concert allow efficient transcellular amino acid absorption.

TRANSPORT OF FREE AMINO ACIDS ACROSS THE EPITHELIUM

When proteins and peptides undergo complete lumenal hydrolysis, free amino acids are released. The 20 proteinogenic amino acids and their derivatives resemble a heterogeneous group of compounds that differ in polarity, net charge, and molecular mass. It is therefore not surprising that membrane transporters with different substrate specificities are found. Whereas the carriers for neutral amino acids in most cases display a rather broad substrate specificity, other carrier types show much more specific interactions with a preference for either acidic or basic side chains of amino acids or for those with an aromatic structure. In addition, differences in the thermodynamic properties of the transport steps are observed with equilibrative systems and systems that are ion-dependent and show uphill transport capability. Almost all carriers display a pronounced stereoselectivity for transport of the physiologically important L-enantiomers of amino acids. Table I summarizes the main transport pathways for amino acids in mammalian cells subdivided into Na^+-dependent and Na^+-independent processes and the zrepresentative cDNAs that have been cloned and that express the corresponding transport activity in a target cell. Not surprisingly, there are numerous genetic and splice variants within the different transporter classes. However, at the structural level, most amino acid carriers show some common features with a peptide backbone of 350 to 700 amino acid residues that crosses the plasma membrane 10 to 12 times.

Enterocytes contain numerous amino acid transporters, some of which are found throughout the organism and others of which are specific for the apical or basolateral membranes of polarized cells. The transporters in the brush border membrane are primarily responsible for the absorption of amino acids from the intestinal lumen and those located on the basolateral surface of the enterocyte facilitate the transfer of amino acids between the enterocyte and the circulation. Influx across the basolateral membrane provides amino acids for the nutritional needs of the enterocyte in the absence of protein intake, whereas when proteins are digested and amino acids are absorbed, basolateral transporters

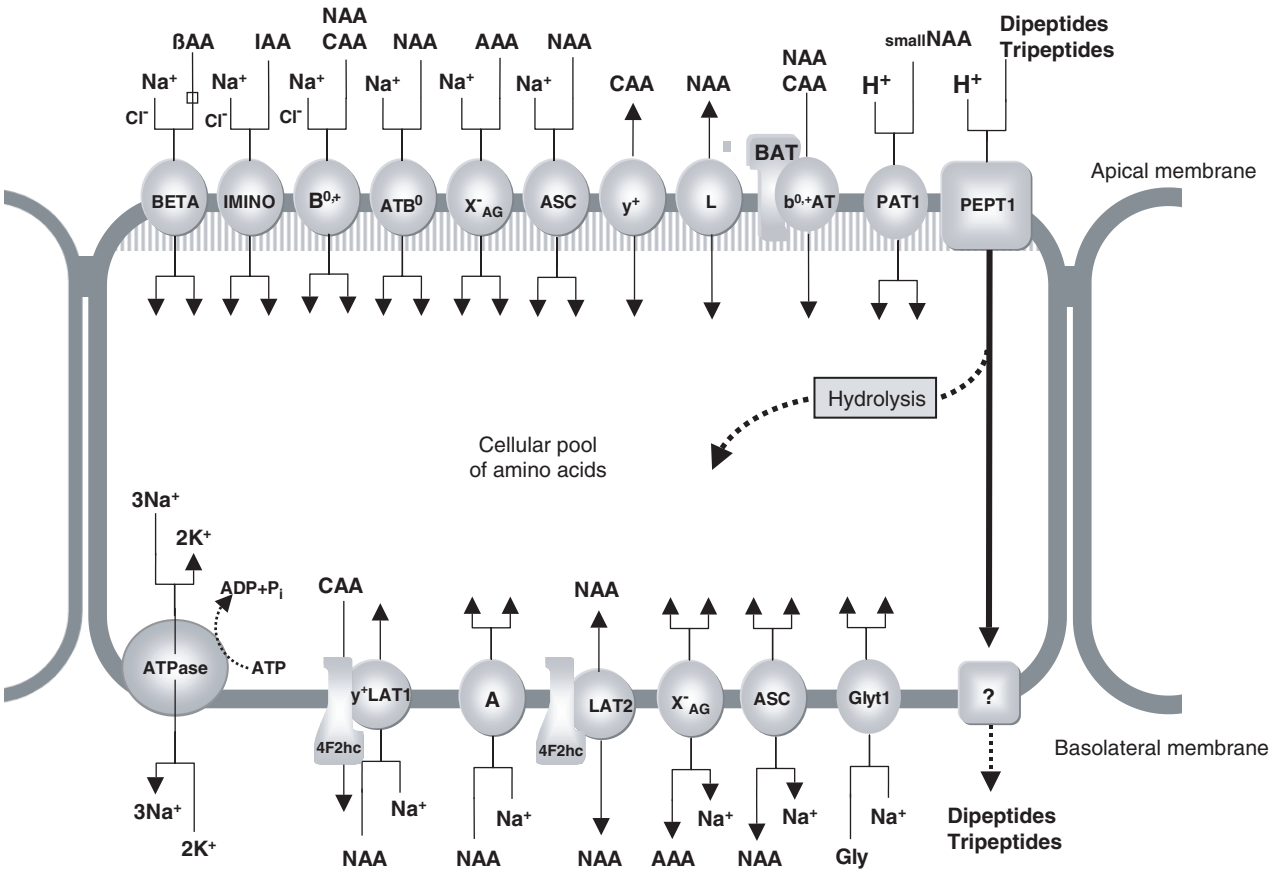


FIGURE 1 Amino acid and peptide transporters in apical and basolateral membranes of intestinal epithelial cells responsible for transepithelial amino acid translocation. Dashed lines indicate the intracellular hydrolysis of di- and tripeptides by cytosolic peptidases. NAA, neutral amino acids; CAA, cationic amino acids; AAA, acidic amino acids; IAA, imino acids; βAA: β-amino acids; Gly, glycine.

mediate net amino acid efflux from the cell into the portal circulation. Figure 1 displays the various carriers in apical and basolateral membranes that are involved in transcellular amino acid transport.

The amino acid transporter class that is dependent on both Na^+ and Cl^- gradients includes system IMINO, which transports imino acids such as proline, system BETA, a transporter of β-amino acids such as β-alanine, and system $B^{0,+}$ ($ATB^{0,+}$), which transports neutral, basic, and some D-enantiomeric amino acids. The second class contains the classical Na^+-dependent "secondarily active" transporters, including system $B^{0,+}$ and ATB^0, which transport a variety of neutral amino acids. Anionic amino acids are transported by system X^-_{AG}. System ASC has a preference for small neutral amino acids including alanine, serine, and cysteine. System A (mainly in basolateral membranes) transports neutral and methylated amino acids. The Na^+-independent transporters that operate in an "equilibrative" mode include system y^+ (the CAT proteins), which carries cationic amino acids, system $b^{0,+}$ ($ATb^{0,+}$), which transports cysteine as well as neutral and cationic amino acids, and system L (the LAT proteins), which recognizes mainly neutral amino acids. Members of the last group are particularly interesting in view of their molecular architecture. The novel structural characteristic of the LAT proteins is that they oligomerize via an extracellular disulfate bridge with a large second subunit to form a complex. In apical membranes, the glycoprotein heavy chain, designated BAT, associates with a LAT protein and the resulting complex possesses sodium-independent amino acid exchange capability. The other heavy chain, 4F2hc, can associate with various light chains (LAT proteins), with the complexes then mediating amino acid exchange across the basolateral membrane of epithelial cells as well as in nonpolarized cells. The heavy chains have a glycosidase-like extracellular domain attached to a single transmembrane

TABLE I Amino Acid Transport Systems in the Plasma Membrane of Mammalian Cells

Transport system	Isolated cDNA(s) encoding this activity
Na^+ dependent	
A	ATA1-3
N	SN1-3
GLY	GlyT1-2
ASC	ASC1-2
BETA	GAT1-3, BGT-1
IMINO	Not identified yet
B^0	ATB^0
$B^{0,+}$	$ATB^{0,+}$
X^-_{AG}	EAAT1-5
Na^+ independent	
L	4F2hc/LAT(X)
y^+	CAT1-3
$b^{0,+}$	BAT/$b^{0,+}$AT
H^+ dependent	
PAT	PAT1 (LYAAT-1), PAT2

domain, whereas the light chains vary in size but all possess 12 membrane-spanning domains, which are typical for polytopic membrane proteins with the amino- and carboxy-termini facing the cytosol. The carrier complexes can transport a wide range of neutral amino acids in an obligatory exchange mode, which means that they mediate the influx of certain amino acids in exchange for intracellular amino acids. The $b^{0,+}$ activity of the BAT-associated complex can in addition transport cationic amino acids in exchange for neutral amino acids, resulting in transport currents, whereas the y^+LAT1/4F2hc complex exchanges neutral amino acids in cotransport with sodium for intracellular cationic amino acids.

A new class of electrogenic proton-dependent amino acid symporters (PAT proteins) has been identified and cloned. From this class, the PAT1 carrier is expressed in the apical membrane of intestinal epithelial cells and mediates the cotransport of amino acids and derivatives that have a short side chain (glycine, alanine, serine, proline, γ-aminobutyric acid) by the coupling of substrate movement to proton movement down an electrochemical proton gradient. This transport activity consequently causes an intracellular acidification that requires apical sodium–proton exchangers for the control of intracellular pH, with proton export enhancing sodium uptake into epithelial cells.

The expression level of intestinal amino acid and peptide transporters can be adapted to dietary needs and a variety of hormones are involved in regulating protein expression and consequently the transport activity of the carriers that are responsible for the absorption of amino acids from dietary proteins.

See Also the Following Articles

Digestion, Overview • Nutrient Transport, Regulation of

Further Reading

Alpers, D. H. (1994). Digestion and absorption of carbohydrates and proteins. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1723–1749. Raven Press, New York.

Chillaron, J., Roca, R., Valencia, A., Zorzano, A., and Palacin, M. (2001). Heteromeric amino acid transporters: Biochemistry, genetics, and physiology. *Am. J. Physiol. Renal Physiol.* **281**, F995–F1018.

Daniel, H. (1996). Function and molecular structure of brush border membrane peptide/H^+ symporters. *J. Membr. Biol.* **154**, 197–203.

Gaudichon, C., Bos, C., Morens, C., Petzke, K. J., Mariotti, F., Everwand, J., Benamouzig, R., Dare, S., Tome, D., and Metges, C. C. (2002). Ileal losses of nitrogen and amino acids in humans and their importance to the assessment of amino acid requirements. *Gastroenterology* **123**, 50–59.

Hausch, F., Shan, L., Santiago, N. A., Gray, G. M., and Khosla, C. (2002). Intestinal digestive resistance of immunodominant gliadin peptides. *Am. J. Physiol. Gastrointest. Liver Physiol.* **283**, G996–G1003.

Meisel, H., and Bockelmann, W. (1999). Bioactive peptides encrypted in milk proteins: Proteolytic activation and thropho-functional properties. *Antonie Van Leeuwenhoek* **76**, 207–215.

Protein-Calorie Deficiency—"Kwashiorkor"

SYED H. TARIQ AND JOHN E. MORLEY

St. Louis University School of Medicine and GRECC Veterans Administration Medical Center, Missouri

cachexia Severe wasting associated with a low serum albumin; due to excess cytokine production.

kwashiorkor Selective protein-calorie malnutrition with edema and a fatty liver.

marasmus Generalized starvation with loss of body fat and protein.

protein-calorie deficiency Occurs either because of deficient protein intake (undernutrition) or because of a relative excess of calorie intake (overnutrition).

sarcopenia Loss of lean body mass associated with the aging process.

undernutrition Defined by one or more of the following conditions: (a) unintentional loss of $>10\%$ of usual body weight in the preceding 3 months, (b) body weight $<90\%$ of ideal for height, or (c) body mass index <18.5. Body weight $<90\%$ of ideal body weight for height represents a risk of malnutrition, $<85\%$ of ideal body weight represents malnutrition, $<70\%$ of ideal body weight constitutes severe malnutrition, and $<60\%$ of ideal body weight is not compatible with life.

Kwashiorkor and marasmus are usually regarded as the extremes of a continuum encompassing protein-calorie malnutrition. Severe protein-energy malnutrition is a leading cause of death in children younger than 5 years of age. Ten to 20% of children die when admitted to the hospital with the diagnosis of kwashiorkor, the highest mortality of any pediatric disease. Protein-calorie malnutrition occurs in infants and children in the developing nations, but it is also reported in older adults and can occur in any country of the world. The prevalence of malnutrition varies in older people; it occurs in 1–15% of ambulatory outpatients, 25–65% of long-term care residents, and 35–65% of hospitalized patients. Malnutrition has been associated with increased mortality, morbidity, and prolongation of hospitalization. It also results in loss of function and adversely affects the quality of life.

INTRODUCTION

Dr. Cecily William first described kwashiorkor in 1935 in the natives of the Krobo–Ga–Adangbe megatribe of South Eastern Ghana. The word "kwashiorkor" (the pronunciation in Krobo is *kwasiorkor*, there being no "sh" in the alphabet) primarily translates to "the disease suffered by the child displaced from the breast." The weaned child is fed a thin gruel of poor nutritional quality or diluted baby formula (compared with mother's milk) and fails to thrive. A child with kwashiorkor tends to be older than one with marasmus and tends to develop the disease after weaning. The marasmic children are associated with early abandonment or failure of breast-feeding and with consequent infections such as gastroenteritis. Such infections result from the improper hygiene and inadequate knowledge of rearing that are often seen in slums in developing nations.

It is said that childhood undernutrition in the developing world represents the final common pathway for the expression of illiteracy, inadequate sanitation, insufficient access to medical care, poverty, poor personal hygiene, overcrowding, poor crop management, droughts, and insufficient or inappropriate use of the available resources. These factors act at national, regional, and village levels as well as at individual family levels. Kwashiorkor has been seen even in developed nations and usually results from poverty, poor nutritional information (restricting milk or formula simply because of parent or child preference, excessive dilution of milk, substituting formula with nondairy creamer, prolonging a liquid diet after hospitalization for gastroenteritis, or eliminating milk from the child's diet), presumed food allergy with avoidance of certain foods, use of diets deficient in proteins, pure vegetarian diets, and, most importantly, poor feeding skills. Chronic malabsorption, resulting from conditions such as cystic fibrosis, is the main cause of protein-calorie deficiency in children in the United States. In developed countries, most protein-calorie malnutrition is seen in older persons with cancer and in patients with HIV infection.

CLASSIFICATION

Clinically, protein-calorie malnutrition has three forms:

1. Marasmus (dry) is the most common form of protein-energy malnutrition in most developing

 Copyright 2004, Elsevier (USA). All rights reserved.

nations. It results from near starvation with deficiency of protein and nonprotein nutrients. The marasmic child consumes less than adequate amounts of food, often because the mother is unable to breast-feed, resulting in loss of fat, loss of muscle mass, and thin appearance.
2. Kwashiorkor (wet) is less common and is usually manifested as marasmic kwashiorkor. It is associated with a low serum albumin and edema.
3. The combined form is marasmic kwashiorkor. Children with this form have some edema and more fat compared to those with marasmus.

PATHOPHYSIOLOGY

Weight loss occurs when there is insufficient intake or absorption of dietary calories and/or increased expenditure of energy compared with daily requirements, and/or increased catabolism, leading to loss of body fat and protein. The healthy human body is composed of fat-free mass (FFM) and body cell mass (BCM). The FFM is composed of extracellular and intracellular water, the bony skeleton, and visceral protein, whereas the BCM is composed of intracellular water, body fat, and energy reserves in the form of intracellular glycogen and proteins. The human body stores between 15 and 25% of its energy as fat, which is greater in women than in men and is available for the production of endogenous fatty acids during starvation. The expenditure of body stores of energy is different in starvation and stress. In starvation, there is a decrease in the size of all body compartments, whereas stress reduces the BCM, increases intracellular water, and has variable effects on body fat. Weight loss is slowed by reducing the metabolic rate of the active tissues, largely the result of loss of some of the body's protein. Such a protein-depleted body requires less dietary protein for maintenance. Muscle proteins are responsible for most of the protein losses, compared to central lean tissues (liver, gastrointestinal tract, kidney, blood cells, and immune cells), which are relatively spared. Most of the physiologic functions and homeostasis are maintained, so long as the starvation energy ration is not too low, allowing for normal adaptation. The clinical consequences of this adaptation are reduced muscle mass (cardiac, respiratory), muscle weakness and functional disability, reduced cardiac and respiratory capacity, mild hypothermia, anemia, and reduced protein reserve.

Patients with severe tissue injury commonly develop systemic inflammatory response syndrome (SIRS), a hypermetabolic response. SIRS is defined as the presence of two or more of the following symptoms: fever (or profound hypothermia), tachycardia, tachypnea, and leukocytosis. Other features include changes in acute-phase serum protein concentration, increased energy expenditure, increased whole-body protein turnover, anorexia, and protein wasting. Similarly, cachexia, or cytokine-induced malnutrition, is seen in patients with inflammatory diseases or malignancy associated with continued weight loss. Classic features include changes in acute-phase serum proteins, (e.g., increased C-reactive protein, fibrinogen, and ferritin, and decreased transferrin, prealbumin, and albumin), the anemia of chronic inflammation, anorexia, and the partial nullification of a previously successful adaptation to starvation. Protein-calorie malnutrition is increasingly recognized as contributing to the protein wasting associated with organ failure in conditions such as chronic renal failure and end-stage heart disease. Protein catabolism is responsible for SIRS, whereas decreased food is the major reason for lean tissue loss in the cachectic syndromes.

ETIOLOGY OF PROTEIN-CALORIE DEFICIENCY

Initially, it was thought that kwashiorkor was due to protein deficiency in the absence of adequate energy intake. However, a number of investigators reported no differences in the nutritional backgrounds of people with kwashiorkor and those with marasmus. This resulted in alternative theories explaining the pathogenesis of kwashiorkor, such as the concept of failed adaptation, which posits that some children adapt appropriately (marasmus) to protein-calorie deficiency whereas others do not (kwashiorkor). In addition, some toxins or nutritional factors have been proposed as etiologic factors, including free radicals, aflatoxins, leukotrienes, essential fatty acids, and zinc deficiency. Mild chronic infection is common in children with kwashiorkor and thus the cytokine excess may result in extravasations of albumin from the intravascular space, producing hypoalbuminemia and edema.

Development of malnutrition is attributed to a number of risk factors, including social and psychological status, diseases producing anorexia or malabsorption, and diseases producing hypermetabolism. These risk factors can easily be remembered by using the mnemonic "Meals on wheels" (Table I). Poor, older adults are especially at increased risk of malnutrition if they live in neighborhoods with high crime rates. Such older adults, fearing for their safety, remain homebound, which limits their ability to shop and thus restricts their chances of optimal nutrition. Older adults

TABLE I Meals on Wheels Acronym: Common Risk Factors for Undernutrition

Medications (polypharmacy, herbal preparations)
Emotional causes (dysphoria, depression, psychosis)
Appetite disorders (anorexia tardive, abnormal eating attitudes, alcholism, abuse)
Late-life paranoia
Swallowing disorders (dysphagia)
Oral factors (tooth loss, periodontal infection, gingivitis, poorly fitted dentures)
No money (poverty), nosocomial infection (tuberculosis, chronic intestinal parasites, *Clostridium difficile*)
Wandering (dementia)
Hyperactivity/hypermetabolism (tremors, movement disorders, thyrotoxicosis, Addison's disease, pheochromocytoma)
Enteral disorders (chronic diarrhea, malabsorption syndromes)
Eating problems (altered food preferences, decreased taste and flavor perception)
Low-nutrient diets (low salt, low cholesterol, antidiabetic, fad diets, dilution of baby formula)
Shopping and food preparation problems (impaired mobility, unsafe environment, inadequate transportation), stones (cholelithiasis)

are further at risk if they have chronic medical conditions (e.g., Parkinson's disease, chronic obstructive pulmonary disease, heart failure, depression, gallstones, hyperthyroidism, hypercalcemia, pheochromocytoma, cancer, and hypoadrenalism). These conditions are complicated by the anorexia of aging, which is often multifactorial, encompassing age-related changes in appetite regulation, systemic diseases, iatrogenesis, and psychological factors. In young children, poverty and ignorance are the major causes of protein-calorie deficiency, and in young adults, AIDS has become the major cause of undernutrition, but in older adults, depression appears to be the major cause. In older men, testosterone deficiency and lack of physical activity are important causes of loss of lean body mass (sarcopenia).

CLINICAL EVALUATION OF PATIENTS

History

In evaluating patients, a history should be obtained from the parents, social workers, and caregivers and medical records should be reviewed. Identification of risk factors will help direct further questioning concerning functional and socioeconomic conditions. A review should include questions about diet habits and dietary restrictions (including those imposed by the patient or parents), religious and cultural beliefs, and use of special diets (heart-healthy, low-cholesterol, renal, diabetic or low-salt diets). Alcoholism is not a rare condition in older adults and should be screened for using the CAGE or Michigan Alcoholism Screening Test (MAST) tools. A history of mouth problems in the older population should be obtained, and referral made to a dentist, if required. A review of medications that can cause anorexia and weight loss is mandatory. These will include cardiovascular drugs (e.g., digoxin, amiodarone, procainamide, quinidine, spironolactone), psychiatric drugs (e.g., phenothiazines, lithium, selective serotonin reuptake inhibitors, tricyclic antidepressants), antiinfective drugs, antineoplastics, antirheumatics, drugs causing malabsorption (e.g., laxatives, cholestyramine, methotrexate), and agents increasing metabolism (e.g., theophylline, thyroid extract, L-thyroxine).

Adults should be screened for depression by history. The degree of symptomatology can be followed using either the Beck Depression Inventory or the Geriatric Depression Scale. Screening tools for dementia are available (e.g., Mini Mental Status Exam or Saint Louis University Mental Status Exam) and can be used if indicated. The Mini Nutritional Assessment (MNA) is a validated tool with a positive predictive value for detecting undernutrition of 97% in older community-dwelling adults. The sensitivity and specificity of this tool are 96 and 98%, respectively. SCALES (an acronym for sadness, cholesterol, albumin, loss of weight, eating problems, and shopping and cooking problems), a well-validated, highly sensitive tool that is simple to administer by nonmedical professionals, can be used in various clinical settings and does not require sophisticated physical examination. The Subjective Global Assessment (SGA) incorporates functional capacity as an indicator of malnutrition and relies mainly on physical signs of malnutrition and malnutrition-inducing conditions. SGA is a validated tool for prognosis (but not nutritional status) in hospitalized patients, with a sensitivity of 82% and specificity of 72%.

Physical Examination

A careful physical examination can characterize and define the extent of malnutrition. Measurements of unclothed weight and height are essential for establishing the severity of malnutrition in all patients, but may be confounded by the effect of edema and ascites. Different classifications, used to determine childhood nutritional status, include those of Gomez, Wellcome, and Waterlow.

Anthropometry

Measurements of subcutaneous fat and skeletal muscle can help to determine the severity of protein-calorie deficiency. Specialized calipers and a tape measure are used to estimate body fat from the thickness of the skin fold of the posterior mid-upper arm. This measurement is not routinely available, and requires some operator experience for accuracy. Table II indicates the severity of malnutrition in children with various upper arm circumferences. More sophisticated tools are available, such as B-mode ultrasound, bioelectrical impedance, underwater weighing, computed tomography, magnetic resonance imaging, and dual-photon absorptiometry. These tests require specialized equipment and are very costly.

Physical Findings

A child with kwashiorkor is characterized by marked muscle atrophy with normal or increased body fat, and anorexia is almost universal. On the other hand, children with marasmus are characterized by wasting of muscle mass and depletion of fat stores. Classically, children have severe constipation and are voraciously hungry. Weight loss can be recognized by decreased temporal and proximal extremity muscle mass, by decreased skin fold thickness, or by a "pinch test," especially in younger adults. Tables III and IV show the clinical findings seen in patients with protein-calorie deficiency.

Laboratory Assessment

Selected laboratory tests, most of which are widely available, are used to characterize and quantify malnutrition. However, recent data showing that most of these tests are altered by cytokine release bring into doubt their specificity and value in quantifying undernutrition.

Serum Proteins

Serum albumin, with a half-life of 2–3 weeks, is a sensitive but nonspecific measure of protein-calorie deficiency. The serum albumin level should be interpreted in the clinical setting, because it can decrease with rapid fluid shifts (as seen in acute trauma, sepsis, or acute inflammation), cirrhosis of liver, AIDS, cancer, inflammatory bowel disease, and nephrotic syndrome. Tumor necrosis factor α, interleukin-2 (IL-2), and IL-6 inhibit the synthesis of albumin and produce extravasation of albumin into the extravascular space. Several serum proteins with short half-lives are also used to measure protein-calorie deficiency; these include transferrin (half-life 9 days), prealbumin (half-life 2 days), retinol binding protein (half-life 12 hours), insulin growth factor (half-life 2–4 hours), fibronectin (half-life 4 hours), and C-reactive protein (half-life 4–6 hours). However, levels of these proteins should be interpreted with caution, because they are affected by shifts in extracellular volume that occur in acute and chronic illnesses.

TABLE II Mid-Upper-Arm Circumference in Children Aged 1–5 Years

Circumference	Level of nutrition
>14 cm	Normal
12.5–14 cm	Mild/moderate malnutrition
<12.5 cm	Severe malnutrition

TABLE III Physical Examination Findings

Kwashiorkor	Marasmus
Normal or nearly normal weight and height for age	Diminished weight and height for age
Anasarca	Emaciated and weak appearance
Moon face (rounded prominence of the cheeks)	Bradycardia, hypotension, and hypothermia
Pursed appearance of mouth	Thin, dry skin
Pitting edema in the lower extremities and periorbitally	Redundant skin folds caused by loss of subcutaneous fat
Dry, atrophic, peeling skin with confluent areas of hyperkeratosis and hyperpigmentation	Thin, sparse hair that is easily plucked
Dry, dull, hypopigmented hair that falls out or is easily plucked	
Hepatomegaly (fatty liver infiltrates)	
Distended abdomen with dilated intestinal loops	

Serum Cholesterol

Serum cholesterol levels lower than 160 mg/dl have been considered a reflection of low lipoprotein and thus of accompanying protein levels. Hypocholesterolemia seems to occur in the late stages of malnutrition, limiting its ability as a screening tool, but it is a useful prognostic indicator. Like serum proteins, its plasma concentration is altered by excess cytokine production.

TABLE IV Physical Findings of Vitamin and Mineral Deficiencies Associated with Protein-Calorie Malnutrition

Deficiency	Findings
Vitamin A	Dry conjunctiva, corneal ulceration, "goose bumps"
Thiamine (B_1)	Opthalmoplegia, congenital hepatic fibrosis, hyoreflexia, confabulation, cerebellar gait, past pointing
Riboflavin (B_2)	Angular stomatitis, cheilosis
Pyridoxine (B_6)	Peripheral neuropathy
Cobalamin (B_{12})	Optic neuritis, loss of vibratory and position sense
Vitamin C	Gingival hypertrophy, easy bruising, perifollicular hemorrhage
Vitamin D	Osteomalacia, muscular hypotonia
Vitamin K	Hemorrhages
Niacin	Dermatitis (hyperpigmentation of sun-exposed areas), diarrhea, dementia, and sometimes death
Zinc	Diminished taste, "flaky rash" on lower extremities

Immune Function

There is mounting evidence that the T lymphocyte count is a useful indicator of nutritional status and outcome. A decrease in total lymphocytes (TLCs) to less than 800/μl reflects severe undernutrition. Undernutrition can also lead to suppression of cellular immunity, manifested by the delayed hypersensitivity reaction. Undernutrition in both young children and older adults results in a marked decrease in CD4+ T cells.

TREATMENT

The guideline for the management of severely malnourished children developed by the World health Organization (WHO) consists of three phases: initial treatment, rehabilitation, and followup. The initial phase is critical, with special emphasis on treating hypoglycemia, hypothermia, infection, and dehydration. Electrolyte and vitamin deficiencies are treated in the initial phase, and treatment is extended to the rehabilitation phase, with the exception that iron is given only in the later phase. Feeding is started in the initial phase but is advanced after the second week in the rehabilitation phase. The rehabilitation phase lasts approximately 2–6 weeks. During this phase, the mother is trained to continue care at home and socioeconomic problems are addressed. In the followup phase, physical, emotional, and mental health issues are monitored and addressed and appropriately treated. It needs to be recognized that prevention by providing adequate food resources, nutritional education, and vaccinations to prevent common diseases are far more cost-effective compared to treating malnutrition in children after it occurs.

In adults, the initial step in treatment of undernutrition is the search for treatable causes. Caloric supplements, when used, should be given between meals and not with them. In some cases, the use of orexigenic drugs such as megestrol or medroxyprogesterone and/or dronabinol (tetrahydrocannibinol) may be useful.

See Also the Following Articles

Malnutrition • Nutrition Assessment • Nutrition in Aging • Pancreas, Nutritional Effects on the

Further Reading

Liu, T., Howard, R. M., Mancini, A. J., and Weston, W. L. (2001). Kwashiorkor in the United States: Fad diets, perceived and true milk allergy, and nutritional ignorance. *Arch. Dermatol.* **137**(5), 630–636.

Morely, J. E. (2003). Anorexia and weight loss in older persons. *J. Gerontol. Med. Sci.* **58A**(2), 131–137.

Ogden, C. L., Kuczmarski, R. J., Flegal, K. M., and Mei, Z. (2002). Centers for Disease Control and Prevention 2000 growth charts for the United States: Improvement to the 1997 National Center for Health Statistics version. *Pediatrics* **109**(1), 45–60.

Omran, M. L., and Morley, J. E. (2000). Assessment of protein energy malnutrition in older persons, Part I: History, examination, body composition, and screening tools. *Nutrition* **16**(1), 50–63.

Omran, M. L., and Morley, J. E. (2000). Assessment of protein energy malnutrition in older persons, Part II: Laboratory evaluation. *Nutrition* **16**(2), 131–140.

Williams, C. D. (1935). Kwashiorkor: A nutritional disease of children associated with maize diet. *Lancet* **ii**, 1151–1152.

William, C. D. (1933). A nutritional disease of childhood associated with a maize diet. *Arch. Dis. Child.* **8**, 423–428.

Protein-Losing Disorders

S. SMALE, I. FORGACS, AND I. BJARNASON
Guy's, King's, and St. Thomas' School of Medicine, London

enteropathy Small intestinal disease.
hypoalbuminemia Abnormally low concentration of albumin in the blood.
hypoproteinemia Abnormally small amounts of total protein in the blood.
intestinal absorption Capacity of the intestinal mucosa to actively absorb digested food.
intestinal permeability Property of the intestinal mucosa to permit the passage (diffusion) of substances across the mucosa.
malabsorption Disorder of the process whereby digested food is actively incorporated or received by the gastrointestinal mucosa.
maldigestion Disorder of the process whereby ingested food is converted into material suitable for assimilation by the gastrointestinal mucosa.

Extragastrointestinal causes of increased protein loss that need to be considered in the investigation of hypoproteinemia include trauma and sepsis. The processes leading to intestinal protein loss may be the result of defects in diet, metabolism, catabolism, absorption, or synthesis or may be related to specific gastrointestinal diseases. The goal of investigation and assessment of gastrointestinal protein loss, whether in the clinical or research setting, is to restore normal protein levels and to treat the underlying cause.

OVERVIEW OF ETIOLOGY

A number of factors can lead to hypoproteinemia and hypoalbuminemia, including dietary factors, maldigestion, malabsorption, reduced protein synthesis, increased catabolism, and protein loss via the gastrointestinal and renal systems or via other tissues in states of injury or inflammation.

Normal Albumin Metabolism

Albumin is quantitatively the most important of the plasma proteins and is widely recognized as an indicator of general health status. The normal serum albumin concentration is usually between 35 and 50 g/liter, the mean being slightly higher in men than in women. Albumin is synthesized from prealbumin in the liver; in normal subjects, hepatic synthesis is equal to degradation, at approximately 1.5 g/day. The half-life of albumin is approximately 15 to 20 days. Increased concentration of albumin (>50 g/liter) is pathognomonic of dehydration. Hypoalbuminemia, however, may occur consequent with a number of factors.

Malnutrition

On a global scale, malnutrition is the most common cause of hypoalbuminemia. Serum levels of prealbumin are exquisitely sensitive to changes in protein intake, reflecting its short half-life of approximately 2 days; however, as with albumin, prealbumin levels may change rapidly in the face of trauma or overwhelming sepsis even in the face of adequate nutritional intake. Albumin is not well suited for the purpose of assessing nutritional status because it may take weeks for levels to decline in response to malnutrition, reflecting its longer half-life in normal conditions; albumin levels may also drop dramatically in the face of acute illness due to increased losses and shifts in albumin distribution. Furthermore, low serum levels of albumin may take 3 or 4 weeks to return to normal, even in the setting of adequate nutrition in a recovering patient.

Protein Maldigestion and Malabsorption

Maldigestion of proteins and malabsorption of amino acids are rare causes of hypoalbuminemia because of the reserve capacity of the functions involved in these processes. Protein digestion begins in the stomach with the action of gastric proteases, which are released as proenzymes that are activated by the low pH within the gastric lumen. Nevertheless, long-standing achlorhydria does not lead to hypoalbuminemia. Pancreatic enzymes, which likewise are secreted as inactive proenzymes, are activated by enterokinase within the microvilli of the brush border in the duodenum. Having been broken down into amino acids, dipeptides, and tripeptides, these are then absorbed by different classes of amino acid transporters on the brush border. Hypoalbuminemia is not, however, a preeminent

 Copyright 2004, Elsevier (USA). All rights reserved.

feature of severe pancreatic insufficiency and is only occasionally seen in celiac disease, in which there may also be substantial loss of albumin to the intestine.

Reduced Synthesis

Albumin synthesis by the liver is dependent on changes in nutritional status, plasma oncotic pressure, cytokines, and hormones. In normal subjects, synthesis may be doubled in response to increased degradation or loss, but the synthesis of albumin may be considerably reduced in patients with severe chronic liver disease. The presence of ascites may significantly increase the total plasma volume and under these circumstances the total body albumin may be normal despite hypoalbuminemia.

Increased Degradation

Hypoalbuminemia is common in seriously ill, injured, and septic patients. This appears not to be related directly to synthesis, which remains either unchanged or increased, but rather to shifts in the distribution into the extravascular space due to increased vascular permeability and changes in the catabolism of albumin in response to proinflammatory cytokines, generated in response to cardiac failure, infection, and other inflammatory stimuli.

PROTEIN LOSS

Loss of protein via the kidney, gut, or other organs in combination with redistribution within extravascular tissues is the major cause of hypoproteinemia in the industrialized world. Such loss may occur in a number of settings.

Renal

Nephrotic syndrome is characterized by proteinuria (greater than 3 g in 24 hours), peripheral edema, and hypoalbuminemia and is the most florid manifestation of renal protein loss. Heavy proteinuria with or without the other features of nephrotic syndrome may occur in a wide variety of renal and systemic disease. Approximately 30% of patients with hypoalbuminemia associated with renal disease have an underlying systemic disease such as diabetes, amyloidosis, or systemic lupus erythromatosis, although increased gastrointestinal loss is also often evident in these cases. The rest is due to primary renal disease in the form of minimal change nephropathy, focal glomerulosclerosis, and membranous nephropathy.

Burns

Hypoalbuminemia is common in the intensive care unit setting, despite administration of albumin and aggressive nutritional support. In part, this relates to the extremely catabolic state of critically ill patients, but there is also a significant increase in leakage of albumin into the interstitial space consequent to endotoxemia and cytokine-induced changes in vascular permeability.

In the case of burns, thermal injury increases microvascular permeability, leading to increased loss of albumin into the extravascular space, into blisters, and as an exudate on the surface of burned skin; severe burns are almost invariably associated with infection leading to albumin displacement into the interstitial space, often resulting in the rapid development of hypoalbuminemia.

Gastrointestinal

Gastrointestinal protein loss should be considered in patients when no other cause for hypoproteinemia can be identified. Whereas renal causes of protein loss are associated with predominant loss of albumin, gastrointestinal loss often leads to low levels of immunoglobulins (IgG, IgA, and IgM), clotting factors (fibrinogen), transferrin, and ceruloplasmin. Unlike liver disease, in which coagulopathy is common, the loss of clotting factors is rarely significant enough to be clinically evident. However, associated fat and carbohydrate malabsorption may lead to deficiency of fat-soluble vitamins. Numerous conditions affecting any part of the gastrointestinal tract can cause protein loss. The mechanisms by which protein loss occurs are as follows:

1. Mucosal injury, with or without inflammation, erosions, or ulcers, leading to a proteinaceous exudate.
2. Increased lymphatic pressure leading to loss of proteinaceous lymphatic fluid across the surface epithelium into the lumen.
3. Increased venous pressure (constrictive pericarditis, severe right heart failure, or portal hypertension) associated with transudation of protein.

Often a single disease may lead to hypoalbuminemia via a number of different mechanisms. Hypoalbuminemia occurs when the net protein loss from the gut exceeds the ability of the liver to synthesize such proteins.

Mucosal Damage Associated with Inflammation, Erosions, or Ulcers

Intestinal inflammation is a ubiquitous feature of most intestinal diseases. The reason for this is that what-

ever causes the initial damage, it results in impaired mucosal barrier function, i.e., increased intestinal permeability. The intestinal luminal contents, whether in the small or large bowel, are exceptionally toxic and proinflammatory. Any breach in the intestinal integrity is therefore associated with intestinal inflammation. If the inflammatory response is sufficiently robust, it will involve (as a passive bystander) the microcirculation with consequent leakage of albumin. In the vast majority of cases this increased loss of albumin is not clinically evident and can only be detected by direct measurement. Even when there is evidence of hypoalbuminemia, this is often not associated with any clinical signs or symptoms (serum albumin of 25−34 g/liter). However, when severe, the peripheral edema is uncomfortable and problematic if cardiac function is compromised. Common intestinal bacteria (*Salmonella*, *Shigella*, *Yersinia*, and *Campylobacter*) or viruses (measles and other enteroviruses and rotavirus) are all associated with significant intestinal inflammation, but clinically significant protein deficiency is rare unless food intake is severely compromised, such as is frequently the case in developing countries. Classical inflammatory bowel disease is the prototype of an inflammatory condition associated with erosions and ulcers. Despite the greater inflammatory intensity in ulcerative colitis, it is Crohn's disease that is more often associated with hypoalbuminemia, due to associated surgically induced malabsorption and decreased food intake. Intestinal protein loss in Crohn's disease correlates with clinical disease activity, suggesting inflammation is the driving force for the protein loss. Hypogammaglobulinemia and cellular immune deficiency may also cause protein loss and often have a presentation similar to that of Crohn's disease.

Gastrointestinal malignancy, carcinoma, lymphoma, Kaposi's sarcoma, and the systemic vasculitides can cause significant ulceration leading to marked protein loss. Nonsteroidal antiinflammatory drug (NSAID) enteropathy may lead to protein loss and hypoalbuminemia, either with or without the presence of ulceration. This is not an uncommon cause of hypoalbuminemia; about 10% of hospitalized patients with rheumatoid arthritis have problematic hypoalbuminemia due to NSAID-induced protein loss. It is of note that NSAID-induced enteropathy is clinically silent. Cytotoxic chemotherapy, via its effects on cell turnover, may cause loss of intestinal epithelial integrity, leading to increased exposure of the mucosal immune system to luminal antigens and thence to mucosal inflammation and protein loss.

A number of diseases affect the integrity of the mucosal cells and the intermediate cell junctions, without causing frank ulceration. These include celiac disease and tropical sprue. Here the protein loss follows disruption of the villous structure and the surface epithelium. Hypoalbuminemia is particularly common and severe in ulcerative jejunitis. A number of other conditions also result in protein loss across unulcerated mucosal surfaces, including allergic gastroenteritis and microscopic colitis.

Infiltration of the mucosa by amyloid protein in patients with amyloidosis can lead to significant protein loss via the gastrointestinal tract. These patients often have concurrent nephrotic syndrome due to renal disease, to which hypoalbuminemia is often attributed.

Gastropathies

The development of giant gastric folds is associated with protein loss in the setting of a several gastric diseases. Menetrier's disease is widely recognized as being most commonly associated with significant protein loss. Hyperplasia of gastric crypts and superficial epithelium occurs in association with replacement of parietal cells with secretory glandular epithelium, leading to hypochlorhydria and increased gastric mucosal permeability.

Gastrointestinal infection has been reported to cause giant gastric folds. In children, cytomegalovirus infection has been associated with significant protein loss, but small intestinal involvement cannot always be ruled out. Infection with *Helicobacter pylori* has also been associated with the development of giant gastric folds and hypoalbuminemia, but this is an exceedingly rare complication. Eradication of *H. pylori* leads to complete resolution. Lymphocytic gastritis, also associated with *H. pylori*, may also present with a picture of giant mucosal folds, protein loss, and hypoalbuminemia.

Lymphatic Obstruction

Obstruction of small bowel lymphatics, termed intestinal lymphangiectasia, leads to dilatation of the lymphatic channels and leakage of proteinaceous lymph fluid into the bowel lumen, resulting in hypoalbuminemia. Intestinal lymphangiectasia can be due to either a primary or a secondary disorder of the lymph vessels. Regardless of the cause, impaired drainage of lymph may result in reduced absorption of fat-soluble vitamins and chylomicrons.

Primary Lymphangiectasia

Primary intestinal lymphangiectasia, a congenital maldevelopment of lymphatics, results in ectatic lymph vessels either focally or diffusely in the gut.

This condition is often associated with lymphoreticular abnormalities in other systems, including the skin. Within the gut, ectatic lymphatics may be found in any of the layers of the bowel wall containing such vessels (mucosa, submucosa, and subserosa).

The condition may be clinically silent or present with a combination of nausea, bloating, and episodic diarrhea. The diagnosis is based on the clinical, laboratory, and pathological findings. In severe cases, laboratory findings typically show reduced levels of albumin, immunoglobulins, clotting factors (although this is usually not clinically significant), and other plasma proteins. Barium follow-through may show thickened nodular mucosal folds, endoscopy characteristically shows a "snowflake" pattern overlying the small bowel mucosa, and histological examination shows dilated lymphatics, most prominently at the tip of the villi. Dilated lymph vessels can also be demonstrated on a lymphangiogram, using magnetic resonance imaging (MRI) or radionuclide imaging. Not uncommonly, the first suspicion of the disease comes from small bowel biopsy done because of malabsorption.

Secondary Lymphangiectasia

Secondary lymphangiectasia may be associated with a wide variety of causes, reflecting either systemic or local disease, including cardiac failure, portal hypertension, infiltration of local and regional lymph nodes by primary or secondary malignancy, and reactive inflammatory changes associated with infection or inflammatory bowel disease.

DIAGNOSIS OF PROTEIN-LOSING ENTEROPATHY

Diagnosis is usually based on the finding of hypoalbuminemia when there is no other cause for decreased protein production or increased loss from obvious sites. Gastrointestinal protein loss can be measured using radioisotopes (^{51}Cr-labeled albumin) or plasma clearance of α1-antitrypsin. In clinical practice, these tests are not often performed because the diagnosis is evident from the setting of hypoalbuminemia in the context of a patient with gastrointestinal symptoms and concurrent histological and radiological findings consistent with the diagnosis. The tests, especially the ^{51}Cr-labeled albumin technique, are, however, frequently used for research purposes to demonstrate therapeutic efficacy or to monitor disease.

TREATMENT

Two principles, diagnostic intervention and nutritive therapy, underlie the treatment. Diagnostically, it is first appropriate to undertake correction of the underlying cause using any available therapy. It is important to ascertain whether the "intestinal inflammation" that leads to protein loss is in part driven by luminal bacteria and their degradation products. In the small bowel, which is predominantly populated by anaerobic bacteria, metronidazole is a logical treatment, in particular in those patients with NSAID enteropathy-induced hypoalbuminemia. If the protein loss is mainly from the colon, a cephalosporin may be beneficial. Second, it is important to maintain appropriate nutrition in order to enable patients to thrive.

Nutrition

For patients with predominant lymph obstruction (lymphangiectasia), a diet that is low in saturated fat and high in protein should be encouraged. Medium-chain triglycerides can be used to supplement lipid intake. These fatty acids bypass the enteric lymphatics and enter the portal system directly. Medium-chain triglycerides have been shown to improve growth and reduce gastrointestinal symptoms in such patients. If the underlying cause for hypoproteinemia cannot be corrected, albumin infusion can be used for symptomatic and supportive treatment of hypoalbuminemia; however, albumin infusion should not be used as a means of nutritional supplementation.

Enteric protein supplementation using commercially available supplements can be used as required, but may not be necessary given sufficient intake of dietary protein in the form of meat and or other foodstuffs. When oral caloric and nutrient intake is insufficient to meet the needs of the patient, supplementary parenteral nutrition may be administered.

See Also the Following Articles

Lymph, Lymphatics, and Lymph Flow • Malabsorption • Malnutrition • NSAID-Induced Injury • Nutritional Assessment • Small Intestine, Absorption and Secretion

Further Reading

Alpers, D. H., Stenson, W. F., and Bier, D. M. (1995). "Manual of Nutritional Therapeutics," 3rd Ed. Little Brown, Boston.

Feldman, M. (1995). "Sleisenger & Fordtran's Gastrointestinal Disease," 5th Ed. W. B. Saunders, Philadelphia.

Heimberger, D. C., and Weinsier, R. L. (1997). "Handbook of Clinical Nutrition." Mosby, St. Louis.

Protein-Losing Enteropathy, Pediatric

MARVIN E. AMENT
University of California, Los Angeles

ascites Excessive accumulation of fluid in the peritoneal cavity.
enteropathy Disease of the small intestine.
hypoalbuminemia An abnormally low concentration of albumin in the blood.

Protein-losing enteropathy in pediatric patients is a pathophysiologic process and not a specific disease. Treatment requires diagnosis of the underlying condition. In many instances, making the specific diagnosis and treating the condition totally resolve the process of protein-losing enteropathy; in other instances, they do not. The symptoms of protein-losing enteropathy may include vomiting, diarrhea, abdominal pain, and allergic symptoms and the signs include positional edema, facial and orbital swelling, ankle or leg edema, abdominal distension, and ascites. In patients with congenital or primary intestinal lymphangiectasia, asymmetric lymphedema may be a clue to an underlying disorder of the lymphatic channels. Patients with protein-losing enteropathy are typically hypoalbuminemic. Therefore, the initial steps in evaluation are to determine whether or not the infant or child is suffering from malnourishment, which could cause the hypoalbuminemia. Routine urinalysis should rule out proteinemia as a source of the hypoalbuminemia, and normal transaminases, bilirubin, and prothrombin time should exclude liver disease as the cause.

DEMONSTRATION OF PROTEIN-LOSING ENTEROPATHY

Radiolabeled macromolecules, such as ^{131}I-labeled polyvinylpyrrolidone and ^{51}Cr-labeled albumin, were the tools originally used to detect protein-losing enteropathy (PLE) in adults and children. However, because of their hazards and inconvenience, mainly in the form of ionizing radiation, they are prohibited from being used in children. ^{99m}Tc-labeled human serum albumin for scintigraphy has demonstrated that 0.5% of an intravenous dose normally appears in stool or urine within 24 h. Scintigraphic imaging has been used to demonstrate gastrointestinal protein loss.

Measurement of fecal α1-antitrypsin (Fα1AT) in random stool specimens has been shown to be an easily performed test for protein-losing enteropathy and is the test most widely used for infants, children, and adults. Fα1AT is independent of serum levels of α1-antitrypsin and fecal water content. It fulfills the criterion as a marker for protein-losing enteropathy and has several other advantages. It is a serum protein and is not found in the diet. Fecal levels demonstrate only protein levels entering the intestine from the intravascular space. Because its molecular weight of 50,000 Da is similar to that of albumin, Fα1AT should mimic the behavior of albumin. Because it is a protease inhibitor, it is excreted without degradation into the stool, and urine contamination of the stool will not invalidate the result. The greatest value of the study is that it can be performed on random specimens and is not affected by urine contamination.

Antitrypsin is excreted in two forms: alone and complexed with the enzyme. Levels of Fα1AT in stool are stable. Ninety-three percent of the Fα1AT remained after 72 h of incubation at 37°C, allowing transport prior to assay. Fα1AT is not found in gastric juice and is destroyed *in vitro* after 1 h of incubation at 37°C in gastric juice, but not in duodenal juice.

Because meconium contains higher levels of α1AT than does stool, this technique is not recommended in infants younger than 1 week of age.

This simple accurate test has broadened the number of disorders in which protein-losing enteropathy is found. Furthermore, its determination has made clinicians aware that a normal serum albumin does not preclude the presence of protein-losing enteropathy, since in one series 24% of patients with PLE had normal serum albumin levels.

Those patients with hypoalbuminemia and/or edema generally have 5 to 10 times the rate of protein excretion found in those with normal serum albumin levels and PLE.

Protein-losing enteropathy in infants and children can be divided into two broad categories: those conditions associated with mucosal erosion or ulceration and those conditions associated with lymphatic obstruction.

 Copyright 2004, Elsevier (USA). All rights reserved.

Disorders associated with mucosal erosion or ulceration may be further divided into those caused by infections and those arising from noninfectious causes.

INFECTIOUS CAUSES OF PROTEIN-LOSING ENTEROPATHY

Gastrointestinal infections that typically are associated with mucosal erosions or ulceration or with damage to enterocytes or colonocytes display protein-losing enteropathy (see Table I).

Fα1AT is transiently increased in rotavirus diarrhea and its increase is related to the severity of diarrhea and its duration. Other viral infections, such as cytomegalovirus, can cause profound hypoalbuminemia and edema by injury to both small intestinal epithelium and colonic epithelium.

Parasites such as *Giardia lamblia* and *Strongyloides stercoralis* can also cause severe PLE. These infections can persist unless the diagnosis is made by examining stools for the parasite or its antigen and appropriate treatment is given to eradicate them.

Clostridium difficile and *Clostridium perfringens* can also cause PLE. Either identifying their toxins or growing the organisms will help to identify them as the cause.

Shigellosis, Salmonellosis, and *Campylobacter* infections all can cause PLE because they produce toxins that damage small intestinal and/or colonic mucosa. Although these infections may be self-limited, those patients with more severe infections are likely to display persistent signs and symptoms.

TABLE I Pathologic Mechanisms and Their Manifestations for Protein-Losing Enteropathy in Infants and Children

Mucosal erosion or ulceration
Infectious
Clostridium difficile
Clostridium perfringens
Cytomegalovirus
Rotavirus
Measles
Giardia lamblia
Strongyloides stercoralis
Salmonellosis
Shigellosis
Campylobacteria
Escherichia coli
Non infectious
Allergic gastroenteropathy
Eosinophilic gastroenteritis
Anastomotic ulcerations/ischemia
Atopic dermatitis
Burns
Gastroesophageal reflux
Gluten-sensitive enteropathy
Graft-versus-host disease
Henoch-Schoenlein purpura
Inflammatory bowel disease
Multiple polyposis
Necrotizing enterocolitis
Systemic lupus erythematosus
Lymphatic obstruction
Intestinal lymphangiectasia
Arsenic poisoning
Familial
Heart disease
Nephrotic syndrome
Noonan's syndrome
Primary
Cirrhosis with portal hypertension
IVC thrombosis post-OLT

NONINFECTIOUS INFLAMMATORY DISORDERS THAT CAUSE PROTEIN-LOSING ENTEROPATHY

Allergic gastroenteropathy or eosinophilic gastroenteritis is an entity that typically presents after the first 6 months of life. Although most cases manifest during infancy, it can develop at almost any age. These patients are typically characterized as having hypoalbuminemia, peripheral eosinophilia, the presence of Charcot-Leyden crystals in stool, iron deficiency anemia, and elevated immunoglobulin E levels. They may have associated asthma, eczema, and allergic rhinitis. Gastrointestinal symptoms may include vomiting, diarrhea, and abdominal pain.

In infants, milk protein may be the predominant protein that causes the injury. However, in some individuals, multiple proteins may be responsible for the condition and dietary restriction may be very difficult to accomplish. Usually, if the protein responsible for the allergy is removed from the diet, the injury is reversed and the markers of injury return to normal values. Corticosteroids can be used to block the injury and reverse the protein-losing enteropathy but corticosteroid use is not a good long-term solution.

Anastomotic ulcers following distal small bowel resection cause recurrent iron deficiency and PLE with or without hypoalbuminemia.

Twenty-five percent of patients undergoing bone marrow transplantation develop graft-versus-host disease involving the intestine. It is typically associated with a severe watery diarrhea. One must make certain that there are no associated infections, such as

cytomegalovirus enterocolitis, in this type of patient; this infection is curable with ganciclovir.

Inflammatory bowel disease is an important cause of protein-losing enteropathy. The severity of hypoalbuminemia and the degree of protein loss are greater in patients with Crohn's disease than in those with ulcerative colitis and greater in those with diffuse small bowel disease than in those with limited small bowel or colonic disease.

At least half of all patients with Crohn's disease have hypoalbuminemia. Many more patients with normal serum albumin levels also have protein-losing enteropathy.

Celiac disease or gluten-sensitive enteropathy is also associated with protein-losing enteropathy because of the damage to the villous structure of the small intestine. It may be mild or severe depending on the extent of the damage to the small intestine and the length of time the damage has persisted. It is reversible with adherence to a strict gluten-free diet. Within 14 days of the patient initiating the diet, the protein-losing enteropathy may be reversed.

Necrotizing enterocolitis may be associated with PLE because of the damage to the small intestine and colon caused by the disease process.

Multiple juvenile polyposis syndrome with chronic blood loss and protein-losing enteropathy has been described. These children typically present in infancy. They have a poor prognosis because of the hundreds of polyps occurring throughout the colon, small intestine, and stomach.

It is unclear why burn victims have increased Fα1AT levels. The level of elevation even in severely burned children is not great. The clinical importance is unclear.

There are a variety of other conditions rarely associated with PLE but in all of these conditions the characteristic feature is the presence of an inflammatory process.

LYMPHATIC OBSTRUCTION

Intestinal lymphangiectasia describes a group of conditions in which dilation of the lacteals, the fine, thin-walled lymphatic channels extending up into the small bowel villi, results from obstruction of the flow of lymph through the thoracic duct and into the superior vena cava. Fat intake leads to further distension and rupture of the lacteals, resulting in steatorrhea and drainage of lymph into the intestine. The resulting loss of protein, lymphocytes, and immunoglobulins ultimately leads to hypoalbuminemia, lymphopenia, and hypogammaglobulinemia.

Intestinal lymphangiectasia may be primary or secondary to other causes of lymphatic obstruction (see Table II). A variety of familial forms with intestinal lymphangiectasia have been described. Some have been described with asymmetrical lymphedema in addition to intestinal lymphangiectasia.

The hypoalbuminemia and lymphopenia may be reversed if the patient responds to dietary therapy. Patients with lymphangiectasia may have an increased risk of infection when untreated secondary to the loss of gammaglobulin and immunocytes. Other patients with Noonan's syndrome have been described with multifocal lymphatic dysplasia. Some but not all of these patients respond to a diet low in long-chain triglycerides (LCT) that provides less than 10% of the lipid as LCT and receives it as medium-chain triglycerides. However, in rare cases, dietary restrictions of fat even below that level did not help reverse the hypoalbuminemia and lymphopenia. A rare case of arsenic poisoning has also been associated with PLE.

TABLE II Intestinal Lymphangiectasia—Disorders Associated with Enteric Protein Loss, Lymphopenia, and Hypoalbuminemia

Primary intestinal lymphangiectasia
Isolated
Associated with lymphatic abnormalities elsewhere in the body
Secondary intestinal lymphangiectasia
Cardiovascular anomalies
Congestive heart failure
Constrictive pericarditis
Budd–Chiari syndrome
Glenn shunt
Fontan procedure
Superior vena cava thrombosis
Inferior vena cava thrombosis
Mesenteric lymphatic involvement
Lymphomas
Tuberculosis
Sarcoidosis
Radiation therapy
Volvulus
Intestinal inflammatory diseases
Systemic lupus erythematosus
Tuberculosis
Behçet's syndrome
Crohn's disease
Drugs
Arsenic
Chemotherapeutic agents
Thoracic duct obliteration
Iatrogenic
Mediastinal tumor

Cardiac disorders or surgical procedures resulting in transmission of elevated pressure from the right atrium into the superior vena cava and thoracic duct, including clinically silent constrictive pericarditis, have been associated with the development of intestinal lymphangiectasia.

Early survival after the Fontan operation for a single ventricle has improved substantially since its inception; however, late-term complications continue to be problematic. One such complication of PLE is seen in 3 to 15% of patients with Fontan procedures. Recently some investigators have shown that increased mesenteric vascular resistance is characteristic of those with Fontan's and PLE.

Recent studies have shown that portal hypertension in some patients with chronic liver disease may result in PLE. The elevated portal pressure may lead to secondary intestinal lymphangiectasia. Reversal of this phenomenon has been observed after liver transplantation.

Obstruction to hepatic venous outflow as occurs in Budd-Chiari syndrome can also cause PLE as can inferior vena cava occlusion following orthotopic liver transplantation.

Children with the rare condition of hypertrophic gastropathy or Menetrier's disease (also known as transient hypertrophic gastropathy) present with abdominal pain and vomiting. They gradually develop edema and ascites. The cause of some of these cases has been suggested to be allergy and other cases appear to be caused by cytomegalovirus infection. Cases of transient hypertrophic gastropathy have been described in children with *Helicobacter pylori* infections.

See Also the Following Articles

Celiac Disease, Pediatric • Colitis, Ulcerative (Pediatric) • Cow Milk Protein Allergy • Crohn's Disease, Pediatric • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Lymph, Lymphatics, and Lymph Flow • Parasitic Diseases, Overview

Further Reading

Catassi, C., Cardinali, E., D'Angelo, G., *et al.* (1986). Reliability of random fecal α-antitrypsin determination on nondried stools. *J. Pediatr.* **109**, 500–502.

Chang, J. M., Hwang, S. J., Chen, H. C., and Lai, Y. H. (2002). Edema due to protein-losing enteropathy—A disorder rarely considered by nephrologists. *Clin. Nephrol.* **57**, 392–397.

Chaubal, N., Shah, M., Dighe, M., Ketkar, D., and Joshi, A. (2002). Juvenile polyposis of the colon: Sonographic diagnosis. *J. Ultrasound Med.* **21**, 1311–1314.

Cheung, Y. F., Tsang, H. Y., and Kwok, J. S. (2002). Immunologic profile of patients with protein-losing enteropathy complicating congenital heart disease. *Pediatr. Cardiol.* **23**, 587–593.

Cohen, H. A., Shapiro, R. P., Frydman, M., *et al.* (1991). Childhood protein losing enteropathy associated with *Helicobacter pylori* infection. *J. Pediatr. Gastroenterol. Nutr.* **13**, 201–203.

Hamada, A., Kondoh, T., Kamei, T., Tominaga, N., Tsuru, A., Matsumoto, T., Matsuzaka, T., and Moriuchi, H. (2002). Protein-losing enteropathy complicated with recurrent convulsions and developmental delay in a 4-month-old boy. *Pediatr. Int.* **44**, 686–689.

Harms, H. K., Zimmer, K. P., Kurnik, K., Bertele-Harms, R. M., Weidinger, S., and Reiter, K. (2002). Oral mannose therapy persistently corrects the severe clinical symptoms and biochemical abnormalities of phosphomannose isomerase deficiency. *Acta Pediatr.* **91**, 1065–1072.

Holzknecht, N., Helmberger, T., Beuers, U., Rust, C., Wiebecke, B., and Reiser, M. (2002). Cross-sectional imaging findings in congenital intestinal lymphangiectasia. *J. Comput. Assist. Tomogr.* **26**, 526–528.

Lee, W. S., John, P., McKiernan, P., de Goyet, J. D. V., and Kelly, D. A. (2002). Inferior vena cava occlusion and protein-losing enteropathy after liver transplantation in children. *J. Pediatr. Gastroenterol. Nutr.* **34**, 413–416.

Marino, B. S. (2002). Outcomes after the Fontan procedure. *Curr. Opin. Pediatr.* **14**, 620–626.

Rychik, J., and Gui-Yang, S. (2002). Relation of mesenteric vascular resistance after Fontan operation and protein-losing enteropathy. *Am. J. Cardiol.* **90**, 672–674.

Rychik, J., and Spray, T. L. (2002). Strategies to treat protein-losing enteropathy. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* **5**, 3–11.

Seok, J. W., Kim, S. J., Lee, S. H., Kim, I. J., and Kim, Y. K. (2002). Protein-losing enteropathy detected on Tc-99m HAS and TC-99m MDP scintigraphy. *Clin. Nucl. Med.* **27**, 431–433.

Silverman, A., and Roy, C. C. (1983). "Pediatric Clinical Gastroenterology," 3rd Ed., p. 377. Mosby, St. Louis, MO.

Tokatli, A., and Ozsoylu, S. (1991). Ménétrier disease in a child. *J. Pediatr. Gastroenterol. Nutr.* **12**, 404–406.

Proton Pump Inhibitors

JAI MOO SHIN AND GEORGE SACHS
University of California, Los Angeles and Veterans Administration Greater Los Angeles Healthcare System

gastric H^+,K^+-ATPase A P_2-type ion-motive ATPase enzyme that carries out an electroneutral exchange of cytoplasmic protons for extracytoplasmic potassium.

omeprazole The prototypical, clinically used proton pump inhibitor.

proton pump inhibitor An inhibitor of gastric H^+,K^+-ATPase activity.

Proton pump inhibitors make up a class of compounds that inhibit gastric H^+,K^+-ATPase activity and thereby inhibit gastric acid secretion. Controlling acid secretion is important in healing gastric ulcer, peptic ulcer, and related diseases. Gastric acid is secreted from the parietal cell on stimulation by histamine, acetylcholine, and gastrin. These stimulants change the morphology of the parietal cell from the resting state to the stimulated state with relocation of the gastric H^+,K^+-ATPase from the tubular vesicles to the apical canalicular membranes. The gastric H^+,K^+-ATPase transports H^+ (H_3O^+) ion from the cytoplasmic region to the lumen with the exchange of K^+ from the lumen to the cytoplasmic side. The gastric H^+,K^+-ATPase consists of two subunits: one is the α-subunit, composed of approximately 1034 amino acids, and the other is the β-subunit, which has approximately 290 amino acids and six or seven N-linked glycosylation sites depending on species. The H^+,K^+-ATPase α-subunit has 10 transmembrane segments and the β-subunit has 1 transmembrane segment. Inhibition of this acid pump enzyme is known to be the most effective therapy for controlling gastric acid secretion.

INTRODUCTION

The proton pump inhibitors (PPIs) can be classified into two groups: irreversible and reversible inhibitors. Irreversible covalent inhibitors are either substituted 2-(pyridinemethylsulfinyl)benzimidazoles or a similar structure, pyridylmethylsulfinyl pyrido-imidazole, which inhibit the pump enzyme by covalently binding to the α-subunit of the H^+,K^+-ATPase. Reversible proton pump inhibitors are mostly K^+-competitive inhibitors, which inhibit the gastric H^+,K^+-ATPase activity by competing with potassium ions.

Since a substituted benzimidazole was first reported to inhibit the H^+,K^+-ATPase, many PPIs have been synthesized and are now in clinical use. Typical proton pump inhibitors in clinical use are listed in Fig. 1.

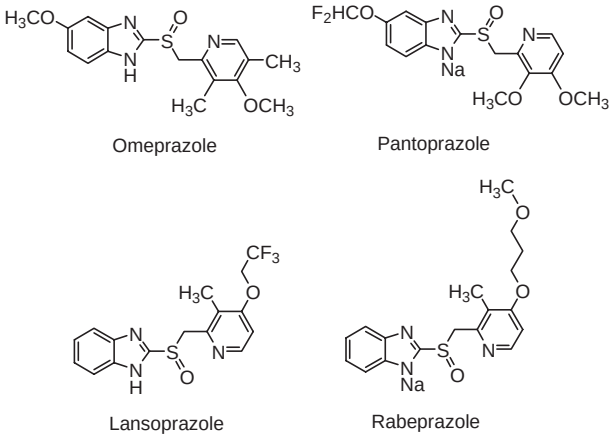
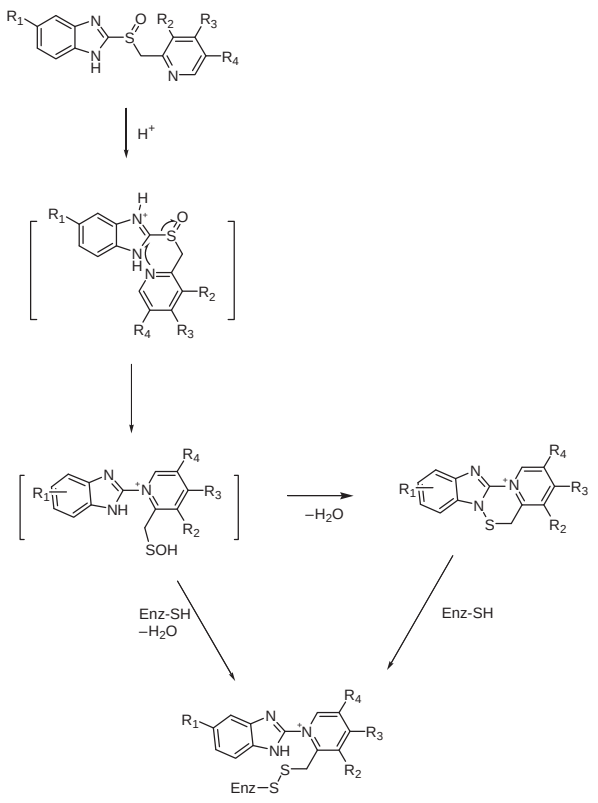


FIGURE 1 Chemical structure of irreversible proton pump inhibitors in clinical use.

IRREVERSIBLE COVALENT BINDING INHIBITORS

Timoprazole was the first compound that was found to inhibit the gastric H^+,K^+-ATPase by covalent binding. This compound is 2-(pyridylmethyl)sulfinylbenzimidazole. The first pump inhibitor used clinically was omeprazole (2-[[3,5-dimethyl-4-methoxypyridin-2-yl]methylsulfinyl]-5-methoxy-1H-benzimidazole). Compounds in this class are acid-activated prodrugs. For example, omeprazole, due to being a weak base, accumulates in the acidic space of the parietal cell and, by acid-catalyzed rearrangement, becomes a thiol-reactive cationic sulfenic acid and/or sulfenamide that binds to cysteinyl-SH groups to form disulfides as shown in Fig. 2. The activation is initiated by the protonation of the pyridine nitrogen, which is followed by transfer of this proton to the benzimidazole nitrogen, which then increases the electrophilic reactivity of the C-2 of the benzimidazole. Now pyridine is ready to attack this 2-position carbon of benzimidazole to form sulfenic acid, which then rearranges to form

 Copyright 2004, Elsevier (USA). All rights reserved.



	R1	R2	R3	R4
O z	–O H 3	– H 3	–O H 3	– H 3
n z	–O H 2	–O H 3	–O H 3	–H
n z	–H	– H 3	–O H 2 3	–H
R z	–H	– H 3	–O H 2 H 2 H 2O H 3	–H

FIGURE 2 Mechanism of action of irreversible proton pump inhibitors. Proton pump inhibitor, a substituted benzimidazole compound, accumulates in the acidic lumen and is converted to active sulfenic acid and (or) sulfenamide, which bind(s) extracytoplasmic cysteines of the gastric H^+,K^+-ATPase.

sulfenamide. The three other proton pump inhibitors, lansoprazole, pantoprazole, and rabeprazole, also undergo similar acid-catalyzed rearrangement to form active sulfenic acids and/or sulfenamides. Although all of the proton pump inhibitors accumulate in the secretory canaliculus of the stimulated parietal cell by virtue of being protonatable weak bases, they show variation in the rate of acid activation. The rate of acid activation is fastest for rabeprazole, equal for omeprazole and lansoprazole, and slowest for pantoprazole. Also, substituted benzimidazole inhibitors showed slightly different effects depending on the inhibitor structure. For instance, omeprazole-bound enzyme is favored in the E_2 form. Another inhibitor, rabeprazole (E3810), 2-[[4-(3-methoxypropoxy)-3-methylpyridin-2-yl]methylsulfinyl]-1H-benzimidazole, stabilized the E_1 form of the enzyme after binding. It is claimed that the K^+-dependent dephosphorylation from the phosphoenzyme was inhibited in the rabeprazole-bound enzyme but not in the omeprazole-bound enzyme, whereas phosphoenzyme formation in the absence of K^+ was inhibited in both the E3810- and the omeprazole-bound enzymes.

Omeprazole binds to cysteines in the extracytoplasmic regions of M5/M6 (Cys-813) and M7/M8 (Cys-892). Pantoprazole binds only to the cysteines in M5/M6, Cys-813 and Cys-822, and lansoprazole binds to Cys-321 in M3/M4, to Cys-813 in M5/M6, and to Cys-892 in M7/M8. These data suggest that, of the 28 cysteines in the α subunit, only the cysteines present in the M5/M6 domain are important for inhibition of acid secretion by the PPIs.

Interestingly, the half-time of recovery of acid secretion in rats following omeprazole treatment was measured to be ~15 h, whereas pump protein half-life is 54 h. If omeprazole binding to the gastric H^+,K^+-ATPase is irreversible, recovery of acid secretion after omeprazole inhibition should be similar to *de novo* synthesis of pump protein, i.e., approximately 54 h. The shorter half-life of acid secretion recovery compared to the half-life of pump enzyme suggests that there must be disulfide cleavage between the pump enzyme and the bound omeprazole moiety and reactivation of inhibited pump. In humans, the half-life of the inhibitory effect on acid secretion is ~28 h for omeprazole and ~46 h for pantoprazole. Only pantoprazole showed a half-life of a duration similar to that of a protein half-life. It was shown that recovery of acid secretion following inhibition by all PPIs other than pantoprazole may depend on both protein turnover and reversal of the inhibitory disulfide bond. In contrast, recovery of acid secretion after pantoprazole may depend entirely on new protein synthesis.

Another type of irreversible proton pump inhibitor is the pyridinylmethylsulfinyl imidazopyridines, such as TU-199 and anilinomethylsulfinyl benzimidazole. Neither is yet available for clinical use.

K^+-COMPETITIVE INHIBITORS

K^+-competitive inhibitors can be thought of as acid pump antagonists. These antagonists contain protonatable nitrogens but have a variety of core structures. One type is represented by the imidazopyridine derivatives such as SCH28080; others are piperidinopyridines, substituted 4-phenylaminoquinolines, pyrrolo[3,2-c]quinolines, guanidino-thiazoles, 2,4-diaminopyrimidine derivatives, and scopadulcic acid.

SCH28080, a substituted imidazo[1,2α] pyridine, is the best defined compound among other reversible proton pump inhibitors. SCH 28080, 3-cyanomethyl-2-methyl-8-(phenylmethoxy) imidazo[1,2α]pyridine, inhibited the H^+,K^+-ATPase competitively with K^+. It binds to free enzyme extracytoplasmically in the absence of substrate to form E_2(SCH 28080) complexes. SCH 28080 occupies the same space in the lumen where omeprazole binds. SCH28080 inhibits ATPase activity with high affinity in the absence of K^+. SCH 28080 has no effect on spontaneous dephosphorylation but inhibits K^+-stimulated dephosphorylation, presumably by forming a E_2-P · [I] complex. Hence, SCH 28080 inhibits K^+-stimulated ATPase activity by competing with K^+ for binding E_2P. Steady state phosphorylation is also reduced by SCH 28080, showing that this compound also binds to the free enzyme. At present, no acid pump antagonist is available for clinical use but some are in development.

CLINICAL USE

Proton pump inhibitors are orally active and used for the therapy of gastric ulcer, duodenal ulcer, gastroesophageal reflux disease, Zollinger-Ellison syndrome, and, combined with antibiotics, for eradication of *Helicobacter pylori*. The primary effect of these proton pump inhibitors is gastric acid suppression. The degree of acid suppression correlates with healing rates for reflux esophagitis and peptic ulcer. Rabeprazole is fast-acting with pain relief. Pantoprazole shows a longer half-time of restoration of acid secretion. Omeprazole and pantoprazole show an increase in acid inhibitory effect over several days of repeated administration. Lansoprazole shows maximal inhibition after the first day. There is poor correlation between maximal plasma drug concentration (C_{max}) and the degree of acid suppression. Instead, the area under the plasma concentration−time curve (AUC) correlates with acid suppression. Among proton pump inhibitors, omeprazole 20 mg and rabeprazole 20 mg showed a significantly lower AUC than pantoprazole 20 and 40 mg and lansoprazole 30 mg. However, all proton pump inhibitors used clinically, omeprazole, lansoprazole, pantoprazole, and rabeprazole, show approximately equivalent potency for gastric acid suppression. Pantoprazole is available for intravenous use and is used to suppress acid secretion in intensive care situations.

All proton pump inhibitors are extensively metabolized in liver by P450 cytochromes. People with hepatic impairment have shown a 7- to 9-fold increase in the AUC with a prolongation of the plasma half-life to 4−8 h. Approximately 3% of the population with a genetic polymorphism are poor metabolizers and show a 5- to 10-fold increase of AUC. Elderly people showed a 50−100% increase in the AUC since hepatic metabolism was poor. All four proton pump inhibitors are metabolized mainly by CYP 2C19 and CYP 3A4. CYP 2C19 generates hydroxylation of proton pump inhibitor, which is responsible for 80% of clearance in the case of omeprazole. CYP 3A4 generates sulfonylation of proton pump inhibitors.

Recently, S-omeprazole became available for clinical use. Omeprazole is a racemate consisting of S- and R- enantiomers. The R-form of omeprazole is sensitive to CYP 2C19 and CYP 3A4 enzymes and the S-form is less sensitive to these CYP enzymes. S-omeprazole has a longer plasma half-life than omeprazole, providing longer acid suppression.

Acid suppression correlates with healing rates for reflux esophagitis and peptic ulcer. Good healing for reflux esophagitis was achieved when the intragastric pH was greater than 4 for 16 h/day and healing for peptic ulcer was best achieved when the intragastric pH was greater than 3. In patients with reflux esophagitis, lansoprazole 30 mg provided faster symptom relief than omeprazole 20 mg; however, no significant difference was observed compared to omeprazole 40 mg in terms of healing rates and symptom relief. Rabeprazole 20 mg and pantoprazole 40 mg provided equivalent healing rates and symptom relief compared to omeprazole 20 mg. In peptic ulcer disease and duodenal ulcers, all four proton pump inhibitors showed very similar efficacy, whereas rabeprazole and lansoprazole claimed a little fast symptom relief.

H. pylori has been successfully eradicated by triple-therapy regimens: clarithromycin, amoxycillin, and proton pump inhibitor. There are no significant differences among four proton pump inhibitors, omeprazole, lansoprazole, pantoprazole, and rabeprazole, when used for this purpose.

See Also the Following Articles

Duodenal Ulcer • Gastric Acid Secretion • Gastric H^+,K^+-ATPase • Gastric Ulcer • Gastrinoma • Gastroesophageal Reflux Disease (GERD) • *Helicobacter pylori* • Pharmacology, Overview

Further Reading

Abelo, A., Andersson, T. B., Antonsson, M., Naudot, A. K., Skanberg, I., and Weidolf, L. (2000). Stereoselective metabolism of omeprazole by human cytochrome P450 enzymes. *Drug Metab. Dispos.* **28**, 966−972.

Sachs, G., Briving, C., Shin, J. M., Hsersey, S., and Wallmark, B. (1995). The pharmacology of the gastric acid pump, the H^+,K^+-ATPase. *Annu. Rev. Pharmacol. Toxicol.* **35**, 277–305.

Shin, J. M., Bayle, D., Bamberg, K., and Sachs, G. (1998). The gastric H^+,K^+-ATPase. *In* "Advances in Molecular and Cell Biology," Vol. 23A, pp. 101–142. JAI Press, Greenwich, CT.

Shin, J. M., and Sachs, G. (2002). Mechanism of restoration of acid secretion following treatment with proton pump inhibitors. *Gastroenterology* **123**(5), 1588–1597.

Stedman, C. A. M., and Barclay, M. L. (2000). Review article: Comparison of the pharmacokinetics, acid suppression and efficacy of proton pump inhibitors. *Aliment. Pharmacol. Ther.* **14**, 963–978.

Pruritus of Cholestasis

NORA V. BERGASA

College of Physicians and Surgeons, Columbia University

nociceptive Unpleasant or painful sensation.

central opioidergic neurotransmission Signals in the central nervous system carried out by opioid peptides and receptors.

opioids Endogenous peptides with an affinity for opioid receptors.

Pruritus is a complication of liver disease, in particular when associated with cholestasis. Its etiology is unknown; most treatments have been empirical and unsatisfactory. Intractable pruritus is an indication for liver transplantation. The pruritus of cholestasis does not correlate with serum markers of liver disease.

ETIOLOGY OF THE PRURITUS OF CHOLESTASIS

It is inferred that the substance(s) that mediates pruritus in cholestasis is made in the liver and excreted in bile. In support of this inference is the finding that patients with cholestasis report disappearance of pruritus after liver transplantation and after removal of blockage in cases of extrahepatic biliary obstruction.

Bile Acids

Bile acids accumulate in the plasma of patients with cholestasis; however, there is no scientific evidence that demonstrates their role in the mediation of the pruritus of cholestasis.

In the context of bile acids as pruritogens in cholestasis, three observations must be considered: (1) the accumulation of bile acids in the skin or interstitial fluid may not have any relevance to the pruritus (2) there are patients with cholestasis and high serum concentrations of bile acids who do not report pruritis, and (3) spontaneous relief of pruritus does not correlate with decreases in serum bile acids.

Histamine

Histamine is pruritogenic. Histamine-mediated lesions, such as skin erythema and edema, however, are not skin findings in patients who experience pruritus secondary to cholestasis. The lack of specific antipruritic effect of antihistamines in patients with the pruritus of cholestasis does not support a role of histamine in this type of pruritus.

The Endogenous Opioid System

Three lines of evidence suggest that patients with cholestasis have alterations in the endogenous opioid system: (1) an opioid withdrawal-like reaction can be experienced by patients with cholestasis after the administration of opiate antagonists, (2) the concentration of opioid peptides in the serum of patients with cholestasis is higher than that of control subjects, and (3) the immunoreactivity of Met-enkephalin, one of the endogenous opioid peptides, is enhanced in the liver of patients with primary biliary cirrhosis, a liver disease associated with cholestasis, in contrast to that of the disease control livers. The opiate withdrawal-like reaction suggests that in cholestasis there is increased central opioidergic neurotransmission.

Copyright 2004, Elsevier (USA). All rights reserved.

Sachs, G., Briving, C., Shin, J. M., Hsersey, S., and Wallmark, B. (1995). The pharmacology of the gastric acid pump, the H^+,K^+-ATPase. *Annu. Rev. Pharmacol. Toxicol.* **35**, 277–305.

Shin, J. M., Bayle, D., Bamberg, K., and Sachs, G. (1998). The gastric H^+,K^+-ATPase. *In* "Advances in Molecular and Cell Biology," Vol. 23A, pp. 101–142. JAI Press, Greenwich, CT.

Shin, J. M., and Sachs, G. (2002). Mechanism of restoration of acid secretion following treatment with proton pump inhibitors. *Gastroenterology* **123**(5), 1588–1597.

Stedman, C. A. M., and Barclay, M. L. (2000). Review article: Comparison of the pharmacokinetics, acid suppression and efficacy of proton pump inhibitors. *Aliment. Pharmacol. Ther.* **14**, 963–978.

Pruritus of Cholestasis

NORA V. BERGASA

College of Physicians and Surgeons, Columbia University

nociceptive Unpleasant or painful sensation.

central opioidergic neurotransmission Signals in the central nervous system carried out by opioid peptides and receptors.

opioids Endogenous peptides with an affinity for opioid receptors.

Pruritus is a complication of liver disease, in particular when associated with cholestasis. Its etiology is unknown; most treatments have been empirical and unsatisfactory. Intractable pruritus is an indication for liver transplantation. The pruritus of cholestasis does not correlate with serum markers of liver disease.

ETIOLOGY OF THE PRURITUS OF CHOLESTASIS

It is inferred that the substance(s) that mediates pruritus in cholestasis is made in the liver and excreted in bile. In support of this inference is the finding that patients with cholestasis report disappearance of pruritus after liver transplantation and after removal of blockage in cases of extrahepatic biliary obstruction.

Bile Acids

Bile acids accumulate in the plasma of patients with cholestasis; however, there is no scientific evidence that demonstrates their role in the mediation of the pruritus of cholestasis.

In the context of bile acids as pruritogens in cholestasis, three observations must be considered: (1) the accumulation of bile acids in the skin or interstitial fluid may not have any relevance to the pruritus (2) there are patients with cholestasis and high serum concentrations of bile acids who do not report pruritis, and (3) spontaneous relief of pruritus does not correlate with decreases in serum bile acids.

Histamine

Histamine is pruritogenic. Histamine-mediated lesions, such as skin erythema and edema, however, are not skin findings in patients who experience pruritus secondary to cholestasis. The lack of specific antipruritic effect of antihistamines in patients with the pruritus of cholestasis does not support a role of histamine in this type of pruritus.

The Endogenous Opioid System

Three lines of evidence suggest that patients with cholestasis have alterations in the endogenous opioid system: (1) an opioid withdrawal-like reaction can be experienced by patients with cholestasis after the administration of opiate antagonists, (2) the concentration of opioid peptides in the serum of patients with cholestasis is higher than that of control subjects, and (3) the immunoreactivity of Met-enkephalin, one of the endogenous opioid peptides, is enhanced in the liver of patients with primary biliary cirrhosis, a liver disease associated with cholestasis, in contrast to that of the disease control livers. The opiate withdrawal-like reaction suggests that in cholestasis there is increased central opioidergic neurotransmission.

Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Therapies for the Pruritus of Cholestasis

Drug	Rationale for use in the treatment pruritus of cholestasis	Dose	Potential side effects
Cholestyramine	Decrease in enterohepatic circulation of bile acids	4 g pre- and postbreakfast and after other meals, not to exceed 16 g per day	Bloating, constipation, malabsorption
Antihistamines	Unknown	Variable	Sedation, dry mouth
Phenobarbital	Enhancement of pruritogen excretion	Variable	Sedation
Rifampicin	Unknown	300 to 450 mg po per day in divided doses	Hepatotoxicity
Opiate antagonists (naloxone, naltrexone)	Antagonism of endogenous opioids	Naloxone: 0.4 mg iv bolus followed by infusions of 0.2 μg/kg/min Naltrexone: 50 mg/day	Opiate withdrawal-like reaction, hepatotoxicity
Serotonin type 3 receptor antagonist (ondansetron)	Interference with mechanisms of nociception	4 to 8 mg po per day	Constipation, headache

Drugs with agonist properties at opioid receptors are pruritogenic, in particular, when centrally administered. Opiate-induced pruritus is effectively treated with opiate antagonists. The pruritogenic property of opiate drugs, the suggestion of increased opioidergic tone in cholestasis, and the anecdotes reporting that opiate antagonists decreased the pruritus of cholestasis suggest that endogenous opioids mediate this type of pruritus, at least in part. A central mechanism has been proposed. Various clinical trials of opiate antagonists for the treatment of the pruritus of cholestasis were conducted. These studies included objective methodology, which allowed for the recording of scratching activity, the behavior that specifically results from pruritus, independent of gross body movements. Opiate antagonists were associated with a decrease in the perception of pruritus and scratching activity. These results support a role of endogenous opioids in the pruritus of cholestasis.

Serotonin System

The serotonin system is involved in the mediation of nociceptive stimuli. Ondansetron, an antagonist of type 3 serotonin receptors, which are found both in the central nervous system and on peripheral nerves, was reported to decrease the pruritus of cholestasis in studies that used subjective methodology.

TREATMENT OF THE PRURITUS OF CHOLESTASIS

Drugs

Table I lists some of the drugs used to treat the pruritus of cholestasis. Some drugs appear to have a rationale for their use and some do not. The doses listed are summarized from published studies; they should be individualized. For a complete review of side effects, the reader is referred to original sources.

Invasive Procedures to Treat the Pruritus of Cholestasis

The need to provide relief to patients with the pruritus of cholestasis is underscored by the use of invasive procedures that aim to remove hypothetical pruritogens from the circulation. These procedures include charcoal hemoperfusion, plasmapheresis, partial external diversion of bile, and ileal diversion. The nature of any relevant substance(s) removed by these interventions is not known.

See Also the Following Articles

Biliary Tract, Developmental Anomalies of the • Bile Formation • Cholestatic Diseases, Chronic • Cirrhosis • Histamine • Liver Transplantation

Further Reading

Bergasa, N. V., and Jones, E. A. (1993). The pruritus of cholestasis. *Semin. Liver Dis.* **13**, 319–327.

Bergasa, N. V., Mehlman, J. K., and Jones, E. A. (2000). Pruritus and fatigue in primary biliary cirrhosis. *Bailliere Best Pract. Res. Clin. Gastroenterol.* **14**, 643–655.

Bergasa, N. V., Liau, S., Homel, P., and Ghali, V. (2002). Hepatic Met-enkephalin immunoreactivity is enhanced in primary biliary cirrhosis. *Liver* **22**, 107–113.

Jones, E. A., and Bergasa, N. V. (1995). The pruritus of cholestasis: Potential pathogenic and therapeutic implications of opioids. *Gastroenterology* **108**, 1582–1588.

Thornton, J. R., and Losowsky, M. S. (1988). Opioid peptides and primary biliary cirrhosis. *Br. Med. J.* **297**, 1241–1242.

Psychiatric Issues, Overview

CAROL S. NORTH AND ELIZABETH TERRY WESTERHAUS
Washington University School of Medicine, St. Louis, Missouri

Crohn's disease Inflammatory bowel disease of part of or the entire gastrointestinal tract and that extends all the way through the intestinal wall.
dysmenorrhea Painful menstruation.
dyspepsia Indigestion, upset stomach, or pain or discomfort centered in the upper abdomen.
dysuria Painful urination.
fibromyalgia Disorder of pain and tenderness of muscle and adjacent connective tissue. Synonyms include fibrositis and fibromyositis.
functional gastrointestinal disease Collection of persistent or recurrent gastrointestinal symptoms not known to be caused or produced by any specific morphologic, biochemical, or physiologic abnormalities.
globus Sensation of a lump or something stuck in the throat, tightness of the throat, or inability to swallow.
inflammatory bowel disease Group of chronic gastrointestinal diseases, including ulcerative colitis and Crohn's disease, involving inflammation of the gastrointestinal tract.
irritable bowel syndrome Group of functional gastrointestinal disorders of the bowel in which abdominal discomfort or pain is associated with defecation, or a change in bowel habit and bowel movements, not due to known structural gastrointestinal disease.
nocturia Excessive urination at night.
psychoform Psychological symptoms suggesting psychiatric disorders that the individual does not have.
psychosomatic Bodily symptoms presumed to arise from psychological origins.
somatization disorder Psychiatric illness that occurs predominantly in women, characterized by multiple physical complaints throughout the body's organ system without medical explanation.
somatoform Physical symptoms suggesting a medical basis but without medical explanation.
structural gastrointestinal disease Well-documented associations of disease of the gastrointestinal tract, with observable structural or morphologic changes.
ulcerative colitis Inflammatory bowel disease characterized by inflammation of the more superficial layers, the mucosa and submucosa, of the colon.

The sheer prevalence of psychiatric illness is of considerable significance to all medical practice. In the general population, one in five people suffer from a diagnosable psychiatric illness in any given year, but most do not find their way to treatment. In medical treatment settings, a current psychiatric diagnosis can be made in approximately one out of every three patients, but, in reality, most psychiatric illness goes undetected.

SCOPE AND PREVALENCE OF PSYCHIATRIC COMORBIDITY IN GASTROINTESTINAL DISEASE

Recognition of psychiatric illness in gastrointestinal disease is important not only because psychiatric illness is a source of much suffering in itself, but also because it is associated with significant disability, worse medical outcomes, and higher medical care costs. Recognition and treatment of psychiatric disease can significantly reduce medical care costs and can be expected to improve the management of comorbid gastrointestinal disease.

Different gastrointestinal diseases vary in both the prevalence and types of associated psychiatric disorders. In the following discussions, the gastrointestinal diseases are examined separately for their associations with particular psychiatric disorders, focusing on how the different psychiatric comorbidities present special problems of management.

PSYCHIATRIC DISORDERS IN STRUCTURAL AND FUNCTIONAL GASTROINTESTINAL DISEASE

Functional gastrointestinal (GI) disorders are defined as persistent or recurrent groups of gastrointestinal symptoms not linked to known morphologic, biochemical, or other physiologic abnormalities. Structural gastrointestinal disease, in contrast, presents with objective diagnostic features on physical examination or on organ-specific tests, such as GI imaging. Psychiatric comorbidity in gastrointestinal disease seems to comply with the functional/structural dichotomy, with higher rates of psychiatric disorders—in particular, somatoform disorders and syndromes—co-occurring

 Copyright 2004, Elsevier (USA). All rights reserved.

with the functional disorders. Studies of psychiatric illness in gastrointestinal disease have found strong associations with functional gastrointestinal disorders, especially irritable bowel syndrome and chest pain of presumed esophageal origin, but weak associations with structural gastrointestinal disease. Structural and functional gastrointestinal diseases appear to have different psychiatric comorbidity patterns.

Psychiatric Illness in Structural Gastrointestinal Diseases

Historically, many of the structural gastrointestinal diseases have been considered to have psychosomatic origins, despite clear evidence of morphologic abnormalities in the gastrointestinal organs affected. Although the term "psychosomatic" has fallen from accepted usage, previously entrenched ideas persist even with empirical data contradicting them. In part, reluctance to move away from previously established ideas may be because hypotheses of psychosocial origins in medical disease cannot be definitively tested to confirm or disprove them.

In particular, peptic ulcer disease and inflammatory bowel disease have a long tradition of presumed psychosomatic origins, and have accumulated the largest literature supporting these ideas. A large psychoanalytically based literature has accumulated that describes the psychological mechanisms thought to produce these diseases, along with other work that describes how these psychological factors might interact physiologically to yield structural changes in the bowel. The research literature supporting the conceptualization of these structural gastrointestinal diseases as psychosomatic, however, is seriously flawed. The work most strongly supporting this view is the most methodologically flawed.

Possibly feeding the assumption of psychosomatic origins to inflammatory bowel disease and peptic ulcer disease is their potential to be mistaken for, or confused with, functional gastrointestinal diseases with clear psychiatric associations, particularly irritable bowel disease and functional dyspepsia. Although the symptom presentations of peptic ulcer disease and functional dyspepsia are identical, somewhere between 50 and 90% of patients presenting with suspected peptic ulcer disease do not show evidence of an ulcer on physical examination. Because these disorders have such different psychiatric comorbidity patterns, observations made in functional disorders cannot be considered applicable to structural disorders of the same organs.

Methodological problems with this literature start with sampling problems, such as sampling bias, lack of confirmation of the diagnosis, and nonseparation of diseases (especially Crohn's disease and ulcerative colitis, which are actually quite different in their associations with psychopathology). Other serious methodological problems in this literature are lack of a comparison group or use of an inappropriate comparison group, such as healthy individuals in the community, and failure to match or control for important confounders, especially gender (because of higher rates of many psychiatric disorders among women). This latter issue is illustrated by the demonstrated prevalence of psychiatric illness in clinical populations with medical disease in general, which runs around one-third. Therefore, to consider a disease to have specific psychiatric associations, one must demonstrate that the associated psychopathology occurs in rates significantly greater than one-third in Western cultures. Measurement problems in these studies include use of assessment instruments without established validity or reliability, failure to apply diagnostic standards, and lack of comparability of methods across studies. Studies using subjective self-report of symptoms or well being by patients confound medically based symptoms of the disease with symptoms originating in functional overlay that subgroups of patients in all populations demonstrate. Therefore, studies documenting a reduction of symptoms with behavioral treatments do not prove psychological origins of the disease, because the results may merely reflect treatment of the functional overlay and not of the disease. Finally, published studies have often drawn conclusions not warranted by the data, and the classic error is assumption of causality from mere association. When two entities statistically occur together, causal relationship may go either direction, or the apparent association may be only indirect through association of both entities with a third variable.

The idea that ulcerative colitis may be a psychosomatic disease first appeared in the scientific literature in 1930, and over several decades a voluminous literature in support of this notion has accumulated. An exhaustive review of the subject found that studies lacking a comparison group were significantly more likely than controlled studies to conclude that ulcerative colitis is psychiatrically based. Looking past the morass of flawed studies, this review proceeded to examine the seven studies with the best methodology. These seven studies all concluded that ulcerative colitis was not distinguishable from other serious medical illnesses in its association with psychopathology.

Crohn's disease is another structural gastrointestinal disease with an established following that holds it to be a disorder rooted in psychological origins. Review of the research literature on this disease reveals the same

kinds of major methodological flaws resulting in unwarranted conclusions based on little evidence. Of 50 original studies examined, only 12 were based on samples of 10 or more subjects. Unlike the best literature on ulcerative colitis, however, the best literature on Crohn's disease overwhelmingly concluded that Crohn's disease was significantly associated with psychiatric illness, on the order of 50% lifetime prevalence. The psychiatric illness consisted largely of major depressive and anxiety disorders. The psychiatric disorders were no more likely to precede than to follow the onset of the Crohn's disease, thus supporting no causal pathway for psychiatric disease in the generation of the gastrointestinal disorder. A study directly comparing Crohn's disease and ulcerative colitis found that the patients with Crohn's disease described poorer psychosocial adjustment, generally reduced well being, and more GI symptoms. Thus, studies suggest that Crohn's disease may be the only structural bowel disease to be associated with psychiatric disorders.

Peptic ulcer disease, probably more than any other structural gastrointestinal disorder, has long been assumed to be a psychosomatic or stress-related condition. Despite this widely held conviction, research shows important distinctions in patterns of psychiatric comorbidity between peptic ulcer disease and other functional gastroduodenal disease. Direct comparison of functional dyspepsia with duodenal ulcer patients found functional dyspepsia to be associated with significantly greater psychopathology (especially anxiety and depression), multiple somatic complaints (especially including dyspepsia symptoms and musculoskeletal symptoms), worse general health, reduced functioning, lower quality of life, and less patient satisfaction with health care received. The patients with duodenal ulcer were older and smoked more often, and almost all were infected with *Helicobacter pylori*. Among duodenal ulcer patients, those with the fewest classic historical risk factors (sex, age, seasonality, family history, smoking, alcohol use, coffee consumption, nonsteroidal antiinflammatory drug use, blood type, serum pepsinogen I, and *H. pylori* antibody titers) had the greatest psychopathology. Thus, it is suggested that clinicians evaluating patients not matching the usual patient profile for duodenal ulcer should be alert for psychologic factors and features suggestive of functional disease instead of, or comorbid with, the peptic ulcer disease. It is important for researchers and clinicians to appreciate that the inevitable occasional psychiatric comorbidity with peptic ulcer disease is not proof of causality from one condition to the other.

Another gastrointestinal disease said to be associated with psychiatric illness is pancreatic cancer. Compared to patients with advanced gastric carcinoma, pancreatic carcinoma patients were found to have higher rates of self-reported depression. The psychiatric findings may be a presenting feature, preceding the diagnosis of cancer. Systematic studies have not been carried out to determine whether it is simple dysphoria or the fully diagnosable syndrome of major depression that precedes the diagnosis of pancreatic cancer. Speculation holds systemic effects of neurotransmitters of the pancreas to be the vehicle for the generation of depressive symptomatology in pancreatic cancer.

Psychiatric Illness in Functional Gastrointestinal Diseases

Functional gastrointestinal disorders are the most prevalent conditions in gastroenterology practice, constituting up to 50% of presenting problems. Functional gastrointestinal disorders are associated with significant disability, reduced quality of life, and increased medical care costs. Functional gastrointestinal disease is also apparently quite prevalent in the general population, identified in nearly two-thirds (62%) of people assessed by a random telephone survey. In this study, the most prevalent class of functional gastrointestinal disease observed was functional bowel disease (42%), with functional esophageal disease (29%) ranking second and functional anorectal syndromes (23%) ranking third. The most prevalent individual disorders were functional heartburn (22%), functional anorectal pain (17%), functional constipation (15%), abdominal bloating (13%), and irritable bowel syndrome (12%). Even though it ranked only fifth in prevalence among functional gastrointestinal diseases, irritable bowel syndrome is the functional disorder that seems to get the most press.

Functional bowel diseases may overlap with one another diagnostically. This is demonstrated by the documented co-occurrence of functional dyspepsia and irritable bowel syndrome. The functional gastrointestinal syndromes have not been differentiated from one another based on their association with psychopathology. The functional gastrointestinal disease best studied for psychiatric comorbidity is irritable bowel syndrome. Therefore, the following discussion treats the functional gastrointestinal disorders collectively, using irritable bowel syndrome as a model and including research from other functional disorders as relevant to the discussion.

Specific psychiatric disorders have not often been examined together in a single study of irritable bowel syndrome, most studies either focusing on a single disorder or reporting a combined diagnostic rate. The

lifetime prevalence of psychiatric disorders in patients with irritable bowel syndrome has been reported in 72–93%, the most prevalent individual diagnoses being major depression (8–61%), anxiety disorders (4–61%), hysteria (17–28%) or somatization disorder (32–48%), and undiagnosed psychiatric disorder (31–32%). The onset of psychiatric illness precedes the onset of the irritable bowel syndrome in four out of five cases, suggesting that the association does not routinely signify the generation of psychopathology from the bowel disease. Although less thoroughly studied, psychiatric comorbidity in other functional disorders has been found in the ranges reported for irritable bowel syndrome. For functional dyspepsia, psychiatric comorbidity has been reported to be 87%, with anxiety disorders diagnosed in 67%. Psychiatric comorbidity in esophageal spasm was reported as 84%, consisting largely of anxiety disorders, major depression, and somatization disorder.

A specific psychiatric disorder singled out in studies of functional gastrointestinal disease is panic disorder. Panic disorder has been reported in association with four functional disorders: globus (25% lifetime panic disorder), noncardiac chest pain of presumed esophageal origins (24–59% prevalence of panic disorder), functional dyspepsia (no panic disorder), and irritable bowel syndrome (23–29% lifetime panic disorder). Panic disorder was identified as being associated with functional gastrointestinal symptoms in a general population study, but more comprehensive analysis found no special association of panic disorder with functional gastrointestinal symptoms. All other psychiatric disorders were found to be similarly associated with the functional gastrointestinal symptoms. Panic disorder was diagnosed in only 4% of individuals with functional gastrointestinal symptoms, which is two to four times the rates (1–2%) in the general population. This nonetheless represented only a fraction of the overall psychopathology, identified in 48% of individuals with functional gastrointestinal symptoms.

Functional gastrointestinal disease is abundant among patients sampled from psychiatric treatment settings. Irritable bowel syndrome has been diagnosed in 17–42% of panic disorder patients, 37% of generalized anxiety disorder patients, 27% of major depression patients, and 42% of patients with alcohol abuse or dependence. These rates are considerably higher than the 12% prevalence of irritable bowel syndrome reported for the general population, suggesting that the association of psychiatric illness with irritable bowel syndrome in patient populations may possibly reflect treatment-seeking bias. Studies of treatment settings specializing in gastrointestinal disease report greater psychiatric comorbidity in association with functional bowel disease compared to data collected from primary care sources. Examination of the relationship of irritable bowel syndrome and psychiatric illness in the general population, in which treatment bias does not apply, does not uniformly show these conditions to be associated, although the reports have been mixed.

The relatively high rates of somatization disorder reported in patients with functional gastrointestinal disease may well represent underestimates of the prevalence of somatization disorder in relation to functional gastrointestinal disease. The source of underestimation lies with the self-report method of obtaining the medical history through cross-sectional patient interviews. Medical histories of patients with this disorder are prone to inaccuracies stemming from misrepresentation of somatic complaints as medically based and failure to report many previous symptoms for which these patients had sought medical intervention.

Somatization disorder was named for its characteristic multiple complaints of symptoms in multiple organ systems. Patients complain that everything is wrong with virtually all the organ systems in their bodies, yet no medical explanation for the symptoms can be found. The multiple complaints offered by patients with somatization disorder include not just physical symptoms but also psychological symptoms. Patients with somatization disorder attending a university psychiatry clinic were found to complain of more depressive symptoms compared patients with a diagnosis of major depression attending the same clinic, as many manic symptoms compared to patients with bipolar disorder, and as many psychotic symptoms compared to patients with schizophrenia. The psychological profile of patients with somatization disorder on the Minnesota Multiphasic Personality Inventory is characterized by a style of exaggeration in reporting symptoms and high rates of endorsement of symptoms on all clinical scales. Therefore, somatization disorder is not only a somatoform disorder in which patients complain of physical symptoms of medical disorders they do not have, but also a "psychoform" disorder in which they also complain of symptoms of psychiatric disorders they do not have. Based solely on their complaints of multiple psychiatric symptoms, additional psychiatric diagnoses may be attributed to them. It could be that the apparent association of irritable bowel syndrome with psychiatric illness is based simply on the comorbidity of irritable bowel syndrome with somatization disorder, the actual source of the many complaints generating psychiatric diagnoses in these patients.

Because irritable bowel syndrome is a disorder defined completely by subjective patient symptom report,

it is possible that the symptom complaints establishing the diagnosis of irritable bowel syndrome in patients with somatization disorder are merely a part of the many symptom complaints of the somatization disorder. The irritable bowel syndrome identified in these patients may not represent the same condition as the irritable bowel syndrome of patients without somatization disorder.

Patients with somatization disorder describe their symptoms as more severe and complain of them more vocally than do other patients. Patients with this disorder are notoriously difficult to manage in the treatment setting. They also report more medication side effects and encounter more treatment complications than other patients. They may be highly demanding and dramatic. Similar observations have been made about the presentation of irritable bowel syndrome. The amount of overlap of somatization disorder with irritable bowel syndrome in patient populations, however, makes it impossible to separate characteristics of the somatization disorder from those of the irritable bowel syndrome. In the irritable bowel syndrome population, much that is ascribed to irritable bowel syndrome may actually represent manifestations of somatization disorder. The extreme nature of the complaints of patients with somatization disorder embedded in an irritable bowel sample may skew findings in irritable bowel syndrome toward poor outcomes and comorbid psychopathology. Therefore, studies of irritable bowel syndrome must separate patients with somatization disorder from those without, to determine what characteristics are driven by the irritable bowel syndrome rather than by somatization disorder.

Irritable bowel syndrome has been described in association with a number of other functional disorders, both within the gastrointestinal tract and in other organ systems. Functional gastrointestinal disorders such as irritable bowel syndrome or functional dyspepsia have been reported in association with fibromyalgia. The more severe the irritable bowel syndrome, the more likely it is associated with fibromyalgia. Irritable bowel syndrome has also been associated with irritable bladder, functional headaches, backaches, muscle aches, dysmenorrhea, urinary frequency and urgency, nocturia, dysuria, sensation of incomplete bladder emptying, chronic pelvic pain, painful sexual intercourse, other sexual dysfunction, dizziness, sleep disturbances, and chronic fatigue. It also shares overlapping features with temporomandibular joint syndrome, premenstrual syndrome, and mitral valve prolapse. Patients with functional bowel disorders visit primary care physicians for symptoms outside the gastrointestinal tract three times more often compared to healthy individuals. Somatization disorder among patients with functional gastrointestinal disease may well account for the apparent overlap of irritable bowel syndrome with other functional disorders.

CAUSAL DIRECTIONALITIES IN COMORBIDITY OF GASTROINTESTINAL AND PSYCHIATRIC DISORDERS

In functional gastrointestinal disease, comorbid psychiatric disease is clearly associated. Although it has long been assumed that this comorbidity reflects a psychiatric contribution to the development of the gastrointestinal disorder, there is no empirical database to support this assumption, and the temporal sequence (gastrointestinal disorder first, psychopathology later) in the majority is not consistent with it. Another consideration is a possibility of opposite causal directionality in which the gastrointestinal disorders lead to psychopathology. Although it may seem intuitive that irritable bowel syndrome might engender anxiety and dysphoria that could be construed to be part of major depressive and anxiety syndromes, the somatization disorder associated with functional bowel disease characteristically starts early in life, typically in the decade following puberty, and therefore is an unlikely outcome of irritable bowel syndrome. Further, it is not intuitive that irritable bowel syndrome would generate major psychiatric disorders when the same cannot be demonstrated in severe structural gastrointestinal diseases with significant morbidity and mortality not found in irritable bowel syndrome.

A remaining source of the link between psychiatric illness and functional gastrointestinal disease is that a third variable associated with both the functional gastrointestinal disease and the psychiatric disorder provides an indirect link between them. One candidate to represent this connection is stressful life events, which are speculated to contribute to the development of functional gastrointestinal disease. In particular, sexual abuse has been described as associated with symptoms of irritable bowel syndrome, dyspepsia, and heartburn, is and implicated in the generation of irritable bowel disorder. However, causality of sexual abuse in functional medical disorders has not been documented, and sexual abuse is unlikely to represent a specific etiological factor. A more likely causal connection is somatization disorder, which occurs in a significant proportion of functional bowel disease cases and which is associated with other psychopathology. More research is needed to determine to what degree the characteristics attributed to functional

gastrointestinal disorders actually represent features of the associated somatization disorders, and what remains of functional gastrointestinal disease when somatization disorder cases are removed and the functional gastrointestinal disease population is reexamined.

A third variable link between functional gastrointestinal disease and psychopathology is thought to reside within the neurologic wiring of the brain and gut, which share common neurotransmitter substances and are connected in a brain–gut communication network. The role such systems may play in the association of psychopathology with functional gastrointestinal disease is theoretical, however, and empirical validation is needed.

Unlike functional gastrointestinal diseases, structural bowel diseases are not generally associated with psychopathology beyond the nonspecific associations of psychiatric disorders and chronic medical disease in general. One exception appears to be Crohn's disease, which has been seen to be significantly associated with psychopathology. Although the associated psychopathology in both Crohn's disease and functional gastrointestinal disorders includes major depressive and anxiety disorders, irritable bowel syndrome differs from Crohn's disease in its frequent association with somatization disorder, which in turn may be a large source of the associated anxiety and depression. Therefore, although these two conditions share an association with psychiatric illness, the specific psychiatric comorbidities are very different. Although long-standing in medical lore, assumptions of psychological etiologies of structural bowel disease lack empirical support in the research literature.

TREATMENT IMPLICATIONS

In consideration of comorbid psychiatric disease in the treatment of gastrointestinal illness, it is important to differentiate major psychiatric disorders, such as major depression, from psychological symptoms, such as dysphoria. Major depression and anxiety disorders constitute serious psychiatric illness associated with significant suffering, disability, morbidity, mortality, poor medical outcome, and increased medical costs. Documented success in satisfactory treatment of major psychiatric disorders dictates active vigilance for psychiatric disorders in the management of medical disorders. Not to be confused with major psychiatric illness, dysphoria and anxiety symptoms may be understandable responses to serious medical illness, or they may be part of many "psychoform" complaints among patients with somatization disorder who are by definition polysymptomatic. It is important to approach both research and clinical management in psychopathology in the context of gastrointestinal disease diagnostically, to make these distinctions, which will drive treatment decisions.

Dysphoria and anxiety associated with gastrointestinal disease in medical practice may be managed by supportive psychotherapy and medical education. Major depression and anxiety disorders respond to treatment with antidepressant and antianxiety medications and psychotherapy. Depressive and anxiety complaints arising from somatization disorder rather than primary major depressive or anxiety disorders are best managed in the context of somatization disorder. Management of somatization disorder involves avoiding the potential for iatrogenic harm to the patient through surgical and invasive diagnostic procedures and abusable medications such as narcotics and benzodiazepines, which these patients often receive inappropriately in response to the magnitude of subjective complaints that are not substantiated by objective findings of disease. Interpretation of stressful life events is best considered in the context of psychiatric diagnosis, with caution in ascribing causal attribution, because association of such events with functional gastrointestinal disorders and psychiatric disorders does not demonstrate etiologic origins. Psychiatric disorders may increase risk of the occurrence of negative life events, and somatization disorder is associated with increased reporting of traumatic events.

Although recognition of the associations of psychopathology, stressful life events, and functional gastrointestinal disease can aid the recognition of functional gastrointestinal disease and psychiatric disorders, treatment should be based on diagnostic assessments rather than on assumption of causality. For example, psychotherapy to solve psychological conflicts surrounding a sexual abuse history based on assumptions of the origins of irritable bowel in the history of abuse is not a logical approach to treatment of the disorder based on empirical research.

CONCLUSIONS AND SUMMARY

Functional and structural gastrointestinal diseases have been seen to have very different associations with psychopathology. Despite a long history of assumption of psychological origins in many structural gastrointestinal diseases, such as inflammatory bowel disease and peptic ulcer disease, empirical data do identify associations supporting such a relationship. In contrast, functional bowel diseases have been seen to be highly associated with psychopathology, but evidence does not specify a causal relationship. Management of

psychopathology in the context of gastrointestinal illness should be aimed at diagnosis of major psychiatric illness and application of treatment approaches appropriate to the disorders identified.

See Also the Following Articles

Colitis, Ulcerative • Crohn's Disease • Duodenal Ulcer • Functional (Non-Ulcer) Dyspepsia • Gastric Ulcer • Irritable Bowel Syndrome • Psychosociology of Irritable Bowel Syndrome

Further Reading

Clouse, R. E., and Lustman, P. J. (1983). Psychiatric illness and contraction abnormalities of the esophagus. *N. Engl. J. Med.* **309**, 1337–1342.

Guze, S. B. (1984). Psychosomatic medicine: A critique. *Psychiatry Dev.* **2**, 23–30.

Helzer, J. E. (1985). Psychiatric aspects of inflammatory bowel disease. *In* "Colon, Rectal, and Anal Surgery: Current Techniques and Controversies" (I. J. Kodner, R. D. Fry, and J. P. Roe, eds.), pp. 329–355. Mosby, St. Louis.

Lenze, E. L., Miller, A., Munir, Z., Pornoppadol, C., and North, C. S. (1999). Psychiatric symptoms endorsed by somatization disorder patients in a psychiatric clinic. *Ann. Clin. Psychiatry* **11**, 73–79.

Liss, J. L., Alpers, D. H., and Woodruff, R. A. (1973). The irritable colon syndrome and psychiatric illness. *Dis. Nerv. Syst.* **34**, 151–157.

Lubrano, E., Iovino, P., Tremolaterra, F., Parsons, W. J., Ciacci, C., and Mazzacca, G. (2001). Fibromyalgia in patients with irritable bowel syndrome: An association with the severity of the intestinal disorder. *Int. J. Colorectal Dis.* **16**(4), 211–215.

Magni, G., diMario, F., and Bernasconi, G. (1987). DSM-III diagnoses associated with dyspepsia of unknown cause. *Am. J. Psychiatry* **144**, 1222–1223.

Mayer, E. A., Fass, R., and Fullerton, S. (1998). Intestinal and extraintestinal symptoms in functional gastrointestinal disorders. *Eur. J. Surg.* **583**, 29–31.

Miller, A., North, C. S., Clouse, R. E., Alpers, D. H., and Wetzel, R. D. (2001). Irritable bowel syndrome, psychiatric illness, personality, and abuse: Is somatization disorder the missing link? *Ann. Clin. Psychiatry* **13**, 25–30.

Nimnuan, C., Rabe-Hesketh, S., Wessely, S., and Hotopf, M. (2001). How many functional somatic syndromes? *J. Psychosom. Res.* **51**(4), 549–557.

North, C. S., Alpers, D. H., Thompson, S. J., and Spitznagel, E. L. (1996). Gastrointestinal symptoms and psychiatric disorders in the general population: Findings from the NIMH Epidemiologic Catchment Area project. *Dig. Dis. Sci.* **41**, 633–640.

North, C. S., Clouse, R. E., Spitznagel, E. L., and Alpers, D. H. (1990). The relation of ulcerative colitis to psychiatric factors: A review of findings and methods. *Am. J. Psychiatry* **147**(8), 974–981.

North, C. S., and Guze, S. B. (1998). Somatoform disorders. *In* "Washington University: Adult Psychiatry" (S. B. Guze, ed.), pp. 269–283. Mosby, St. Louis.

Smith, G., Rost, K., and Kashner, T. (1995). A trial of the effect of a standardized psychiatric consultation on health outcomes and costs in somatizing patients. *Arch. Gen. Psychiatry* **52**, 238–243.

Wilhelmsen, I., Haug, T. T., Ursin, H., and Berstad, A. (1995). Discriminant analysis of factors distinguishing patients with functional dyspepsia from patients with duodenal ulcer. Significance of somatization. *Dig. Dis. Sci.* **40**, 1101–1111.

Young, S. J., Alpers, D. H., Norland, C. C., and Woodruff, R. A. (1976). Psychiatric illness and the irritable bowel syndrome: Practical implications for the primary physician. *Gastroenterology* **70**, 162–166.

Yunus, M. B., Masi, A. T., and Aldag, J. C. (1989). A controlled study of primary fibromyalgia syndrome: Clinical features and association with other functional syndromes. *J. Rheumatol. Suppl.* **19**, 62–71.

Psychosociology of Irritable Bowel Syndrome

BRENDA B. TONER AND TARYN N. TANG
University of Toronto

abuse Threats or actions of an emotional, sexual, or physical nature in which a power differential exists between the perpetrator and the victim.

antidepressant A class of drugs, the primary effect of which is to correct neurotransmitter imbalance in the central nervous system occurring in a major depression. Antidepressants tend to be useful in the functional gastrointestinal disorders, both to treat concomitant anxiety and depression and to reduce pain.

anxiety disorder Excessive anxiety and worry that cannot be controlled and is persistent with a range of symptoms. Mild forms include phobias; more severe forms include panic disorder.

biofeedback The use of electronic or mechanical devices to provide visual and/or auditory information (feedback) on a biological process for the purpose of teaching an individual to control the biological process.

bulking agents Macromolecular substances that increase stool bulk and soften feces by water binding. They may be of plant origin (e.g., bran, *Plantago*) or synthetic (e.g., polyethylene glycol). They cannot be split by the enzymes of the human gut, but may be partially digested by the colonic flora.

depressive disorders Depression accompanied by reduced activity, reduced appetite, changes in sleep pattern, feelings of fatigue or loss of energy, and feelings of guilt or worthlessness. Suicidal ideas occur in severe forms.

dualism A concept, first proposed by Descartes, that separates mind and body. Cartesian dualism (the biomedical model) is the dominant model of illness in Western society and is challenged by the biopsychosocial model.

health-related quality of life The impact that illness has on quality of life, including the individual's perception of his or her illness.

Irritable bowel syndrome (IBS) is defined as a group of functional bowel disorders in which abdominal discomfort or pain is associated with defecation or a change in bowel habit. The recently revised Rome II diagnostic criteria for IBS include abdominal pain or discomfort persisting for at least 12 weeks or more that has two or three of the following features: (1) relief with defecation; and/or (2) a change in frequency of stool; and/or (3) a change in form (appearance) of stool. Symptoms associated with IBS include the following: abnormal stool frequency (more than three bowel movements per day or less than three bowel movements per week), abnormal stool form (hard or loose/watery stool), abnormal stool passage (straining or urgency, feeling of incomplete evacuation), passage of mucus, and bloating or feeling of abdominal distension. IBS is estimated to affect 9–22% of the Western population. However, of those affected, very few seek medical consultation or treatment for their gastrointestinal (GI) symptoms. IBS accounts for 28% of gastroenterological practice and 12% of primary care in Western societies. Recent surveys have suggested that individuals with GI symptoms first present to a physician between the ages of 30 and 50 years and there appears to be a decrease in reporting frequency among older adults. Prevalence rates have been found to be similar among Caucasians and African Americans. There have been only a limited number of studies of non-Western ethnic groups but the available reports suggest that IBS appears to be as common in India, China, Japan, and South America as in Western societies. Surveys in the United States indicate that IBS is associated with unnecessary procedures and surgeries and results in over 2.2 million prescriptions per year. In Canada, it is estimated that at least 1.3 million people have IBS, which accounts for over $1.3 billion per year in direct and indirect health care costs. The economic impact of IBS is considerable because in addition to these medical costs, IBS results in decreased work productivity. Next to the common cold, it ranks as the second most common cause of work absenteeism. Furthermore, difficulties in diagnosis and treatment produce uncertainty, frustration, and dissatisfaction within the patient–physician relationship, which can affect patient satisfaction, adherence to treatment, and the clinical outcome.

BIOPSYCHOSOCIAL MODEL OF IRRITABLE BOWEL SYNDROME

Within a biopsychosocial framework, it is no longer imperative that researchers try to assess whether pain or bowel symptoms are caused by physiological, psychological, or social factors. Rather, the goal of investigators is to determine the extent to which

 Copyright 2004, Elsevier (USA). All rights reserved.

these multiple factors contribute to irritable bowel syndrome (IBS). The research and clinical challenge faced by investigators and clinicians is to determine for each individual the degree to which each of these interacting factors is present and modifiable using a multitude of therapeutic options. Before turning to the specific psychosocial issues that have been identified in the literature on IBS, this article will discuss the stigma associated with IBS and the fact that most of the research to date has focused on one gender when describing psychosocial factors and interventions in IBS.

STIGMA

Nearly every medical specialty has identified a functional somatic syndrome (FSS). These syndromes are usually defined by physical symptoms unexplained by organic disease. The term "functional" implies a disturbance of physiological function rather than anatomical structure. Functional is often contrasted with organic and is often conceptualized as psychogenic and less "real." The stigma associated with the term functional has resulted in a variety of labels used to describe FSSs, including somatic disorders, health anxiety, physical symptoms unexplained by organic disease, unexplained medical symptoms, and psychophysiological disorders. One of the most common FSSs that have received increased research and clinical attention during the past decade is IBS.

In Western societies in general, and in medicine in particular, there is a moral implication to having a FSS. Underlying the dualistic metaphysics of Western medicine, illness is attributed to impersonal causes or viewed as an accident that befalls the patient as victim or else is viewed as psychologically caused and mediated and potentially under the person's voluntary control. The morally pejorative connotations of a FSS often leave patients believing that their problems are not being treated as real but instead are due to a psychological or moral defect or weakness. Research indicates that disorders disproportionately prevalent in women, such as IBS, are often trivialized or described as psychological in origin. Thus, women may be especially attentive to the possibility that their illness symptoms are not being taken seriously. Accordingly, it is important to highlight that, when a person with IBS is referred to a health care professional, he or she may come into the office with the expectation that the caregiver does not think his or her symptoms are real or serious, but are "all in his or her head." Validating the reality of the person's symptoms and challenging society's view of the artificial dualism of functional/organic components of illness can enhance the therapeutic alliance. The stigma associated with IBS further highlights the need for patients and health professionals to conceptualize IBS within a biopsychosocial framework.

GENDER

IBS is a disorder that is diagnosed mostly in women. Although men and women are affected by IBS, studies consistently demonstrate that women outnumber men within the nonpatient population, within primary care settings, and within tertiary care settings. To date, most of the information about IBS has been drawn from women research participants. A review of the literature indicated that the majority of studies investigating IBS used only women in their samples. Moreover, among the few studies that did include men, a gender difference analysis was rarely performed. Studies that examined sex differences did so only in a descriptive manner, did not test for statistically significant differences, and sampled only a small number of people. Of those studies that investigated sex differences and included both male and female participants in the sample, the focus was in the areas of frequency of physician visits, psychological symptoms, physical symptoms, and abuse histories. Thus far, the literature suggests that there are few consistent sex differences. However, since a significant percentage of patients with IBS are women, the issue of gender must be integrated into the conceptualization and treatment of this disorder.

ROLE OF PSYCHOSOCIAL FACTORS

There is an increasing consensus in the literature that specific psychosocial factors are not characteristic of the disorder and thus are not considered as a part of a diagnosis. Nonetheless, psychosocial factors are important to identify in order to help to understand their role in clinical presentation and in planning relevant interventions. It is important to keep in mind that the specific psychosocial factors described below are not unique to patients with IBS but have been reported to occur in other patients with chronic medical conditions. Research on the role of psychosocial factors in IBS has focused on four general areas: anxiety and depression; stress; abuse; and quality of life.

Anxiety and Depression

A large proportion of patients with IBS manifest concurrent anxiety and/or depression. Research using standardized interviews indicates that among IBS patients in tertiary health centers, the prevalence of a

psychiatric disorder (mainly anxiety and depressive disorders) ranges from 40 to over 90%. Three possible, interrelated, explanations for the association between IBS and anxiety and depression are suggested here. The first is that the co-occurrence of IBS with anxiety and depression may simply be due to overrepresentation of these disorders among women in the general population. A second possible explanation is that people with IBS and an associated anxiety or depressive disorder seek more specialized help than people with IBS without anxiety and/or depression. Thus, the former group may be more likely to enter the health care system because they have more difficulty coping with their IBS symptoms. Finally, it is also possible that IBS may have such a debilitating effect on individuals' lives that anxiety and depression manifest. It is important to note that although the prevalence of anxiety and depressive disorders is overrepresented among the subset of patients seeking health care for their gastrointestinal (GI) symptoms, these disorders are not associated with IBS per se.

Stress

Stressful events produce GI symptoms in most people but patients with IBS may be particularly susceptible and have a greater reactivity to stress. Among patients with IBS, stress is associated with symptom onset and severity. Moreover, research suggests that patients with IBS report more lifetime and daily stressors when compared with other medical populations or healthy controls. Thus, identifying specific types of stressors, including psychological, social, physical, dietary, and hormonal stressors, that are related to an exacerbation of GI symptoms may be helpful in devising a treatment plan that utilizes appropriate and relevant interventions.

Abuse

Patients diagnosed with functional bowel disorders are significantly more likely than patients with organic bowel disorders to have had a history of forced intercourse and to have experienced frequent physical abuse. Moreover, female patients with functional GI disorders are more likely to have experienced life-threatening physical abuse and to have been victims of rape in their lifetime when compared with control group female patients. Studies indicate that a history of physical or sexual abuse among patients with functional GI disorders contributes to poor health status. Specifically, women who have experienced abuse are more likely than women who have not experienced abuse to complain about pelvic pain, headaches, backaches, fatigue, and joint pain to their physicians.

Sexual abuse, in particular, may act as a nonspecific but severe psychological stressor, increasing physiological arousal and thereby triggering or exacerbating a patient's GI symptoms. However, despite the strong evidence regarding the prevalence and ramifications of physical and sexual abuse, few studies to date have examined the impact of emotional abuse among patients with IBS. In one recent study, emotional abuse was found to be significantly more prevalent in the women with IBS seen at tertiary centers than in the women with organic bowel disorders. Using a qualitative, semistructured interview, it was found that a large number of women with IBS perceived that past physical, sexual, or emotional abuse played an important role in the precipitation and/or exacerbation of their GI symptoms.

Quality of Life

Research suggests that patients with IBS have significantly poorer health-related quality of life than the general population or other patient groups (such as diabetes and end-stage renal disease). When compared with patients having other GI conditions, physical, emotional, and social role functions and energy were poorer among patients with IBS. However, this general finding is qualified such that the degree of impairment among people with IBS relates to the target population studied. Specifically, nonpatients with IBS have health-related quality of life scores that are intermediate between referred IBS patients and nonpatient control groups. When using questionnaires specific to IBS, patients fare the worst in terms of food avoidance, activity interference, and health worry concern. It was also found that quality of life improves in relation to changes in pain severity and daily function after psychological or antidepressant treatment.

PSYCHOSOCIAL INTERVENTIONS

In general, controlled studies have demonstrated the effectiveness of treating IBS with cognitive-behavioral therapy, relaxation training, hypnosis, and dynamic/interpersonal therapy. Since most of the research to date has focused on cognitive-behavioral therapies, this section is discussed in more detail relative to the other therapies.

Cognitive-Behavioral Therapy

Cognitive-behavioral techniques consist of a wide range of strategies and procedures designed to bring

about alterations in patients' perceptions of their situation and their ability to control their GI symptoms. The focus in cognitive-behavioral therapy (CBT) is on exploring how certain cognitions and behaviors may affect GI symptoms and associated psychosocial distress.

There have been 12 controlled studies including cognitive-behavioral or cognitive techniques in the treatment of IBS. Most of these techniques have been used within a multicomponent cognitive-behavioral treatment package. Treatment packages have included various combinations of cognitive therapy, stress management training, contingency management, relaxation techniques, psycho-educational components, assertiveness training, pain management, and bowel habit training. A significant amount of evidence supports the efficacy of CBT in relieving IBS symptoms and psychological distress (namely, depression and anxiety) relative to control conditions, such as waiting list control group, antispasmodics and bulking agents control group, symptom-monitoring control group, attention placebo control group, and psycho-educational control groups. A brief summary of findings from CBT packages will be presented by control group. Compared to a waiting list control group, one study found that CBT improved abdominal symptoms, coping strategies, and avoidance behavior. Several studies have found that CBT resulted in superior or similar improvement in GI symptoms when compared to the use of antispasmodics and/or bulking agents. Studies that compared CBT to symptom-monitoring controls found that CBT treatment packages improved GI symptoms and psychological distress relative to controls. One study indicated that CBT group therapy improved depressive symptoms and bowel symptom diary scores. Similar improvements were not found in the psycho-educational group. Finally, studies using cognitive therapy found significant improvement in IBS symptoms, depression, and anxiety relative to symptom monitoring and significant improvement in IBS symptoms and depression relative to an attentional-placebo control.

Relaxation Training

Relaxation or arousal reduction techniques encompass a variety of different methods used to teach patients how to counteract the physiological sequelae of stress or anxiety. The rationale for these techniques is premised on the belief that if muscle tension or autonomic arousal decreases, subjective anxiety or tension will also decrease as a consequence. The most common arousal reduction techniques include progressive muscle relaxation training, biofeedback for striated muscle tension, skin temperature, or electrodermal activity, autogenic training, and transcendental or Yoga meditation.

These techniques are typically combined with other treatments, making it difficult to determine the precise contribution of relaxation training. Moreover, support for the efficacy of any one relaxation training technique for the treatment of IBS is inconclusive. One study that did evaluate progressive muscle relaxation alone found greater reductions in IBS symptoms when compared with the symptom-monitoring control group. However, other research indicates that specific biofeedback to modify colon contractions was not effective for IBS even though generalized biofeedback with other relaxation techniques has been used successfully.

Hypnosis

The hypnotic "state" is one of heightened suggestibility. Following induction, the hypnotherapist uses progressive muscular relaxation plus suggestions of relaxation to reduce striated muscle tension. "Gut-directed" imagery and suggestions are used to relax GI smooth muscle. Hypnotherapy sessions end with the patient being told that he or she will feel positive and good about himself or herself. Patients are also asked to practice autohypnosis at home with an audiotape, with the ultimate goal of being able to administer suggestions of relaxation to themselves.

Studies on hypnotherapy have found that hypnosis results in a reduction of abdominal pain and altered bowel habits that can be maintained for at least 18 months. Improvements in quality of life, psychological symptoms, and rectal pain sensitivity have also been found.

Dynamic/Interpersonal Therapy

Dynamic or interpersonal psychotherapy is derived from psychodynamic principles and integrates humanistic and interpersonal concepts. This approach differs, however, from that of traditional psychoanalysis by moving away from the asymmetrical relationship between the therapist and the patient. Dynamic/interpersonal psychotherapy is most suitable for patients with problems stemming from difficulties in interpersonal relationships.

There are few studies on this form of therapy. The available evidence on dynamic/interpersonal therapy is supportive. Compared to the control conditions, dynamic/interpersonal psychotherapy led to greater reductions in bowel symptoms and psychological symptoms. Improvements were sustained at long-term follow-up.

As a final note on psychosocial interventions, regardless of the approach, a good working partnership between the person with IBS and his or her health care

professional is integral to success. Health care professionals must recognize that individuals with IBS can provide expertise on the factors that aggravate their bowel symptoms. Moreover, patients with IBS can rightly expect health care professionals to collaborate with them in understanding and managing their condition. As with many other illnesses, IBS can be influenced by psychosocial factors interacting with biological processes. Educating health care professionals and general society about the multidimensional nature of IBS can help toward conceptualizing multiple treatment approaches and developing a collaborative partnership between the patient and the health care professional.

FUTURE DIRECTIONS

This article has reviewed evidence pointing to the importance of psychological and sociological factors in IBS; however, further research is still required using a biopsychsocial perspective. Some possible avenues that this research might take are as follows: More research is needed to fully understand the influence of gender and sociocultural factors and the influence of clinical setting (e.g., nonpatients, primary care, GI referral, psychiatric referral) on IBS. Studies that will standardize current measures and develop new instruments for functional GI disorders are needed in order to examine interactions between psychosocial variables and bowel symptoms. Moreover, outcome measures that focus on clinically meaningful responses, such as satisfaction with treatment, health-related quality of life, global well−being, and coping with symptoms, should be further refined and incorporated into clinical trials.

Much research is still needed in the domain of treatment for IBS. A recent review of the literature on psychosocial treatment for IBS pointed to some key issues that should be addressed in the future. First, more attention should be focused on the distress associated with living with a chronic, debilitating illness. Society has continued to stigmatize patients with these disorders, trivializing their symptoms and treating them with a lack of empathy. To date, little theoretical or empirical work has been directed toward identifying and integrating the concerns of IBS patients into treatment plans. As a result, few psychosocial approaches have been tailored to the specific needs of people suffering from IBS. Second, studies must be designed to overcome the methodological limitations of previous investigations in this area. Psychosocial intervention studies should improve upon previous methodology by: including specific selection criteria for IBS; stratifying patients by symptom severity; enrolling sufficient numbers of female and male patients; including sufficient documentation of treatment plans to allow standardization and replication of treatment protocols; including session-by-session treatment manuals and measures of therapist adherence to treatment protocols; using appropriate placebo conditions to address expectancy and attention; and measuring credibility to treatment condition.

Further studies are required to determine the characteristics of patients, which predict response to specific psychosocial treatments and the specific components of psychosocial treatment packages (e.g., relaxation and cognitive restructuring), which account for their effectiveness. In addition, well-designed, randomized, controlled trials that offer a more holistic biopsychosocial approach are needed. In particular, there is growing evidence in this field that certain combination treatments that address both biological and psychosocial aspects may have synergistic effects (CBT plus antidepressant medication). Treatment studies can help clinicians to understand the effect of physician communication skills on patient satisfaction with care, adherence to treatment, and outcome.

See Also the Following Articles

Irritable Bowel Syndrome • Psychiatric Issues, Overview • Stress

Further Reading

Ali, A., Toner, B. B., Stuckless, N., Gallop, R., Diamant, N. E., Gould, M. I., and Vidins, E. I. (2000). Emotional abuse, self-blame, and self-silencing in women with irritable bowel syndrome. *Psychosomatic Med.* **62**, 76−82.

Drossman, D. A., Corazziari, E., Talley, N. J., Thompson, W. G., and Whitehead, W. E. (eds.) (2000). "Rome II: The Functional Gastrointestinal Disorders," 2nd Ed. Degnon Associates, McLean, VA.

Drossman, D. A., Creed, F. H., Olden, K. W., Svedlund, J., Toner, B. B., and Whitehead, W. E. (1999). Psychosocial aspects of the functional gastrointestinal disorders. *Gut* **45**, II25−II30.

Drossman, D. A., Mayer, E. A., Camilleri, M., and Whitehead, W. E. AGA technical review on irritable bowel syndrome. *Gastroenterology*, in press.

Toner, B. B., and Akman, D. E. (2000). Gender issues in irritable bowel syndrome. *Am. J. Gastroenterol.* **95**, 11−16.

Toner, B. B., Segal, Z. V., Emmott, S. D., and Myran, D. (2000). "Cognitive-Behavioral Treatment of Irritable Bowel Syndrome: The Brain−Gut Connection." Guilford, New York.

Toner, B. B., Segal, Z. V., Emmott, S. D., Mryan, D., Ali, A., DiGasbarro, I., and Stuckless, N. (1998). Cognitive-behavioral group therapy for patients with irritable bowel syndrome. *Int. J. Group Psychother.* **38**, 215−243.

Pylephlebitis

ALLISON LIDDELL
Presbyterian Hospital of Dallas

pylephlebitis Suppurative endophlebitis of the portal venous system.
pylethrombosis Thrombosis of the portal venous system.

Acute suppurative pylephlebitis is septic thrombophlebitis of the portal venous system, a rare entity first described in the late nineteenth century in association with appendicitis. With early surgical intervention and antibiotics in patients with appendicitis, the overall incidence has decreased, and other pyogenic processes draining into the portal circulation or arising in close proximity have become more prominent as predisposing conditions. Original reports noted 100% mortality, which has decreased with modern therapy. Despite these improvements, suppurative pylephlebitis remains a serious life-threatening development. Most case series in the literature are small subgroups of larger series of either pyogenic liver abscess or pylethrombosis, making conclusions regarding proper therapy difficult.

ETIOLOGY

Most cases of pylephlebitis are due to a primary infectious process in the abdomen or pelvis. Diverticulitis is the most common cause, accounting for 32% of 19 cases in a recent review of the literature since 1979 by Plemmons. Malignancy, particularly of the biliary tree, has accounted for a larger number of cases in more recent reviews, and some researchers suggest that more aggressive treatment of advanced biliary cancers may account for the increase. Appendicitis still accounts for a significant portion of pediatric cases, and inflammatory bowel disease, pancreatitis, cholangitis, endometritis, hemorrhoidal disease, peptic ulcer disease, and Behçet's disease have all been reported causes. Suppurative pylephlebitis may also develop secondarily as a complication of bacteremia with coexistent portal vein thrombosis.

MICROBIOLOGY

Gut flora, particularly *Escherichia coli, Proteus mirabilis,* and *Bacteroides fragilis*, are the predominant pathogens isolated from blood cultures and aspirations. A significant portion of infections are polymicrobial. In Plemmons' series, *B. fragilis* was the single most common species isolated from blood. Less common reported isolates include *Fusobacterium nucleatum, Gardnerella*, and *Candida albicans*.

CLINICAL MANIFESTATIONS

Symptoms are nonspecific, including fever, abdominal pain, chills and rigors, jaundice, and anorexia. Symptoms relating to the primary infectious process may be absent. Diagnosis relies on a high index of suspicion given the protean nature of the history and physical examination findings. Hepatomegaly may be present, and jaundice occurs less often and later than in cholangitis, and becomes more likely if liver abscesses develop. Leukocytosis is a prominent, albeit nonspecific, feature. Blood cultures are positive in 88% of cases. Radiology is the mainstay of diagnosis. Plain films are insensitive but may reveal air in the portal tree. Ultrasound with Doppler may indicate obstruction to portal venous flow. Computer tomography (CT) with intravenous contrast injection offers the advantage of evaluating the pylephlebitis and diagnosing the primary infectious focus, as well as complications such as liver abscess. Characteristic findings include lack of the expected contrast-opacified portal veins, intravascular air, and clots. Cavernous transformation, well described in subacute nonsuppurative pylethrombosis, may not be present in acute cases. Angiography has been used but has significant complications. Magnetic resonance angiography may in time provide a low-risk tool.

COMPLICATIONS

Pyogenic liver abscesses occur in about 50% of cases of pylephlebitis, and about 15% of pyogenic liver abscess cases demonstrate pylephlebitis. Bacteremia is present in 88% of cases, and sepsis, occurring in 21% of patients, remains the most common and serious threat. Mesenteric ischemia is rare but catastrophic. Reported mortality in the older literature in the antibiotic era was 50%; for more modern cases, 32% mortality is reported for all

 Copyright 2004, Elsevier (USA). All rights reserved.

etiologies, but in one review it is 80% for nonappendiceal cases. Several researchers have noted higher mortality for cases attributed to diverticulitis.

TREATMENT

No controlled studies exist, and limited case series differ widely in treatment modalities and recommendations. All agree that the mainstay of therapy is immediate antibiotics with activity against coliforms, enterococci, and anaerobes, which may be modified based on the results of search for the primary infectious source and on culture data. Typical recommendations are for 4–6 weeks of therapy at a minimum, with an initial period of intravenous therapy followed by a number of weeks of oral therapy. Followup CT may be helpful, but complete resolution of CT findings will lag behind microbiologic cure. Some researchers recommend followup imaging to document resolution of the thrombus, inasmuch as late presentations of complications of portal vein thrombosis have been reported.

Anticoagulation remains controversial due to reports of recanalization of the portal vein without intervention, and no clear evidence of benefit. One case series of 44 heterogeneous patients suggested improved outcomes in the few patients with underlying hypercoagulable states who received anticoagulation, and no adverse effects secondary to the anticoagulant therapy. However, in the absence of a documented hypercoagulable state, many researchers do not recommend anticoagulation. An important point is the paucity of reported deaths due to progression of mesenteric ischemia or embolization of clot. Most deaths occur due to infection-related complications. Several reports exist of CT-guided percutaneous drainage in selected patients, and surgical drainage may be required. Thrombolytic therapy has been considered contraindicated by some researchers due to the possible need for emergent surgical intervention, but a few reports of success exist.

See Also the Following Articles

Appendicitis • Computed Tomography (CT) • Liver Abscess • Portal Vein Thrombosis

Further Reading

Balthazar, E. (2000). Septic thrombophlebitis of the mesenteric and portal veins: CT imaging. *J. Comput. Assist. Tomogr.* **24**(5), 755–760.

Pelsang, R. (2001). Management of suppurative pylephlebitis by percutaneous drainage: Placing a drainage catheter into the portal vein. *Am. J. Gastroenterol.* **96**(11), 3192–3194.

Plemmons, R., Dooley, D., and Longfield, R. (1995). Septic thrombophlebitis of the portal vein (pylephlebitis): Diagnosis and management in the antibiotic era. *Clin. Infect. Dis.* **21**(5), 1114–1120.

Saxena, R., Adolph, M., Ziegler, J., Murphy, W., and Rutecki, G. (1996). Pylephlebitis: A case report and review of outcome in the antibiotic era. *Am. J. Gastroenterol.* **91**(6), 1251–1253.

Setto, R., and Rockey, D. (1996). Pyogenic liver abscess. Changes in etiology, management, and outcome. *Medicine* **75**(2), 99–113.

Pyloric Stenosis

JULIE R. FUCHS AND MORITZ M. ZIEGLER
Children's Hospital Boston and Harvard Medical School

gastric outlet obstruction Near-complete or complete blockage of the pyloric channel connecting the stomach to the duodenum, manifesting as early satiety and vomiting of undigested food in adults and projectile, nonbilious vomiting in children. Serum electrolytes may reveal a hypochloremic, hypokalemic alkalosis due to loss of H^+ and Cl^- in the vomitus, with compensatory renal secretion of K^+ in exchange for H^+. In adults, peptic ulcer disease is the most common etiology; in children, pyloric stenosis is most common, with other less frequent causes being pyloric atresia, gastric duplication, and ectopic pancreatic tissue.

gastroesophageal reflux disease Clinical syndrome that includes a variety of symptoms and tissue injury associated with abnormal esophageal exposure to regurgitated gastric contents. Long-term complications of consist of erosion ulcers, stricture, Barrett's metaplasia, and esophageal adenocarcinoma. In children, gastroesophageal reflux disease is a significant entity in the differential diagnosis of pyloric stenosis.

hyperplasia Growth characterized by an increase in the number of cells.

hypertrophy Growth characterized by an augmentation of cell volume.

pyloric stenosis Acquired condition involving the thickening of the circumferential muscle of the pyloric sphincter, which results in elongation and obliteration of the pyloric channel. The condition is the most common cause of gastric outlet obstruction in children and is one of the most frequent conditions requiring operation in the first month of life.

pyloric traumamyoplasty Alternative operative approach to pyloric stenosis involving the use of a Babcock clamp to grasp and pinch the hypertrophied pyloric muscle, creating two lateral slits on the superior and inferior edges. Results with this technique have been shown to be similar to results with the traditional Ramstedt procedure.

Ramstedt extramucosal pyloromyotomy Standard operative approach to pyloric stenosis; involves grasping the pylorus, incising the serosa longitudinally, and spreading or dividing the thickened pyloric muscle until the mucosa is bulging between the separated halves of the pylorus.

Pyloric stenosis, or hypertrophic pyloric stenosis, is an acquired condition involving the thickening of the circumferential muscle of the pyloric sphincter, which results in elongation and obliteration of the pyloric channel. A near-complete gastric outlet obstruction is produced with secondary dilation, hypertrophy, and hyperperistalsis of the stomach. The observed thickening of the smooth muscle is a result of hypertrophy, not hyperplasia. Pyloric stenosis is the most common cause of gastric outlet obstruction in children and is one of the most frequent conditions requiring operation in the newborn.

INTRODUCTION

Sabricius Hildanus first described pyloric stenosis in 1627. Subsequently, Blair described an infant with clinical as well as postmortem findings consistent with hypertrophic pyloric stenosis. Although sporadic reports of children with gastric outlet obstruction in Europe and the United States followed the initial description, the disease was not accepted as a true entity until the description in 1888 of two cases by Hirschsprung, who described it as a congenital disease representing the failure of the involution of the fetal pylorus and named it *angeborener pylorusstenose* (congenital pyloric stenosis). At this time, the preferred treatment was medical, using a combination of gastric lavage, antispasmodic drugs, dietary modifications, and the local application of heat, secondary to a 100% surgical mortality rate. Lobker performed the first successful surgical procedure to treat an infant with pyloric stenosis, using a gastrojejunostomy to bypass the obstructed pylorus. Unfortunately, the overall mortality for this procedure at that time remained high, approximately 50–60%. Nicoll and Fredet, in 1906 and 1907, each independently described a technique of longitudinal submuscosal division of the thickened pyloric muscle with transverse suturing of the defect (pyloroplasty). This type of extramucosal pyloroplasty was unsatisfactory, due to excessive hemorrhage, which occurred when sutures tore through the edematous muscle that had been closed. In 1912, Ramstedt simplified the Fredet procedure by omitting the transverse suturing, which leaves the mucosa exposed in the longitudinal seromuscular defect. Not only was the procedure successful, but its essential elements have remained generally unmodified and remain the surgical standard.

 Copyright 2004, Elsevier (USA). All rights reserved.

Pyloric stenosis is the most common cause of gastric outlet obstruction in children. The prevalence of pyloric stenosis ranges from 1.5 to 4 in 1000 live births among Caucasians but is less prevalent in Asians, Hispanics (1.8 per 1000 live births), and African-Americans (0.7 per 1000 live births). Multiple reports have suggested that the incidence may be increasing. For instance, a United Kingdom population-based study has documented a rise in incidence from 0.1–0.2% up to 0.3–0.8% during the past several decades. Additionally, at the Mayo Clinic, a large population-based study has documented an overall incidence of 0.26% in Olmsted County, Minnesota, from 1950 to 1984, but showed that the rate approached 0.5% by the end of the study period. It is well known that pyloric stenosis is more common in males than females, with a ratio of 2 : 1 to 5 : 1, although the long-held belief that it primarily affects first-born males has not been confirmed.

The development of pyloric stenosis has been associated with several variables, including both environmental and familial factors. A genetic contribution is supported by the fact that 19% of boys and 7% of girls whose mothers had pyloric stenosis also have the disease. Pyloric stenosis occurs in only 5% of boys and 2.5% of girls whose fathers have the disease, suggesting some type of variable maternal transmission. Additionally, the risk of pyloric stenosis is lower with older maternal age, higher maternal education, and low birth weight.

Generally, a majority of patients with pyloric stenosis are felt to have an acquired defect, with or without a preexisting genetic predisposition. One supportive study examined a series of 1000 males with barium swallow immediately after birth, finding no abnormalities of the pylorus. Subsequently, 5 of those infants went on to develop pyloric stenosis. In a second study, 1400 randomly selected newborn infants underwent ultrasonographic measurements of the pylorus, revealing normal pyloric dimensions; 9 infants (0.65%) later developed pyloric stenosis. On the other hand, as many as 7% of infants with pyloric stenosis have associated malformations, including intestinal malrotation, obstructive uropathy, and esophageal atresia. Other anomalies associated with pyloric stenosis include hiatal hernia and a deficiency in hepatic glucuronyl transferase activity.

ETIOLOGY/PATHOPHYSIOLOGY

The cause of pyloric stenosis remains poorly understood, but several hypotheses have emerged. Family history, sex, and maternal feeding patterns, among others, have all been deemed potential risk factors. In addition to the variability among races and the clear male predominance, there appears to be an increased risk with a positive family history and certain ABO blood types. Environmental factors associated with pyloric stenosis include the feeding method (breast vs. formula feeding), seasonal variability, and transpyloric feeding in premature infants. An association between systemic erythromycin in infants and subsequent pyloric stenosis has also been investigated. Although numerous theories have been proposed, none has received generalized acceptance.

Some investigators have suggested that milk curds in the stomach could obstruct the pyloric channel or produce edema of the pyloric mucosa and submucosa, obstructing the pyloric channel, leading to compensatory hypertrophy of the pyloric muscle. Others have focused on pyloric muscle innervation and relaxation. Different investigators have found that the numbers of ganglion cells in the pylorus and/or their maturity have been abnormal, although these results have not consistently been reproduced. Specifically, a markedly decreased number of glial-derived growth factor-positive nerve fibers has been found in patients with pyloric stenosis, as well as a reduced production of neurotropins. Confocal microscopy studies have shown the presence of abnormally thick contorted nerve bundles in the pyloric muscle of infants with pyloric stenosis. Others have investigated a lack of nitric oxide synthase in pyloric tissue as being a potential contributor to pylorospasm, possibly leading to pyloric stenosis, because nitric oxide is a known muscle relaxant.

Further hypotheses involve levels of gastrointestinal endocrine or paracrine factors. Various investigators have shown that gastrin is elevated in patients with pyloric stenosis and may be a stimulus toward muscle hypertrophy. On the other hand, hypergastrinemia and hyperacidity are known to occur secondary to gastric outlet obstruction from any cause. Furthermore, these patients have been shown to have elevated prostaglandins as well. Infants receiving infusions of certain prostaglandins have been shown to develop gastric outlet obstruction, but have not gone on to develop pyloric stenosis. Substance P has been shown to produce chronic pylorospasm, leading to muscle hypertrophy, and has also been found in higher concentrations in the pyloric muscle of patients with pyloric stenosis. Other peptides, such as secretin, enteroglucagon, neurotensin, and vasoactive intestinal peptide (VIP), have also been linked to pyloric stenosis, although their exact roles have yet to be elucidated. Furthermore, elevated levels and expression of transforming growth factor-α (TGF-α) and epidermal growth factor (EGF) mRNA have been

found in infants with pyloric stenosis, although the significance is unclear. Currently, the cause of pyloric stenosis has not been definitively elucidated and appears to be multifactorial, although considerable basic science investigations are ongoing.

CLINICAL PRESENTATION

Typically, an infant with pyloric stenosis presents with a normal feeding history and new onset of nonbilious, emesis at 2 to 8 weeks of age, with a peak incidence of 3 to 5 weeks. Initially the emesis may not be frequent or forceful, but becomes progressively worse over days until nearly every feeding is forcefully vomited in a "projectile" fashion. Less frequently, the emesis may be blood tinged or have a coffee-ground appearance due to gastritis or esophagitis. Generally, infants appear hungry immediately after vomiting and do not appear ill early in the disease course. Typically, stool frequency is concomitantly diminished. If there is a significant delay in diagnosis, severe dehydration and lethargy can occur. Some children have diarrhea, complicating the diagnosis. Others (2–5%) may have jaundice secondary to glucuronyl transferase deficiency. In premature infants, the presentation is commonly delayed and occurs about 2 weeks later compared to full-term infants; often there is a slower progression of emesis.

The differential diagnosis for pyloric stenosis includes gastroesophageal reflux and formula intolerance, although these usually present with a more gradual onset of emesis. However, a more frequent cause of nonbilious vomiting in the first several weeks of life is either overzealous volume or frequency of feedings (especially formula) offered to the infant. Other causes of nonbilious vomiting include medical disorders such as sepsis, hydrocephalus, and metabolic diseases, as well as entities often requiring surgical intervention, such as antral webs, pyloric atresia, gastric duplication, microgastria, and ectopic pancreatic tissue.

DIAGNOSIS

The cardinal features of pyloric stenosis include nonbilious projectile vomiting, visible peristaltic waves in the left upper abdomen, a palpable "olive" or enlarged pylorus, and an associated hypochloremic or hypokalemic metabolic alkalosis. These features are more prominent when the diagnosis is delayed, and more subtle earlier in the course. Generally, a definitive diagnosis can be made on the basis of clinical presentation and careful physical examination in as many as 80% of patients, although an increasing reliance on imaging modalities has been seen in the current era of managed care. A recent study has shown that the number of patients diagnosed solely by clinical examination has been decreasing (from 74%, decreasing to 28%), and that the use of diagnostic tests has increased (ultrasonography increasing from 16 to 65% and upper gastrointestinal imaging increasing from 12 to 28%). Several maneuvers can increase the sensitivity of physical examination in pyloric stenosis. The infant must be calm and cooperative and use of a pacifier and warm blanket may be helpful. Additionally, decompression of the stomach with a nasogastric tube may aid in the palpation of the "olive." The examiner standing on the infant's left side should flex the baby's hips with the left hand and palpate the liver edge from above with the right hand. Gentle pressure is applied deep to the liver, and eventually the palpating fingers are moved distally to find a palpable pylorus that can be rolled under the fingertips, making the diagnosis. Examination by a surgeon before imaging studies can be cost effective, with a specificity of 90%, but a sensitivity of only 50%.

In the absence of a palpable mass, an upper gastrointestinal contrast study or ultrasound can usually confirm or dispute the diagnosis. In most pediatric centers, abdominal ultrasound is the study of choice. This technique has a sensitivity of 97%, a specificity of 100%, and a positive and negative predictive value of 100 and 98%, respectively, and has been studied extensively by numerous groups. Its main limitations are operator dependence and the inability to diagnose other causes of nonbilious vomiting. Measurements found to have greater than 90% positive predictive value include a muscle diameter of 17 mm or more, a muscular wall thickness of 4 mm or greater, and a channel length of 17 mm or greater. One group concluded that a muscle thickness of 3 mm should be considered a positive finding for pyloric stenosis in children less than 30 days of age. Last, when measurements are equivocal, calculation of the pyloric volume has been reported to be more accurate.

In the past, an upper gastrointestinal tract series was the gold standard diagnostic study for pyloric stenosis. The test is sensitive and is also helpful in indicating other causes of nonbilious emesis, such as gastroesophageal reflux disease, gastric atony, and delayed gastric emptying. It will also diagnose the dreaded malrotation. The classic radiographic contrast findings are the "string sign," produced by contrast medium outlining the narrowed pyloric channel, and the "shoulder sign," caused by the hypertrophied muscle protruding into the gastric lumen. The pyloric channel may also appear as two parallel threads resembling railroad tracks. A potential disadvantage is that the study exposes the patient to ionizing radiation and involves filling the obstructed

stomach with barium before induction of general anesthesia, increasing the risk of vomiting and aspiration. Therefore, it is recommended that as much of the contrast agent as possible be removed prior to induction of anesthesia.

One suggested means to select the optimal imaging test for pyloric stenosis is to measure the volume of the gastric aspirate obtained by a nasogastric tube. If less than 10 ml is obtained, an upper gastrointestinal tract series may be warranted, because 86% of such infants have been shown to have gastroesophageal reflux disease (GERD). If more than 10 ml is suctioned, an ultrasound examination of the pylorus should be obtained, because 92% of such infants will have pyloric stenosis.

PREOPERATIVE MANAGEMENT

Once the diagnosis has been made, preoperative preparation is essential, because pyloric stenosis is not a surgical emergency. The child with an early presentation without dehydration, with normal serum electrolyte and glucose concentrations, and normal urine output may be operated on at the earliest convenience. Infants who present with clinical dehydration, including dry mucous membranes, depressed fontanels, increased skin tenting, and varying degrees of malnutrition, require more extensive resuscitation. They characteristically have a hypochloremic, hypokalemic metabolic alkalosis with some degree of hyponatremia and hypoglycemia. One group has defined three levels of severity based primarily on the carbon dioxide content (slight, <25 mEq/liter; moderate, 26–35 mEq/liter; and severe, >35 mEq/liter). Most infants require resuscitation for less than 24–48 hours prior to operation. Initial resuscitation often begins with normal saline or lactated Ringer's solution in boluses of 10–20 ml/kg. A continuous infusion of 5 or 10% dextrose in 0.45% saline is then started at 1.5 times maintenance. Potassium is added after urine output is established. Once the correction of dehydration and restoration of near-normal serum potassium and chloride levels are achieved, as well as correction of the alkalosis to a serum bicarbonate level of below 30 mEq/liter, the operation can be safely conducted. Failure to correct the alkalosis adequately preoperatively can result in postanesthetic apnea and respiratory arrest.

Most infants with pyloric stenosis do not have a complete gastric outlet obstruction and can handle their gastric secretions. Oral feedings are discontinued, but a nasogastric tube is not routinely placed other than temporarily for diagnostic purposes, because it removes additional fluid and acid from the stomach, exacerbating the metabolic alkalosis.

OPERATIVE MANAGEMENT

The Ramstedt extramucosal pyloromyotomy has long been the classic surgical approach to pyloric stenosis. The standard approach is a right upper quadrant transverse incision of 2.5–3 cm over the right rectus muscle at or above the liver edge. The rectus muscle is either divided or split longitudinally, and the peritoneal cavity is entered through the posterior rectus sheath. The pylorus may be identified by bringing the greater curvature of the stomach through the incision and using it to externalize the pylorus. The serosa on the anterior wall of the pylorus is incised from just proximal to the pyloric vein to the antrum just proximal to the area of hypertrophied muscle. The duodenal end is usually identified by the color change from the pale pylorus to the pink duodenal wall, as well as by the prominent pyloric vein. The myotomy is performed using the back of a scalpel handle to split the hypertrophied muscle bluntly down to the submucosa, or by using a spreading clamp, such as that described by Benson. Once the submucosa is exposed, the overlying muscle fibers should be spread more widely, allowing the submucosa to herniate or bulge out. When the two halves of the pylorus can be moved independently back and forth in opposite directions by a rocking motion, the extent of the pyloromyotomy is complete (Fig. 1). Venous congestion caused by delivering the pylorus through a relatively small incision can result in bleeding from the muscle and submucosa, but generally resolves when the pylorus is returned to the abdominal cavity. Inspection to assure hemostasis should occur before closing the incision.

If the submucosa or mucosa is violated, management is individualized by one of two fundamental techniques. If perforation occurs early in the myotomy, the mucosa is often closed with fine absorbable sutures and the muscle is closed. A second myotomy can then be performed by rotating the pylorus 180°. If the injury occurs when the myotomy is finished, the perforation can be closed and the injury site covered with omentum. Postoperatively, some surgeons decompress the stomach with a nasogastric tube whereas others withhold feedings for a few days. At the end of an uncomplicated pyloromyotomy, some surgeons opt to check for leaks by filling the stomach with 60–100 cm^3 of air.

Some surgeons prefer a supraumbilical, curvilinear incision, due to its superior cosmetic results, although there have been reports of a higher incidence of wound

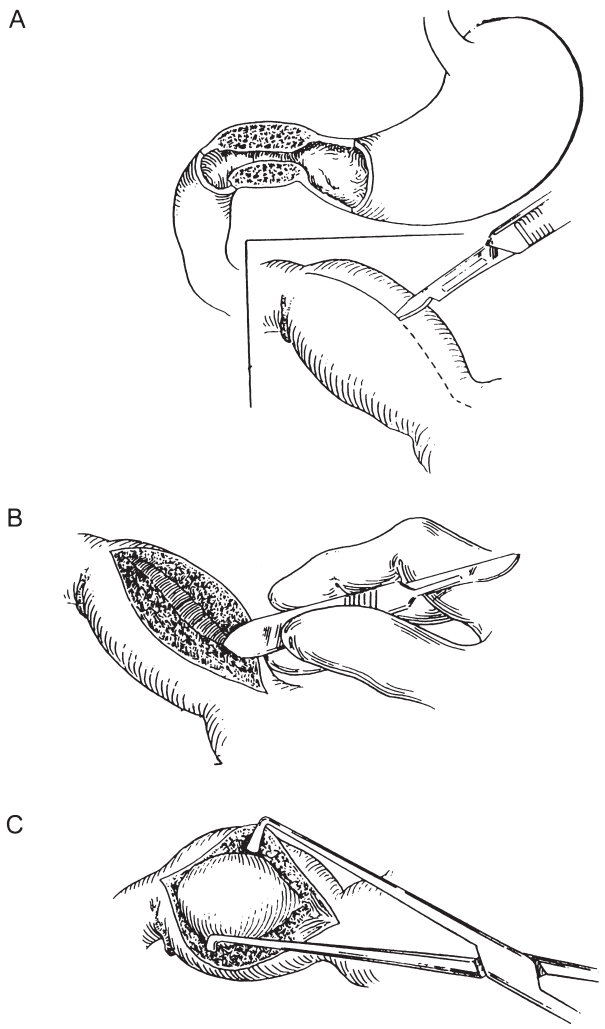


FIGURE 1 The Ramstedt extramucosal pyloromyotomy, the standard operative technique. (A) Cross-section of the pylorus, with inset showing the incision of the serosa on the anterior wall from just proximal to the pyloric vein to the antrum, just proximal to the area of hypertrophied muscle. (B) The myotomy is performed using the back of a scalpel handle to split the hypertrophied muscle bluntly down to the submucosa. (C) After exposure of the submucosa, the overlying muscle fibers are spread relatively widely, allowing the submucosa to herniate or bulge out. Reproduced with permission of the publisher from Ziegler, M. M. *et al.* "Operative Pediatric Surgery," Ch. 52. McGraw-Hill.

complications. Additionally, this incision is contraindicated in the presence of an umbilical remnant, a "wet" umbilicus, or periumbilical erythema. Laparoscopic pyloromyotomy has been gaining popularity since its first description by Alain and colleagues in 1991. This technique may have the advantage of improved appearance, but it has the disadvantage of longer operative times and a variable learning curve. No significant difference between the time the infant can resume full feedings or the length of hospital stay has been shown. A recent retrospective study found that outcomes between the open and laparoscopic techniques were similar, but the laparoscopic approach incurred a greater expense and decreased general surgery resident operative experience. Last, an additional technique, pyloric traumamyoplasty (either open or laparoscopic), has been introduced by various groups; this involves the used of a Babcock clamp to grasp and pinch the hypertrophied muscle, creating two lateral slits on the superior and inferior edges. Results with this technique have been shown to be similar to results with the traditional Ramstedt procedure, but experience to date has been limited.

Nonoperative management has not gained general acceptance in North America but has been practiced in some European countries. Infants can be managed with frequent small feedings or even temporary total parenteral nutrition, but this requires a prolonged hospital stay. Successful endoscopic balloon dilatation for pyloric stenosis has been reported in Japan, being used selectively in patients who have undergone previous extensive abdominal operative procedures. Botulinum toxin A has been attempted without success, whereas others have successfully used intravenous and oral atropine sulfate.

POSTOPERATIVE MANAGEMENT

Most infants can be fed within 6 hours postoperatively. Many feeding regimens have been used, but most generally start with a small volume of sugar water, advancing volume and osmolarity every 2–3 hours until the child is taking formula or milk without significant vomiting. Data have shown that either a standardized feeding regimen or a more rapid, *ad libitum* feeding schedule will lead to earlier discharge, and the amount of vomiting with either technique is similar.

OUTCOMES AND COMPLICATIONS

Mortality after pyloromyotomy has been extremely low, with reported rates of less than 0.5%. With appropriate management of fluids and electrolytes preoperatively and intraoperative management by a pediatric surgeon, most infants can be discharged from the hospital within 1–2 days. Similarly, morbidity is low; the major complications include wound infection or dehiscence, mucosal perforation, and inadequate pyloromyotomy. The incidence of wound infections has been variably reported from 1 to 5%. The reported rates of duodenal perforation range from 1 to 30% but generally remain in the 1–3% range in pediatric surgery centers. Although

many (30–90%) infants will have some degree of postoperative emesis, this usually resolves spontaneously within the first week. If prolonged emesis occurs, the possibilities of an unrecognized perforation, gastroesophageal reflux, or an incomplete myotomy should be considered. Contrast studies are of little value other than to diagnose a leak, because the radiologic and ultrasonographic appearances of the hypertrophied pylorus before and after pyloromyotomy are similar, and such an appearance may persist for many weeks to months. Therefore, the decision to reoperate for presumed incomplete myotomy is typically delayed for at least 2–3 weeks postoperatively.

Infrequent long-term effects of pyloromyotomy have been reported. One group evaluated the presence of gastrointestinal symptoms, gastric emptying, and pyloric measurements in adults after and without pyloromyotomy and found no differences. A second group found higher pyloric tone and force of gastric contraction, but no difference in clinically relevant gastric emptying in the same two treatment groups.

SUMMARY

Pyloric stenosis is the most frequent cause of gastric outlet obstruction in children. Currently, the cause has not been definitively elucidated and appears to be multifactorial, although considerable basic science investigations are ongoing. The cardinal features of pyloric stenosis include nonbilious projectile vomiting, visible peristaltic waves in the left upper abdomen, a palpable olive (enlarged pylorus), and an associated hypochloremic, hypokalemic metabolic alkalosis. Diagnosis is made by history and physical examination alone or in combination with ultrasonography or gastrointestinal contrast studies. The Ramstedt extramucosal pyloromyotomy has long been the classic surgical approach to pyloric stenosis, although laparoscopic pyloromyotomy and traumamyoplasty are new additions to the operative armamentarium. Mortality and morbidity after pyloromyotomy have been extremely low and long-term sequelae of the procedure have not been reported.

See Also the Following Articles

Gastric Outlet Obstruction • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Pyloroplasty • Pylorus • Webs

Further Reading

Campbell, B. T., McLean, K., Barnhart, D. C., Drongowski, R. A., and Hirschl, R. B. (2002). A comparison of laparoscopic and open pyloromyotomy at a teaching hospital. *J. Pediatr. Surg.* **37**, 1068–1071.

Dillon, P. W., and Cilley, R. E. (2000). Lesions of the stomach. *In* "Pediatric Surgery" (K. W. Ashcraft, J. P. Murphy, R. J. Sharp, D. L. Sigalet, and C. L. Snyder, eds.), 3rd Ed., pp. 391–405. W. B. Saunders, Philadelphia.

Leinwand, M. J., Shaul, D. B., and Anderson, K. D. (2000). A standardized feeding regimen for hypertrophic pyloric stenosis decreases length of hospitalization and hospital costs. *J. Pediatr. Surg.* **35**, 1063–1065.

Letton, R. W. (2001). Pyloric stenosis. *Pediatr. Ann.* **30**, 745–750.

Lobe, T. E. (1995). Pyloromotomy. *In* "Rob & Smith's Operative Surgery, Pediatric Surgery" (L. Spitz and A. G. Coran, eds.), 5th Ed., pp. 320–327. Chapman & Hall, London.

Magnuson, D. K., and Schwartz, M. Z. (1997). Stomach and duodenum. *In* "Surgery of Infants and Children" (K. T. Oldham, P. M. Colombani, and R. P. Foglia, eds.), pp. 1133–1162. Lippincott-Raven, Philadelphia.

Poon, T. C., Zhang, A., Cartmill, T., and Cass, D. T. (1996). Changing patterns of diagnosis and treatment of infantile hypertrophic pyloric stenosis: A clinical audit of 303 patients. *J. Pediatr. Surg.* **31**, 1611–1615.

Schwartz, M. Z. (1998). Hypertrophic pyloric stenosis. *In* "Pediatric Surgery" (J. A. O'Neill, M. I. Rowe, J. L. Grosfeld, E. W. Fonkalsrud, and A. G. Coran, eds.), 5th Ed., Vol. 2, pp. 1111–1116. Mosby-Year Book, Inc., St. Louis.

Sun, W. M., Doran, S. M., Jones, K. L., Davidson, G., Dent, J., and Horowitz, M. (2000). Long-term effects of pyloromyotomy on pyloric motility and gastric emptying in humans. *Am. J. Gastroenterol.* **95**, 92–100.

Teele, R. L., and Smith, E. H. (1977). Ultrasound in the diagnosis of idiopathic hypertrophic pyloric stenosis. *N. Engl. J. Med.* **296**, 1149–1150.

Pyloroplasty

GEORGE A. SAROSI, JR.
University of Texas Southwestern Medical Center, Dallas

pyloroplasty Surgical procedure that incises and divides the normal pyloric muscle, destroying it as a sphincter, and then reconstructs the pyloric channel to facilitate gastric emptying.

vagotomy Surgical procedure in which the vagus nerves are cut, to reduce innervation of the stomach.

Pyloroplasty is typically performed in conjunction with a truncal or selective vagotomy. Vagal denervation of the stomach is used to decrease gastric acid production in various acid-peptic disorders, but also results in failure of the antral–pyloric pump mechanism and markedly delays gastric emptying. Originally described in the late nineteenth century for use in the treatment of gastric outlet obstruction, pyloroplasty did not become a common procedure until vagotomy became the main treatment of peptic ulcer disease in the 1940s.

PHYSIOLOGY

The physiologic role of the normal pylorus remains controversial. Although it is thought to provide resistance to gastric emptying, to allow for complete mixing of gastric contents by the antral–pyloric pump, pyloroplasty (or pylorectomy) in dogs with normal gastric innervation causes little change in gastric emptying. After pyloroplasty alone, the gastric antrum can fulfill many of the functions of a normal pylorus. However, after vagotomy of the gastric antrum, the pylorus provides sufficient resistance to gastric emptying to cause markedly delayed gastric emptying of solids in more than one-third of patients. After vagotomy and pyloroplasty, the loss of proximal gastric receptive relaxation results in more rapid emptying of liquids despite a loss of antral pump function, presumably because of loss of gastric capacitance and loss of pyloric resistance to gastric emptying secondary to pyloroplasty.

SURGICAL TECHNIQUE AND INDICATIONS FOR PROCEDURE

There are two different techniques for performing a pyloroplasty, the Heineke–Mikulicz/Weinberg variant and the Finney variant. A third procedure, Jaboulay pyloroplasty, is really a gastroduodenostomy without division of the pyloric ring and is therefore not technically a pyloroplasty. In the Heineke–Mikulicz/Weinberg variant, a longitudinal incision is made from the distal gastric antrum across the pylorus onto the first portion of the duodenum. This incision is then closed in a transverse fashion in either one or two layers of sutures, creating a wide gastric outlet (Fig. 1A). The advantage of this procedure is ease of performance. In the Finney variant, the duodenum is first mobilized and then approximated in a side-to-side fashion to the greater curvature of the stomach. A long incision is then made from the gastric antrum across the pylorus onto the duodenum. A side-to-side anastomosis is then constructed in two layers between the stomach and duodenum (Fig. 1B). The advantage of this procedure is that it produces a very large gastric outlet. However, significant scarring of the duodenum can make performance of a Finney pyloroplasty impossible. The main indication for a pyloroplasty is to facilitate gastric emptying after vagotomy. Historically, the most common use of pyloroplasty was in combination with truncal vagotomy in operations for peptic ulcer disease. With the marked decrease in elective surgery for peptic ulcer disease, the main modern use of pyloroplasty is in treating bleeding duodenal ulcers, to provide access to the ulcer and the gastroduodenal artery. It is now also commonly a component of reconstruction procedures following esophageal resection when the stomach is used as an esophageal conduit, in order to prevent gastric stasis in the intrathoracic stomach.

COMPLICATIONS

The major complications associated with pyloroplasty and vagotomy include dumping syndrome, which is seen in up to 20% of patients. Postvagotomy diarrhea is seen in approximately 10% of patients after the procedure. Early satiety and epigastric fullness are seen in many patients in the early postoperative period, but usually resolve spontaneously over time. Bile reflux and bilious vomiting are seen in 2–5% of patients after pyloroplasty. In one study with long-term followup

 Copyright 2004, Elsevier (USA). All rights reserved.

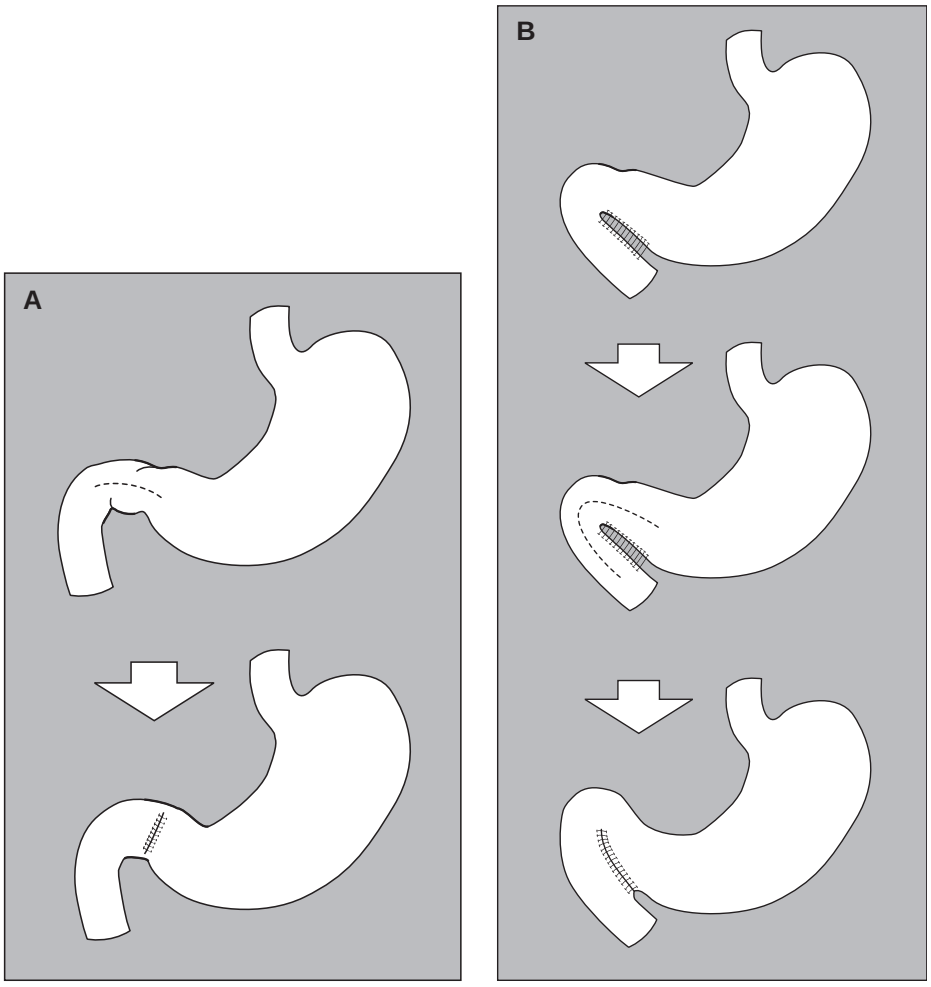


FIGURE 1 (A) In a Heineke–Mikulicz pyloroplasty, a horizontal incision is made across the pylorus and then is closed vertically, to widen the pyloric orifice. (B) In a Finney pyloroplasty, a back row of seromuscular sutures is first placed to approximate the duodenum and the greater curvature of the antrum. A U-shaped incision is then made from the antrum across the pylorus onto the duodenum. This is then closed in two layers, to construct a side-to-side gastroduodenal anastomosis. Illustrations by Jim Hardy, VA North Texas Health Center.

after vagotomy and pyloroplasty, 40% of patients experienced some gastrointestinal disturbance, and 6% of patients experienced severe disturbances.

See Also the Following Articles

Dumping Syndrome • Gastric Outlet Obstruction • Gastric Surgery • Pylorus

Further Reading

Dayton, M. (2002). Vagotomy and drainage (2002). *In* "Shackelford's Surgery of the Alimentary Tract" (D. Dempsey, ed.), 5th Ed., Vol. 2, pp. 117–130. W. B. Saunders, Philadelphia.

Richter, H. (1994). Physiologic consequences of vagotomy and gastric resection. *Gastroenterol. Clin. North Am.* **23**(2), 193–213.

Skellenger, M., and Jordan, P. (1983). Complications of vagotomy and pyloroplasty. *Surg. Clin. North Am.* **63**(6), 1167–1178.

Pylorus

ANJANA KUMAR AND KONRAD S. SCHULZE
University of Iowa

bile reflux Movement of bile from the small intestine (duodenum) into the stomach.
dumping syndrome Abnormally rapid emptying of stomach contents into the small intestine, associated with symptoms of dizziness, rapid heart rate, sweating, nausea, and diarrhea occurring mainly after a meal.
pyloric stenosis Narrowing of the pyloric (outlet) region of the stomach.
pyloric tone Variable level of ongoing contraction of the musculature in the pyloric region of the stomach.
pyloroplasty A surgical procedure that enlarges the opening separating the stomach from the small intestine.

Pylorus means gatekeeper in Greek. The term aptly reflects the fact that the pylorus controls the resistance to flow across the narrow gastric outlet. The diameter of the pylorus is set by the baseline tension of the pyloric muscle (pyloric tone). Forceful pyloric closure is part of the terminal antral contraction that leads to retropulsion of gastric contents. The pyloric sphincter is a complex muscular structure that is enforced by dense connective tissue.

ANATOMY

The pyloric segment extends from the proximal pyloric muscle loop (PPL) to the distal pyloric muscle loop (DPL) (see Fig. 1). The PPL is a broad band of circular muscle that fans out from the lesser curvature close to the duodenal bulb and reaches the greater curvature several centimeters upstream. The DPL is a thick bundle of circular muscle at the base of the duodenal bulb. Reinforced by collagen-rich connective tissue, this pyloric ring surrounds the narrow lumen of the pyloric orifice. PPL and DPL converge on the lesser curvature in a knot of connective tissue and fat, known as pyloric torus. On contraction, the torus wedges into the groove between the two muscle loops on the greater curvature. The pyloric canal refers to the lumen inside the contracted pyloric segment.

PHYSIOLOGY AND INNERVATION

The pylorus differs from other gastrointestinal sphincters in that it does not occlude the lumen at rest. As gastric contractions reach the PPL, the entire pyloric segment closes in rapid sequence, the terminal antral contraction. Antral folds prolapse into the pylorus and form a mucosal plug, which reduces the pyloric lumen to a star-like slit. Simultaneously, the pylorus shortens and moves orad and toward the left. As the contraction passes, the pyloric segment flares open from its proximal end.

Pyloric activity is controlled by enteric nerves in the myenteric plexus. Tonic inhibitory nervous input maintains the pylorus in a relaxed state at rest. Inhibitory nerves release nitric oxide and/or vasoactive intestinal peptide. Nitridergic inhibition of the pylorus occurs with vagal and antral stimulation. Excitatory cholinergic and enkephalinergic neurons stimulate pyloric muscle tone and contractions.

Enteric neurons synapse with sympathetic neurons from the celiac plexus and parasympathetic (vagal) neurons via the nerve of Latarjet. The pyloric musculature contains fewer interstitial cells of Cajal than the gastric and duodenal musculature.

FUNCTION

Pyloric activity adapts to the properties of the luminal contents. A widely patent pylorus (diameter of approximately 1 cm in humans) allows for rapid outflow of isotonic solutions. Acid increases pyloric tone through cholinergic stimulation and slows flow. Fat triggers intermittent pyloric contractions in the absence of antral or duodenal activity (isolated pyloric contractions). Contractions of proximal stomach and antrum drive boluses of liquids and small particles through the pylorus. Larger particles and fat are trapped by the terminal antral contraction before reaching the pyloric orifice and are dispersed by a powerful retrograde jet. The separation of liquid and solid phases at the gastric outlet is known as sieving and occurs in part through decanting. Phasic contractions of pylorus and antrum generate the shear forces that break down particles and mix particles, fat droplets, and secretions to the slurry known as gastric chyme.

 Copyright 2004, Elsevier (USA). All rights reserved.

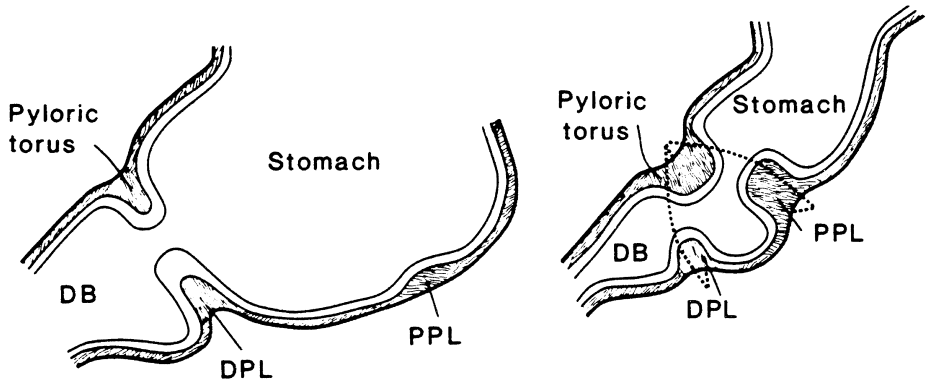


FIGURE 1 Scheme of the relaxed (left) and contracted (right) pylorus. DB, duodenal bulb; DPL, distal pyloric loop; PPL, proximal pyloric loop. (Right) The contraction of the pyloric muscle loops has led to closure of the pyloric canal. Reprinted from Keet, A.D., Jr. (1957). The prepyloric contractions in the normal stomach. *Acta Radiol.* **48**, 413–424, with kind permission of the author and editor of *Acta Radiologica*.

ABNORMAL FUNCTION

Pyloric function may be affected by operative intervention (gastric resection, pyloroplasty), by denervation (vagotomy, diabetic autonomic neuropathy), by muscular hypertrophy, or by mucosal inflammation (peptic ulcer disease). Pyloric incompetence leads to precipitous gastric emptying or dumping, particularly of fluids. Diarrhea and maldigestion may result. Pyloric stenosis leads to gastric retention, vomiting, and weight loss.

Infantile Hypertrophic Pyloric Stenosis

Infantile hypertrophic pyloric stenosis refers to gastric outlet obstruction from muscular hypertrophy. It affects 1 of 150 males and 1 of 750 females at birth. A deficiency of nitric oxide-containing neurons is invoked.

Peptic Disease

Peptic disease can cause pyloric deformation. The disruption of the torus and DPL leads to a short, wide, and incompetent pylorus. A "keyhole" deformity of the pylorus and foreshortening of the duodenal bulb may be seen. Scars involving the PPL produce an antral web and a long stenotic pylorus.

Denervation

Vagotomy may result in a functional pyloric stenosis (pylorospasm). An autovagotomy is considered to be responsible for disrupted gastroduodenal motor activity and pylorospasm in autonomic neuropathy, particularly in diabetes mellitus.

See Also the Following Articles

Dumping Syndrome • Electrogastrography • Gastric Motility • Pyloric Stenosis • Sphincters

Further Reading

Allescher, H.-D., and Daniel, E.-E. (1994). Role of NO in pyloric, antral and duodenal motility and its interaction with other inhibitory mediators. *Digest. Dis. Sci.* **39**, 73S–75S.

Biancani, P., Zabinski, M.-P., Kerstein, M.-D., and Behar, J. (1979). Mechanical characteristics of the cat pylorus. *Gastroenterology* **77**, 301–309.

Ehrlein, H.-J. (1988). Motility of the pyloric sphincter studied by the inductograph method in conscious dogs. *Am. J. Physiol.* **254**, G650–G657.

Lu, C.-C., and Schulze-Delrieu, K. (1990). Pyloric deformation from peptic disease—Radiographic evidence for incompetence rather than obstruction. *Digest. Dis. Sci.* **35**, 1459–1467.

Mearin, F., Camilleri, M., and Malagelada, J.-R. (1986). Pyloric dysfunction in diabetics with recurrent nausea and vomiting. *Gastroenterology* **90**, 1919–1925.

Torgensen, J. (1942). The muscular build and movements of the stomach and duodenal bulb. *Acta Radiol.* **45**, 1–191.

Radiology, Interventional

ROY L. GORDON
University of California, San Francisco

Interventional radiology (IR) is a subspecialty of radiology, which is based on the use of imaging techniques for diagnostic purposes and to guide minimally invasive therapeutic procedures. For example, in a patient with bleeding into the gastrointestinal tract, a small catheter is inserted percutaneously into the femoral artery in the groin. The catheter is then manipulated up the aorta and selectively into a mesenteric artery, using fluoroscopic guidance. An X-ray contrast agent is injected through the catheter and radiographs are taken. These identify and locate a bleeding site within the bowel, providing an excellent diagnostic test. Then, an embolic agent, such as a metallic coil, can be injected through the same catheter to block the bleeding artery and provide the definitive treatment to stop the bleeding. These image-guided minimally invasive techniques are not only used intravascularly, but are valuable in many other gastrointestinal applications, such as in the lumen of the alimentary tract, percutaneously into the liver, biliary tract, and other viscera, and for the drainage of abscesses or fluid collections. The instruments employed are typically needles, wires, catheters, balloons, snares, collapsible baskets, emboli, and other small-caliber devices. Guidance may be provided by X-ray image-intensified fluoroscopy, computed tomography, ultrasonography, or less commonly, magnetic resonance imaging. In most cases, the procedures are accomplished under local anesthesia with moderate sedation, eliminating the need for general anesthesia, which is often needed for conventional surgical interventions. Recovery time is typically much shorter than for surgery and many procedures can be performed on an outpatient basis. No surgical incisions are needed. The use of IR in the gastrointestinal tract as it pertains to the different parts of the gastrointestinal tract will be described.

ALIMENTARY CANAL

Feeding Tube Manipulation

Patients often are unable to eat or to feed themselves because of impaired consciousness or other temporary problems. A simple temporary solution is to pass a small tube into the nose, through the esophagus, and into the stomach.

This can be carried out by a nurse at the bedside. Some patients do not empty their stomachs well, so that the liquid nutrients injected down the tube, called a feeding tube or nasogastric tube, accumulate in the stomach and may cause reflux into the lungs and aspiration pneumonia. Using fluoroscopic guidance and guide-wire and catheter techniques, the interventional radiologist can maneuver these tubes deep into the third part of the duodenum. This stops reflux and delivers the nutrients to the jejunum, where they can be absorbed.

Percutaneous Gastrostomy

Nasogastric tubes are not suitable for long-term use because they are uncomfortable for patients and may become dislodged. Long-term tube feeding to the stomach or the jejunum can best be achieved by placing the tube percutaneously through the anterior abdominal wall directly into the lumen of the stomach. This can be done during open surgery, by laparoscopic techniques, by combined gastric endoscopy and surgery, and most simply by direct puncture under fluoroscopic guidance by the interventional radiologist. The interventional radiology technique is to inflate and distend the stomach with air injected down a nasogastric tube. The stomach pushes up against the anterior abdominal wall and displaces all other structures. The entry site is just below the left costal edge in line with the nipple. The procedure is performed using local anesthesia and moderate sedation. A needle is passed percutaneously into the distended air-filled stomach, which is easily seen fluoroscopically. T-fasteners are deployed through three separate needle punctures and are used to fix the stomach to the anterior abdominal wall. A fourth puncture is used to place a stiff guide wire into the stomach. Serially enlarging dilators are used over the wire to enlarge the tract. Finally, a 14-French self-retaining catheter is passed over the wire and provides access to feed the stomach.

 Copyright 2004, Elsevier (USA). All rights reserved.

Diagnosis of Bowel Ischemia

The arterial blood supply to the bowel comes from three main arteries arising from the aorta. These are the celiac trunk, the superior mesenteric artery, and the inferior mesenteric artery. An additional arterial collateral supply can be developed via branches of the internal iliac arteries. Placement of a catheter in the aorta allows the injection of contrast agent and the filming of an angiogram to show any blockage of the arteries or veins that may cause the bowel to be ischemic or have an impaired blood supply. With rapid progress in noncatheter angiography techniques such as computed tomography (CT) or magnetic resonance imaging (MRI), catheter angiography may be needed infrequently and will be largely supplanted by the noninvasive techniques. It is important to demonstrate the site of acute arterial occlusion at an early stage so that treatment can be instituted before lack of blood causes irreversible ischemic necrosis and perforation of the bowel.

Diagnosis and Treatment of Intestinal Bleeding

Catheter angiography is an important and accurate method of demonstrating the actual intestinal bleeding site. An angiogram is performed, opacifying the arterial tree supplying the bowel. Bleeding is demonstrated by the leakage of contrast from the lumen of the artery, causing a persistent stain or collection of extravascular contrast agent. Once the site of active bleeding is demonstrated, it is possible in many cases to pass a small catheter selectively down the bleeding artery to the point of the bleeding. The artery is then sealed off at this point, by the trans-catheter injection of an embolic agent such as a metal coil or polylvinyl alcohol particles. This treatment is rapid and effective and eliminates the need for conventional surgical operative treatment. The interventional treatment of variceal bleeding from portal hypertension is described below under transjugular intrahepatic portosystemic shunt (TIPS). X-ray fluoroscopy provides the visualization necessary to guide these procedures.

Stricture Dilation

Strictures at each end of the gut can be traversed by catheter and guide-wire techniques allowing for the coaxial passage of dilating balloons or, in addition, expandable metallic stents. Malignant and benign strictures of the esophagus, duodenum, and recto-sigmoid colon have been successfully treated and stricture dilations have been useful as palliation in terminal malignancy or as preparation for surgery on the colon.

LIVER

Tumor Ablation

Tumors or metastatic masses in the liver that cannot be surgically excised can be completely or partially destroyed by percutaneous alcohol injection directly into the tumor, by percutaneous placement of a radiofrequency probe into the tumor, which delivers heat energy to destroy the tumor mass, or, alternatively, by destroying the tumor via freezing, using a cryoprobe placed into the tumor mass. A different approach is selective infusion of chemo-embolic agents through an arterial catheter. This technique delivers high concentrations of agents specifically targeted at the tumor cells and is less toxic than intravenous administration of these agents. Blocking the feeding artery with emboli reduces the oxygenation to the tumor cells, making them more vulnerable to the chemotherapeutic drugs. All these procedures are dependent on image guidance techniques such as fluoroscopy, ultrasound, CT, MRI, or a combination of two modalities. The effectiveness of these procedures varies greatly from occasional complete cure to minimal palliation. The results depend on the size of the mass, the number of masses, the location of the masses, and the etiology of the masses.

Diagnosis and Treatment of Bleeding

Bleeding can occur in the liver due to trauma, iatrogenic injury such as biopsy, or tumors. Angiography can be used to identify the site of bleeding and then the bleeding can frequently be stopped using selective trans-catheter embolization.

Percutaneous Cholecystostomy

The gallbladder can be drained percutaneously in patients who have inflammation of the gallbladder (cholecystitis), who have an obstruction of the cystic duct draining the gallbladder, or who have stones in the gallbladder. These patients would usually undergo surgical removal of their gallbladders, but they may require percutaneous drainage as a temporary measure until they are well enough to have surgery. The drainage catheter is placed into the gallbladder via a percutaneous needle puncture and coaxial guide-wire technique using ultrasound guidance.

Percutaneous Transhepatic Cholangiography

In a similar manner, a thin 22-gauge needle can be passed percutaneously into a bile duct. Contrast material is injected through the needle, thus opacifying the bile ducts. X-ray films can then be taken showing the anatomy and any pathological changes of the bile ducts. The procedure is performed under fluoroscopic guidance and in a number of cases ultrasound can also be used to assist in duct puncture, especially on the left side of the liver. Percutaneous transhepatic cholangiography (PTC) is of particular value when retrograde endoscopic cholangiography is not possible. PTC is the first step in percutaneous transhepatic biliary drainage, described in the next section.

Percutaneous Transhepatic Biliary Drainage

Once a bile duct has been punctured, as described above, a guide wire can be passed through the needle into the lumen of the punctured bile duct. The needle is removed and a catheter can be passed over the retained guide wire into the duct, establishing access from the skin to the biliary tree. This percutaneous, skin-to-duct tract provides a route for drainage of a blocked system, access for balloon dilation of strictures, stone removal, or endolumenal biopsy. Expandable metallic stents can be placed via this route to keep narrowed ducts open. These techniques have found wide application in both the treatment of benign disease and the palliation of malignant disease.

Transjugular Intrahepatic Portosystemic Shunt

One of the most dramatic procedures highlighting the scope of interventional radiology is the TIPS procedure, in which a shunt is created by making a tract through the substance of the liver to connect the portal vein directly to the hepatic vein (Fig. 1). TIPS is used as a treatment for the serious condition of high pressure in the portal vein (portal hypertension). The consequences of portal hypertension are life-threatening gastrointestinal bleeding and ascites. A long needle is passed via a percutaneous puncture of the internal jugular vein in the neck, down through the superior vena cava and the right atrium, and into the right hepatic vein. The needle

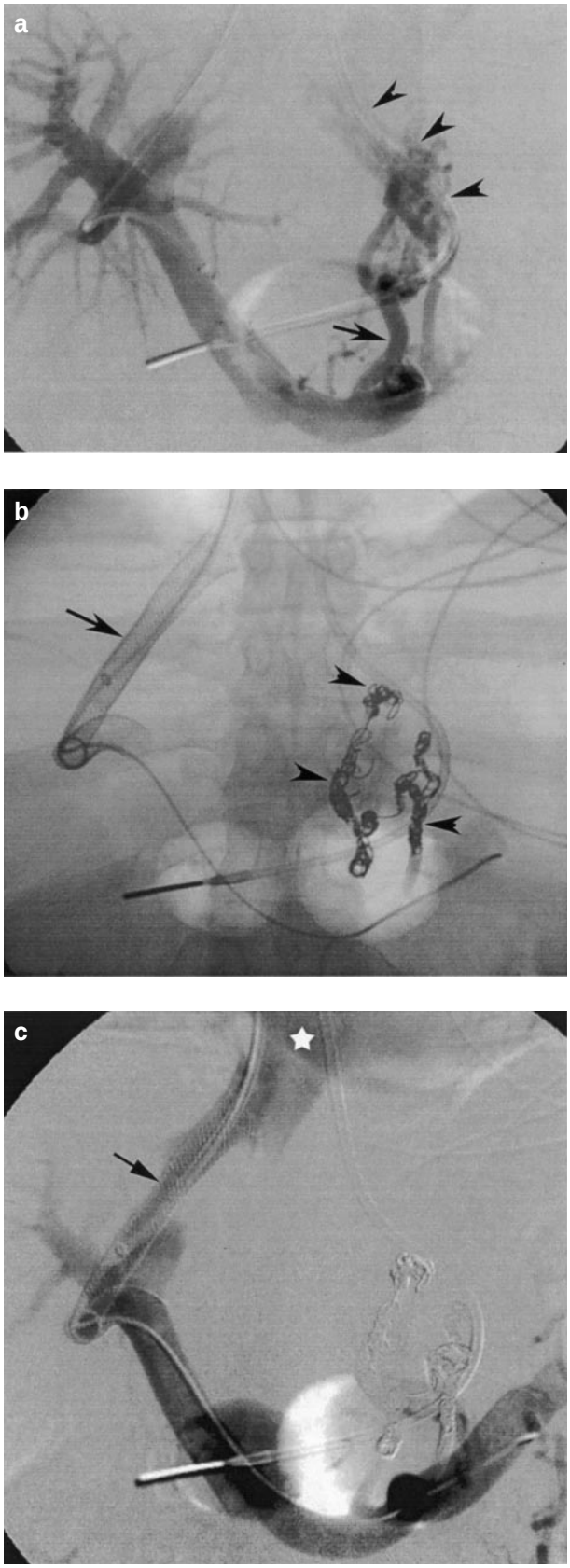


FIGURE 1 (a) Transjugular portal venogram showing abnormally large coronary vein (arrow) and varices (arrowheads). (b) The TIPS has been constructed and the metal stents are easily seen (arrow). Metal coils have been used to block the varices (arrowheads). (c) Completion venogram shows that blood flows from the portal vein through the TIPS (arrow) to the right atrium (star). The varices no longer fill with blood.

is then passed through the substance of the liver to puncture the right portal vein within the liver. This passage is then dilated with a balloon catheter and it is kept open by means of an expandable metallic sent. Blood is shunted through this newly created tract, bypassing the diseased liver and flowing directly from the portal system to the hepatic vein and then to the inferior vena cava. The portal system is decompressed, which stops portal bleeding and allows ascites to resolve.

ABSCESS DRAINAGE

Imaging techniques such as ultrasound, CT, and MRI provide an excellent demonstration of any fluid accumulations or abscesses within a patient's body. In most cases, a needle can be inserted percutaneously under local anesthesia into the collection, using ultrasound, CT, or possibly MRI guidance. Then, via the needle lumen, a guide wire is inserted and the needle is exchanged for serially enlarging dilators and finally a self-retaining catheter to provide drainage. Occasionally it is sufficient to just aspirate the contents of the collection via the needle. Percutaneous abscess drainage is in very frequent use and has greatly improved and simplified the scope of care of patients with primary abscesses and especially of patients with postoperative abscesses.

BIOPSY

The same image-guided techniques allow the needle biopsy of suspicious lumps in all parts of the body. Often, aspirating some tissue cells through a thin needle is enough to allow the cytologist to make a diagnosis. In other instances, a large-caliber needle is used to provide a core biopsy for full histological examination. Percutaneous biopsy is safe, simple, and effective and is widely used.

See Also the Following Articles

Alimentary Tract, MRI of the • Cholecystectomy • Computed Tomography (CT) • Gastrostomy • Lower Gastrointestinal Bleeding and Severe Hematochezia • Magentic Resonance Imaging (MRI) • Percutaneous Transhepatic Choliangiography (PTC) • Upper Gastrointestinal Bleeding

Further Reading

Backal, C. W. (2002). "Vascular and Interventional Radiology: Principles and Practice." Thieme, Stuttgart, Germany.

LaBerge, J. M., Gordon, R. L., Kerlan, R. K., and Wilson, M. W. (2000). "Interventional Radiology Essentials." Lippincott Williams & Wilkins, Philadelphia, PA.

Valji, K. (1999). "Vascular and Interventional Radiology." W. B. Saunders, Philadelphia, PA.

Rectal Ulcers

LISA M. YERIAN
University of Chicago

erosion A partial-thickness mucosal defect.
neoplasia The pathological process that results in the formation and growth of a tumor.
prolapse The descent of a body part from its usual anatomical position.
tenesmus A physical symptom described as urgent, painful, and ineffective straining at stool.
ulcer A full-thickness mucosal defect.

Rectal ulcers are full-thickness defects of the rectal mucosa. They typically present clinically with rectal bleeding or pain. As in other portions of the colon, various pathologic processes and injurious agents can lead to rectal mucosal injury and resultant pathologic changes ranging from focal erythema or erosion to ulceration. The injury may be localized to the rectum or the rectum may be affected by disease processes that also affect other portions of the colon. The diagnosis of rectal ulcers is based on a typical clinical presentation combined with characteristic endoscopic and histologic findings.

INTRODUCTION

Rectal mucosal ulceration represents a nonspecific effect of mucosal injury. The injury may be confined to the rectum or it may be diffuse or multifocal, also affecting other portions of the gastrointestinal tract. The type of injury varies, but common causes of rectal mucosal injury include ischemia, physical trauma, and infection. The degree of mucosal injury depends on the timing, type, and severity of the injury, and the resultant pathologic changes range from mild mucosal erythema to erosion and, ultimately, ulceration. This article focuses on the clinicopathologic features of common diseases in which rectal ulceration can occur. These entities are divided into those processes that typically are limited to the rectum and those in which rectal involvement may occur as part of a more diffuse colonic disease.

ULCERS UNIQUE TO THE RECTUM

Mucosal Prolapse Syndromes

The mucosal prolapse syndromes [solitary rectal ulcer syndrome (SRUS), rectal prolapse, proctitis cystica profunda (PCP), and inflammatory cloacogenic polyp] are a group of entities characterized by rectal mucosal injury related to mucosal prolapse. The mucosal prolapse syndromes often coexist and exhibit overlapping clinical, endoscopic, and histologic features. The endoscopic and histologic features depend on the frequency, location, and duration of prolapse and the underlying cause. All of these entities are united by a similar process of injury resulting in ulceration followed by mucosal regeneration, leading to the development of a polypoid lesion. The syndromes have been referred to by the various names listed above, depending on which histologic feature is most prominent.

Patients affected by mucosal prolapse typically present in the third or fourth decade with symptoms of anorectal disease. Patients may complain of rectal bleeding, diarrhea, anorectal pain, abdominal cramps, difficulty defecating (constipation, straining, rectal prolapse, or a sense of incomplete rectal emptying), or fecal incontinence. Endoscopic examination reveals ulcerated, polypoid, and/or indurated areas, most often located on the anterior or anterolateral wall. Ulcers, if present, often straddle a rectal fold and vary in size from a few millimeters to several centimeters in diameter. Ulcers are not always present, however, and some patients only have an erythematous area with or without polypoid projections.

Solitary Rectal Ulcer Syndrome

Solitary rectal ulcer syndrome is characterized by the presence of ulcers or polypoid inflammatory lesions in the rectum (Fig. 1). The descriptive name of this condition is misleading, however, as the lesions are frequently multiple and in almost half of the cases there is no ulceration. This entity affects both children and adults, mostly those between 20 and 40 years of age, and is more common in women. The most common symptom is rectal bleeding during defecation, followed by mucus discharge, anorectal or abdominal pain, and tenesmus. Most patients report having had symptoms for many years and many cases are initially misdiagnosed.

 Copyright 2004, Elsevier (USA). All rights reserved.

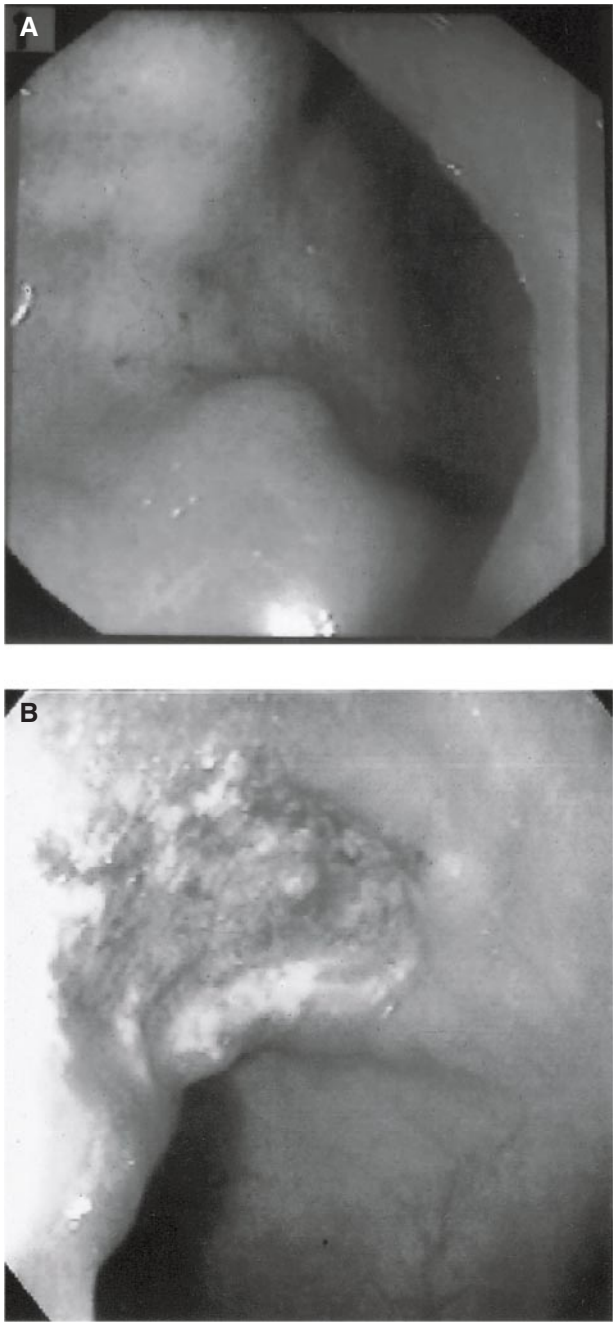


FIGURE 1 Endoscopic images of solitary rectal ulcer syndrome. (A) Typical case of solitary rectal ulcer syndrome, exhibiting polypoid projection of rectal mucosal but no ulceration. (B) Solitary rectal ulcer with ulceration overlying prolapsed rectal mucosa.

The pathogenesis of SRUS is related to abnormal colonic motility. Most patients with SRUS have abnormal contraction of the puborectalis muscle, increased external sphincter tone, increased intrarectal pressure during evacuation, and overt or occult prolapse of the rectal mucosa. The pathogenesis is related to inadequate relaxation of the internal anal sphincter or the overall rectal musculature, resulting in excessive straining during defecation. The combination of intermittent mucosal prolapse with vascular compromise and mechanical trauma to the mucosa results in mucosal erosions. With repeated episodes, there is progressive injury and mucosal regrowth, leading to the formation of a polyp.

The diagnosis of SRUS is based on symptoms in conjunction with the characteristic endoscopic and histologic findings. In some cases a cystic mass, polyp, or a sessile villous lesion may be seen. The lesions occur most frequently on the anterior or anterolateral rectal wall, but can be more extensive and may even be circumferential. Histologic sections often show features of mucosal hyperplasia, often exhibiting architectural distortion with serration of glandular lumens. There may be surface erosion or ulceration. There is often fibrosis or neovascularization of the lamina propria and thickening of the muscularis mucosae, with upward extension of smooth muscle fibers into the lamina propria.

The prognosis for SRUS varies. The lesions may regress, remain stable, or progress and become disabling. Although there is no definitive therapy for SRUS, symptoms may be reduced by conservative measures including dietary modifications, local agents to promote tissue regeneration, or biofeedback behavioral training to correct the abnormal defecation process. In severe cases, surgical treatment may be considered.

Rectal Prolapse

Rectal prolapse is the descent of some or all of the layers of the rectal wall through the anal sphincter. Rectal prolapse occurs in infants, is uncommon in children and young adults, and increases in frequency after age 40. Patients with rectal prolapse often complain of straining or pain during defecation, fecal incontinence, mucus discharge, pruritus, rectal bleeding, a sense of obstruction of incomplete rectal evacuation, and perineal or intervaginal pressure. The presence of reddened, protruding rectal mucosa is characteristic of rectal prolapse or a palpable mass may be detected on digital rectal examination (Fig. 2). There may be surface erosion or ulceration. Because the lesions can have an endoscopic appearance and a clinical presentation similar to rectal cancer, histologic analysis is important for diagnosis. Histologic features are similar to those seen in SRUS.

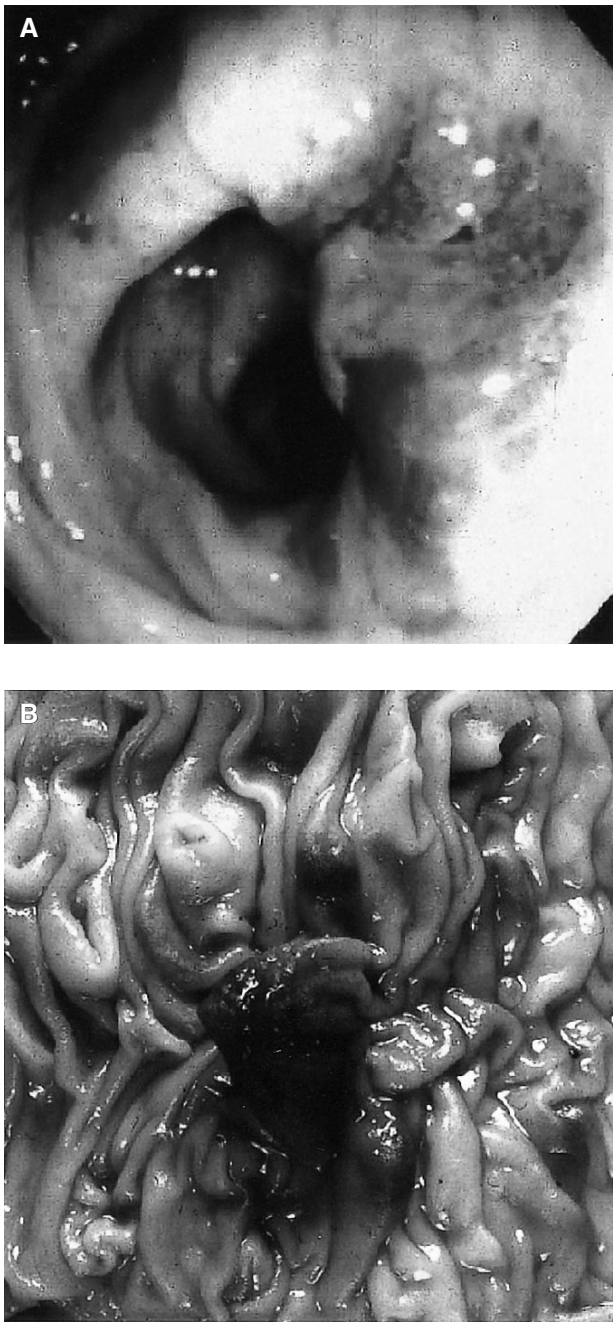


FIGURE 2 Endoscopic (A) and gross resection specimen (B) images of rectal mucosal prolapse. Clinically and endoscopically the lesion was suspected to represent rectal cancer. However, histologic analysis of the endoscopic biopsies and evaluation of the resection specimen confirmed the diagnosis of rectal mucosa prolapse.

Proctitis Cystica Profunda

Proctitis cystica profunda is the deep displacement of rectal mucosa through the muscularis mucosae and into the submucosa or deeper parts of the intestinal wall. This entity represents regenerative changes following deep ulceration of the colon from any cause. Patients may complain of blood and mucus in stools, but PCP may not be symptomatic. PCP most commonly occurs as an isolated lesion in the late stages of SRUS but multiple small lesions may be seen in chronic inflammatory conditions such as inflammatory bowel disease. PCP is rarely observed in an otherwise apparently normal colon.

PCP may not be visible endoscopically, frequently presenting as an incidental histologic finding. When observed grossly, PCP may appear as a raised or polypoid lesion, as focal mucosal edema or erythema, or with obvious cysts from which thick mucus may exude when compressed. Histologic sections reveal cystically dilated colonic glands within the submucosa or deeper in the intestinal wall. The glandular epithelium is usually normal or hyperplastic, but may show regenerative features.

Inflammatory Cloacogenic Polyp

Inflammatory cloacogenic polyp is a polypoid prolapse of the anorectal transitional zone mucosa. It typically presents in the fifth to seventh decades as a small, sessile polyp at the anorectal junction. It was originally defined by its surface lining of transitional-type epithelium, but this feature is now thought to represent a metaplastic change in SRUS. This inflammatory polyp has histologic features similar to those seen in mucosal prolapse and SRUS, including fibrosis and smooth muscle fibers extending into the lamina propria. These similarities suggest that prolapse may be important in the pathogenesis of these polyps and have led some authors to include this entity in the mucosal prolapse syndromes.

Stercoral Ulcer

Stercoral ulcers are longitudinal mucosal tears or perforations that result from fecal impaction. They occur most frequently in the distal colon and rectum. Patients typically present with severe chronic constipation, pain, and rectal bleeding. The hard, impacted fecal material causes localized pressure, ischemia, and subsequent necrosis of the mucosal surface. The ulcers may be single or multiple and usually have sharply defined edges and there is congestion of the adjacent mucosa. Most lesions are confined to the submucosa, but deeper ulceration and perforation can occur. Biopsy may be obtained to exclude other inflammatory causes of ulceration or neoplasia. Histologic sections of the early lesions reveal ischemic injury with extensive necrosis and entrapped fecal material,

vascular congestion, and patchy hemorrhage. Chronic ulcers show reparative changes, fibrosis, and inflammation and may exhibit a granulomatous response to the fecal matter.

RECTAL INVOLVEMENT IN DISEASES ALSO AFFECTING OTHER PORTIONS OF THE GASTROINTESTINAL TRACT

In addition to mucosal prolapse syndromes and stercoral ulcer, which typically affect the rectum, rectal involvement may also occur as part of a more diffuse colonic disease. Inflammatory bowel disease (Crohn's disease and ulcerative colitis) can present with localized rectal disease or with rectal involvement by more diffuse disease. Ischemic bowel disease can involve the rectum and may cause mucosal ulceration. Localized rectal ulcers can occur with the use of nonsteroidal anti-inflammatory drugs, especially when used as a suppository. Radiation injury to the rectum occurs when high doses of radiation are used to treat tumors arising in the pelvis. Ulceration is a common feature of many infectious processes. Those that typically affect the rectum include syphilis and herpes infections, both of which are spread by direct inoculation and cause ulceration.

See Also the Following Articles

Colitis Cystica Profunda and Solitary Rectal Ulcer Syndrome • Colitis, Ulcerative • Crohn's Disease • Solitary Rectal Ulcer Syndrome

Further Reading

Cooper, H. (1998). Benign epithelial polyps of the intestines. *In* "Pathology of the Gastrointestinal Tract" (S. Ming and H. Goldman, eds.), 2nd Ed., pp. 819–853. Williams & Wilkins, Baltimore, MD.

Fenoglio-Preiser, C., Noffsinger, A., Stemmermann, G., Lantz, P., Listrom, M., and Rilke, F. (eds.) (1999). "Gastrointestinal Pathology: An Atlas and Text." Lippincott-Raven, Philadelphia, PA.

Goldman, H. (1998). Other inflammatory disorders of the intestines. *In* "Pathology of the Gastrointestinal Tract" (S. Ming and H. Goldman, eds.), 2nd Ed., pp. 719–753. Williams & Wilkins, Baltimore, MD.

Levine, D. (1987). "Solitary" rectal ulcer syndrome: Are "solitary" rectal ulcer syndrome and "localized" colitis cystica profunda analogous syndromes caused by rectal prolapse? *Gastroenterol.* **92**, 243–253.

Poon, K., Mills, S., Booth, I., and Murphy, S. (1997). Inflammatory cloacogenic polyp: An unrecognized cause of hematochezia and tenesmus in childhood. *J. Pediatr.* **130**, 327–329.

Tendler, D., Aboudola, S., Zacks, J., O'Brien, M., and Kelly, C. (2002). Prolapsing mucosal polyps: An underrecognized form of colonic polyp: A clinicopathological study of 15 cases. *Am. J. Gastroenterol.* **97**, 370–376.

Tsiaoussis, J., Chrysos, E., Glynos, M., Vassilakis, J., and Xynos, E. (1998). Pathophysiology and treatment of anterior rectal mucosal prolapse syndrome. *Br. J. Surg.* **85**, 1699–1702.

Rectum, Anatomy

CAROLYN COMPTON
McGill University, Montreal

mucosa Surface epithelial layer.
rectum Most distal portion of the colon.
TNM staging International classification analyzing the extent of cancer spread; assessment of tumor, nodal, and metastatic status.

The rectum is the distal-most aspect of the large intestine, located distal to the sigmoid colon and proximal to the anal canal. The wall of the rectum has four layers: mucosa, submucosa, muscularis propria, and an adventitia or subserosa and visceral peritoneum (present anteriorly along the proximal portion of the rectum at the peritoneal reflection). The rectum is entirely lined by glandular mucosa that is identical to that of the colon. Although these basic characteristics are universally accepted, more precise definitions of the location, extent, and boundaries of the human rectum are surprisingly controversial. Recently, increased interest in rectal anatomy has provoked a great deal of debate; consensus has been sought for standardization of surgical techniques for rectal resection and for data comparisons among clinical studies of rectal disease.

CLINICAL IMPORTANCE OF RECTAL ANATOMY

A precise anatomic definition of the rectum is clinically important for at least three reasons. A precise anatomic definition (1) identifies the boundary between the rectum and the anal canal, which determines the appropriate staging system (i.e., colorectal vs. anal canal) for cancers of this region; (2) delineates the lymph nodes that are included in the regional lymphatic drainage and are assigned to the N category, versus those that are nonregional and assigned to the M category in the TNM staging system of the American Joint Committee on Cancer (AJCC) and the Union Internationale Contre le Cancer (UICC); and (3) describes the histologic layers of the viscus, which are directly relevant to the pathologic evaluation of surgical resection specimens and to analysis of circumferential (nonperitonealized, radial) margins.

PRACTICAL ASPECTS OF RECTAL ANATOMY

The origin of the rectum is marked by several distinctive structural features. At the point of transition from sigmoid colon to rectum, the tenia coli of the sigmoid fuse to form the continuous circumferential layer of the rectal muscularis propria, and the epiploic appendages disappear. It also is at the point of origin of the rectum that the free peritoneal mesentery of the sigmoid colon terminates and the mesorectum begins. The mesorectum, a subperitoneal layer of fibroadipose tissue containing all of the nerves, blood vessels, and regional lymph nodes, is enveloped by a fascia propria. The mesorectum surrounds the rectum for most of its length, ending at the level of the pelvic floor a few centimeters proximal to the termination of the rectum and the beginning of the anal canal. The mesorectal collar is asymmetric, with the bulk of the mesorectal soft tissue lying posterior to the rectum. Anteriorly, the areolar plane that lies external to the mesorectum condenses and the fascia propria of the mesorectum merges with the retrovesical fascia (in the male) or the retrovaginal septum (in the female). Laterally, the merged fascias appear as discrete ligaments, often called the lateral ligaments of the rectum.

The point of termination of the rectum, widely recognized by surgeons as a palpable landmark, is called the anorectal ring. This represents the site of merger of the muscles that form the levator ani (pubococcygeus, ileococcygeus, and puborectalis) with the muscle of the superior aspect of the anal sphincter. Despite these generally accepted anatomic definitions of the origin and termination of the rectum, the boundaries of the rectum are more commonly defined by the more clinically convenient measurements from the anal verge. Given the wide anatomic variation among individual patients of different sexes and body habitus, it is not surprising that definitions based on measurements have not concurred.

CONTROVERSIES IN DEFINING RECTAL ANATOMY

In *Guidelines for Colon and Rectal Cancer Surgery*, a surgical consensus statement published in 2001, the

 Copyright 2004, Elsevier (USA). All rights reserved.

rectum was defined as being 12 cm or less from the anal verge by rigid proctoscopy, but the distal rectal margin (at the point of transition to the anal canal) was not defined. This definition of the proximal rectal border was considered justifiable on a biologic rather than anatomic basis, because clinical observations indicate that the patterns of recurrence of tumors above 12 cm are more consistent with colonic cancers than rectal cancers. The multidisciplinary Colorectal Common Data Elements Task Force sponsored by the National Institutes of Health and the National Cancer Institute defined the rectum as beginning 12 cm above the perianal skin as measured endoscopically, but the distal boundary was defined as 2 cm above the distal-most aspect of the dentate line. Both of these definitions emphasized measurement from the anal verge, the point of transition from the hair-bearing perianal epidermis to the squamous mucosa of the anal canal, to identify the origin of the rectum. Measurement from the anal verge is itself inherently imprecise due to the considerable anatomic variation of this landmark.

Clinically, a precise definition of the border between the distal rectum and the proximal anal canal is essential because adenocarcinomas of these adjacent regions are staged differently. Specifically, in the TNM staging system for colorectal cancer of the AJCC and UICC, the T category of colorectal cancer is defined by the degree of extension through the wall, and the N category of colorectal cancers is defined by the number of involved nodes. In contrast, the T category (i.e., T1 and T2) of anal canal cancer is defined by tumor size, and the N category is defined by the location of involved nodes. Despite the importance of defining anatomic location for proper staging of distal rectal versus anal canal cancers, the staging literature has been rife with confusing statements. The fifth edition of the AJCC staging manual, on which TNM staging was based from 1998 until January 2003, offered two contradictory descriptions of the rectum. On the one hand, it defined the rectum as the distal 10 cm of large intestine as measured from the anal verge with a sigmoidoscope, but it also stated that the rectum is approximately 12 cm in length, making no allowance for the anal canal.

The revised definition of the rectum offered in the sixth edition of the AJCC staging manual allows for more anatomic variation but is still in conflict with other currently published definitions. In the sixth edition of the manual, the origin of the rectum is described as being variably located from 12 to 15 cm from the dentate line. The definition contrasts with that from the *Guidelines for Colon and Rectal Cancer Surgery* and the *Colorectal Common Data Elements*, in which the origin of the rectum is defined as 12 cm or less from the anal verge or the perianal skin, respectively. Thus, consensus is still lacking, and for the purposes of data comparison, these conflicting definitions are problematic.

By any of these definitions, the border between the rectum and anal canal is not a clear-cut anatomic landmark on visual inspection. In practice, however, physicians may ignore this controversial issue altogether and tend to regard the readily identifiable dentate line as the anorectal border. If it is accepted that the border between the anal canal and rectum is ledge of the anorectal ring, the most proximal aspect of the anal canal is lined by rectal-type glandular mucosa. At the dentate line, a narrow zone of transitional mucosa similar to urothelium also may be present. This proximal zone of the anal canal (i.e., from the top of the panaorectal ring to the dentate line, including the transitional zone) measures approximately 1–2 cm. Thus, the termination of the rectum is actually located 1–2 cm above the dentate line, and tumors with an epicenter located up to 2 cm above the dentate line are staged as anal canal cancers, not rectal cancers.

CONCLUSION

The rectum begins at the point of fusion of the tenia coli and the termination of the mesosigmoid at the termination of the sigmoid colon. It ends where the anal canal begins, about 1–2 cm above the dentate line. The length of the rectum, measured between these two landmarks, varies among individuals of different sex and body habitus.

See Also the Following Articles

Anal Canal • Anal Cancer • Colon, Anatomy • Colorectal Adenocarcinoma • Colorectal Adenomas • Gastrointestinal Tract Anatomy, Overview

Further Reading

Fleming, I. D., Cooper, J. S., Henson, D. E., *et al.* (eds.) (1997). Digestive system. *In* "AJCC Manual for Staging of Cancer," 5th Ed., pp. 57–109. Lippincott Raven, Philadelphia.

Greene, F. L., Balch, C. M., Fleming, I. D., *et al.* (eds.) (2002). Digestive system. *In* "AJCC Manual for Staging of Cancer," 6th Ed., pp. 89–164. Springer Verlag, New York.

Nelson, H., Petrelli, N., Carlin, A., *et al.* (2001). Guidelines for colon and rectal cancer surgery. *J. Natl. Cancer Inst.* **93**, 583–596.

Stocchi, L., Nyam, D. C. N. K., and Pemberton, J. H. (2002). The anatomy and physiology of the rectum and anus including applied anatomy. *In* "Surgery of the Alimentary Tract" (G. D. Zuidema and C. J. Yeo, eds.), 5th Ed., Vol. IV, pp. 332–356. W. B. Saunders, Philadelphia.

Recurrent Abdominal Pain (RAP)

JOHN TIMOTHY BOYLE
University of Alabama, School of Medicine

cognitive behavioral therapy Therapeutic approach of adding to behavioral interventions (e.g., relaxation and behavior management techniques) strategies such as cognitive restructuring; for example, a therapist may evaluate a patient's cognitive interpretation of bodily sensations and teach how cognition impacts affective experience and behavior.

functional abdominal pain Most common cause of recurrent abdominal pain in children (unknown etiology); no specific structural, infectious, inflammatory, or biochemical cause for the abdominal pain can be determined.

recurrent abdominal pain Common term in the pediatric literature; refers to paroxysmal abdominal pain that persists for greater than 3 months' duration and affects normal activity.

visceral hypersensitivity Increased visceral perception; perceived as pain.

Recurrent abdominal pain (RAP), a common term in the pediatric literature, refers to paroxysmal abdominal pain in children that persists for longer than 3 months and affects normal activity. RAP has been reported to occur in 10–15% of children between the ages of 4 and 16 years. At least as many children experience chronic pain, but maintain normal activity and rarely come to the attention of the physician. Recurrent abdominal pain is not a singular diagnosis. The differential diagnosis of RAP includes organic pain (anatomical, infectious, inflammatory, and biochemical causes), psychogenic pain, and functional abdominal pain. Although no incidence data are available, clinical experience suggests that by far, the most common cause of RAP is functional abdominal pain. The modifier "functional" is used in gastroenterology if no specific structural, infectious, inflammatory, or biochemical cause for the abdominal pain can be determined. Yet, in clinical practice, functional abdominal pain should not be a diagnosis of exclusion. Primary-care physicians should be able to make a primary diagnosis of functional abdominal pain without resorting to a large battery of biochemical or X-ray tests. Management of functional pain is facilitated by early diagnosis, parental education and reassurance, and clear delineation of goals of therapy. The major outcome variable in management of functional abdominal pain in children is lifestyle, not cure of the pain.

DIAGNOSIS OF FUNCTIONAL ABDOMINAL PAIN

One reason why primary-care physicians have difficulty making a positive diagnosis of a functional abdominal pain is that there is rarely a clear distinction between acute and chronic abdominal pain. Primary caregivers must often deal with the evolution of pain from the initial acute presentation to a chronic or recurring problem. A stepwise series of diagnostic studies is often initiated during early stages of the pain when an organic etiology is considered to be more likely. Empiric therapy with nonopioid analgesic medications, antispasmotic/anticholinergic agents, and gastric acid-reducing agents may be tried before time criteria for RAP are met. Parents tend to become more frustrated and anxious, particularly if they perceive a serious disorder is being missed, or if the physician implies that the primary factors that influence the perception of pain are cognitive and emotional. Parental uncertainty only increases the stressful environment, which provokes or reinforces the pain behavior. Thus, the concept of functional abdominal pain must be introduced into the differential diagnosis of abdominal pain in children before the 3-month time criteria for duration of pain are met.

The key variables that point toward a functional diagnosis are a normal physical exam, other than abdominal pressure tenderness, and absence of signs and symptoms that, despite evidence-based verification, are generally accepted to be alarm signals for an organic disorder. None of the following criteria has been shown to discriminate between organic, functional, or psychosomatic disorders: frequency of pain, character of pain, location of pain, pain awakening patient at night, associated gastrointestinal (GI) symptoms (including anorexia, nausea, episodic vomiting, increased gas, or altered bowel pattern), or associated extraintestinal symptoms (including fatigue, headache, and arthralgia). Even with a normal physical exam, further diagnostic testing is definitely indicated in the presence of the following alarm signals: involuntary weight loss, growth retardation, significant vomiting, significant diarrhea, GI blood loss, associated fever, arthritis, rash, symptoms of a psychiatric disorder, or family history of inflammatory

 Copyright 2004, Elsevier (USA). All rights reserved.

bowel disease. Alarm signals in the physical examination include evidence of linear growth deceleration, localized tenderness in the right upper or lower quadrants, localized fullness or mass effect, hepatomegaly, splenomegaly, back or costo-vertebral angle (CVA) tenderness, perianal fissure, fistula, soiling, and guaiac-positive stools.

Diagnostic testing is indicated when alarm signals or abnormal physical findings suggest a high possibility of organic disorder. Diagnostic testing may be considered to reassure the parent, patient, or physician when the most likely diagnosis is functional pain. The physician may also need to do testing to rule out organic disease in the patient when pain continues to severely affect lifestyle despite a functional diagnosis. Establishing a working diagnosis of functional pain and initiating conservative therapy before time criteria are achieved do not preclude an ongoing focused diagnostic workup. Although there are no evidence-based data, clinical experience suggests that subclassifying pain presentations may facilitate choice of testing by narrowing zdifferential diagnosis. Children with abdominal pain may be subclassified by one of three clinical presentations: (1) abdominal pain associated with symptoms of upper abdominal distress, (2) abdominal pain associated with altered bowel pattern, and (3) isolated paroxysmal abdominal pain. The frequent occurrence of upper and lower bowel symptoms in the same patient is not uncommon. Functional abdominal pain should be presented as the most common cause of all three clinical presentations. Synonyms of functional pain that may be useful for individualizing diagnosis in a given patient are functional dyspepsia for pain with upper abdominal symptoms, and irritable bowel syndrome for pain associated with altered bowel pattern. Abdominal migraine is a variant of isolated functional abdominal pain. Diagnostic criteria include three or more episodes of intense, acute midline pain during a 3-month period lasting several hours to days with intervening symptom-free intervals lasting weeks to months. Two of the following features are required for diagnosis: (a) headache during episodes, (b) photophobia during episodes, (c) associated classical unilateral migraine headaches, which may or may not be associated with abdominal pain, (d) family history of migraine, and (e) visual, sensory, or motor aura antedating acute pain.

DIFFERENTIAL DIAGNOSIS OF SUBCATEGORIES OF ABDOMINAL PAIN

Abdominal Pain Associated with Symptoms of Upper Abdominal Distress

Symptoms of upper abdominal distress include pain or discomfort localized in the upper abdomen, and pain related to eating, nausea, episodic vomiting, bloating, early satiety, and occasional heartburn and oral regurgitation. Table I lists the differential diagnosis of abdominal pain associated with symptoms of upper abdominal distress. Alarm signals such as anorexia, vomiting, weight loss, and evidence of GI bleeding (hematemesis, melena, occult bleeding) suggest an upper GI inflammatory, infectious, structural, or biochemical disorder. Focused laboratory evaluation should be performed in any patient with historical or physical alarm signals, including complete blood count with differential, erythrocyte sedimentation rate (ESR), *Helicobacter pylori* serology and/or stool antigen, hepatic panel, and pancreatic enzyme measurement. In cases in which recurrent vomiting is a significant part of the history, an upper GI series with small bowel follow-through and abdominal ultrasound should be considered to rule out gastric outlet disorder, malrotation, partial small bowel obstruction, small bowel Crohn's disease, gallstones, pseudocyst, hydronephrosis secondary to ureteropelvic junction (UPJ) obstruction, and retroperitoneal mass. Gastroesophageal reflux disease should be suspected when heartburn or oral regurgitation of sour or bitter gastric contents is prominent parts of the history. Biliary pain is typically episodic, severe, constant pain in the right upper quadrant or epigastrium that persists for 20 minutes to 2 hours, usually triggered by eating. In relapsing pancreatitis, recurrent severe epigastric pain persistent for days and may radiate to the back. Recurrent epigastric or right upper quadrant pain associated with tender hepatomegaly suggests chronic hepatitis. Continuous pain, especially in the context of

TABLE I Major Differential Diagnosis of Recurrent Abdominal Pain Associated with Symptoms of Upper Abdominal Distress

Functional dyspepsia
Gastroesophageal reflux disease
Drug-associated GI disorder (NSAID, iron, antibiotic)
Helicobacter pylori gastritis
Peptic ulcer
Gastroparesis
Eosinophilic gastroenteritis
Crohn's disease
Obstructive disorders (e.g., malrotation, partial small bowel obstruction)
Biliary tract disease (choledocholithiasis, chronic cholecystitis, biliary dyskinesia)
Hydronephrosis
Chronic hepatitis, sclerosing cholangitis
Relapsing pancreatitis/pancreatic pseudocyst
Parasitic infection
Celiac disease

multisystem complaints, is an alarm signal for possible psychiatric disease. Eating disorder should also be considered in any young patient with significant weight loss.

In the absence of peptic ulcer disease, the relationship between *H. pylori* infection and abdominal pain remains unclear. Although there are no evidence-based data to establish a clear link *H. pylori* gastritis alone, and abdominal pain associated with symptoms of upper abdominal distress, most gastroenterologists will treat a symptomatic child who has been identified as *H. pylori* positive. The rationale is that *H. pylori* may act as a physical trigger of functional dyspepsia in selected patients. Some clinicians have concluded that the most cost-effective approach is to test serologically for *H. pylori* and to treat all infected cases. However, many investigators have pointed out that commercially available serological assays do not appear to have the necessary sensitivity or specificity to screen pediatric patient populations. Empirical treatment of *H. pylori* should only be considered in patients with elevated immunoglobulin G (IgG) antibody, and is not recommended for patients with positive IgM or IgA antibody. It is not unreasonable to avoid antibody testing completely, and consider treatment only in patients with endoscopically proved infection.

Upper endoscopy should be considered in untreated patients with alarm signals, in patients who fail to respond to time-limited antisecretory therapy, and in patients in whom symptoms recur after the end of treatment. Upper endoscopy is the gold standard to rule out inflammatory disorders in the upper GI tract and to establish a firm diagnosis of functional dyspepsia. Recognizable objective findings by gross endoscopic examination include superficial erosions, ulcer, stricture, antral nodularity associated with *H. pylori* gastritis, gastric rugal hypertrophy associated with Menetrier's and cytomegalovirus (CMV) gastritis, and the small heaped-up, volcanic-like mounds, pocked with a central crater, associated with chronic varioliform gastritis. Objective histologic findings may help to diagnose reflux esophagitis, eosinophilic gastroenteritis, CMV gastritis, *H. pylori* gastritis, Crohn's disease, and celiac disease. In the absence of gross ulcer or histologic evidence of *H. pylori*, superficial antral gastritis or duodenitis are of questionable clinical significance, and should not dissuade a diagnosis of functional abdominal pain. There is no evidence in children that nonspecific superficial antral gastritis or duodenitis progresses to peptic ulcer. Evaluation of gastric emptying by scintigraphy to rule out gastroparesis, and gallbladder function assessment by hepatobiliary scan with ejection fraction, to rule out chronic cholecystitis and biliary dyskinesia, should be considered only after upper endoscopy and consultation by a pediatric gastroenterologist. Endoscopic retrograde cholangiopancreatography is indicated only if there is biochemical or radiological evidence of recurrent pancreatitis, or biliary-type abdominal pain following cholecystectomy.

Abdominal Pain Associated with Symptoms of Altered Bowel Pattern

Altered bowel pattern may include change in frequency and/or consistency of stools (diarrhea or constipation), pain relief with defecation, straining or urgency, feeling of incomplete evacuation, passage of mucus, or a feeling of bloating or abdominal distension. Table II lists the differential diagnosis of abdominal pain associated with symptoms of altered bowel pattern. In patients with diarrhea, focused laboratory evaluation should include complete blood count with differential, erythrocyte sedimentation rate, stool for ovum parasites, and stool for *Clostridium difficile* toxin. Lactose intolerance should be considered as a potential primary etiology of chronic abdominal pain in the presence of diarrhea. A trial of a lactose-free diet or performance of a lactose breath hydrogen test is prudent in children with pain associated with loose bowels, bloating, and increased flatulence. Alarm signals including evidence of GI bleeding, tenesmus, pain or diarrhea repeatedly wakening the patient from a sound sleep, involuntary weight loss, linear growth deceleration, extraintestinal symptoms (fever, rash, joint pain, recurrent aphthous ulcers), positive family history of inflammatory bowel disease, iron deficiency anemia, and elevated ESR are indications to pursue a diagnosis of inflammatory bowel disease by colonoscopy and upper gastrointestinal (UGI) study with small bowel follow-through. Diarrhea associated with encopresis suggests chronic fecal retention and megacolon. At present, although there are no evidence-based data, serological testing for celiac disease should be considered in patients with pain and altered bowel pattern. Serological testing should

TABLE II Major Differential Diagnosis of Recurrent Abdominal Pain Associated with Altered Bowel Pattern

Irritable bowel syndrome
Chronic infection (parasitic, *Clostridium difficile*)
Inflammatory bowel disease (Crohn's disease, ulcerative colitis, microscopic colitis, e.g., lymphocytic colitis, eosinophilic colitis, collagenous colitis)
Lactose intolerance
Fecal retention/megacolon
Drug-associated GI disorder

definitely be performed in all patients with iron deficiency anemia or secondary amenorrhea. Chronic watery diarrhea is also an indication to pursue colonoscopy to rule out microscopic inflammation, which may alter colonic motility and absorptive function. The large volume of diarrhea (400−1200 g/day) distinguishes patients with microscopic lymphocytic, collagenous, or eosinophilic colitis from those with irritable bowel, for whom stool weight in excess of 300 g/day is rare.

The accuracy of colonoscopy in diagnosing inflammatory conditions of the colon is superior to barium enema because of the direct visualization of the mucosal surface and the ability to obtain biopsy and culture specimens. Intubation of the terminal ileum can also aid in the diagnosis of Crohn's disease. Recognizable objective findings by gross examination with a flexible endoscope include edema, erosions, ulceration, pseudomembranes (discrete yellow plaques on the colonic mucosa), and polyps. Subjective gross endoscopic findings, including erythema, increased vascularity, and spontaneous friability, become meaningful only in the context of histology, because they are subject to more interobserver variation in interpretation. Objective histologic findings include (1) cryptitis, crypt abscesses, and crypt distortion with branching and drop out, suggesting ulcerative colitis or Crohn's disease, (2) noncaseating granuloma specific for Crohn's disease, (3) fibrosis and histiocyte proliferation in the submucosa, suggesting Crohn's disease, and (4) epithelial and intraepithelial lymphocytes or eosinophils, with or without subepithelial collagen thickening in lymphocytic colitis, eosinophilic colitis, and collagenous colitis, respectively. The latter should be considered specific findings only in patients with profuse diarrhea. Mild superficial increases in interstitial lymphocytes or eosinophils in the absence of crypt distortion or significant diarrhea are nonspecific, and should not dissuade the physician from making a positive diagnosis of irritable bowel syndrome.

Isolated Paroxysmal Recurrent Abdominal Pain

Table III lists the major differential of recurrent paroxysmal periumbilical abdominal pain in children. It is important to try to see the patient during an attack, because it is frequently the only opportunity to assess alarm signals. The Carnett test may help to determine whether pain is arising from the abdominal wall or has an intraabdominal origin. The site of maximum tenderness is found through palpation. The patient is then asked to cross arms and assume a partial sitting position (crunch), which results in tension of the abdominal wall. If there is greater tenderness on repeat palpation in this position, abdominal wall disorders (such as cutaneous nerve entrapment syndromes, abdominal wall hernia, myofascial pain syndromes, rectus sheath hematoma, or costochondritis) should be suspected. Discitis, which is really an osteomyelitis of the vertebral end plate, may present as a combination of back and abdominal pain, The condition is usually associated with intermittent fever, elevated peripheral white blood cell count, and elevated erythrocyte sedimentation rate. Unrecognized constipation should be suspected if a left lower quadrant or suprapubic fullness or mass effect is appreciated on abdominal exam, and rectal exam reveals evidence of firm stool in the rectal vault, or soft stool in a dilated rectal vault with evidence of perianal soiling. Often, a history of constipation or encopresis is unknown to the parent. Parasitic infections, particularly *Giardia lamblia*, *Blastocystis hominis*, and *Dientamoeba fragilis*, may present with chronic pain in children in the absence of altered bowel pattern. Alarm signals, including pain repeatedly awakening the patient from a sound sleep, anorexia, involuntary weight loss, linear growth deceleration, evidence of GI bleeding, and extraintestinal symptoms (fever, rash, joint pain), are also indications to evaluate for Crohn's disease, or rare disorders such as polyarteritis nodosa, intestinal ischemia, eosinophilic gastroenteritis, and angioneurotic edema, which can be indistinguishable from Crohn's disease on clinical grounds. Suspicion of polyarteritis nodosa rests on evidence of extraintestinal disease, particularly renal involvement. Mesenteric vein obstruction should be considered in adolescents using oral contraceptives. Clinically, it can present gradually with progressive abdominal pain over a period of weeks. Pneumatosis is usually a late finding. The clinical presentation of eosinophilic gastroenteritis depends on the depth of the infiltration by the eosinophilic process. Submucosal disease can become manifest with abdominal pain and signs of obstruction.

TABLE III Major Differential Diagnosis of Isolated Recurrent Abdominal Pain

Functional abdominal pain
Abdominal migraine
Intermittent intestinal obstruction (Crohn's disease, malrotation w/wo volvulus, intussusception with lead point, postsurgical adhesions, small bowel lymphoma, eosinophilic gastroenteritis, angioedema)
Unrecognized constipation
Appendiceal colic
Dysmenorrhea (endometriosis, ectopic pregnancy, adhesions from pelvic inflammatory disease)
Cystic teratoma of ovary
Musculoskeletal disorders (muscle pain, linea alba hernia, discitis)
Vascular disorders (mesentertic thrombosis, polyarteritis nodosa)
Acute intermittent porphyria

Any region of the GI tract can be involved. Angioneurotic edema can be heralded by recurrent episodes of pain in the absence of cutaneous or oropharyngeal edema. Family history is usually positive for allergy. Recurrent fever associated with generalized abdominal pain and peritoneal signs suggests the possibility of familial Mediterranean fever. Appendiceal colic is a controversial cause of chronic abdominal pain. Appendiceal spasm has been postulated to be caused by inspissated casts of fecal material within the appendix. A number of anecdotal surgical reports have described complete resolution of pain symptoms following elective appendectomy. Appendiceal colic should be suspected in patients with recurrent acute episodes of well-localized abdominal pain and tenderness, most commonly in the right lower quadrant, demonstrated on several examinations. Dull, midline, or generalized lower abdominal pain at the onset of a menstrual period suggests dysmenorrhea. The pain may coincide with the start of bleeding or may precede the bleeding by several hours. Gynecological disorders associated with secondary dysmenorrhea include endometriosis, partially obstructed genital duplications, ectopic pregnancy, and adhesions following pelvic inflammatory disease. Cystic teratoma has been described in prepubertal patients presenting with right or left lower quadrant pain. The vast majority of such patients have a palpable abdominal mass. Benign ovarian cysts in adolescent females do not cause recurrent abdominal pain. Acute intermittent porphyria (AIP) is a rare disorder characterized by the temporal association of paroxysmal abdominal pain and a wide variety of central nervous system symptoms, including headache, dizziness, weakness, syncope, confusion, memory loss, hallucinations, seizures, and transient blindness. AIP is often precipitated by low intake of carbohydrate or by specific drugs such as barbiturates or sulfonamides.

Laboratory evaluation might include complete blood count (CBC) with differential and ESR to screen for occult systemic inflammatory condition. The decision to do stool ova and parasite exams is dependent on the incidence of *G. lamblia*, *B. hominis*, and *Dientamoeba fragilis* within the community. The most valuable diagnostic test in a patient with symptoms suggesting obstruction is an upper GI series and small bowel follow-through (SBFT). Rare conditions such as lymphoma, angioneurotic edema, mesenteric vein thrombosis with ischemia, eosinophilic gastroenteritis, and pseudo-obstruction will also be suggested by the UGI series. Abdominal ultrasound and abdominal computed tomography (CT) scans have low diagnostic yield for picking up appendiceal abnormalities with recurrent right lower abdominal pain. Colonoscopy and ileoscopy should be performed to rule out Crohn's disease in such patients if blood work or UGI–SBFT suggest the possibility of inflammatory disease. Elective laparoscopy with planned appendectomy should be considered in patients with chronic right lower quadrant pain and negative infectious, inflammatory, and anatomical evaluation. Head CT scan to rule out intracranial space-occupying lesions should be considered in patients with recurrent abdominal pain and headache.

In the absence of historical or physical alarm signals, the diagnosis of functional abdominal pain should be introduced into the differential diagnosis of abdominal pain persisting a month beyond the usual course of an acute disease (e.g., gastroenteritis, urinary tract infection). Parents should be told that a diagnosis of functional pain can be made if duration of pain goes on to exceed 3 months. It is important to provide a brief explanation of visceral hypersensitivity and altered motility, the concept of stress factors, and natural history. Parents should also be told early on that functional pain is difficult to eradicate, and some continuing pain will often have to be accepted by the patient.

TREATMENT OF FUNCTIONAL ABDOMINAL PAIN

Management of functional abdominal pain begins with a positive diagnosis and an explanation of suspected pathophysiology, natural history, and goals of therapy. Specific treatments include dietary modification, drug therapy, and psychological support. Hospitalization is rarely indicated for patients with functional abdominal pain.

Positive Diagnosis

Positive diagnosis is based on normal physical examination and absence of alarm signals in the history, as previously described.

Explanation of Suspected Pathophysiology and Natural History

The exact etiology and pathogenesis of functional abdominal pain in children are unknown. There is general agreement that functional pain is genuine. The prevailing viewpoint is that the pathogenesis of the pain involves visceral hypersensitivity and altered conscious awareness of gastrointestinal sensory input, with or without disordered gastrointestinal motility. Painful sensations may be provoked by physiologic phenomena or concurrent physical and psychological stressful life events. Examples of physiologic phenomena that may trigger pain include postprandial gastric or intestinal

distension, gastric emptying, intestinal contractions of the migrating motor complex, intestinal gas, or gastroesophageal reflux. Intraluminal physical stress factors that may trigger pain include aerophagia, simple constipation, lactose intolerance, minor noxious irritants such as spicy foods, *H. pylori* gastritis, celiac disease, or drug therapy. Systemic physical or psychological stress factors may also provoke or reinforce the pain behavior by altering the conscious threshold of GI sensory input in the central nervous system. Acute or chronic physical illness may unmask functional pain. Psychological stress factors may include death or separation of a significant family member, physical illness or chronic handicap in parents or sibling, school problems, altered peer relationships, family financial problems, or recent geographical relocation.

It is not clear whether the different clinical presentations of functional abdominal pain result from a heterogeneous group of disorders, or represent variable expressions of the same disorder. The frequent occurrence of upper and lower bowel symptoms in the same patient suggests that the latter scenario may indeed be the case. There appears to be a genetic vulnerability because of the high frequency of functional disorders in family members. The fact that most children "outgrow" pain symptoms also suggests that variation of neuroendocrine development may also be a factor in the pathophysiology. In some patients, associated symptoms, including headache, dizziness, motion sickness, pallor, temperature intolerance, and nausea, suggest a generalized dysfunction of the autonomic nervous system. Sex, intelligence, and personality traits do not distinguish patients with functional pain from those with organic pain. The majority of patients are of average intelligence. The generalization that patients with functional abdominal pain are superintellects, perfectionists, over-achievers, bad social mixers, or constant worriers is without foundation. However, there are some data suggesting that the incidence of functional gastrointestinal disorders may be higher in patients with mental diagnoses, such as attention-deficit disorder/hyperactivity, anxiety, depression, school phobia, posttraumatic stress, bipolar disorder, autism, and eating disorders.

The morbidity associated with functional abdominal pain is rarely physical, but results from interference in normal school attendance and performance, peer relationships, participation in organizations and sports, and personal and family activities. Only 1 in 10 children with functional abdominal pain attend school regularly, and absenteeism is greater than 1 day in 10–28% of patients. A common misconception is that pain is the direct cause of the morbidity. In fact, focus on symptom relief by parents, school, and physicians reinforces the pain behavior with attention at the time of pain, rest periods during pain, tactile stimulation and medication to alleviate pain symptoms, and absence from school on days of pain. This approach fails to reinforce nonpain responses, such as normal activity. Although pain does not originate from its consequences, ongoing pain behavior is often accounted for and modified by its consequences.

Goals of Therapy

The focus of treatment is not "cure" of pain, but rather management of symptoms and adaptation to illness. Goals of treatment include regular school attendance, school performance to the child's ability, participation in desired extracurricular activities, normal weight gain and growth, and normal sleep pattern.

Dietary Modification

The role of dietary modifications in the management of functional pain disorders is not established. Postprandial symptoms in functional dyspepsia may be improved by eating low-fat meals or by ingesting more frequent but smaller meals throughout the day. High-fiber diet is recommended for both diarrhea-predominant and constipation-predominant irritable bowel and isolated functional pain. The goal for fiber intake in grams is calculated by adding the patient's age +5. Excessive fiber in the diet may result in increased gas and distension, actually provoking pain. Malabsorption of dietary carbohydrates may act as provocative stimuli in functional abdominal pain. Most often, the patient does not perceive a temporal association between ingestion of a particular sugar and the abdominal pain. Avoidance of excessive intake of milk products (lactose), carbonated beverages (fructose), dietary starches (corn, potatoes, wheat, oats), or sorbitol-containing products (vehicle for oral medication, sugar substitute in gum and candy, ingredient in toothpaste, and a plasticizer in gelatin capsules) is not unreasonable. Conformation of lactose intolerance by a lactose breath hydrogen test should be considered before recommending prolonged lactase enzyme replacement therapy or commercial milk products that have been pretreated with lactase enzyme. Excessive gas in patients with irritable bowel syndrome (IBS) can be managed by advising the patient to eat slowly, to avoid chewing gum, and to avoid excessive intake of

carbonated beverages, legumes, foods of the cabbage family, and foods or beverages sweetened with aspartame.

Drug Therapy

There are no evidence-based data on the effects of pharmacological therapy in pediatric patients with functional abdominal pain. Antispasmodic agents, including hyoscyamine, dicyclomine, glycopyrrolate, peppermint oil, and calcium-channel blockers, are commonly used in clinical practice to treat visceral abdominal pain, although efficacy is unproved. Time-limited (8 weeks) empirical medical therapy with a histamine-2 (H2) blocker or proton pump inhibitor (PPI) is an acceptable diagnostic test of self-limiting upper GI inflammation in patients with abdominal pain and symptoms of dyspepsia. If possible, it is prudent to stop nonsteroidal antiinflammatory drugs (NSAIDs), iron preparations, and antibiotics such as erythromycin or tetracyclines in a patient complaining of upper abdominal discomfort. After a firm diagnosis of functional dyspepsia is established by upper endoscopy, it is not unreasonable to continue acid inhibition therapy in patients who initially responded to shot-term empiric treatment, but had recurrence of pain symptoms with attempts at step-down therapy. Short-term step-up to a PPI may be tried in patients who previously did not respond to an H2 blocker. In adults, sucralfate, a drug that stimulates mucosal prostaglandin synthesis and release of cytokines and has cytoprotective properties, has been reported to be superior to placebo and H2 blocker in alleviating symptoms of dyspepsia. Prokinetic therapy has also been reported in adults to provide superior symptom improvement compared to placebo, especially in patients with dysmotility-like dyspepsia in which the predominant complaint is an unpleasant discomfort in the upper abdomen characterized by upper abdominal fullness, nausea, early satiety, or bloating. At present, metoclopramide is the only option for treating pediatric patients. Metoclopramide has a significant side-effect profile, including drowsiness, dystonic reactions, and increased prolactin levels. As stated previously, although *H. pylori* eradication therapy is not established to be effective in adults with functional dyspepsia, the available data clearly do not rule out the possibility. Thus, most pediatric gastroenterologists still recommend treating documented *H. pylori* in conjunction with endoscopic-established functional dyspepsia.

There are also no evidence-based data on the effects of pharmacological therapy in pediatric patients with irritable bowel syndrome. Synthetic opioids such as loperamide and diphenoxylate are effective in treating IBS-associated diarrhea. Loperamide is preferred over diphenoxylate because it does not traverse the blood–brain barrier. Nonstimulating laxatives such as polyethylene glycol solution, mineral oil, milk of magnesia, and lactulose are effective adjuncts in treating constipation-predominant IBS.

Although formal randomized placebo-controlled trials are lacking, there has been a recent surge in using antidepressant and psychotropic agents to treat both diarrhea-predominant IBS and functional dyspepsia in adults. Anecdotally, this class of drugs appears to be effective in adults with or without psychiatric abnormalities, especially low-dose tricyclic antidepressants. These drugs may act as "central analgesics" to raise perception threshold for abdominal pain to or down-regulate pain receptors in the intestine. There are as yet no data on treatment of pediatric patients. For adults, there has been a recent surge in the development of novel drugs for IBS, including 5-hydroxytryptamine isotype 3 (5-HT3) receptor antagonists, 5-hydroxytryptamine isotype 4 (5-HT4) receptor agonists, and κ-opioid agonists, aimed at restoring normal visceral sensation. Significant beneficial effect of the 5-HT3 antagonist alosetron has been reported in diarrhea-predominant adult women with IBS. Significant beneficial effect of the 5-HT4 agonist tegaserod has been reported in constipation-predominant adult women with IBS. None of these drugs has been studied in children.

Psychological Support

The first goal is to identify, clarify, and possibly reverse psychological stress factors that may have an important role in onset, severity, exacerbations, or maintenance of pain. Equally important is to reverse environmental reinforcement of the pain behavior. Parents and school must be engaged to support the child rather than the pain. Regular school attendance is essential regardless of the continued presence of pain. In many cases, it is helpful for the physician to communicate directly to school officials to explain the nature of the problem. School officials must be encouraged to be responsive to the pain behavior but not to let it disrupt attendance, class activity, or performance expectations. Within the family, less social attention should be directed toward the symptoms. Consultation with a child psychiatrist or psychologist may be indicated when there is concern about maladaptive family coping mechanisms, or if attempts at environmental modification do not result in return to a normalized lifestyle.

Many patients and parents are unable or unwilling to report emotional states or acknowledge a relationship between psychogenic stresses and pain symptoms. It is best to limit discussion of psychological issues to what the patient and family can accept and let the physician/patient/family relationship evolve by continuing to listen actively, provide empathy, and educate about potential benefit of relaxation techniques and coping strategies. Referral for psychological treatment can be proposed as part of a multispecialty treatment package to help the patient manage the pain symptoms better. It is critical that the psychologist or psychiatrist initially focus on illness behavior and expand psychotherapeutic treatments as indicated only as the patient or parents begin to see the benefits of referral.

Cognitive behavioral therapies add strategies such as cognitive restructuring to behavioral interventions such as teaching relaxation and behavior management techniques. For example, a therapist would evaluate a patient's cognitive interpretation of bodily sensations and teach how cognition impacts affective experience and behavior. The perception that abdominal pain is a sign of impending physical disease must be countered, both to address functional disability and to reassure the family that a functional diagnosis is credible. Attribution styles can also be examined for distortions. Patients are taught to treat their beliefs as hypotheses to be tested, rather than accept their beliefs as inherently valid. Cognitive behavioral interventions targeting children's competence in social roles may be a useful adjunct to other medical treatment in reducing illness behavior. In addition, parents are trained to behaviorally reinforce appropriate coping behavior. Evidence-based data show that cognitive behavioral treatment helps to reduce pain and improve functioning. Cognitive behavioral treatment has been compared to standard pediatric care, and both groups demonstrate reductions in pain at 3 months; however, those receiving cognitive behavioral treatment are more likely to be pain free at 6-month (55.6 vs. 23.8%) and 12-month followup (58.8 vs. 36.8%) evaluations. These findings are very encouraging, although replication by different investigators is still needed.

Hospitalization

During hospitalization, 50% of patients experience relief of symptoms. However, no data have been presented that the natural history of the pain is affected. Hospitalization does not enhance the fundamental goals of environmental modification. More commonly, it will reinforce pain behavior.

PROGNOSIS OF RECURRENT FUNCTIONAL ABDOMINAL PAIN IN CHILDREN

There are no prospective studies of the outcome of any of the various presentations of functional abdominal pain. Once functional abdominal pain is diagnosed, subsequent followup rarely identifies an occult organic disorder. Interestingly, pain resolves completely in 30–50% of patients by 2–6 weeks after diagnosis. This high incidence of early resolution suggests that child and parent accept reassurance that the pain is not organic and that environmental modification is effective treatment. Nevertheless, more long-term studies suggest that 30–50% of children with functional abdominal pain in childhood experience pain as adults, although in 70% of such individuals, the pain does not limit normal activity. Of patients with functional abdominal pain, 30% develop other chronic complaints as adults, including headaches, backaches, and menstrual irregularities. Based on a small number of patients, Apley and Hale have described several factors that adversely influence prognosis for a lasting resolution of pain symptoms during childhood, including male sex, age of onset less than 6 years, strong history of a "painful family," and greater than 6 months elapsed time from onset of pain symptoms to established functional diagnosis.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Celiac Disease, Pediatric • Colitis, Ulcerative (Pediatric) • Cow Milk Protein Allergy • Crohn's Disease, Pediatric • Gastritis and *Helicobacter pylori*, Pediatric • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Intussusception • Irritable Bowel Syndrome • Malrotation • Parasitic Diseases, Overview • Volvulus

Further Reading

Apley, J. (1975). The child with abdominal pains. Blackwell Scientific Publ., London.

Boyle, J. T. (1997). Recurrent abdominal pain: An update. *Pediatr. Rev.* **18**, 310–321.

Rasquin-Weber, A., Hyman, P. E., Cucchiara, S., Fleisher, D. R., Hyams, J. S., Milla, P. J., and Staiano, A. (1999). Childhood functional gastrointestinal disorders. *Gut* **45**(Suppl. II), II60–II68.

Walker, L. S., Guite, J. W., Duke, M., Barnard, J. A., and Greene, J. W. (1998). Recurrent abdominal pain: A potential precursor of irritable bowel syndrome in adolescents and young adults. *J. Pediatr.* **132**, 1010–1015.

Sanders, M. R., Shepard, R. W., Cleghorn, G., and Woolford, H. (1994). The treatment of recurrent abdominal pain in children: A controlled comparison of cognitive behavioral family intervention and standard pediatric care. *J. Consult. Clin. Psychol.* **62**, 306–314.

Rotavirus

ROBERT D. SHAW
State University of New York at Stony Brook and Northport VA Medical Center

hemagglutinin The viral protein that binds to erythrocytes.
Jennerian Regarding vaccines, the use of naturally attenuated but antigenically similar animal strains as human vaccines, as Jenner used cowpox virus to protect against smallpox.
positive and negative strands The coding sequence and the complementary antisense sequence, respectively.
RNA polymerase The viral enzyme needed to replicate the double-stranded RNA viral genome in the mammalian host.

In 1972, a virus was first implicated as a cause of human gastroenteritis. Rotaviruses are now identified as the leading cause of severe dehydrating gastroenteritis in infants and children.

CLASSIFICATION

The rotavirus genus is contained within the family *Reoviridae*. Viral particles are 1020 Å in diameter and consist of two protein capsids surrounding a central protein core that contains the genome. Ten monocistronic segments and one bicistronic genomic segment form the double-stranded RNA genome. Rotavirus appears in electron microscopy as a sharply defined rim to which spokes radiate from a large central hub, thus suggesting the name from the Latin *rota*, meaning wheel.

HOST RANGE AND VIRUS PROPAGATION

Animal strains rarely infect humans, although human isolates that probably are derived from feline, canine, bovine, or porcine rotaviruses have been described. The potential of animal reservoirs as a source of genetic diversity for the evolution of new human strains is unknown.

GENETICS

The double-stranded RNA genes distribute by size into four classes that produce a characteristic pattern when separated by polyacrylamide gel electrophoresis. The RNA itself is not infectious; rotaviruses contain an endogenous RNA-dependent RNA polymerase that transcribes the gene segments into mRNA. Transcripts are full-length positive strands from which negative-strand synthesis occurs following the formation of replicase particles in the cytoplasm.

Genes 5, 7, 8, 10, and 11 code for nonstructural proteins known as NSP1–5, but functional roles are incompletely understood. Genes 1–4, 6, and 9 code for structural proteins VP1–4, 6, and 7, respectively. VP4 and VP7 are the two surface proteins of the virion and VP6 is the major constituent of the second protein layer.

SEROLOGIC RELATIONSHIPS

VP6 bears most of the common group antigens. Group A rotaviruses cause most human disease. The glycoprotein VP7 is the viral tinin, an important determinant of virulence, and VP4 is the cell attachment protein. VP7 and VP4 are used in the serologic classification of rotaviruses. The VP7 serotype is classified as a G (glycoprotein) type and the VP4 type as a P (protease) type. Fourteen G types and 20 P types have been distinguished, many in humans. However, the relationship of serotype to protective immunity is not entirely clear. For instance, immunoglobulin A (IgA) anti-VP6 may mediate intracellular viral neutralization and particles containing only VP2 and VP6 have induced protective immunity in animals.

EPIDEMIOLOGY

Group A rotaviruses are the principal cause of severe gastroenteritis in infants and young children, accounting for one-third of all diarrheal episodes requiring hospitalization in children under the age of 2 years. In developing countries, the annual toll includes roughly 18 million cases of severe diarrhea and nearly a million deaths. Infants in the first 2 to 3 months of life are relatively protected from severe disease, probably because of residual maternal transplacental antibodies. Rotavirus infections occur beyond 3 years of age and

 Copyright 2004, Elsevier (USA). All rights reserved.

into adult life but are typically mild or asymptomatic. Human illness occurs in the cooler months in developed countries, peaking in January and February. This seasonality does not occur in tropical climates where rotavirus infections occur throughout the year.

Rotaviruses are transmitted by the fecal–oral route. Rapid appearance of antibodies to rotaviruses is noted by 3 years of age in all areas of the world regardless of hygiene. Asymptomatic infection occurs frequently in newborn nurseries and day-care centers. The virus is quite stable on environmental surfaces for prolonged periods. These factors complicate the control of hospital outbreaks.

CLINICAL FEATURES OF INFECTION

The incubation period is 24–72 h. Malnutrition may increase the severity of the symptoms. Diarrhea, vomiting, and fever are usually noted. Symptoms related to severe volume depletion such as lethargy, irritability, confusion, and eventually vascular collapse and death can be seen.

PATHOLOGY AND HISTOPATHOLOGY

Villus blunting and vacuolation of columnar intestinal epithelial cells occur within hours after infection. Distended endoplasmic reticulum, mitochondrial swelling, and denuded microvilli are also seen together with mononuclear cell infiltration of the lamina propria. Viral particles may be seen within columnar epithelial cells, goblet cells, phagocytic cells, and M cells. Intestinal morphology returns to normal in approximately 7 days.

Diarrhea is not clearly related to histologic damage. NSP4 may have toxin-like effects that induced diarrhea in mice by stimulation of chloride secretion. Water absorption is impaired, but is ameliorated by glucose–salt solutions. Abnormal motility may contribute to vomiting and diarrhea. Carbohydrate malabsorption and secondary osmotic diarrhea may occur. An integrated view of pathogenesis of diarrhea has not yet been achieved.

IMMUNE RESPONSE

The immune response to infection includes serum and mucosal antibodies that are thought to be important in the prevention of subsequent infections. Cytotoxic T cells have been identified in the intestinal mucosa. The rapid resolution of acute infection occurs before the immune response is fully developed, so some of the factors responsible for resolution of the illness appear to be unrelated to acquired T- and B-cell immunity.

TREATMENT AND PREVENTION

Effective treatment of rotavirus diarrhea is accomplished with oral rehydration with glucose and electrolyte solutions. Logistical and educational difficulties limit this treatment in underdeveloped areas. Vaccination strategies have utilized the host range restrictions of animal rotaviruses in a "Jennerian" approach to disease prevention. Unfortunately, increased rates of intussusception immediately after vaccination caused withdrawal of the only commercially marketed attenuated rotavirus vaccine.

See Also the Following Articles

Diarrhea, Infectious • Gastroenteritis • Nosocomial Infections

Further Reading

Desselberger, U., Iturriza-Gomara, M., *et al.* (2001). Rotavirus epidemiology and surveillance. *Novartis Found. Symp.* **238**, 125–147, discussion 147–152.

Lundgren, O., and Svensson, L. (2001). Pathogenesis of rotavirus diarrhea. *Microbes Infect.* **3**, 1145–1156.

Offit, P. A. (2002). The future of rotavirus vaccines. *Semin. Pediatr. Infect. Dis.* **13**, 190–195.

Patton, J. T. (2001). Rotavirus RNA replication and gene expression. *Novartis Found. Symp.* **238**, 64–77, discussion 77–81.

Roux Stasis Syndrome

RUSSELL J. NAUTA AND ALLISON C. NAUTA
Harvard Medical School and Mount Auburn Hospital

denervation Cutting of a nerve.
vagotomy Division of the vagus nerve.

The Roux stasis syndrome is a motility disorder of the residual stomach and proximal jejunum following partial gastric resection and Roux-en-Y gastrojejunostomy. Characterized by abdominal pain, epigastric fullness, nausea, and vomiting of food (rather than bile), the syndrome develops in up to 30% of patients with Roux-en-Y gastrojejunostomy. Careful history and diagnostic imaging distinguish these postgastrectomy complaints from those of efferent loop obstruction, bile reflux gastritis, or mechanical obstruction.

ALTERATIONS IN ANATOMY AND PHYSIOLOGY

The Roux-en-Y reconstruction following partial gastric resection is named for the Finnish surgeon Cesar Roux. Duodenal and proximal jejunal secretions join the fecal stream through a jejunojejunostomy made at least 40 cm distal to the gastrojejunostomy. The iatrogenic denervation abnormalities of both the stomach and Roux limb may contribute to development of the syndrome.

Vagotomy is often done with gastric resection to avoid complications of ulceration at the anastomosis, but may impair gastric compliance, gastric emptying, and jejunal contractions. Duodenal pacing of proximal jejunal contractions is interrupted by the creation of the Roux limb, resulting in fewer depolarizations in the limb, with resultant stasis. Ectopic pacemakers in the limb propagate in both directions, compounding the motility disorder. Phase III contractions, which normally are intense motor impulses thought to be helpful in clearing indigestible material from the lumen of the jejunum, are not as effective or as regular in the Roux limb as in the nonoperated jejunum.

TREATMENT OF THE ROUX STASIS SYNDROME

Endoscopy, barium studies, and scintigraphic measure of gastric emptying are used to exclude mechanical obstruction. The treatment of choice is resection of all but a small portion of the gastric cardia and creation of a new gastrojejunostomy. At the time of revision, the Roux limb is measured and shortened if longer than 40 cm (Fig. 1). To make the Roux limb shorter than this predisposes to reflux of bile and pancreatic secretions.

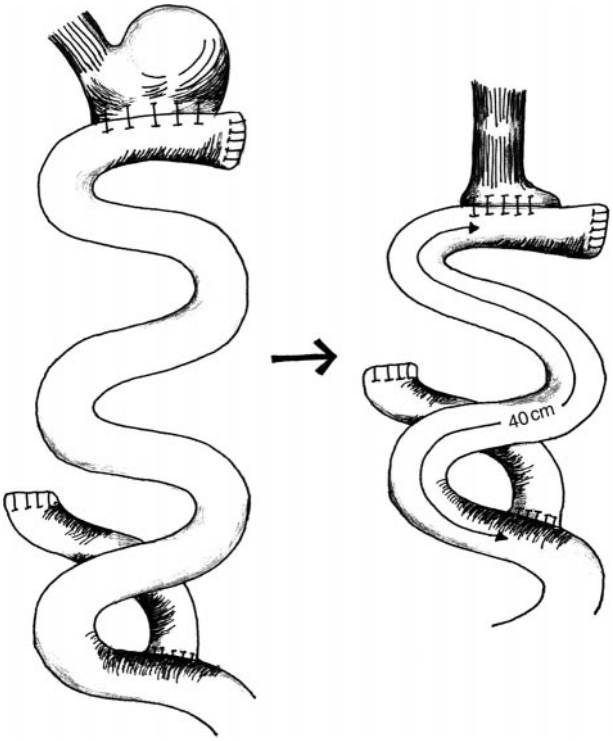


FIGURE 1 Roux-en-Y reconstruction after resection of a portion of the stomach is revised such that only a small portion of proximal stomach remains. The jejunal Roux limb between the two anastomoses is shortened such that it is less than 40 cm long.

More than 80% of patients gain weight after the revisional procedure; failures are likely due to sustained and irreversible abnormalities of the Roux limb pacing and mechanical clearing mechanisms.

See Also the Following Articles

Gastric Emptying • Gastroenterostomy

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Eagon, J., Miedama, B., and Kelly, K. (1992). Post-gastrectomy syndromes. *Surg. Clin. North Am.* **72**(2).

Gustavsson, S., Ilstrup, D. M., Morrison, P., and Kelly, K. A. (1988). Roux-Y stasis syndrome after surgery. *Am. J. Surg.* **155**, 490–494.

Miedema, B. (1992). The Roux stasis syndrome: Treatment by pacing and prevention by use of an "uncut Roux limb." *Arch. Surg.* **127**, 295–300.

Sawyers, J. (1990). Management of post-gastrectomy syndromes. *Am. J. Surg.* **159**, 8–14.

Rumination Syndrome

RICHARD MCCALLUM AND JENNIFER GRAFF
University of Kansas Medical Center

pathognomonic Characteristic or indicative of a disease; denoting especially one or more typical symptoms, findings, or pattern of abnormalities specific for a given disease and not found in any other condition.

Rumination syndrome is characterized by the regurgitation of undigested food and liquids, usually within 15 minutes after eating. Once the regurgitant reaches the mouth, the person can consciously choose to reswallow or expel it. Rumination syndrome is believed to be psychological in origin, which is supported by the absence of organic disease and the link seen between a traumatic or stressful "triggering event" and the onset of symptoms.

INTRODUCTION

Historically, rumination syndrome, the spontaneous regurgitation of undigested food and liquids shortly after ingestion, has not been associated with nausea, abdominal pain, heartburn, or nocturnal symptoms; however, more recent studies have found that these symptoms may be present to varying degrees in this syndrome. Due to lack of familiarity with the disease, physicians often confuse rumination syndrome with gastroparesis, gastroesophageal reflux disease, functional esophageal motility disturbances, and eating disorders. Rumination syndrome can be diagnosed solely by taking a careful clinical history, though performing a 4-hour gastric emptying test to rule out gastroparesis is appropriate. Although rumination syndrome has no cure, relaxation therapy is thought to be the most effective form of treatment at this time.

EPIDEMIOLOGY

Rumination syndrome can affect both sexes at any age. Researchers disagree about whether it is more prevalent in men or women or if it is seen equally in both. Though rumination syndrome occurs in children and adults of normal intelligence, it has been widely studied among infants and people with mental retardation. In fact, it has been estimated that 6–10% of mentally challenged persons in institutions ruminate. The prevalence of the syndrome among people of normal intelligence is difficult to estimate because these patients often do not seek medical attention or are misdiagnosed due to physician unawareness of the disorder.

DIAGNOSIS

The diagnostic criteria for rumination syndrome are based on evaluation of the patient for the following symptoms, which occurred for at least 12 weeks, not necessarily consecutive, in the preceding 12 months:

1. Persistent or recurrent regurgitation of recently ingested food into the mouth with subsequent remastication and reswallowing or expulsion.
2. Absence of nausea and vomiting.
3. Cessation of the process when the regurgitated material becomes acidic.
4. Absence of pathologic gastroesophageal reflux, achalasia, or other motility disorder with a recognized pathologic basis as the primary disorder.

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Eagon, J., Miedama, B., and Kelly, K. (1992). Post-gastrectomy syndromes. *Surg. Clin. North Am.* **72**(2).

Gustavsson, S., Ilstrup, D. M., Morrison, P., and Kelly, K. A. (1988). Roux-Y stasis syndrome after surgery. *Am. J. Surg.* **155**, 490–494.

Miedema, B. (1992). The Roux stasis syndrome: Treatment by pacing and prevention by use of an "uncut Roux limb." *Arch. Surg.* **127**, 295–300.

Sawyers, J. (1990). Management of post-gastrectomy syndromes. *Am. J. Surg.* **159**, 8–14.

Rumination Syndrome

RICHARD MCCALLUM AND JENNIFER GRAFF
University of Kansas Medical Center

pathognomonic Characteristic or indicative of a disease; denoting especially one or more typical symptoms, findings, or pattern of abnormalities specific for a given disease and not found in any other condition.

Rumination syndrome is characterized by the regurgitation of undigested food and liquids, usually within 15 minutes after eating. Once the regurgitant reaches the mouth, the person can consciously choose to reswallow or expel it. Rumination syndrome is believed to be psychological in origin, which is supported by the absence of organic disease and the link seen between a traumatic or stressful "triggering event" and the onset of symptoms.

INTRODUCTION

Historically, rumination syndrome, the spontaneous regurgitation of undigested food and liquids shortly after ingestion, has not been associated with nausea, abdominal pain, heartburn, or nocturnal symptoms; however, more recent studies have found that these symptoms may be present to varying degrees in this syndrome. Due to lack of familiarity with the disease, physicians often confuse rumination syndrome with gastroparesis, gastroesophageal reflux disease, functional esophageal motility disturbances, and eating disorders. Rumination syndrome can be diagnosed solely by taking a careful clinical history, though performing a 4-hour gastric emptying test to rule out gastroparesis is appropriate. Although rumination syndrome has no cure, relaxation therapy is thought to be the most effective form of treatment at this time.

EPIDEMIOLOGY

Rumination syndrome can affect both sexes at any age. Researchers disagree about whether it is more prevalent in men or women or if it is seen equally in both. Though rumination syndrome occurs in children and adults of normal intelligence, it has been widely studied among infants and people with mental retardation. In fact, it has been estimated that 6–10% of mentally challenged persons in institutions ruminate. The prevalence of the syndrome among people of normal intelligence is difficult to estimate because these patients often do not seek medical attention or are misdiagnosed due to physician unawareness of the disorder.

DIAGNOSIS

The diagnostic criteria for rumination syndrome are based on evaluation of the patient for the following symptoms, which occurred for at least 12 weeks, not necessarily consecutive, in the preceding 12 months:

1. Persistent or recurrent regurgitation of recently ingested food into the mouth with subsequent remastication and reswallowing or expulsion.
2. Absence of nausea and vomiting.
3. Cessation of the process when the regurgitated material becomes acidic.
4. Absence of pathologic gastroesophageal reflux, achalasia, or other motility disorder with a recognized pathologic basis as the primary disorder.

Copyright 2004, Elsevier (USA). All rights reserved.

The key to diagnosing rumination syndrome is to look for the characteristic regurgitation of food and liquids within 15 minutes of eating and the absence of any organic disease. The ability to regurgitate water within minutes of drinking is almost pathognomonic. Though the above criteria state otherwise, nausea can be an associated symptom. However, this feeling is usually brief or comes as a "wave" before the regurgitation. Patients do not awake with nausea or remain chronically nauseated between regurgitation events. Additional symptoms may include abdominal pain (typically in the epigastrium), bloating, fatigue, dehydration, heartburn, belching, and weight gain or loss (though many ruminators maintain a stable weight). The abdominal pain may actually represent the effects of retching or vomiting on the rectus abdominus muscle. Rumination syndrome is distinguished from gastroparesis by the fact that patients with gastroparesis bring up food that is hours or days old, whereas rumination syndrome patients bring up completely undigested food within minutes. In addition, patients with gastroparesis should have a delayed gastric emptying test (GET), whereas rumination syndrome patients should have a normal GET. A full 4-hour GET must be employed to obtain a reliable reading. Rumination syndrome is differentiated from gastroesophageal reflux disease (GERD) by the absence of nocturnal symptoms frequently seen in GERD. Furthermore, patients with rumination syndrome may discontinue ruminating once the stomach contents become acidic, whereas GERD patients may experience acidic reflux for hours. In addition, during the actual postprandial events, rumination patients do not experience heartburn because the acidic contents of the stomach are well buffered by accompanying foods and liquids.

PATHOPHYSIOLOGY

The physiological mechanism underlying rumination in animals is centered on reverse peristalsis of the esophagus. Because humans are incapable of this activity, three theories have been proposed to explain the rumination phenomena in humans. The first theory states that rumination is due to the simultaneous increase in abdominal pressure and relaxation of the lower esophageal sphincter. This theory has been supported by upper gastrointestinal manometry studies that show a characteristic manometry pattern linking rumination events with simultaneous pressure waves caused by an increase in intraabdominal pressure. The second theory similarly describes a relaxation of the lower esophageal sphincter, but states that it is a voluntary action. The third theory attributes the mechanism of ruminating as an adaptation of the belch reflex.

TREATMENT

Though rumination syndrome has no cure, various treatment therapies exist that may provide substantial relief. The most commonly recommended treatment is relaxation therapy. Patients should attend several sessions with a clinical psychologist to learn effective relaxation techniques. This involves thinking about a relaxing topic or image, diaphragmatic breathing and meditation, or maintaining a "pseudohypnotic" concentration state for the minutes immediately following eating or drinking. Once the patient learns these techniques, he or she must practice them independently, oftentimes after every meal to control consciously the urge to regurgitate. Other therapies include psychological therapy, behavioral therapy, biofeedback therapy, and a change of diet. These treatments, especially when combined with relaxation therapy, may provide significant relief. Although the rumination events may not be completely eradicated, they may be reduced to occurring only occasionally, or only during certain stressful situations. Finally, education of the patient, family, friends, and school or work personnel is key. Associates should be informed that rumination is a "reflex" or habit that is being addressed. They can be reassured that the patient is not pregnant, does not have a malignancy, or an undiagnosed disease.

See Also the Following Articles

Achalasia • Anorexia Nervosa • Belching • Bulimia Nervosa • Emesis • Gastric Emptying • Gastroesophageal Reflux Disease (GERD) • Nausea

Further Reading

Clouse, R., and Richter, J. (2000). "Rumination Syndrome. Rome II: The Functional Gastrointestinal Disorders" (D. Drossman, E. Corazziari, N. Talley, W. Thompson, and W. Whitehead, eds.), 2nd Ed., pp. 258–263. Degnon Associates, McLean, Virginia.

Malcolm, A., Thumshirn, M., Camilleri, M., and Williams, D. (1997). Rumination syndrome. *Mayo Clin. Proc.* **72**, 646–652.

O'Brien, M., Bruce, B., and Camilleri, M. (1995). The rumination syndrome: Clinical features rather than manometric diagnosis. *Gastroenterology* **108**, 1024–1029.

Soykan, I., Chen, J., Kendall, B., and McCallum, R. (1997). The Rumination syndrome: Clinical and manometric profile, therapy, and long-term outcome. *Dig. Dis. Sci.* **42**(9), 1866–1872.

Tougas, G., Eaker, E., McCallum, R., *et al.* (2000). Assessment of gastric emptying using a low fat meal: Establishment of international control values. *Am. J. Gastroenterol.* **95**(6), 1456–1462.

S

Salivary Glands, Anatomy and Histology

ARTHUR R. HAND
University of Connecticut Health Center, Farmington

acini Secretory end-pieces that have an approximately spherical shape.
demilunes Crescent-shaped groups of serous secretory cells at the ends of mucous end-pieces.
excretory ducts Largest ducts in salivary glands running in the interlobular connective tissue and conveying saliva to the mouth.
intercalated ducts Small ducts in salivary glands connecting the secretory cells to larger striated ducts.
intercellular canaliculi Finger-like projections of the end-piece lumen extending between adjacent secretory cells.
lumen Central space of an end-piece into which saliva is secreted; continuous with the lumina of the duct system.
mucous cells Salivary gland secretory cells that produce a viscous saliva containing highly glycosylated mucins.
myoepithelial cells Contractile cells with branching processes surrounding end-pieces; their contractions force saliva from the end-pieces into the ducts.
serous cells Salivary gland secretory cells that produce a watery saliva rich in proteins with enzymatic or antimicrobial functions.
striated ducts Intralobular salivary ducts with a striated appearance due to membrane infoldings and aligned mitochondria; active in electrolyte secretion and absorption.

The salivary glands are a collection of three paired major and many minor glands located in and around the oral cavity. Together they produce saliva, a watery fluid that lubricates and protects the oral hard and soft tissues and facilitates taste reception, mastication, swallowing, and speech. The initial, or primary, saliva, including most of the organic components and essentially all of the fluid, is formed by secretory cells organized into secretory end-pieces. The primary saliva is modified by duct cells constituting a series of convergent tubes that eventually form a single large excretory duct that empties into the oral cavity. Secretory end-pieces may be composed solely of serous cells that produce watery saliva rich in proteins and glycoproteins, many of which have enzymatic, antimicrobial, or protective functions, or mucous cells that produce a viscous saliva containing highly glycosylated mucins. Serous secretory end-pieces, or acini, usually are spherical in shape, whereas mucous secretory end-pieces typically are tubular in shape and usually are larger than serous end-pieces. The blind ends of mucous end-pieces frequently are capped by a crescent of serous cells, termed a demilune. Considerable variability exists in the histology and cellular secretory products among glands and species. This article focuses on human salivary glands and rodent salivary cells as the latter are used most often in studies of salivary function.

HUMAN SALIVARY GLANDS

In humans, there are three paired major salivary glands, located extraorally, and several hundred smaller minor salivary glands, located in the lips, cheeks, tongue, palate, fauces, and retromolar areas. The parotid gland is located subcutaneously, lying over the masseter muscle, just in front of the ear, with a deeper portion extending behind the ramus of the mandible. Its duct, Stensen's duct, runs anteriorly, crossing the masseter muscle and entering the oral cavity at the parotid papilla on the buccal mucosa, opposite the maxillary second molar. The blood supply of the parotid comes from branches of the external carotid artery and the parasympathetic nerve supply is mainly from the glossopharyngeal nerve (cranial nerve IX) via the otic ganglion and the auriculotemporal nerve. The sympathetic innervation of all of the salivary glands is provided by postganglionic fibers from the superior cervical ganglion, traveling with the blood supply.

The submandibular gland is located in the submandibular triangle, below the mylohyoid muscle, with its posterior portion wrapped around the posterior border of the mylohyoid and extending anteriorly for a short distance. Its duct, Wharton's duct, travels anteriorly below the mucosa of the floor of the mouth, opening at the sublingual caruncle. The blood supply of the submandibular gland comes from the facial and lingual arteries and the parasympathetic nerve supply is mainly from the facial nerve (cranial nerve VII), through the lingual nerve and submandibular ganglion.

The sublingual gland, the smallest of the major glands, is located in the floor of the mouth, medial to the mandible and just above the mylohyoid muscle. Its

 Copyright 2004, Elsevier (USA). All rights reserved.

main duct, Bartholin's duct, opens with the duct of the submandibular gland at the sublingual caruncle. Several smaller ducts of the sublingual gland, the ducts of Rivinus, open separately along the sublingual fold in the floor of the mouth. The blood supply of the sublingual gland comes from the sublingual and submental arteries and the parasympathetic innervation is derived from the facial nerve (cranial nerve VII), via the lingual nerve and submandibular ganglion.

The parotid gland contains only serous secretory end-pieces. The intercalated ducts typically are long and the striated ducts are prominent. The submandibular gland (Fig. 1A) is a mixed gland, with both serous and mucous secretory end-pieces; however, the serous end-pieces predominate. The mucous end-pieces are capped by serous demilune cells. The intercalated ducts also are relatively long and the striated ducts are prominent. The sublingual gland (Fig. 1B) also is a mixed gland, consisting predominantly of mucous end-pieces and serous demilunes; few, if any, serous end-pieces are present. The intercalated ducts are short and relatively few striated ducts are present.

The minor salivary glands consist of small aggregates of secretory end-pieces and ducts, located in the submucosal layer of the oral mucosa or between muscle fibers of the tongue. The ducts typically open directly onto the oral mucosal surface. Most of the minor glands are mucous and some include a serous cell component arranged as occasional demilunes. The one exception is the lingual serous (von Ebner's) gland, located in the posterior part of the tongue. Von Ebner's gland is a pure serous gland and its ducts open into the troughs surrounding the circumvallate papillae and at the rudimentary foliate papillae on the sides of the tongue.

SALIVARY GLAND HISTOLOGY

Serous Cells

Serous cells (Figs. 1A and 2A) are pyramidal in shape, with the basal surface forming an interface with the extracellular matrix and the apical surface forming a portion of a small lumen that is continuous with the first part of the duct system. Intercellular canaliculi, small finger-like projections of the lumen that increase the lumenal surface area, extend along the lateral cell surfaces toward the bases of the cells. Adjacent cells are held together at their apical ends and along the intercellular canaliculi by junctional complexes consisting of a tight junction, an adhering junction, and one or more desmosomes. Occasional gap junctions also are present on the lateral cell surfaces. The nucleus usually is spherical and is located in the basal half of the cell.

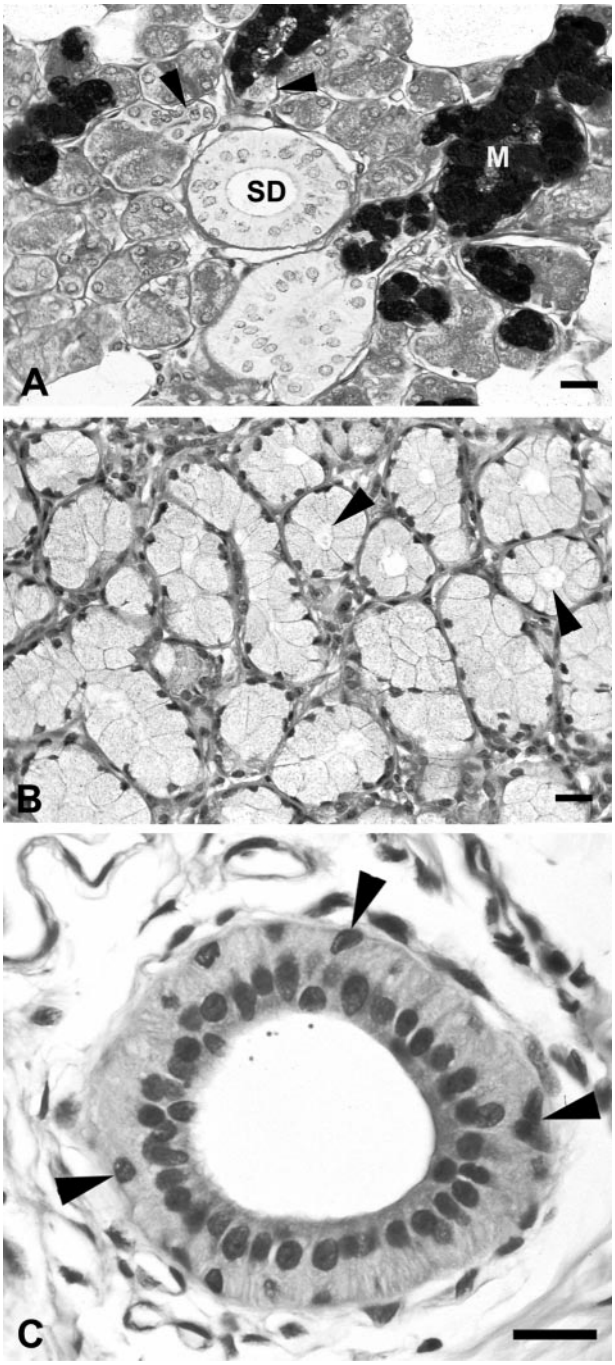


FIGURE 1 Light micrographs of human salivary glands. Scale bars, 20 μm. (A) Submandibular gland, stained with periodic acid-Schiff (PAS)/Alcian blue/hematoxylin. Serous cells, with spherical basal nuclei, contain magenta-staining secretory granules. Mucous cells (M), arranged in tubular secretory end-pieces, are filled with dark red- to purple-staining mucin. Two intercalated ducts (arrowheads), with small cuboidal cells at their junction with the secretory end-pieces, and two striated ducts (SD), with pale-staining columnar cells, are present. Basement membranes around the end-pieces and ducts are stained with PAS. Fat cell

The basal cytoplasm is filled with rough endoplasmic reticulum (RER), a well-developed Golgi complex lies apical or lateral to the nucleus, and the apical cytoplasm is filled with membrane-bound secretory granules, approximately 1 μm in diameter. The granules of human serous cells typically exhibit a bi- or tripartite structure, with an electron-dense core that may be eccentrically located in the granule and one or more regions of lower electron density constituting the remainder of the content. Immunolabeling studies have demonstrated that at least some secretory proteins are differentially distributed in the granule content. Mitochondria, lysosomes, and peroxisomes also are found scattered throughout the cell. Actin filaments usually are associated with the tight and adhering junctions and form a web beneath the apical cell membrane and intermediate filaments are associated with desmosomes as well as with hemidesmosomes that attach the cells to the basal lamina. Microtubules often are present in the Golgi region and the supranuclear cytoplasm. On stimulation, the secretory granules fuse with the lumenal membrane, releasing their contents, which are dissolved in the aqueous fluid transported into the lumen by the cells.

Mucous Cells

Mucous cells (Figs. 1B and 2B) are large cells shaped like a truncated pyramid. The apex of the cell has a larger lumenal surface than serous cells, but intercellular canaliculi usually are not present between mucous cells. They are joined to their neighbors by junctional complexes and gap junctions. The nucleus is oval in shape, contains denser chromatin than serous cells, and usually is located adjacent to the basal plasma membrane. The RER is present mainly in the basal cytoplasm and a large Golgi complex is located apical or lateral to the nucleus. The apical cytoplasm contains mucous secretory granules that typically are irregular in shape and have a pale content, with some fine granular or filamentous material. The membranes of the granules often are disrupted and/or fused with those

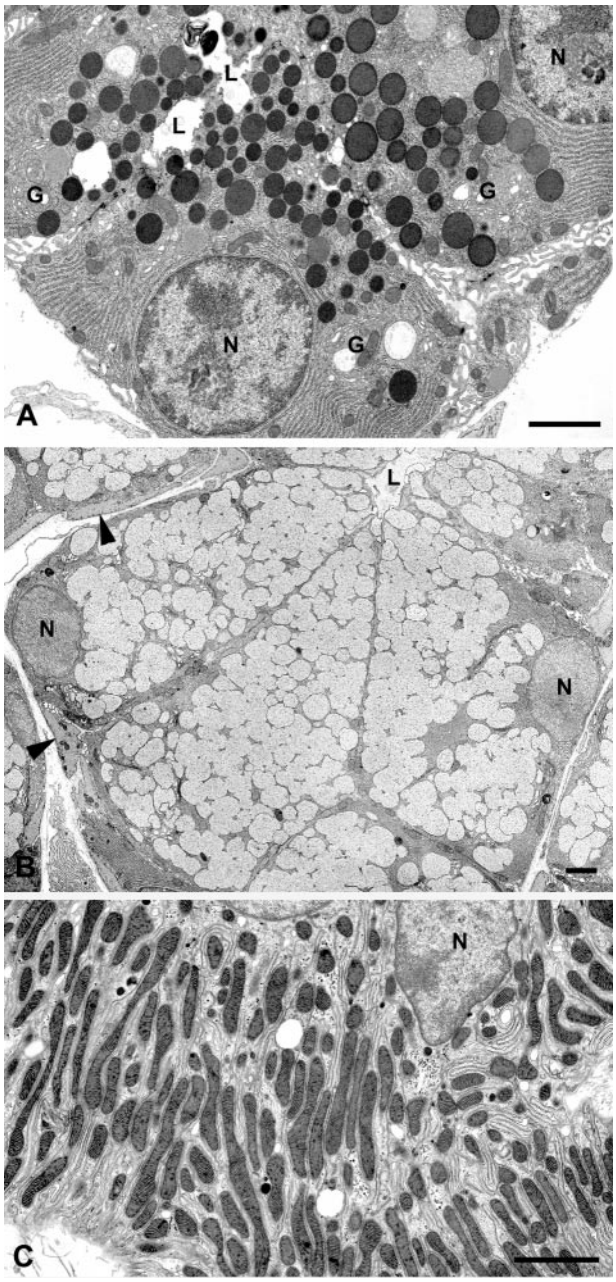


FIGURE 2 Electron micrographs of rodent salivary glands. Scale bars, 2.0 μm. (A) Rat parotid gland end-piece. Serous secretory cells have a basally located spherical nucleus (N), abundant RER, prominent Golgi complexes (G), and electron-dense secretory granules. L, lumen. (B) Mouse sublingual gland end-piece. Several mucous secretory cells are filled with electron-lucent mucous granules, which compress the nuclei (N) against the basal cell membrane. Myoepithelial cell processes (arrowheads) are located along the bases of the secretory cells. L, lumen. (C) Mouse parotid gland striated duct cell. The basal cell membranes of the striated duct cells are extensively folded and numerous elongated mitochondria are present in the cytoplasm between the membranes. Glycogen deposits and a few small dense lysosomes also are present. N, nucleus.

spaces are seen at the edges of the field. From Hand, A. R. (1986). "Oral Histology: Inheritance and Development," (D. V. Provena and W. Seiebel, eds.). 2nd Ed., Lea & Febiger. Copyright Lippincott Williams & Wilkins. (B) Sublingual gland, stained with hematoxylin and eosin (H&E). Mucous cells, with dense basally located nuclei and pale-staining mucin, are arranged in tubular end-pieces with large lumina (arrowheads). Few serous cells are visible. (C) Submandibular gland, stained with H&E. A small excretory duct, with a large lumen, is lined by a pseudostratified epithelium of tall columnar cells and several small basal cells (arrowheads). Numerous small blood vessels are present in the surrounding interlobular connective tissue.

of adjacent granules. This appearance of the mucous granules is believed to be an artifact induced by chemical fixation, which results in a loss of Ca^{2+}, disaggregation of the content, an increase in osmotic pressure, and an influx of water that causes granule swelling. Mucins constitute the main product of these cells; few other macromolecules are known to be secreted by mucous cells and the rate of fluid secretion is much lower than for serous cells.

Myoepithelial Cells

Myoepithelial cells are contractile cells associated with the secretory end-pieces. They are branching or stellate-shaped cells with processes containing actin and myosin that embrace the secretory cells (Fig. 2B). Although their structural and functional characteristics are similar to those of smooth muscle cells, myoepithelial cells are derived from epithelium and reside on the epithelial side of the basal lamina. The myoepithelial cells provide support for the secretory cells and their contraction helps to expel saliva from the end-pieces into the ductal system. Myoepithelial cells also provide signals to the secretory cells that help to maintain cell polarity and the organization of the end-piece and they produce proteins with tumor suppressor activity, such as proteinase inhibitors and anti-angiogenesis factors.

Ducts

The ductal system consists of three main subdivisions. The first type of duct is the intercalated duct, which connects the secretory end-pieces to the second part of the duct system, the striated duct. Intercalated ducts are the smallest ducts; the cells are low cuboidal (Fig. 1A) and have a relatively simple structure with a few RER cisternae, a small Golgi complex, and, in cells close to the end-pieces, a few secretory granules. The ducts draining several end-pieces typically converge to form a larger intercalated duct that connects to the striated duct. Myoepithelial cells may be located at the basal side of the intercalated ducts; often they have a fusiform shape and are oriented lengthwise along the duct and their processes extend onto the end-piece. Intercalated ducts are thought to participate in the formation of primary saliva, including the fluid component as well as specific organic products such as lysozyme and lactoferrin. In some species, intercalated duct cells have a relatively high mitotic rate and it is thought that they may house a stem cell capable of differentiation into other gland cell types.

Striated ducts (Figs. 1A and 2C) constitute the main intralobular component of the duct system. The duct cells are columnar, with centrally placed nuclei, abundant mitochondria, lysosomes, peroxisomes, and frequently some small apical vesicles and/or secretory granules. They have only small amounts of RER but may have some smooth endoplasmic reticulum in the apical cytoplasm. Their most characteristic feature is the presence of extensive infoldings of the basal and lateral plasma membranes that interdigitate with similar folds of neighboring cells (Fig. 2C). Elongated mitochondria are present in the cytoplasmic partitions between the folds. Striated duct cells function to modify the primary saliva secreted by the end-pieces, principally by reabsorption and secretion of electrolytes. The apical and basolateral membranes contain a number of ion channels and transporters and Na^+,K^+-ATPase is abundant in the basolateral membrane. The duct cells also secrete kallikrein, which is stored in the apical granules, and the cells are capable of endocytosing foreign proteins introduced into the ductal lumen.

As the ducts leave the gland lobules and enter the interlobular connective tissue, they are called excretory ducts (Fig. 1C). These ducts typically have a pseudostratified epithelium, with columnar cells that resemble the striated duct cells in morphology and function and small basal cells that presumably are undifferentiated cells capable of division and that serve to replace the columnar cells. Tuft cells, or brush cells, which probably have some type of sensory function, and dendritic (antigen-presenting) cells, which function in the immune response, also are found in the excretory ducts. Mucous goblet cells also may be present. As the excretory ducts increase in size, the epithelium may become stratified cuboidal or stratified columnar. Near the oral opening, the main excretory duct frequently is lined by a stratified squamous epithelium.

Connective Tissue, Vessels, and Nerves

The glands are surrounded by a connective tissue capsule, which extends into the gland as septa, dividing it into lobes and lobules. Finer partitions of connective tissue surround the ducts and end-pieces within the lobule. Collagen and elastic fibers, along with glycoproteins and proteoglycans typical of other connective tissues, constitute the extracellular matrix components, whereas fibroblasts, mast cells, plasma cells, macrophages, and dendritic cells, as well as occasional polymorphonuclear leukocytes and lymphocytes, constitute the cellular components. The blood vessels and nerves that supply the secretory end-pieces and ducts also are present in the connective tissue. The vessels and nerves enter at the hilus of the gland, branching to follow the excretory ducts to the lobules. Within the lobules,

the vessels break up into capillary plexuses around the striated ducts and end-pieces. The nerves, which are unmyelinated postganglionic sympathetic and parasympathetic axons that are enveloped by Schwann cells, exhibit two different types of relationships with the secretory cells. In the most common pattern, termed extraparenchymal, as the nerve bundle approaches a secretory cell, one or more axons exhibit a swelling or varicosity containing a few mitochondria and a cluster of neurotransmitter vesicles. The Schwann cell covering usually is absent at these sites, but the axon remains separated from the secretory cells by approximately 0.1–0.2 μm and by the basal laminae surrounding the nerve bundle and the secretory cell. In the other pattern, termed intraparenchymal, an axon leaves the nerve bundle, penetrates the basal lamina around the secretory cell, and makes close contact (10–20 nm) with the secretory cell. Both types of innervation are effective in stimulating secretion. In most glands, the secretory cells are innervated by both sympathetic and parasympathetic nerves. It is unknown, however, whether every cell is contacted by a nerve. Because the secretory cells are joined by gap junctions, it seems likely that the secretory stimulus spreads from cell to cell and that each secretory endpiece is a functional unit.

See Also the Following Articles

Digestion, Overview • Gastrointestinal Tract Anatomy, Overview • Salivary Glands, Physiology

Further Reading

Hand, A. R. (1990). The secretory process of salivary glands and pancreas. *In* "Ultrastructure of the Extraparietal Glands of the Digestive Tract" (A. Riva and P. M. Motta, eds.), pp. 1–17. Kluwer Academic, Boston, MA.

Hand, A. R. (2003). Salivary glands. *In* "Ten Cate's Oral Histology" (A. Nanci, ed.), 6th Ed., Mosby, St. Louis, MO, in press.

Hand, A. R., Pathmanathan, D., and Field, R. B. (1999). Morphological features of the minor salivary glands. *Arch. Oral Biol.* **44**, S3–S10.

Lantini, M. S., Proto, E., Puxeddu, P., Riva, A., and Testa Riva, F. (1990). Fine structure of excretory ducts of human salivary glands. *J. Submicrosc. Cytol. Pathol.* **22**, 465–475.

Pinkstaff, C. A. (1979). The cytology of salivary glands. *Int. Rev. Cytol.* **63**, 141–261.

Riva, A., and Riva-Testa, F. (1973). Fine structure of acinar cells of human parotid gland. *Anat. Rec.* **176**, 149–165.

Riva, A., Motta, G., and Riva-Testa, F. (1974). Ultrastructural diversity in secretory granules of human major salivary glands. *Am. J. Anat.* **139**, 293–298.

Riva, A., Tandler, B., and Testa Riva, F. (1988). Ultrastructural observations on human sublingual gland. *Am. J. Anat.* **181**, 385–392.

Riva, A., Testa-Riva, F., Del Fiacco, M., and Lantini, M. S. (1976). Fine structure and cytochemistry of the intralobular ducts of the human parotid gland. *J. Anat.* **122**, 627–640.

Riva, A., Valentino, L., Lantini, M. S., Floris, A., and Testa Riva, F. (1993). 3D-structure of cells of human salivary glands as seen by SEM. *Microsc. Res. Tech.* **26**, 5–20.

Tandler, B., and Erlandson, R. A. (1972). Ultrastructure of the human submaxillary gland. IV. Serous granules. *Am. J. Anat.* **135**, 419–433.

Young, J. A., and van Lennep, E. W. (1978). "The Morphology of Salivary Glands." Academic Press, London.

Salivary Glands, Physiology

JAMES E. MELVIN AND DAVID J. CULP
University of Rochester, New York

acetylcholine The major neurotransmitter mediating the secretion of saliva, which is initiated by the parasympathetic nervous system.
acinus The proximal grape-shaped end-piece in exocrine salivary glands composed of secretory cells that produce essentially all of the fluid and the major portion of the protein content of saliva.
mucus cells Salivary gland cells that synthesize, store, and secrete large amounts of mucins.
muscarinic receptors Receptors for acetylcholine present in salivary gland cells that act to increase intracellular Ca^{2+}.
myoepithelial cells Contractile cells that surround acinar structures in salivary glands; their contraction accelerates the expulsion of saliva.
$Na^+/K^+/2Cl^-$ cotransporter The major membrane protein mediating uptake of Cl^- across the salivary acinar cell basolateral membrane.
parotid glands The major serous salivary glands located bilaterally under the ear.
serous acinar cells Salivary acinar cells, which secrete a high volume of water and enzymes.
xerostomia The subjective feeling of dryness of the mouth, usually associated with diminished or arrested salivary secretion.

Normal salivary gland function is essential for providing lubrication during chewing and swallowing of food, for maintaining the hydration of hard and soft oral tissues, and for protecting against mechanical and bacterial insults. Correspondingly, salivary gland dysfunction is associated with an increased incidence of oral disease, which subsequently is often linked to systemic disease. Several million Americans suffer from some form of oral dryness (xerostomia). Most xerostomia is a consequence of head and neck irradiation to treat tumors, Sjögren's syndrome (an autoimmune disease), or drug therapy. Hundreds of prescribed medications interfere with normal salivary gland function, frequently acting at the receptor level. In approximately 10% of dry mouth cases, the underlying etiology is idiopathic. In these cases, xerostomia is likely related to a genetic defect in either water and ion transporter proteins or in a key signaling pathway. An initial step in the development of treatments for xerostomia requires an appreciation of salivary gland physiology.

INTRODUCTION

The secretion of saliva involves the coordinated activation of a diverse array of glandular structures, each of which delivers its own exocrine and fluid secretory constituents into the oral cavity. Three "major" pairs of salivary glands (parotid, submandibular, and sublingual) are linked to the oral cavity through relatively long excretory ducts. In addition, localized just under the oral epithelium are numerous smaller "minor" glands, named according to their anatomical locations (i.e., labial, lingual, palatal, buccal, and minor sublingual). This article will focus on highlighting the current understanding of regulatory processes mediating secretion of fluid and exocrine products from these various glands. The focus will mainly be on results from human and rodents, as the latter are the current major animal models in use.

VARIATIONS AMONG GLANDS IN INNERVATION AND THE RELATIONSHIP TO SECRETION

General Concepts of Central and Autonomic Control

Salivary glands form a network of secretory units that, depending on the stimulus or insult, react to provide an appropriate combination of secretory products. In general, secretory responses are controlled by complex interactions between the central and autonomic nervous systems. Glands receive postganglionic autonomic fibers to activate specific constituent secretory cells and/or to regulate blood flow. Preganglionic parasympathetic nerve fibers emanate from salivatory centers in the medulla oblongata (i.e., the inferior and superior salivatory nuclei) to synapse in multiple ganglia with postganglionic fibers that innervate the different salivary glands. Postganglionic fibers feeding the submandibular and sublingual glands emanate primarily from the submandibular ganglion. Parotid glands receive postganglionic fibers from the otic ganglion and minor glands are innervated via fibers from the

 Copyright 2004, Elsevier (USA). All rights reserved.

sphenopalatal ganglion or submandibular ganglion. There are no distinct sympathetic salivary centers, although preganglionic fibers emerge from the initial two thoracic segments of the spinal cord and ascend via the sympathetic trunk to synapse in the superior cervical ganglion. Postganglionic fibers follow along with arteries to supply each gland.

Regions within the central nervous system can influence the salivary centers in a positive or a negative manner. For example, the negative influences of central regions on salivary centers are responsible for a dry mouth during periods of emotional stress. Conversely, positive influences on salivary centers are likely responsible for increased salivation in response to appropriate smells or to the anticipation of feeding or vomiting. Reflex secretion in response to gustatory (taste) or mechanical (mastication and touch) stimuli appears to be controlled more by localized neural feedback loops that involve input from lingual taste receptors and periodontal mechanoreceptors, respectively. The broad distribution of diverse glands supplying the oral cavity provides for different reflex stimuli to influence the nature of the final secretion, in both protein and fluid content. Secretion can also be localized more to a general site within the oral cavity, such as in the case of a unilateral chewing stimulus and the resultant dominant parotid secretory response on the same side.

The pattern of autonomic innervation to cellular elements is not necessarily similar between glands or between species for the same gland. In general, postganglionic parasympathetic fibers are abundant in all major and minor salivary glands examined to date. In addition, acetylcholine (ACh) efferent fibers near blood vessels and ducts also contain vasoactive intestinal peptide (VIP), peptide with N-terminal histidine and C-terminal isoleucine (PHI), substance P (SP), and calcitonin gene-related peptide (CGRP). Only VIP and possibly enkephalins are co-localized with Ach in fibers around acini, although SP is also present in the rat. In addition to norepinephrine, sympathetic fibers near blood vessels contain neuropeptide Y. Sympathetic fibers to vascular, ductal, and acinar elements may also contain enkephalins.

Parotid and Submandibular Glands

In parotid glands, adrenergic and cholinergic postganglionic fibers are associated with serous acinar exocrine cells, intercalated ducts, large vascular structures, myoepithelial cells, and striated ducts. Submandibular glands have a similar pattern of innervation, despite marked species differences in acinar cell types. Human submandibular acini are primarily serous with a small proportion of mucous tubuloacini, whereas in rodents, submandibular acini are uniformly seromucous and secrete a low-molecular-weight mucin. Parasympathetic stimulation to either parotid or submandibular glands results in limited acinar cell degranulation accompanied by profuse fluid secretion. Conversely, sympathetic activation produces little fluid output and extensive acinar cell degranulation.

Mucous Glands

The major sublingual glands of mice, rats, and humans as well as the minor sublingual, buccal, and labial glands in humans have a rich supply of cholinergic nerves to acinar elements and blood vessels. These glands have a more sporadic cholinergic innervation of ducts. Adrenergic innervation is localized primarily to blood vessels and is absent or barely detectable within acini. Because all these glands contain abundant mucous tubuloacini usually capped by serous (seromucous) demilune cells, other minor mucous glands (palatal and lingual mucous glands) are likely innervated in a similar fashion. Mucous glands also undergo reflex secretion (e.g., gustatory stimulation, chewing, and speaking) as demonstrated in human labial glands. Both fluid secretion and exocrine secretion by mucous acinar cells are controlled primarily by muscarinic cholinergic receptors. Exocrine secretion by mucous cells, *in vitro*, is also responsive to VIP but to a maximal response that is less than half of the muscarinic-induced response. Unlike mucous cells, the exocrine response of demilune cells appears unresponsive to parasympathetic stimulation or muscarinic agonist but is instead stimulated by β-adrenergic agonists. Whether demilune cells contribute a fluid component to assist in the flow of viscous mucins has been speculated upon but has yet to be demonstrated. The serous lingual glands of von Ebner also undergo exocytosis in response to muscarinic agonist or parasympathetic nerve activation. Sympathetic fibers may function in a similar manner, although results are conflicting.

Myoepithelial Cells

Myoepithelial cells that surround all acinar structures (rat parotid is an exception) as well as intercalated ducts contract on stimulation of parasympathetic fibers (via muscarinic receptors) and, when present, sympathetic fibers (via α-adrenergic receptors). Contraction initially accelerates the expulsion of saliva and is then thought to provide support against increased intralumenal pressures during secretion. Gap junctions between myoepithelial and acinar cells likely function

in intracellular communications and may help to propagate a neurostimulus.

Innervation of Blood Vessels and Control of Blood Flow

There is a complex integration between signals to activate secretion and blood flow to specific glands. The volume of secretion during parasympathetic salivation is related directly to the glandular venous pressure through the opening of arteriovenous anastomoses. Simultaneous arteriolar dilation further assists to elevate capillary hydrostatic pressure. Both acetylcholine and VIP are vasodilatory and function synergistically, whereas sympathetic vasomotor fibers can counteract these effects through α-adrenergic mechanisms. Moreover, nitric oxide synthase (NOS) is localized to postganglionic parasympathetic nerves feeding vascular elements in submandibular glands and functions to enhance the release of VIP and reduce the degradation of VIP-induced cyclic AMP (cAMP) in vascular smooth muscle. In contrast, NOS is mostly absent in sympathetic postganglionic nerves. Afferent sensory fibers of unknown function are also localized to blood vessels and occasionally to ducts within all three major glands. These fibers originate in the trigeminal ganglion and display immunoreactivity to CGRP, neurokinin A, and SP. Because not all elements within a gland are thought to be activated simultaneously during normal reflex responses, perhaps afferent fibers provide feedback of blood flow and/or pressure to help regulate localized reflex secretory responses.

G-PROTEIN-COUPLED RECEPTORS AND INTRACELLULAR SIGNALING IN SALIVARY ACINAR CELLS

General Aspects

Salivary glands possess an abundance of receptor types that, in general, reflect the pattern of autonomic innervation to each gland. Two signaling pathways, calcium and cyclic AMP, and the receptors responsible for their initiation have historically been the primary focus of studies to elucidate secretory responses of salivary cells to various neurotransmitters. However, in the past decade multiple isoforms of receptors, intracellular signaling molecules, and components of fluid/exocrine secretory pathways have been identified. As the annotation of mammalian genomes progresses, many more isoforms and/or splice variants will likely be revealed. As specific isoforms of these molecules are identified in salivary cells, the task of characterizing functional signaling pathways becomes even more complicated, especially when multiple and operationally equivalent isoforms are expressed. Furthermore, it is becoming increasingly apparent that cells utilize macromolecular complexes for the temporal–spatial localization, efficiency, and integration of signaling pathways. Therefore, delineating such complexes and their functions represents a formidable quest in current and future salivary research.

Calcium Signaling

Acetylcholine from postganglionic parasympathetic nerves stimulates acinar cell muscarinic cholinergic receptors. There are five subtypes of muscarinic receptors (M1–5) and acinar cells of the major glands (probably all minor glands as well) express abundant M3 muscarinic receptors. Additionally, mucous cells of sublingual and possibly minor glands also express significant M1 muscarinic receptors (equivalent to M3 receptor levels) and recent data raise the possibility of the M5 receptor subtype in salivary glands. M1 and M3 receptors couple predominantly to G_q and G_{11} G-proteins to activate (via the G_α subunit) phosphatidylinositol 1,4-bisphosphate (PIP2)-specific phospholipase C (PLC). The β_3 isoform of PLC is present in rat parotid glands. PLC hydrolyzes plasma membrane PIP2 to release diacylglycerol and inositol 1,4,5-trisphosphate (IP_3). In rodent parotid and submandibular acinar cells, IP_3 initiates a rapid increase in the intracellular free calcium ($[Ca^{2+}]_i$) by opening IP_3-sensitive Ca^{2+} channels (260 kDa IP_3 receptor glycoproteins in tetrameric form) within components of the endoplasmic reticulum (ER) localized in the apical cytoplasm. Depletion of internal Ca^{2+} stores activates the influx of extracellular Ca^{2+} through a storage-operated Ca^{2+} entry mechanism in the plasma membrane that has yet to be defined, although evidence suggests a role for *trp* (transient receptor potential) gene products. These Ca^{2+} release and influx mechanisms function in concert with Ca^{2+} pumps to regulate $[Ca^{2+}]_i$ in a temporal–spatial manner to produce either oscillation or waves of increases in $[Ca^{2+}]_i$ in response to muscarinic receptor activation. These Ca^{2+} pumps are localized in the plasma membrane to extrude intracellular Ca^{2+} as well as SERCA (sarco-endoplasmic reticulum Ca^{2+}) pumps in the ER to refill Ca^{2+} stores. The pattern of oscillatory or wave-like changes in $[Ca^{2+}]_i$ is the result of multiple factors that are the focus of intense investigation. Such factors likely include (1) the spatial distribution of SERCA pumps (isoforms 3 and 2b); (2) expression levels and distribution of IP_3 receptor isoforms that display various sensitivities to IP_3 and are also regulated by Ca^{2+} levels in both the ER stores

and cytosol; and (3) contributions from ryanodine receptors that also function as Ca^{2+}-release channels within ER membranes and are controlled by multiple mechanisms including $[Ca^{2+}]_i$ and phosphorylation by serine/threonine kinases. All five IP_3 receptor isoforms and ryanodine type III receptors are expressed in mouse parotid glands.

Acinar cells of rodent parotid and/or submandibular glands also express α_1-adrenergic receptors ($\alpha_{1a} > \alpha_{1b}$ isoforms) and substance P receptors (two isoforms of the tachykinin NK_1 receptor) that on receptor activation produce an increase in $[Ca^{2+}]_i$ presumably through coupling to $G_{q/11}$ proteins and to downstream mechanisms similar to M3 receptors. Possible receptor-specific patterns of oscillatory or wave-like changes in $[Ca^{2+}]_i$ have yet to be defined. It must be noted that extrapolating results of rodent tachykinin receptors to human receptors may not be warranted given the low levels of substance P innervation to acinar cells in human major and minor glands. Furthermore, studies of rodent sublingual and human labial glands suggest the absence of α_1-adrenergic receptor in mucous cells.

Four different P2 nucleotide receptor subtypes ($P2X_7$, $P2X_4$, $P2Y_1$, and $P2Y_2$) are present in salivary glands. $P2X_7$ and $P2X_4$ receptors are ATP-gated non-selective cation channels that increase $[Ca^{2+}]_i$ when activated. $P2X_7$ receptors can also form pores for molecules up to 900 Da. $P2Y_1$ receptors (ADP-selective) and $P2Y_2$ receptors (UDP-selective) couple to $G_{q/11}$ proteins to increase $[Ca^{2+}]_i$ and are expressed at very low levels in normal glandular tissues. Interestingly, glandular $P2Y_2$ receptors are up-regulated in response to injury, suggesting a role in cell renewal.

Signaling via Cyclic AMP

Acinar cells of parotid and submandibular glands express β-adrenergic receptors (β_1 and β_2 isoforms) coupled via G_s-proteins to activate adenylyl cyclase, increasing intracellular cAMP and subsequently stimulating cAMP-dependent protein kinase A (PKA). In contrast, only serous demilune cells, not mucous cells of sublingual glands, express β-adrenergic receptors. Multiple isoforms of adenylyl cyclase (3, 5/6, and 8) and PKA regulatory subunits (isoforms I and II) are present in rodent parotid and submandibular glands. Although PKA type II is predominantly expressed, both isoforms are associated with identical catalytic subunits and are activated on β-adrenergic receptor stimulation. Acinar cells of all three major glands express VIP receptors also linked to G_s-proteins to activate adenylyl cyclase. Adrenergic α_{2A} receptors are present both pre- and postsynaptically in rodent major salivary glands. These receptors are coupled to G_i G-proteins and may either inhibit adenylyl cyclase or activate PLC, although their function in salivary glands has not been extensively explored.

ACINAR CELL FLUID SECRETION

Fluid and Electrolyte Secretion Mechanisms

Fluid secretion is driven by transepithelial Cl^- movement (Fig. 1). Salivary gland acinar cells secrete an isotonic, plasma-like fluid. This primary fluid is subsequently modified in a gland-specific manner as it passes through the duct system (e.g., NaCl reabsorption, K^+ secretion, additional protein secretion). The opening of K^+ and Cl^- channels in the basolateral and apical membranes, respectively, of acinar cells initiates the secretion process. Simultaneous activation of these two types of channels permits Cl^- to exit across the lumenal membrane and loss of K^+ into the interstitial fluid, creating a lumen negative transepithelial potential difference. This transepithelial electrical potential difference leads to passive Na^+ passage across tight junctions. The resulting lumenal NaCl accumulation and transepithelial osmotic gradient drive the movement of water, creating a plasma-like primary secretion. Water channels expressed in acinar cells are important for transcellular water transport and the generation of saliva.

Sustained Cl^--dependent fluid secretion requires a robust Cl^- reuptake mechanism. The primary Cl^- uptake pathway is the $Na^+/K^+/2Cl^-$ cotransporter located in the basolateral membrane. This mechanism concentrates intracellular Cl^- four to five times above its electrochemical gradient and is up-regulated many fold during sustained stimulation. Paired Cl^-/HCO_3^- and Na^+/H^+ exchangers are also located in the basolateral membrane of acinar cells and significantly contribute to saliva formation. The energy required for Cl^- uptake via the cotransporter and the paired exchangers is stored in the inward-directed Na^+ chemical gradient created by Na^+ pumps.

Muscarinic Cholinergic Stimulation of Fluid Secretion

Parasympathetic stimulation of fluid secretion has been studied mostly in parotid acinar cells. The dominant mechanism producing fluid secretion is the muscarinic-induced increase in acinar $[Ca^{2+}]_i$ with subsequent activation of Ca^{2+}-gated K^+ and Cl^- channels. The fluid secretion model shown in Fig. 1 places the Ca^{2+}-dependent K^+ and Cl^- channels in the

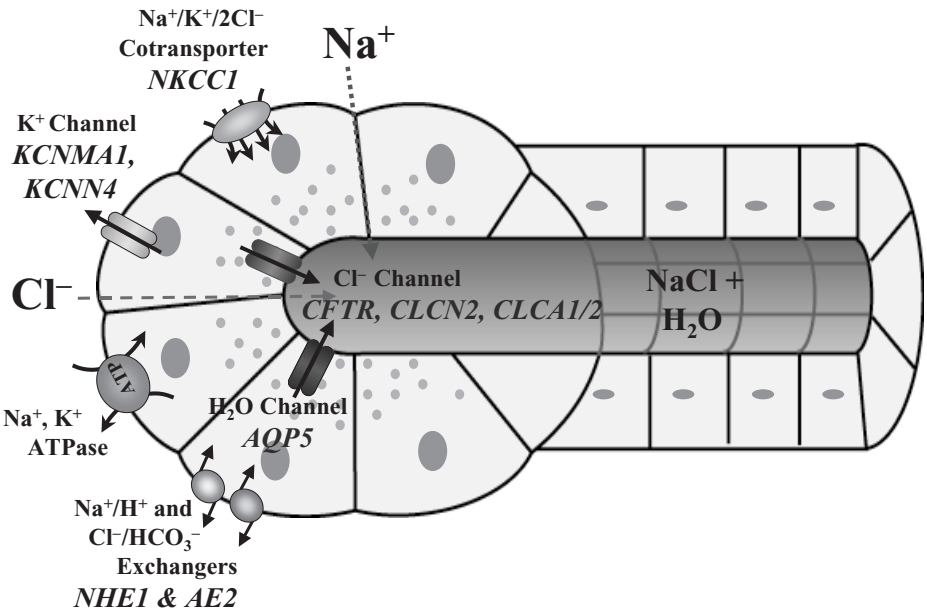


FIGURE 1 Fluid secretion model. Ion transport pathways in a Cl^--secreting acinar cell. Transepithelial Cl^- movement drives the fluid and electrolyte secretion process. This cell shows the seven essential water and ion transport mechanisms involved in fluid and electrolyte movement in Cl^--secreting epithelia: the basolateral Na^+,K^+-ATPase with a stoichiometry of 3 Na^+ : 2 K^+; the electroneutral $Na^+/K^+/2Cl^-$ cotransporter; the basolateral K^+ channel; paired basolateral Na^+/H^+ and Cl^-/HCO_3^- exchangers; the apical water channel; and apical Cl^- channels. Cl^- is concentrated in acinar cells by the $Na^+/K^+/2Cl^-$ cotransporter and paired Na^+/H^+ and Cl^-/HCO_3^- exchangers. K^+ and Cl^- exit when the K^+ and Cl^- channel open in response to an increase in the intracellular $[Ca^{2+}]$. The accumulation of Cl^- in the acinar lumen is neutralized by Na^+ movement across tight junctions and water follows osmotically. See text for details.

basolateral and apical membranes of acinar cells, respectively. A sustained increase in $[Ca^{2+}]_i$, which is dependent on external Ca^{2+}, is required to generate a prolonged secretion. Up-regulation of the $Na^+/K^+/2Cl^-$ cotransporter and the Na^+/H^+ exchanger by the increase in $[Ca^{2+}]_i$ ensures that the intracellular Cl^- concentration remains above its electrochemical gradient.

Cl^- efflux mediated by the Ca^{2+}-activated Cl^- channel is likely the rate-limiting step in the fluid secretion process. Thus, regulation of Ca^{2+}-dependent Cl^- channel activity is the key element in the determination of flow rate. An important mechanism regulating this channel is its sensitivity to the intracellular pH. Following muscarinic activation, the intracellular pH of acinar cells typically drops due to HCO_3^- efflux via Ca^{2+}-activated Cl^- channels. Channel activation is regulated by pH such that as the intracellular pH decreases, the channel is inhibited. Consequently, HCO_3^- efflux via the Ca^{2+}-activated Cl^- channel is blunted and thus the magnitude of the resulting intracellular pH drop is decreased. The importance of the pH sensitivity of the Ca^{2+}-activated Cl^- channels is in sustaining fluid secretion during prolonged stimulation. Activation of Na^+/H^+ exchange raises the intracellular pH of acinar cells 0.1–0.3 units higher than the pH in unstimulated cells. As the intracellular pH rises, the Ca^{2+} sensitivity of the Cl^- channel increases (and consequently the channel activity increases), evoking continued fluid and electrolyte movement even as the cytosolic Ca^{2+} concentration decreases to near resting levels.

Molecular Identity of Ion and Water Transport Pathways

Fluid secretion is dependent on the activation of Ca^{2+}-gated Cl^- channels. Although four different classes of Cl^- channels have been identified in salivary acinar cells, only one of these is activated by an increase in the intracellular free Ca^{2+} concentration. The molecular identity of the Ca^{2+}-activated Cl^- channel is unknown; however, *CLCA1* and *CLCA2*, members of a putative Ca^{2+}-gated Cl^- channel gene family, are expressed in salivary glands. Activation of the other three Cl^- channels is dependent on an increase in cAMP,

hyperpolarization of the plasma membrane, or cell swelling. The cAMP-dependent and the hyperpolarization-activated Cl^- channels are due to the expression of *CFTR* and *CLCN2*, respectively, but the identity of the cell swelling-dependent channel remains unclear. Mutations in functional domains of the *CFTR* gene result in cystic fibrosis, whereas targeted disruption of the *Clcn2* gene leads to degeneration of the retina and testis.

At least two types of Ca^{2+}-dependent K^+ channels are present in acinar cells, a large-conductance voltage- and Ca^{2+}-dependent K^+ channel (125–250 pS), the so-called maxi-K^+ channel, and an intermediate conductance Ca^{2+}-dependent K^+ channel (22–35 pS). Evidence suggests that the large-conductance voltage- and Ca^{2+}-dependent K^+ channel is encoded by *KCNMA1*, whereas the intermediate-conductance Ca^{2+}-dependent K^+ channel is *KCNN4*. The two Cl^- uptake pathways, the $Na^+/K^+/2Cl^-$ cotransporter and the paired Cl^-/HCO_3^- and Na^+/H^+ exchangers, are encoded by the *NKCC1* (*SLC12A2*), *NHE1* (*SLC9A1*), and *AE2* (*SLC4A2*) genes, respectively. Although several different water channel genes are expressed in salivary glands, only the aquaporin *AQP5* has been shown to play a major role in salivation.

SECRETORY GRANULE DISCHARGE

Pattern of Innervation versus Signaling of Exocrine Secretion

As described above, both parotid and submandibular acinar elements receive innervation from both branches of the autonomic nervous system. Sympathetic innervation through the activation of β-adrenergic receptors potently stimulates exocrine secretion in these glands mediated by an increase in cAMP and the subsequent activation of PKA. Signaling components downstream of PKA that may function directly in exocrine secretion have been difficult to discern, although a 26 kDa protein phosphorylated by PKA in response to isoproterenol, and having phosphorylation–dephosphorylation kinetics consistent with exocytosis, has been identified in rat parotid and submandibular acinar cells.

Parasympathetic nerves have a much less direct effect on exocrine secretion in these glands. Each of the cotransmitters VIP and acetylcholine may have a positive effect on exocrine secretion, mediated by VIP activation of the cAMP/PKA pathway and acetylcholine stimulation of muscarinic M3 receptors to activate protein kinase C (PKC). PKC consists of three families of isoforms, conventional (α, β_I, β_{II}, γ), novel (δ, ϵ, η, θ), and atypical (ζ, λ/τ), that differ in expression, subcellular localization, substrate specificity, and activation mechanisms. In response to muscarinic agonist, acinar cells from rat parotid glands demonstrate activation of both PKCα and PKCδ, whereas submandibular acinar cells undergo stimulation of PKCϵ but not PKCα. It is still unclear which PKC isoform(s) may be coupled to the exocrine pathways in these cells.

Contrary to parotid and submandibular glands, exocrine secretion as well as fluid secretion by sublingual mucous acinar cells is primarily under the control of muscarinic receptors with activation of both M1 and M3 subtypes required for a maximal exocrine response. The muscarinic exocrine response is totally dependent on activation of a Ca^{2+}-dependent PKC isoform(s), of which PKCα is the likely candidate based on expression levels. The function of muscarinic receptor redundancy is unclear. Because minor mucous glands also have a dominant parasympathetic innervation, it is reasonable to speculate that muscarinic receptors also function in a similar manner to stimulate both exocrine and fluid secretion. VIP can also activate exocrine secretion by sublingual mucous cells, but through activation of PKA and to a much lower maximal effect than the muscarinic pathway.

Pathways for the Sorting and Release of Secretory Proteins

The major regulated pathway for exocrine secretion of salivary proteins/glycoproteins involves the synthesis and transport of proteins/glycoproteins through the ER and Golgi, the budding of condensing vacuoles from the *trans*-Golgi to form immature secretory granules, maturation to secretory granules, and the exocytotic release of granule contents in response to agonist. This pathway functions primarily in the extensive secretion of secretory material during periods of eating. Three additional pathways for the release of secretory proteins have been identified in rat parotid serous acinar cells. Two of these pathways are independent of agonist stimulation; one pathway is associated with the unstimulated exocytosis of mature secretory granules and is apparent after cells are replete with granules. The second unstimulated pathway, the constitutive-like pathway, and a minor regulated pathway are both initiated at a mutual step. This step involves secretory proteins that are sorted inefficiently and enter vesicles that then bud off from both condensing vacuoles and immature secretory granules. Vesicles destined for the constitutive-like vesicles then undergo passage to the apical membrane through recycling endosomes. These endosomes also contain secretory proteins internalized by endocytosis of contents released previously in the apical lumen. During a

4 to 6 h period, from 10 to 15% of newly synthesized secretory proteins are released via the two unstimulated pathways.

Vesicles destined for the minor regulated pathway do not pass through recycling endosomes but instead are sequestered in the apical cytoplasm until induced to release their contents via exocytosis at the apical membrane. This pathway represents a small but significant pool of secretory proteins (up to 10% of newly synthesized amylase) and is responsive to very low doses of either β-adrenergic ($\leq$5 n*M* isoproterenol) or muscarinic (40 n*M* carbachol) agonist. In contrast, induction of the major regulated pathway requires higher concentrations of isoproterenol ($\geq$1 μ*M* isoproterenol). The minor regulated pathway may thus contribute more significantly to basal or resting secretions during periods between meals when the parotid receives only low-frequency stimulation from parasympathetic and possibly sympathetic nerves.

Elucidation of distinct pathways for the release of secretory proteins adds additional levels of complexity in efforts to define mechanisms responsible for the differential sorting of vesicles as well as for the control of exocytosis by Ca^{2+}- and cAMP-mediated signaling. In other exocrine systems, especially with excitable cells, proteins termed SNARE (soluble *N*-ethylmaleimide-sensitive attachment protein receptor) have been identified and function in a reciprocal recognition between v-SNAREs (vesicular membrane-associated proteins) and t-SNAREs (plasma membrane-associated proteins). Studies with rat parotid acinar cells suggest a role for the v-SNARE protein VAMP2 (vesicle-associated membrane protein 2) in cAMP-mediated exocrine secretion. Also implicated in trafficking of parotid vesicles are the small GTP-binding proteins Rab3, Rap1, and ARF1 (ADP-ribosylation factor 1) as well as SCAMP1 (secretory carrier membrane protein 1). Moreover, the recent demonstrated expression in parotid glands of additional SNAREs as well as associated regulatory proteins known to function in membrane fusion events further manifests the complexity of defining the key downstream events regulating salivary exocrine secretion.

INTERACTIONS BETWEEN SECRETORY STIMULI

Crosstalk between Signaling Pathways

Glandular elements normally receive multiple signals, either from co-localized neurotransmitters or from the simultaneous firing of both sympathetic and parasympathetic fibers (at least in parotid and submandibular glands). In general, synergistic fluid and exocrine effects are observed in response to mixed low-level nerve stimulation in parotid and submandibular glands. Mechanisms of synergistic interactions may involve contributions from myoepithelial cells, augmentation of calcium or cAMP signaling by different receptors, or convergence of two distinct pathways at a common secretory mechanism. For example, norepinephrine from sympathetic nerves is likely to stimulate a maximal exocrine response from serous cells as both the cAMP and Ca^{2+} pathways are activated via β- and α_1-adrenergic receptors, respectively. Moreover, α_1-adenergic receptors mediate the small sympathetic-derived fluid component of norepinephrine-induced secretion. In mouse parotid acini, the observed synergism between the muscarinic and β-adrenergic pathways in fluid and exocrine secretion is likely related to (1) Ca^{2+} activation of calmodulin (CaM) with subsequent CaM stimulation of the type 8 isoform of adenylyl cyclase and (2) cAMP potentiation of Ca^{2+} release from intracellular stores due to PKA-mediated phosphorylation of IP_3 receptors. In addition, a functional Ca^{2+}–nitric oxide–cGMP pathway in acinar cells may also serve to integrate Ca^{2+} and cAMP levels through actions to promote or amplify increased $[Ca^{2+}]_i$ and inhibit adenylyl cyclase isoforms.

See Also the Following Articles

Autonomic Innervation • Salivary Glands, Anatomy and Histology • Sjögrens Syndrome • Substance P • Taste and Smell • Vasoactive Intestinal Peptide (VIP) • Xerostomia

Further Reading

Ahn, J. S., Camden, J. M., Schrader, A. M., Redman, R. S., and Turner, J. T. (2000). Reversible regulation of $P2Y_2$ nucleotide receptor expression in the duct-ligated rat submandibular gland. *Am. J. Physiol. Cell Physiol.* **279**, C286–C294.

Ambudkar, I. S. (2000). Regulation of calcium in salivary gland secretion. *Crit. Rev. Oral Biol. Med.* **11**, 4–25.

Begenisich, T., and Melvin, J. E. (1998). Regulation of chloride channels in secretory epithelia. *J. Membr. Biol.* **163**, 77–85.

Bockman, C. S., Bradley, M. E., Dang, H. K., Zeng, W., Scofield, M. A., and Dowd, F. J. (2001). Molecular and pharmacological characterization of muscarinic receptor subtypes in a rat parotid gland cell line: Comparison with native parotid gland. *J. Pharmacol. Exp. Ther.* **297**, 718–726.

Bruce, J. I., Shuttleworth, T. J., Giovannucci, D. R., and Yule, D. I. (2002). Phosphorylation of inositol 1,4,5-trisphosphate receptors in parotid acinar cells. *J. Biol. Chem.* **277**, 1340–1348.

Cook, D. I., Van Lennep, E. W., Roberts, M. L., and Young, J. A. (1994). Secretion by the major salivary glands. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., pp. 1061–1117. Raven Press, New York.

Culp, D. J., Luo, W., Richardson, L. A., Watson, G. E., and Latchney, L. R. (1996). Both M1 and M3 receptors regulate exocrine secretion by mucous acini. *Am. J. Physiol.* **271**, C1963–C1972.

Dowd, F. J. (1999). Saliva and dental caries. *Dent. Clin. North Am.* **43**, 579–597.

Garrett, J. R. (1987). Innervation of salivary glands: Neurohistological and functional aspects. *In* "The Salivary System" (L. M. Sreebny, ed.), pp. 69–93. CRC Press, Boca Raton, FL.

Huang, A. Y., Castle, A. M., Hinton, B. T., and Castle, J. D. (2001). Resting (basal) secretion of proteins is provided by the minor regulated and constitutive-like pathways and not granule exocytosis in parotid acinar cells. *J. Biol. Chem.* **276**, 22296–22306.

Lundberg, J. M., Martling, C.-R., and Hökfelt, T. (1988). Airways, oral cavity and salivary glands: Classical transmitters and peptides in sensory and autonomic motor neurons. *In* "Handbook of Chemical Neuroanatomy, Vol. 6, The Peripheral Nervous System" (A. Björklund, T. Hökfelt, and C. Owman, eds.), pp. 391–444. Elsevier Science, Amsterdam.

Melvin, J. E. (1991). Saliva and dental diseases. *Curr. Opin. Dent.* **1**, 795–801.

Nauntofte, B. (1992). Regulation of electrolyte and fluid secretion in salivary acinar cells. *Am. J. Physiol.* **263**, G823–G837.

Quissell, D. O., Watson, E., and Dowd, F. J. (1992). Signal transduction mechanisms involved in salivary gland regulated exocytosis. *Crit. Rev. Oral Biol. Med.* **3**, 83–107.

Turner, R. J., Paulais, M., Manganel, M., Lee, S. I., Moran, A., and Melvin, J. E. (1993). Ion and water transport mechanisms in salivary glands. *Crit. Rev. Oral Biol. Med.* **4**, 399–406.

Salmonella

ANDREW S. NEISH
Emory University

nuclear factor κB Conserved signal transduction pathway critical for the activation of innate immune and inflammatory responses.

Toll-like receptors Transmembrane surface receptors involved in the detection and recognition of pathogen structural determinants.

The *Salmonella*, a large group of common bacterial enteric pathogens, cause a spectrum of food-borne and waterborne diseases of worldwide importance. Generally, clinical syndromes are divided into nontyphoidal salmonellosis and typhoid fever. Nontyphoidal salmonellosis is an acute inflammatory gastroenteritis of great public health importance in industrialized countries and is caused by many *Salmonella* strains. Typhoid fever, a more severe systemic waterborne disease endemic in developing nations, is caused by a single serovar, *Salmonella typhi*.

MICROBIOLOGY

The *Salmonella* are gram-negative, facultative anaerobic, noncapsulated (except *S. typhi/paratyphi*), nonspore-forming, motile rods. They possess a 4.8 Mb genome, variably modified by multiple lysogenic phage genomes, plasmids, and other mobile genetic elements. Interestingly, the *S. typhi* genome contains numerous inactive pseudogenes, which are functional in enteropathogenic *Salmonella* strains. *Salmonella* taxonomy is complex and in a constant state of revision. All *Salmonella* are members of the family Enterobacteriaceae and are closely related to other medically important enteric bacteria including *Escherichia coli, Yersinia* sp., and *Shigella* sp.

By DNA sequence analysis, six subgroups of "*Salmonella enterica*" are recognized. Almost all human (and other warm-blooded animal) pathogens, including *S. typhi*, are in Group 1. Group 1 is further categorized into more than 1000 serovars based on the antigenicity of surface oligosaccharides (O-antigens) and flagellar structures (H-antigens). Roughly 10 Group 1 serovars account for over 70% of human infections. Other subgroups of *Salmonella* are found primarily in cold-blooded animals and the environment, though several serovars of Group 3 can occasionally be pathogenic in humans. In the medical and epidemiological literature, isolates are generally referred to by serovar elevated to species name.

All *Salmonella* are adapted to the intestinal tracts of vertebrates. Some can colonize/infect a broad range of vertebrate hosts; alternatively, strains may have absolute host specificity. *S. typhi*, for example, is strictly host limited to humans. Certain *Salmonella* strains can

 Copyright 2004, Elsevier (USA). All rights reserved.

Garrett, J. R. (1987). Innervation of salivary glands: Neurohistological and functional aspects. *In* "The Salivary System" (L. M. Sreebny, ed.), pp. 69–93. CRC Press, Boca Raton, FL.

Huang, A. Y., Castle, A. M., Hinton, B. T., and Castle, J. D. (2001). Resting (basal) secretion of proteins is provided by the minor regulated and constitutive-like pathways and not granule exocytosis in parotid acinar cells. *J. Biol. Chem.* **276**, 22296–22306.

Lundberg, J. M., Martling, C.-R., and Hökfelt, T. (1988). Airways, oral cavity and salivary glands: Classical transmitters and peptides in sensory and autonomic motor neurons. *In* "Handbook of Chemical Neuroanatomy, Vol. 6, The Peripheral Nervous System" (A. Björklund, T. Hökfelt, and C. Owman, eds.), pp. 391–444. Elsevier Science, Amsterdam.

Melvin, J. E. (1991). Saliva and dental diseases. *Curr. Opin. Dent.* **1**, 795–801.

Nauntofte, B. (1992). Regulation of electrolyte and fluid secretion in salivary acinar cells. *Am. J. Physiol.* **263**, G823–G837.

Quissell, D. O., Watson, E., and Dowd, F. J. (1992). Signal transduction mechanisms involved in salivary gland regulated exocytosis. *Crit. Rev. Oral Biol. Med.* **3**, 83–107.

Turner, R. J., Paulais, M., Manganel, M., Lee, S. I., Moran, A., and Melvin, J. E. (1993). Ion and water transport mechanisms in salivary glands. *Crit. Rev. Oral Biol. Med.* **4**, 399–406.

Salmonella

ANDREW S. NEISH
Emory University

nuclear factor κB Conserved signal transduction pathway critical for the activation of innate immune and inflammatory responses.

Toll-like receptors Transmembrane surface receptors involved in the detection and recognition of pathogen structural determinants.

The *Salmonella*, a large group of common bacterial enteric pathogens, cause a spectrum of food-borne and waterborne diseases of worldwide importance. Generally, clinical syndromes are divided into nontyphoidal salmonellosis and typhoid fever. Nontyphoidal salmonellosis is an acute inflammatory gastroenteritis of great public health importance in industrialized countries and is caused by many *Salmonella* strains. Typhoid fever, a more severe systemic waterborne disease endemic in developing nations, is caused by a single serovar, *Salmonella typhi*.

MICROBIOLOGY

The *Salmonella* are gram-negative, facultative anaerobic, noncapsulated (except *S. typhi/paratyphi*), nonspore-forming, motile rods. They possess a 4.8 Mb genome, variably modified by multiple lysogenic phage genomes, plasmids, and other mobile genetic elements. Interestingly, the *S. typhi* genome contains numerous inactive pseudogenes, which are functional in enteropathogenic *Salmonella* strains. *Salmonella* taxonomy is complex and in a constant state of revision. All *Salmonella* are members of the family Enterobacteriaceae and are closely related to other medically important enteric bacteria including *Escherichia coli, Yersinia* sp., and *Shigella* sp.

By DNA sequence analysis, six subgroups of "*Salmonella enterica*" are recognized. Almost all human (and other warm-blooded animal) pathogens, including *S. typhi*, are in Group 1. Group 1 is further categorized into more than 1000 serovars based on the antigenicity of surface oligosaccharides (O-antigens) and flagellar structures (H-antigens). Roughly 10 Group 1 serovars account for over 70% of human infections. Other subgroups of *Salmonella* are found primarily in cold-blooded animals and the environment, though several serovars of Group 3 can occasionally be pathogenic in humans. In the medical and epidemiological literature, isolates are generally referred to by serovar elevated to species name.

All *Salmonella* are adapted to the intestinal tracts of vertebrates. Some can colonize/infect a broad range of vertebrate hosts; alternatively, strains may have absolute host specificity. *S. typhi*, for example, is strictly host limited to humans. Certain *Salmonella* strains can

 Copyright 2004, Elsevier (USA). All rights reserved.

be effective intestinal commensals in birds and reptiles and can also colonize reproductive organs. This ability allows infection of developing eggs and subsequent vertical transmission, which accounts for human infection via the poultry industry and pet trade. *Salmonella* are hardy organisms and can survive under a variety of environmental conditions outside a vertebrate host. For instance, infective organisms can survive in soil for months and in dried food products for years.

CLINICAL SYNDROMES

Nontyphoidal salmonellosis is caused by many *Salmonella* serovars and is generally manifested as an acute inflammatory gastroenteritis. Symptoms usually begin after an incubation period of 1–4 days with initial nausea and vomiting followed by cramping abdominal pain, diarrhea, headache, and fever. The illness generally lasts 3–7 days and is usually self-limited. Mortality is rare and is usually confined to elderly and immunocompromised patients, such as individuals with acquired immunodeficiency syndrome or neoplastic disease. The diarrhea is usually of moderate volume and without blood. Frankly bloody and purulent "dysentery-like" stools can be seen, though such symptoms are more typical of *Shigella* or enterohemorrhagic *E. coli* infections. In addition, voluminous watery "cholera-like" stools occur occasionally in nontyphoidal salmonellosis. Rarely, an enteric fever-like clinical picture can result from infection with nontyphoidal *Salmonella*.

The diagnosis of nontyphoidal salmonellosis should be considered in any enterocolitis, especially when associated with fever and headache. Definitive diagnosis is achieved by stool culture on selective medium to rule out other enteropathogens, such as *Campylobacter* sp., *Shigella, Yersinia*, or pathogenic *E. coli*, and to differentiate from acute idiopathic ulcerative colitis or toxigenic secretory diarrhea.

Complications include bacteremia and subsequent localized extraintestinal infections, e.g., arthritis, meningitis, osteomyelitis (especially in patients with sickle cell anemia), and endocarditis. Such bacteremic complications are often associated with certain serovars (e.g., *S. choleraesuis* and *S. dublin*) or in the clinical setting of immunodeficiency or hemolytic disorders. Autoimmune sequelae such as Reiter syndrome (joint pain, uveitis, and urethritis) can follow *Salmonella* (and other enteric) infections, generally in HLA-B27-positive males.

Pathological change is limited to the colon and ileum and is marked by mucosal erosions with a variable inflammatory infiltrate in the epithelia and lamina propria. In acute infections, neutrophilic infiltration of the epithelia and lumen is characteristic. The histopathologic picture may be mistaken for acute ulcerative colitis.

Typhoid or enteric fever is caused by *S. typhi* and is a severe systemic febrile infection that involves the reticuloendothelial system. Paratyphoid fever is a similar slightly milder syndrome caused by the closely related *S. paratyphi*. After oral inoculation and an incubation period of 8–14 days, patients present with variable gastrointestinal symptoms (generally milder than those in nontyphoidal salmonellosis), prolonged high fever, constipation, headache, and incapacitating malaise. A characteristic maculopapular rash (rose spots) may be seen and hepatosplenomegaly is common. Untreated, the clinical course is 3–8 weeks, with convalescence (and possible relapse) often extending much longer, resulting in mortality from chronic inanition. Overall, the mortality rate of untreated infection is 10–20%.

Definitive diagnosis is made by culture (the highest yield is obtained from bone marrow; otherwise blood, duodenal fluid, or skin is taken). Pathology is characterized by expansion of the macrophage-like cells in reticuloendothelial and lymphoid tissues, accounting for the hepatosplenomegaly. Massive enlargement, necrosis, and rupture of mucosal Peyer's patches can result in gut perforation, a feared often-lethal complication. As in nontyphoidal salmonellosis, disseminated infection can occur.

EPIDEMIOLOGY

Nontyphoidal salmonellosis is rising in incidence in industrialized nations, with 2–4 million documented cases per year and far more going unreported. The predominant (90%) route of infection is food-borne, especially through beef and poultry products. Person-to-person transmission occurs, such as in day-care centers, but is less common. Infections peak in the summer and fall months and most cases are sporadic, though recognized outbreaks are well known and can often be traced to a common source. Although all ages can be affected, young children and the elderly are diagnosed far more frequently. Immunization is not effective in nontyphoidal salmonellosis and individuals can suffer repeated infections.

Typhoid fever is uncommon in industrialized nations (several hundred cases annually in the United States since the mid-1960s, largely confined to travelers and laboratory workers) but remains a serious issue in the developing world. The World Health Organization

estimates 33 million cases annually worldwide, with >500,000 deaths. As humans are the only host of this organism, the disease is spread in an indirect person-to-person fashion, generally through fecally contaminated water supplies, and thus is endemic in areas with limited sanitation. In developed nations with efficient water purification and sewage disposal capabilities, typhoid fever is rare. Endemic areas tend to be large impoverished urban areas in Central America, South America, South Asia, and Southeast Asia.

In striking contrast to *Salmonella* enteritis, typhoid fever predominantly strikes older children and young adults, with fewer infections in the very young and old. Interestingly, pretechnical societies have a low incidence of infection, probably due to universal exposure during the relatively protected period of infancy, resulting in population immunity.

Typhoid fever confers immunity to subsequent attacks; however, *S. typhi* has a predilection for colonizing the gallbladder and can remain in high numbers even after the patient has clinically recovered. Such chronic carrier individuals (2–5% of cases) continually shed large numbers of organisms into the bile and intestinal tract, presenting a public health challenge. The case of "Typhoid Mary," a 19th century New York cook deemed responsible for over 3000 infections, vividly illustrates the potential for person-to-person transmission from chronic carriers.

PATHOGENESIS

All *Salmonella* are enteric pathogens and are transmitted orally via contaminated food and water. Thus, all enteropathogens must survive transit through the acidic gastric environment to gain access to the distal ileum and colon, affix themselves to the lumenal wall to avoid peristaltic elimination, and engage the host at the intestinal epithelium.

Salmonella are invasive organisms; events leading to disease are elicited by the actual penetration of the epithelial barrier by living organisms. This is in contrast to other enteridites mediated by the action of extracellular secreted toxins, such as in cholera or staphylococcal-mediated food poisoning. It is generally accepted that the process of invasion is mediated by bacterial "effector" proteins injected into the cytoplasm of the host cell by means of a specialized secretory structure (type III secretion system, or TTSS). The TTSS is found in many gram-negative pathogens and its component structural and regulatory genes tend to be physically clustered on "pathogenicity islands" (PAIs). Effector proteins may be encoded in pathogenicity islands either contiguous with TTSS proteins or elsewhere on the chromosome. Effector proteins are thought to usurp host cellular processes to facilitate the bacterial life cycle. The PAI of *Salmonella* shows significant homology with the PAIs of other enteropathogens, in both gene order and gene sequence, suggesting that these determinants of virulence have been disseminated among enteric organisms by horizontal gene transfer. The presence of virulence factors on other mobile genetic elements, such as temperate phage DNA and plasmids, further supports the idea of horizontal transmission of genes involved in pathogenicity.

Translocated effector proteins can induce phagocytosis into intestinal epithelial cells, allowing penetration of the epithelial barrier, occupation of an intracytoplasmic vacuole, and egress through the basolateral aspect of the epithelium into the lamina propria. In addition, lumenal *Salmonella* may be taken up by M cells, which are modified epithelial cells overlying lymphoid tissue that are specialized to sample lumenal particulate matter, and thus gain access to the lamina propria in this manner.

In nontyphoidal salmonellosis, the presence of invading bacteria is detected by the innate immune system, most likely via perception of bacterial surface structures by Toll-like receptors. The subsequent elicitation of the nuclear factor κB and other pro-inflammatory cellular signaling pathways results in activation of a classical acute inflammatory response typified by an intense neutrophilic infiltrate present in the mucosa, submucosa, and lumen. This cellular infiltrate and the resultant increased epithelial permeability directly contribute to the clinical manifestations of the disorder, namely, inflammatory and secretory diarrhea. *Salmonella* are rapidly killed by neutrophils. Thus, the brisk inflammatory response, despite the clinical symptoms it induces, localizes the infection and arrests systemic dissemination.

In typhoid fever, *S. typhi* and *S. paratyphi* are able to adhere, invade, and penetrate the mucosa without eliciting a significant inflammatory response. This property may be mediated by the polysaccharide capsule specific to these serovars. Organisms are phagocytosed by recruited macrophages and ultimately reach and proliferate within the monocytic cells of the reticuloendothelial system, with resultant enlargement of lymph nodes and spleen. These cells are apparently unable to kill *S. typhi*, resulting in protection of the pathogen from host innate and adaptive immunity. The prolonged fever and malaise associated with typhoid fever are likely a reaction to circulating bacterial products and endogenous cytokines produced

chronically from the massive persistent systemic infection.

TREATMENT AND PREVENTION

Nontyphoidal salmonellosis is a self-limited disease in immunocompetent adults. Antimicrobial administration does not reduce symptom severity or duration; indeed, antibiotics prolong asymptomatic passage of organisms during convalescence, presumably due to effects on normal flora that otherwise suppress growth of pathogens. In cases of *Salmonella* bacteremia or extraintestinal infection, quinolones, trimethoprim-sulfamethoxazole, and amoxicillin are generally indicated. Passive immunization against intestinal salmonellosis is not available.

The most effective control mechanisms are directed toward the source of infections. Improved methods of animal husbandry, meat processing, and storage are effective in reducing outbreaks. Hand-washing among workers in day care and other institutional settings and education of individuals about proper food-handling practices and preparation techniques at the commercial and household level are also important.

Typhoid fever generally responds to chloramphenicol, quinolones, and trimethoprim-sulfamethoxazole, though the emergence of multidrug-resistant strains is an ominous problem. In the case of chronic carriers, *S. typhi* can be successfully eradicated by prolonged antimicrobial therapy, underscoring the importance of identifying and treating these individuals. As typhoid is waterborne, efforts in developing countries to guard the purity of water sources and safely process sewage are considered primary methods of control. Reasonably effective vaccines are available and are often used by travelers to endemic areas.

With all forms of *Salmonella*-mediated diseases, domestic and international epidemiological surveillance is necessary to detect outbreaks and identify carriers. Unfortunately, the severity of typhoid fever and environmental hardiness of *S. typhi* could lead to deliberate contamination of water supplies, further emphasizing the need to monitor this disease.

See Also the Following Articles

Cholera • Foodborne Diseases • Food Poisoning • Food Safety • Gastroenteritis • *Shigella* • *Yersinia*

Further Reading

Hornick R. (1994). Nontyphoidal salmonelloses. *In* "Infectious Diseases" (P. Hoeprich, M. Jordan, and A. Ronald, eds.), 5th Ed., pp. 741–747. Lippincott, Philadelphia, PA.

Hornick R. (1994). Typhoid fever. *In* "Infectious Diseases" (P. Hoeprich, M. Jordan, and A. Ronald, eds.), 5th Ed., pp. 747–753. Lippincott, Philadelphia, PA.

Levine, M. (1998). Typhoid fever. *In* "Bacterial Infections of Humans: Epidemiology and Control" (A. Evans and P. Brachman, eds.), 3rd Ed., pp. 839–858. Plenum, New York.

Pegues, D., Ohl, M., and Miller, S. (2002). *Salmonella*, including *Salmonella typhi*. *In* "Infections of the Gastrointestinal Tract" (M. Blaser, P. Smith, J. Ravdin, H. Greenberg, and R. Guerrant, eds.), 2nd Ed., pp. 669–697. Lippincott Williams & Wilkins, Philadelphia, PA.

Tauxe, R., and Pavia, A. (1998). Salmonellosis: Nontyphoidal. *In* "Bacterial Infections of Humans: Epidemiology and Control" (A. Evans and P. Brachman, eds.), 3rd Ed., pp. 613–630. Plenum, New York.

Satiety

MICHAEL M. MEGUID*, ALESSANDRO LAVIANO†, AND AKIO INUI‡

*University Hospital and SUNY Upstate Medical University, Syracuse, New York,
†University "La Sapienza," Rome, and ‡Kobe University, Japan

anabolic pathway/signal Neural circuit/input promoting food intake.

blood–brain barrier Specific configuration of the endothelial cells of the brain blood vessels that does not allow the passage of molecules; only molecules using energy-requiring transport mechanisms can cross this barrier.

catabolic pathway/signal Neural circuit/input reducing food intake.

circumventricular organs Brain area where blood vessels are not equipped with the blood–brain barrier, thus allowing the entry of small peptides into the brain.

hypothalamus Integrative brain area made up of different nuclei; it receives inputs from the periphery and triggers the appropriate behavioral and biochemical responses.

neuropeptide A short chain of amino acids acting as a neurotransmitter.

neurotransmitter A molecule transferring information between neurons.

peptide A short chain of amino acids.

vagus nerve/vagal afferents The parasympathetic component of the autonomic nervous system, conveying metabolic and mechanical information from the periphery to the hypothalamus.

To sustain life and growth, satiety is a critical feeling whose function is to avoid overfeeding. As a consequence, the biological mechanisms regulating the onset of satiety are closely interconnected with those regulating the onset of appetite and ultimately they both control food intake. This article will discuss the currently accepted model of the regulation of food intake, by detailing the peripheral inputs that are conveyed to the brain, which in turn integrates this information and triggers the appropriate behavioral response.

SATIETY AND APPETITE AS REGULATORS OF FOOD INTAKE

In everyday parlance, the words life and energy are often used as synonyms. This common use sounds scientific, because a close biological relationship exists between these two concepts since growth, metabolic processes, physical activity, and reproduction would be impossible without energy. Thus, the ability to maximize the balance between energy expenditure and energy intake represents the primary factor in promoting the survival and evolution of a species. Less efficient processes would ensure the progressive disappearance of a species.

One of the most fascinating biological characteristics of humans is the ability to closely match energy intake with energy expenditure, which should prevent the onset of morbid obesity or malnutrition. This extraordinary level of precision can be illustrated by a few calculations. Over the course of a decade, the weight of an average adult tends to increase slightly whereas over the same period, approximately 10 million kilocalories are consumed. To account for the modest change in weight that is generally observed, energy intake must closely match energy expenditure within 0.17% per decade. An efficient regulatory system must therefore exist.

Behaviorally, energy intake is regulated by a simple mechanism, involving the cyclical occurrence of specific feeling-associated stimuli informing the individual when a meal should be initiated and when it should be stopped. Two anabolic stimuli regulate the start of a meal, each indicating a specific need: hunger, which represents a metabolic feeling since it expresses a general need for calories; and appetite, which results mainly from cognitive inputs because it expresses the need for a specific food and is thus related to the palatability of food, its texture, and one's previous experience regarding that particular food. Similarly, two catabolic stimuli control when a meal should be stopped: satiation, representing a "physical" feeling since it expresses the feeling of abdominal fullness that stops a meal; and satiety, a primarily metabolic feeling since it expresses the interprandial lack of any desire to start a new meal. These basic feeling-related stimuli are controlled and regulated by a very complex network, whose integrating center is located in the brain and principally in the hypothalamus. In health and disease, the control of food intake is based on a complex series of biochemical interactions between the brain and the peripheral organs, usually leading to appropriate food intake

 Copyright 2004, Elsevier (USA). All rights reserved.

behavioral responses. Indeed, the currently accepted model for the control of energy intake postulates that energy intake is modulated mainly within the hypothalamus. This continuously regulates the energy status of the body by directly sensing the presence of nutrients in the bloodstream and by receiving afferent input from the periphery (oronasal, gut, liver, adipose tissue). Also, monoamines, neuropeptides, and cytokines produced in the brain and the gastrointestinal (GI) tract during a meal can directly or indirectly activate vagal afferents and mediate many of the nutrients' effects on appetite, gut functions, anabolism, and catabolism. In the hypothalamus, specific neuronal populations transduce these inputs into neuronal responses and, via second-order neuronal signaling pathways and efferent output, into behavioral responses.

PERIPHERAL SIGNALS INFLUENCING SATIETY

To closely regulate the cyclical recurrence of hunger and satiety, the hypothalamus needs to be continuously informed about adipose tissue status, the activity of the GI tract, and the metabolic status of peripheral tissues. To this end, a number of peripheral signals have evolved to assist in maintaining the homeostasis of energy intake (Fig. 1).

Adiposity Signals

Two main adiposity signals exists, leptin and insulin. Leptin is produced primarily by adipocytes and insulin is secreted by the endocrine pancreas. Plasma concentrations of leptin and insulin are proportionate to body fat mass. Both hormones enter the brain via specific receptors located on the blood–brain barrier. Among them, leptin appears to exert a greater influence on energy intake and a rise in its circulating levels results in inhibition of energy intake and increasing energy expenditure. Similarly, insulin enters the brain from the circulation and acts to reduce energy intake. Leptin and insulin receptors are expressed by brain neurons involved in energy intake. Administration of either peptide directly into the brain reduces food intake, whereas their antagonists or a deficiency of either hormone has the opposite effect. Leptin and insulin act on central effector pathways in the hypothalamus, activating catabolic circuits that inhibit food intake and increase energy expenditure, while simultaneously silencing brain anabolic neural circuits that stimulate eating and inhibit energy expenditure. Conversely, low leptin and insulin concentrations in the brain, which occur during weight loss, stimulate the activity of anabolic neural pathways that enhance eating and suppress energy expenditure; low concentrations of leptin and insulin also inhibit the activity of catabolic pathways that cause anorexia and weight loss.

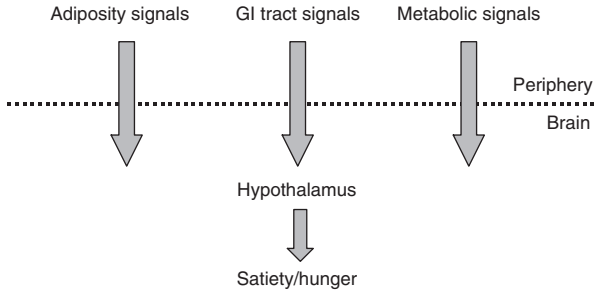


FIGURE 1 Mechanisms controlling food intake. The hypothalamus integrates a large number of inputs relaying to the center the comprehensive status of peripheral tissues. Based on the analysis of these biochemical, mechanical, and metabolic signals, the hypothalamus triggers the appropriate behavioral response, being either the onset of satiety, thus stopping a meal, or the onset of hunger, thus initiating a new meal.

GI Tract Signals

The GI tract produces a number of peptides in response to feeding and fasting, which act directly in the hypothalamus. Among these, the most important are ghrelin and cholecystokinin (CCK).

Ghrelin

Ghrelin is a peptide/neuropeptide released from the stomach in response to fasting and stimulates food intake. It is a ligand for the growth hormone secretagogue (GHS) receptor and when purified has 28 amino acids. Ghrelin and its mRNA as well as the GHS receptors are expressed in the hypothalamus. Both peripheral and central administration of ghrelin stimulated food intake and increased body weight in freely feeding rodents. Ghrelin concentrations in blood and its mRNA in stomach increase by fasting and decrease by refeeding and after Roux-en-Y gastric bypass surgery. The ingestion of sugar, but not gastric distension, decreases circulating ghrelin concentrations. These data suggest that the presence of endogenous ghrelin stimulates the hypothalamus, indicating that ghrelin is the first appetite stimulatory peptide produced by the oxyntic cells of the stomach that act as a neuropeptide in the hypothalamus.

Cholecystokinin

Cholecystokinin is an important signal involved in the regulation of food intake. It is a satiety signal that

acts as a paracrine substance to stimulate pancreatic secretion via vagal cholinergic fibers, but it also reaches the brain to exert its catabolic effect. In addition, CCK sensitizes vagal afferents to mechanical stimuli (e.g., gastric distension) and potentiates the effects of mechanical stimulation on meal termination. Indeed, the mechanical distension of abdominal walls, in particular the gastric wall, is a potent satiety signal that is promptly relayed to the brain via neural afferents. Also, the presence of nutrients in the intestine inhibits eating and gastric emptying.

Peptide YY_{3-36}

The hormone PYY_{3-36} is secreted by endocrine cells lining the distal small bowel and colon in response to food. Its concentrations in the blood remain elevated between meals and thus it suppresses interprandial appetite. It is a member of the neuropeptide Y (NPY) family, acting in the arcuate nucleus of the hypothalamus, where it binds specifically to the Y2 receptor. It modulates the NPY/Agouti-related peptide complex (see below).

Metabolic Signals

The onset of satiety is controlled not only by the extension of adipose tissue and by the presence of food within the GI tract but by the metabolic status of the body and the circulating levels of some nutrients.

Energy Signals

Like changes in fat mass, changes in hepatic energy metabolism influence energy intake in a leptin-independent manner via energy signals relayed to the brain via vagal afferents. A number of studies suggest that a metabolic control of food intake also exists, in which the biochemical partitioning between fatty acid oxidation and synthesis represents a key signal indicating catabolic or anabolic energy status. Although apparently similar, adiposity signals (i.e., leptin and insulin) and energy signals are different and specific. Adiposity signals act as an adipostat and inform the brain about the extension of body fat mass. Energy signals are independent of the leptin pathway and thus independent of body mass extent, but they inform the brain about the metabolic switch occurring at a subcellular level between fatty acid oxidation and synthesis. Further evidence for the involvement of energy signals in the control of energy intake has recently been provided by findings showing that systemic and intracerebroventricular treatment of mice with fatty acid synthase inhibitors leads to inhibition of feeding and weight loss. More specifically, it appears that fatty acids are not the signal determining the cessation of energy intake but rather the intracellular levels of malonylcoenzyme A: its intracellular accumulation inhibits energy intake and its depletion restores energy intake. Fatty acid synthase inhibitors inhibit the expression of the prophagic signal NPY in the hypothalamus, acting in a leptin-independent manner.

Nutrient-Related Signals

Glucose-sensitive cells are present in the endocrine pancreas, liver, and duodenum and are innervated by vagal afferents projecting via the vagus to the nucleus of the solitary tract (NTS). Neurons receptive to concentrations of glucose and other nutrients are present in several regions of the central autonomic network. Central glucose receptors respond to other metabolites (e.g., free fatty acids) as well as changes in concentrations of insulin and glucagon. Glucose concentrations probably alter neuronal firing via ATP-sensitive potassium ion channels similar to those present in the pancreas.

The macronutrient composition of the diet may also promote the onset of satiety by modulating brain monoamines in the hypothalamic sites involved in food intake regulation, based on the competitive uptake of free tryptophan (Trp) with other large neutral amino acids, particularly the branched-chain amino acids. Animals consuming a high-carbohydrate compared to a low-carbohydrate and high-protein diet show increased concentrations of circulating Trp or tyrosine (Tyr). The monoamines include dopamine (DA) and serotonin (5-HT), and the catecholamines norepinephrine (NE) and epinephrine (E). DA and NE are synthesized from Tyr and 5-HT is derived from Trp. The rate at which they synthesize their neurotransmitters is influenced by the precursor amino acid concentration available to the neuron and is controlled by an enzyme that is only partly saturated with substrate at normal brain amino acid concentrations. Therefore, the increase or decrease in concentration of Trp or Tyr can influence the synthesis of 5-HT or DA and NE, respectively. A high-carbohydrate diet promotes the uptake of Trp into the brain and its subsequent conversion to 5-HT, which then terminates the meal and induces satiety. Carbohydrates terminate a meal in two ways, either via a direct effect on 5-HT synthesis or in conjunction with insulin, which also affects 5-HT and its metabolite. However, it must be emphasized that under physiological conditions, several neurotransmitters are involved in food intake initiation and termination. As an example, DA regulates hunger and satiety by acting on corresponding reciprocal hypothalamic areas: DA in

the lateral hypothalamic area (LHA) has a positive stimulatory effect on food intake by modulating the size of a meal via changing gastric compliance, whereas DA in the ventromedial hypothalamus (VMH) inhibits LHA activity. On the other hand, increased intra-LHA and decreased-VMH serotonin levels are associated with the regulation of food intake. These monoamines function in close conjunction with the stimulatory and inhibitory neuropeptides. Thus, the central catabolic initiators interact with numerous factors: psychological and physiological factors, peripheral and central factors, as well as monoaminergic and peptidergic factors.

HYPOTHALAMIC INTEGRATION OF PERIPHERAL SIGNALS

As previously noted, the hypothalamus plays a major role in the sequence of chemical, autonomic, and endocrine events regulating food intake and metabolism. Not only does the hypothalamus contain glucose-sensitive neurons, it also is extensively vascularized and receives hormones including gastrointestinal peptides via the circumventricular organ. The hypothalamus receives viscero-sensory inputs from vagal afferents, via a relay in the NTS and ventrolateral medulla. Connections of the hypothalamus with the limbic cortex, amygdala, and nucleus accumbens allow integration of peripheral information with various cognitive and emotional factors related to memory of the degree of pleasure associated with food. Its autonomic outputs are important for the control of intermediary metabolism, including the concentrations of blood glucose, free fatty acids, and amino acids as well as the regulation of the endocrine pancreas.

Under normal conditions, peripheral signals (adiposity signals, GI tract signals, and metabolic signals) reach the hypothalamus and directly or indirectly interact with two separate neuronal populations: the NPY/Agouti-related peptide (AgRP) neurons and the pro-opiomelanocortin (POMC) neurons. These neurons constitute two pathways, the former stimulating and the latter inhibiting energy intake. As a consequence, when energy intake needs to be initiated, peripheral signals activate the NPY/AgRP pathway, simultaneously inhibiting the POMC pathway. When energy intake needs to be inhibited, the rise in peripheral signals inhibits the NPY/AgRP pathway while simultaneously activating POMC neurons and thus up-regulating the expression of a number of POMC pathway-related factors, including α-melanocyte-stimulating hormone (α-MSH), corticotropin-releasing factor (CRF), and cocaine- and amphetamine-related transcript (CART). These catabolic and anabolic effector systems constitute a series of discrete neurotransmitter systems and axonal pathways in the brain, which are concentrated in the arcuate nucleus (ARC) in the ventral hypothalamus.

Neuropeptide Y

Among the best-described anabolic effector peptides is NPY, a 36-amino-acid peptide. Although the NPY mRNA and the peptide are distributed throughout the central nervous system with a particularly high concentration in the hypothalamus, NPY-containing cell bodies in the ARC are especially important in the control of energy homeostasis. Administration of exogenous NPY into the cerebral ventricle elicits a rapid increase in food intake and a decrease in energy expenditure. Repeated central administration of NPY leads readily to obesity. Inhibition of endogenous NPY synthesis in the ARC by antisense oligonucleotides reduces food intake and body weight.

Melanocortins

The catabolic effector system POMC also resides within the ARC. Melanocortins (MCs) have the opposite effect as NPY. Melanocortins constitute a family of peptides including ACTH, α-MSH, and CART and represent a growing list of peptides that promote negative energy balance. Neuronal synthesis of these peptides increases in response to increased adiposity signaling in the brain. Among these, the MC system is an important catabolic effector system due to its complexity in energy homeostasis. MCs are cleaved from the POMC, a precursor molecule, and exert their effects by binding to members of a family of MC receptors. Two MC receptors, MC3 and MC4, have been identified within the hypothalamus and are involved in the control of energy homeostasis. Administration of α-MSH and other MC receptor agonists into the third ventricle reduces food intake and body weight, whereas MC receptor antagonists (such as SHU-9119) increase food intake and body weight. In addition, the MC system is important in mediating the effects of leptin, which stimulates POMC mRNA. POMC gene expression is reduced in negative energy balance and is concomitantly increased in positive energy balance. Moreover, the MC receptor antagonist blocks the effect of leptin in reducing food intake. These hypothalamic control systems suggest that the endogenous POMC/α-MSH/MC receptors are a key catabolic effector pathway capable of regulating the effects on food intake and body weight that mediate the effect of adiposity signals in the CNS.

SECOND-ORDER NEURONAL SIGNALING PATHWAYS

NPY/AgRP and POMC neurons largely project to other hypothalamic areas, including the paraventricular nucleus (PVN), LHA, and VMH, interacting with a number of neuronal populations. These hypothalamic areas are involved in the regulation of food intake and energy expenditure via efferent pathways, including the control of sympathetic and parasympathetic outputs to the endocrine pancreas and the adrenal gland. These pathways act as the final "effector" mechanism controlling food intake and intermediary metabolism.

The PVN is the site of integration and interaction of multiple influences affecting GI function and food intake. Stimulation of the PVN influences gastric motility and secretion via its connections with the dorsomedial hypothalamic nucleus. Circulating glucocorticoids potentiate norepinephrine, which acts on the PVN to stimulate carbohydrate intake. As described above, carbohydrate intake stimulates the production of 5-HT and is inhibited by 5-HT receptors, thereby explaining the anorexic effect of 5-HT and the orexigenic effect of 5-HT antagonists. The PVN is also the main site where other peptides, including opioids, act to increase the ingestion of fat, which is attenuated by DA. This explains both the anti-orexigenic effects of dopamine-releasing drugs such as amphetamine and the increase of fat intake and body weight gain observed as a side effect of treatment with dopamine receptor-blocking neuroleptics.

Many pathways serving as second-order neuronal signaling pathways, which are important mediators in the energy homeostasis process, have been described. They are also intimately linked with monoaminergic coreceptors, suggesting their combined role in the process of catabolism and anabolism. For example, the PVN has neurons that synthesize CRF, thyrotropin-releasing hormone, and oxytocin and their administration causes a net catabolic effect. On the other hand, the LHA synthesizes melanocyte-concentrating hormone and the orexins and the administration of these neuropeptides into the central nervous system causes a net anabolic response.

Corticotropin-Releasing Factor

Corticotropin-releasing factor, a 41-amino-acid peptide, is a mediator of endocrine, autonomic, and immune responses in stress, and activation of the CRF system is suggested to induce stress-related responses including anorexia and anxiety-like behaviors. Two subtypes of CRF receptors, CRF1 and CRF2 receptors, have been identified and cloned. Centrally administered CRF decreases food intake in both CRF1 receptor null mice and wild-type control mice equally. These results suggest that central CRF2 receptor may mediate the appetite-suppressing effects of CRF and CRF-like peptides.

Urocortin

Urocortin, an edogenous CRF-related peptide that has a much higher affinity for the CRF2 receptor than CRF, induces more potent anorectic effects than CRF after central administration. A CRF2 receptor-selective antagonist, antisauvagine-30, reverses or attenuates the effects of urocortin and CRF on food intake and body weight.

Melanin-Concentrating Hormone and Orexins

Orexins (hypocretins) and melanin-concentrating hormone (MCH), neuropeptides localized to the LHA, have an orexigenic effect after central injection. Expression of both MCH and orexin mRNA is increased in response to fasting. MCH-deficient mice reduce food intake and body weight with an increase in metabolic rate. Orexin knockout mice are also hypophagic but have normal body weight, indicating a difference in metabolic rate. These results suggest that these neuropeptides also play a role in the regulation of energy homeostasis.

SATIETY IN DISEASE

The clinical course of a number of acute and chronic diseases is characterized by the development of a persistent form of satiety, which is called secondary anorexia and is different from anorexia nervosa, a neuropsychiatric disease. Anorexia is usually defined as the loss of the desire to eat and leads to reduced food intake. Unfortunately, anorexia and reduced food intake are often neglected issues in the clinical management of patients, although they have a significant negative impact on morbidity and mortality.

The neurochemical mechanisms responsible for secondary anorexia are still a matter of debate. However, a general consensus exists on at least two issues related to its pathogenesis: (1) multifactoriality and (2) a relationship to disturbances of the previously reviewed central mechanisms controlling food intake under normal conditions. Thus, secondary anorexia might result from defective signals arising from the periphery, errors in the transduction process, or disturbances in the activity of the second-order neuronal signaling pathways. Consistent data suggest that secondary anorexia might be triggered by cytokines, particularly by interleukin-1

(IL-1), IL-6, tumor necrosis factor α, and interferon-γ, which are peptides involved in the modulation of the immune response, exerting biochemical and behavioral effects. They are primarily produced by immune system cells and their mRNA is overexpressed in the hypothalamus of tumor-bearing or septic animals with anorexia.

The mechanisms by which cytokines potentiate satiety during illness are currently under investigation. However, it is likely that cytokines may play a pivotal role in long-term inhibition of feeding by mimicking the hypothalamic effect of excessive negative feedback signaling. This could be achieved by inhibition of the NPY/AgRP orexigenic network, as well as by persistent stimulation of the POMC anorexigenic pathway but also of other networks, including hypothalamic neurotransmission. In this light, the link between cytokines and brain monoaminergic neurotransmission is strengthened by a number of lines of evidence indicating that cytokines, and particularly IL-1, stimulate the release of hypothalamic 5-HT and DA, which directly stimulate melanocortins to induce anorexia. Both monoamines modulate neuronal activity via their action on calcium channels. Neurotransmitters in turn can also affect IL-1-mediated activities. More recently, it has been shown that cytokines potentiate satiety by up-regulating the enzyme cyclooxygenase 2 in the brain, thus resulting in an overexpression of prostaglandin E2 (PGE2). PGE2 may in turn act as a neuromodulator by influencing hypothalamic monoaminergic neurotransmission.

It is also reasonable to speculate that during illness the hypothalamus might influence skeletal muscle wasting, thus synergistically acting with the well-established peripheral cachectic factors. As previously noted, the hypothalamus controls not only energy intake, but also energy expenditure. In pure neurogenic muscular involvement, muscle wasting is secondary to the activation of the same intracellular proteolytic pathways responsible for tumor-induced wasting. In elderly cancer patients, changes in hypothalamic-driven sympatho-vagal balance associated with weight loss have been detected. It is therefore tempting to speculate that, during illness, deranged hypothalamic activity may not only corroborate anorexigenic pathways, but also send proteolytic signals to skeletal muscles.

See Also the Following Articles

Appetite • Brain–Gut Axis • Cholecystokinin (CCK) • Pancreatic Polypeptide Family

Further Reading

Batterham, R. L., Cowley, M. A., Small, C. J., Herzhog, H., Cohen, M. A., Dakin, C. L., Wren, A. M., Brynes, A. E., Low, M. J., Ghatei, M. A., Cone, R. D., and Bloom, S. R. (2002). Gut hormone PYY3–36 physiologically inhibits food intake. *Nature* **418**, 650–654.

Broberg, C., and Hokfelt, T. (2001). Hypothalamic and vagal neuropeptide circuitries regulating food intake. *Physiol. Behav.* **74**, 669–682.

Halford, J. C., and Blundell, J. E. (2000). Pharmacology of appetite suppression. *Prog. Drug Res.* **54**, 25–58.

Inui, A. (2001). Ghrelin: An orexigenic and somatotrophic signal from the stomach. *Nat. Rev. Neurosci.* **2**, 551–560.

Kahler, A., Zimmermann, M., and Langhans, W. (1999). Suppression of hepatic fatty acid oxidation and food intake in men. *Nutrition* **15**, 819–828.

Makarenko, I. G., Meguid, M. M., and Ugrumov, M. V. (2002). Distribution of serotonin 5-hydroxytryptamine 1B [5-HT(1B)] receptors in the normal rat hypothalamus. *Neurosci. Lett.* **328**, 155–159.

McMinn, J. E., Baskin, D. G., and Schwartz, M. W. (2000). Neuroendocrine mechanisms regulating food intake. *Obes. Rev.* **1**, 37–46.

Meguid, M. M., Yang, Z.-J., Bellinger, L. L., Gleason, J. R., Koseki, M., Laviano, A., and Oler, A. (1996). Innervated liver plays an inhibitory role in regulation of food intake. *Surgery* **119**, 202–207.

Meguid, M. M., Yang, Z.-J., and Gleason, J. R. (1996). The gut–brain brain–gut axis in anorexia: Toward an understanding of food intake regulation. *Nutrition*, **12**, S57–S62.

Moran, T. H. (2000). Cholecystokinin and satiety: Current perspectives. *Nutrition* **16**, 858–865.

Nandi, J., Meguid, M. M., Inui, A., Xu, Y., Makarenko, I. G., Tada, T., and Chen, C. (2002). Central mechanisms involved with catabolism. *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 407–418.

Plata-Salaman, C. R. (1996). Anorexia during acute and chronic disease. *Nutrition* **12**, 69–78.

Schwartz, G. J., and Moran, T. H. (1996). Sub-diaphragmatic vagal afferent integration of meal related gastrointestinal signals. *Neurosci. Biobehav. Rev.* **20**, 47–56.

Schwartz, M. W., Woods, S. C., Porte, D. Jr., Seeley, R. J., and Baskin, D. G. (2000). Central nervous system control of food intake. *Nature* **404**, 661–671.

Vergani, A. V., and Bertolini, A. (2000). Role of melanocortins in the central control of feeding. *Eur. J. Pharmacol.* **405**, 25–32.

Secretin

ELAHE MOLLAPOUR AND MICHAEL A SHETZLINE
Duke University Medical Center

desensitization A rapid process that results in the attenuation of G-protein-coupled receptor responsiveness, despite continuing agonist stimulation.

G-protein-coupled receptor (GPCR) An integral membrane-bound protein that consists of a seven-transmembrane α-helical domain. GPCRs are responsible for converting an extracellular signal to an intracellular message via ligand binding and activation of heterotrimeric G-proteins.

G-proteins Heterotrimeric guanine nucleotide-binding proteins consisting of three subunits: α, β, and γ. G-proteins are regulatory molecules that mediate intracellular signaling pathways.

hormone A Greek word meaning "to excite"; it is used to describe a chemical messenger that induces a specific response in target cells distant from the site of synthesis.

receptor A molecular structure within or on the surface of a cell that binds to a specific ligand and initiates a cellular response. The largest class of receptors is the G-protein-coupled receptor family.

second messenger An intracellular mediator produced in response to the binding of agonist to its specific receptor. These signaling molecules activate effector molecules either directly or via activation of protein kinases. Examples of second messengers include cyclic AMP, Ca^{2+}, and diacylglycerol.

signaling A sequential process that results in conversion of an extracellular event into an intracellular message.

In 1902, Bayliss and Starling observed that a chemical substance in upper intestinal extracts, when injected into anesthetized dogs with denervated intestine, stimulated pancreatic secretion. The discovery of this first hormone, named secretin, was the beginning of the search for many other chemical messengers that, when released from one tissue, travel through the bloodstream and exert their effects on distant tissues in the body. Hence, the name hormone, a Greek word meaning arise to activity, was used to describe these chemical substances.

PEPTIDE

Secretin is a basic 27-amino-acid neuroendocrine polypeptide, with a molecular weight of 3055 Da (Table I). The gene encoding human prosecretin is mapped to chromosome 11p15.5. Secretin belongs to a family of peptide hormones with similar amino acid sequences, indicating a common ancestral gene (Table II). The human secretin locus has four exons, which encode a signal sequence, an N-terminal peptide, secretin, and an amidated C-terminal extension peptide. The secretin genes from human, rat, porcine, guinea pig, and canine have been sequenced and demonstrate a high degree of evolutionary conservation.

TISSUE DISTRIBUTION

Secretin is produced and secreted mainly from specialized cells, known as S cells, in the villi of the small intestine. The secretin-producing cells are found along the entire small intestine. Northern blot analysis has detected human secretin mRNA in the testis, small intestine, and spleen, as well as in different areas of the brain, with the highest level in the medulla.

RECEPTOR BIOLOGY

Secretin exerts its effects on its target organs via specific cell surface receptors. The secretin receptor belongs to a unique subfamily of receptors known as

TABLE I Amino Acid Sequence of Secretin Peptides

Human	H S D G T F T S E L S R L R E G A R L Q R L L Q G L V
Dog	H S D G T F T S E L S R L R E S A R L Q R L L Q G L V
Rat	H S D G T F T S E L S R L Q D S A R L Q R L L Q G L V
Cow	H S D G T F T S E L S R L R D S A R L Q R L L Q G L V
Pig	H S D G T F T S E L S R L R D S A R L Q R L L Q G L V
Avian	H S D G L F T S E Y S K M R G N A Q V Q K F I Q N L M

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE II Secretin Family of Gastrointestinal Peptides

Secretin
Glucagon
Vasoactive intestinal peptide
Glucagon-like peptides 1 and 2
Glucose-dependent insulinotropic polypeptide
Pituitary adenylate cyclase-activating polypeptide
Peptide histidine–isoleucine

G-protein-coupled receptors (GPCRs), the largest family of receptors identified to date. The secretin receptor contains seven membrane-spanning α-helices, an extracellular amino-terminus, and an intracellular carboxy-terminus (Fig. 1). This receptor is representative of a receptor family that includes receptors for secretin, glucagon, vasoactive intestinal peptide, and many other gastrointestinal peptide receptors.

The secretin receptor shares a common molecular architecture with other members of this family of GPCRs (Table III). The overall primary structure of the secretin receptor has several interesting features; it consists of a long amino-terminal domain that is important for ligand binding and receptor activation, based on studies carried out with either truncated or chimeric receptors. The N-terminal domain contains sites for possible asparagine-linked glycosylation (positions 72, 100, 106, 128, and 291). There are 10 extracellular cysteine residues, 2 of which are involved in linking the first and second extracellular loops, as well as 7 highly conserved residues. Three exoloops and three cytoplasmic loops plus a hydrophilic C-terminal domain separate the putative seven-transmembrane regions.

Ligand binding to the N-terminal of the secretin receptor results in coupling of heterotrimeric guanine nucleotide-binding proteins (G-proteins) at the C-terminal of the receptor. G-proteins are intermediate regulatory molecules that initiate the intracellular signaling process. They consist of three subunits: α, β, and γ. The binding of secretin to its receptor promotes the exchange of GDP for GTP on the G_α subunit, which allows the separation of G_α from the $G_{\beta\gamma}$ subunit. The G-protein subunits then amplify intracellular signals with subsequent activation of the effector molecules such as adenylyl cyclase, phosphodiesterases, phospholipase A_2, phospholipase C, and ion channels (Fig. 2). These in turn produce second messengers, including cyclic AMP (cAMP), inositol 1,4,5-trisphosphate, and diacylglycerol.

Termination of the secretin receptor signal is just as important for balanced and normal cellular function as the initiation of signaling. In order to prevent receptor overstimulation, cells have evolved a feedback

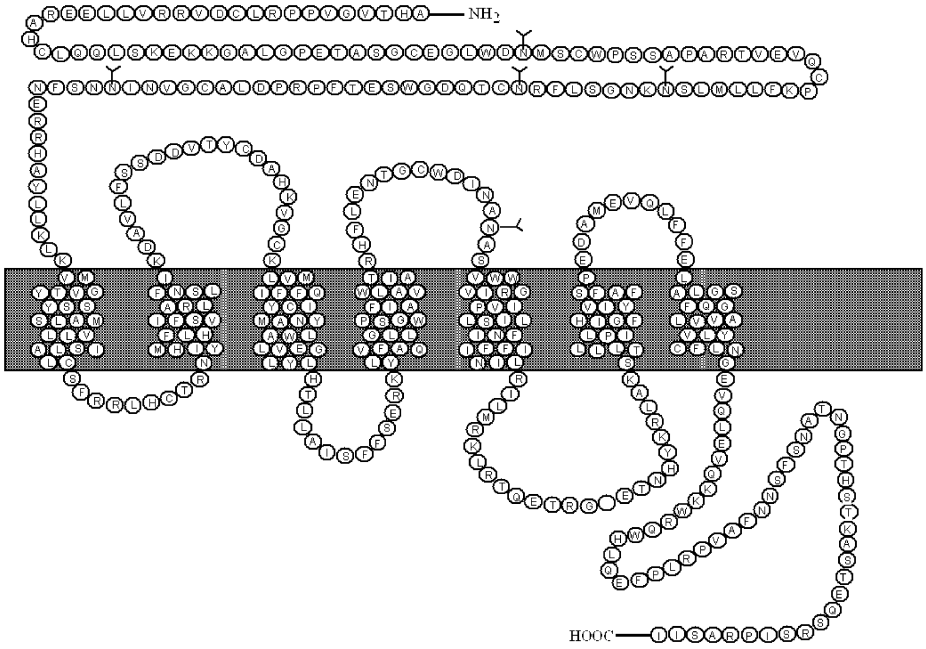


FIGURE 1 Schematic representation of the general structure of the secretin receptor. This receptor contains a seven-transmembrane α-helical region. The three intracellular loops are important for interaction with G-proteins and contain multiple phosphorylation sites. The N-terminal is important for ligand binding and contains N-linked glycosylation sites.

TABLE III The Secretin Family G-Protein-Coupled Receptors

Secretin receptor
Parathyroid hormone receptor
Glucagon receptor
Diuretic hormone receptor
PACAP receptor
Leukocyte antigen CD97
Vasoactive intestinal peptide receptor
Calcitonin receptor
Glucagon-like peptide 1 and 2 receptors
Corticotropin-releasing factor receptor
Growth hormone-releasing factor receptor
Cell surface glycoproteins EMR1 and F4/80

mechanism called desensitization, by which the receptor becomes less responsive to further stimulation. The signal can be terminated either through receptor down-regulation or by a rapid process that involves two different serine/threonine protein kinases: (1) G-protein-coupled receptor kinases (GRKs) or (2) second messenger-dependent protein kinases. In homologous desensitization, GRK-mediated phosphorylation of the secretin receptor facilitates the binding of cytosolic proteins known as β-arrestins to the receptor followed by receptor endocytosis. In contrast, heterologous desensitization involves second messenger-dependent protein kinases such as protein kinase A (PKA) or protein kinase C, acting on both active and unstimulated receptors. It has been shown that GRK-specific phosphorylation is involved in the rapid attenuation of secretin receptor signaling in HEK 293 cells, whereas secretin receptor internalization occurs via PKA-dependent phosphorylation in the same cell line.

The cDNA for the secretin receptor has been cloned from human, rat, and rabbit. The human secretin receptor consists of 440 amino acids with a molecular weight of approximately 49 kDa. The gene for the secretin receptor is localized on human chromosome 2q14.1.

FUNCTIONS

Over the years, secretin has been shown to elicit a variety of physiological actions, as well as functional responses at pharmacological serum levels. These functions of secretin are classified as either stimulatory or inhibitory, affecting organs such as the pancreas, liver, and intestine. The essential role of secretin is to regulate pancreatic fluid and bicarbonate secretion, which results in neutralization of acidic chyme from the stomach. Secretin is released from the small intestine in response to

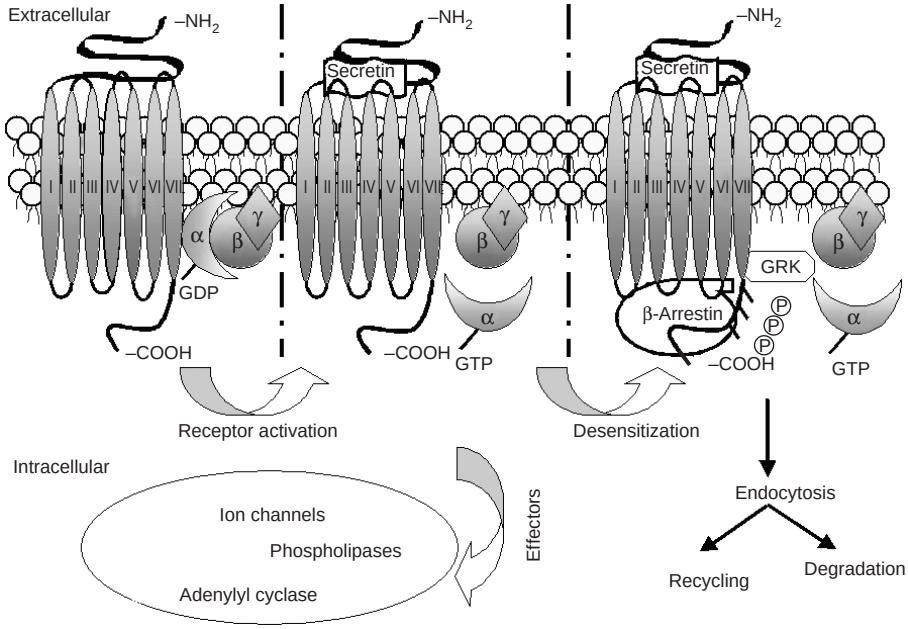


FIGURE 2 Binding of secretin to its receptor results in the exchange of GDP for GTP on the G_α subunit with subsequent dissociation of G_α from the $G_{\beta\gamma}$ subunit, which activates G-protein effectors. β-Arrestin binds to GRK-phosphorylated receptor and uncouples the receptor from its G-protein, initiating receptor endocytosis, which is an important step in receptor dephosphorylation and recycling.

gastric acid and as duodenal pH rises, further release of secretin is curtailed via a negative feedback mechanism. It has been reported that acid-stimulated secretin release is mediated by an endogenous secretin-releasing peptide that is sensitive to trypsin. This peptide stimulates secretin release until sufficient pancreatic proteases degrade secretin-releasing peptide and terminate secretin release. In addition, ingested fats stimulate secretin as they are converted to fatty acids in the gastrointestinal tract.

It has been shown that pharmacological concentrations of secretin have the ability to stimulate gastric pepsin secretion and inhibit gastric acid release, gastrin secretion, and motility of the small intestine. Pharmacological doses of secretin have been shown to increase secretion of a bicarbonate-rich fluid via the biliary tract, increase the lower esophageal sphincter pressure, increase cardiac output, increase renal excretion, increase insulin release, and increase epidermal growth factor production from Brunner's glands in the duodenum. Secretin may also be involved in an early stage in the development of enteroendocrine cells.

Apart from its action on the pancreas, secretin has been found in the central nervous system and may play a role in neurotransmission. For example, secretin increases cAMP levels in different areas of the brain. It has also been suggested that secretin may act as a neurotransmitter through activation of tyrosine hydrolase, an enzyme that is required for the synthesis of catecholamines.

CLINICAL SIGNIFICANCE

Secretin has been used as a diagnostic tool to evaluate digestive and pancreatic function. The most common clinical use of secretin has been in the diagnosis of patients with Zollinger-Ellison syndrome (gastrinoma). In patients with gastrinoma, administration of secretin causes an increase in gastrin release. Elevated serum gastrin levels are the basis for determining the presence of gastrin-producing tumors. In fasting patients, baseline blood samples are taken 5 min prior to and immediately prior to administration of secretin. Secretin [2 units/kg intravenously (iv)] is given over a 30 s interval and serum samples are taken 2 and 5 min after injection and then at 5 min intervals for 20 min. Baseline serum gastrin levels in patients with gastrinoma are usually greater than 150 pg/ml. The iv secretin stimulation test produces a quick and substantial increase in serum gastrin (>200 pg/ml). A positive response (>200 pg/ml) occurs in over 95% of patients with proven gastrinoma. However, achlorhydria or profound hypochlorhydria can result in increased fasting serum gastrin levels and exaggerate the response to iv secretin stimulation. Patients on acid-suppressive therapy should be studied when they are off these medications. Since secretin is a peptide hormone that increases the volume and bicarbonate content of pancreatic juice, patients with acute pancreatitis should not have this test performed.

Secretin is also used to diagnose pancreatic insufficiency and is given during endoscopic retrograde cholangiopancreatography to assist in ductal cannulation.

The behavior of children with autism has been studied following administration of secretin. Initial observations indicating that secretin improved behavior and social interaction were not supported in a double-blind, placebo-controlled study. Therefore, at present, there is no evidence to support a role for secretin in the treatment of autistic children's behavior. There are no known diseases of secretin.

SUMMARY

Secretin was the first hormone to be discovered. It is produced primarily by endocrine cells of the small intestine and its primary action is to regulate pancreatic fluid and electrolyte secretion. Secretin exerts its biological effects through specific cell surface receptors. Binding of secretin to its receptor amplifies signals inside the cell, thus activating effector molecules and increasing cAMP levels. Secretin is used clinically in the diagnosis of gastrinomas and pancreatic insufficiency.

Acknowledgments

This work is supported in part by National Institutes of Health Grants DK02544-04 and DK58828-01 and by the Mal Tyor Scholarship Award in Gastroenterology (to M.A.S).

See Also the Following Articles

Gastric Acid Secretion • Gastrin • Gastrinoma • Pancreatic Bicarbonate Secretion • Pancreatic Function Tests

Further Reading

Bayliss, W. M., and Starling, E. H. (1902). The mechanism of pancreatic secretion. *J. Physiol.* **28**, 325–353.

Holtman, M. H., Ganguli, S., Hadac, E. M., Dolu, V., and Miller, L. J. (1996). Multiple extracellular loop domains contribute critical determinants for agonist binding and activation of the secretin receptor. *J. Biol. Chem.* **271**, 14944–14949.

Ip, N. Y., and Zigmond, R. E. (2000). Synergistic effects of muscarinic agonists and secretin or vasoactive intestinal peptide on the regulation of tyrosine hydrolase activity in sympathetic neurons. *J. Neurobiol.* **42**, 14–21.

Leiter, A. B., Chey, W. Y., and Kopin, A. S. (1994). "Secretin" (J. H. Walshand and G. J. Dockray, eds.), pp. 147–173. Raven Press, New York.

Li, P., Lee, K. Y., Chang, T. M., and Chey, W. Y. (1990). Mechanism of acid-induced release of secretin in rats: Presence of a secretin-releasing peptide. *J. Clin. Invest.* **86**, 1474–1479.

Rindi, G., Ratineau, C., Ronco, A., Candusso, M. A., Tsai, M.-J., and Leiter, A. B. (1999). Targeted ablation of secretin-producing cells in transgenic mice reveals a common differentiation pathway with multiple enteroendocrine cell lineages in the small intestine. *Development* **126**, 4149–4156.

Sandler, A. D., Sutton, K. A., DeWeese, J., Girardi, M. A., Sheppard, V., and Bodfish, J. W. (1999). Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive development disorder. *N. Engl. J. Med.* **341**, 1801–1804.

Shetzline, M. A., Premont, R. T., Walker, J. K., Vigna, S. R., and Caron, M. G. (1998). A role for receptor kinases in the regulation of class II G protein-coupled receptors: Phosphorylation and desensitization of the secretin receptor. *J. Biol. Chem.* **273**, 6756–6762.

Walker, J. K., Premony, R. T., Barak, L. S., Caron, M. G., and Shetzline, M. A. (1999). Properties of secretin receptor internalization differ from those of the β2-adrenergic receptor. *J. Biol. Chem.* **274**, 31515–31523.

Whitmore, T. E., Holloway, J. L., Lofton-Day, C. E., Maurer, M. F., Chen, L., Quinton, T. J., Vincent, J. B., Scherer, S. W., and Lok, S. (2000). *Cytogenet. Cell. Genet.* **90**, 47–52.

Sensory Innervation

WEN JIANG, KIRK HILLSLEY, AND DAVID GRUNDY
The University of Sheffield, United Kingdom

afferent fiber Nerve fibers conducting sensory information from the periphery to the central nervous system. In this context, "sensory" does not necessarily imply sensation.

dorsal root ganglia Cell bodies of spinal afferent neurones are located in sensory ganglia on the dorsal root of each spinal nerve. They project centrally to make synaptic connections in the spinal cord and peripherally to terminate in the viscera.

enteric sensory neurones Located entirely within the wall of the gastrointestinal tract, the enteric sensory neurones have cell bodies in either the submucosal or myenteric plexus and do not project beyond the gut wall; also referred to as intrinsic primary afferent neurones.

nociceptor Sensory nerve endings that are stimulated by tissue injury and can give rise to pain.

nodose ganglia Sensory ganglia containing the cell bodies of afferent neurones that innervate the thoracic and abdominal viscera. These neurones project centrally to make synaptic connections in the nucleus of the tractus solitarius and peripherally to terminate in the viscera.

Sensory neurones are the information superhighway from the gastrointestinal tract to the central nervous system. These neurones have terminations in the gut wall that are specialized to detect changes in the gut environment, and through the generation of trains of action potentials convey the coded information that keeps the central nervous system abreast of events in the bowel. Most of this information goes unperceived but is used to trigger reflexes that coordinate digestive activity according to the needs of the individual. Sensory information also contributes to behavioral mechanisms, particularly those involved in food assimilation, and are integral to satiety and anorexia. Sensory afferents also mediate sensations such as fullness, bloating, nausea, and discomfort. These sensations from the gastrointestinal tract, like visceral pain, generally tend to be vague and poorly localized, often being referred to somatic sites (dermatomes) because of the way visceral and somatic sensory inputs converge in the spinal cord. However, some individuals have heightened visceral sensitivity and this is a hallmark of functional bowel disorders such as irritable bowel syndrome. These patients either generate aberrant sensory signals from the gastrointestinal tract or the normal signals associated with digestion are interpreted inappropriately (or there is a combination of both). Either way, the sensory innervation of the gastrointestinal tract has become a focus for considerable clinical and therapeutic interest.

INTRODUCTION

The gastrointestinal tract has to perform manifestly diverse functions. It is responsible for the digestion and absorption of nutrients that are vital to an organism's

 Copyright 2004, Elsevier (USA). All rights reserved.

Li, P., Lee, K. Y., Chang, T. M., and Chey, W. Y. (1990). Mechanism of acid-induced release of secretin in rats: Presence of a secretin-releasing peptide. *J. Clin. Invest.* **86**, 1474–1479.

Rindi, G., Ratineau, C., Ronco, A., Candusso, M. A., Tsai, M.-J., and Leiter, A. B. (1999). Targeted ablation of secretin-producing cells in transgenic mice reveals a common differentiation pathway with multiple enteroendocrine cell lineages in the small intestine. *Development* **126**, 4149–4156.

Sandler, A. D., Sutton, K. A., DeWeese, J., Girardi, M. A., Sheppard, V., and Bodfish, J. W. (1999). Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive development disorder. *N. Engl. J. Med.* **341**, 1801–1804.

Shetzline, M. A., Premont, R. T., Walker, J. K., Vigna, S. R., and Caron, M. G. (1998). A role for receptor kinases in the regulation of class II G protein-coupled receptors: Phosphorylation and desensitization of the secretin receptor. *J. Biol. Chem.* **273**, 6756–6762.

Walker, J. K., Premony, R. T., Barak, L. S., Caron, M. G., and Shetzline, M. A. (1999). Properties of secretin receptor internalization differ from those of the β2-adrenergic receptor. *J. Biol. Chem.* **274**, 31515–31523.

Whitmore, T. E., Holloway, J. L., Lofton-Day, C. E., Maurer, M. F., Chen, L., Quinton, T. J., Vincent, J. B., Scherer, S. W., and Lok, S. (2000). *Cytogenet. Cell. Genet.* **90**, 47–52.

Sensory Innervation

WEN JIANG, KIRK HILLSLEY, AND DAVID GRUNDY
The University of Sheffield, United Kingdom

afferent fiber Nerve fibers conducting sensory information from the periphery to the central nervous system. In this context, "sensory" does not necessarily imply sensation.

dorsal root ganglia Cell bodies of spinal afferent neurones are located in sensory ganglia on the dorsal root of each spinal nerve. They project centrally to make synaptic connections in the spinal cord and peripherally to terminate in the viscera.

enteric sensory neurones Located entirely within the wall of the gastrointestinal tract, the enteric sensory neurones have cell bodies in either the submucosal or myenteric plexus and do not project beyond the gut wall; also referred to as intrinsic primary afferent neurones.

nociceptor Sensory nerve endings that are stimulated by tissue injury and can give rise to pain.

nodose ganglia Sensory ganglia containing the cell bodies of afferent neurones that innervate the thoracic and abdominal viscera. These neurones project centrally to make synaptic connections in the nucleus of the tractus solitarius and peripherally to terminate in the viscera.

Sensory neurones are the information superhighway from the gastrointestinal tract to the central nervous system. These neurones have terminations in the gut wall that are specialized to detect changes in the gut environment, and through the generation of trains of action potentials convey the coded information that keeps the central nervous system abreast of events in the bowel. Most of this information goes unperceived but is used to trigger reflexes that coordinate digestive activity according to the needs of the individual. Sensory information also contributes to behavioral mechanisms, particularly those involved in food assimilation, and are integral to satiety and anorexia. Sensory afferents also mediate sensations such as fullness, bloating, nausea, and discomfort. These sensations from the gastrointestinal tract, like visceral pain, generally tend to be vague and poorly localized, often being referred to somatic sites (dermatomes) because of the way visceral and somatic sensory inputs converge in the spinal cord. However, some individuals have heightened visceral sensitivity and this is a hallmark of functional bowel disorders such as irritable bowel syndrome. These patients either generate aberrant sensory signals from the gastrointestinal tract or the normal signals associated with digestion are interpreted inappropriately (or there is a combination of both). Either way, the sensory innervation of the gastrointestinal tract has become a focus for considerable clinical and therapeutic interest.

INTRODUCTION

The gastrointestinal tract has to perform manifestly diverse functions. It is responsible for the digestion and absorption of nutrients that are vital to an organism's

 Copyright 2004, Elsevier (USA). All rights reserved.

survival. Yet, the functional adaptations that favor absorption are also a defense liability. Central to these ostensibly conflicting tasks is the ability to monitor the contents of the gastrointestinal lumen and various aspects of the gastrointestinal function in order to orchestrate appropriate patterns of motility, secretion, and blood flow that, on the one hand, facilitate nutrient absorption and, on the other, rapidly dilute and expel potentially harmful antigenic or pathogenic material through diarrhea and vomiting. The gastrointestinal sensory innervation plays a pivotal role in these processes.

SCALE OF THE SENSORY INNERVATION

The gastrointestinal tract has an extensive sensory innervation. These sensory neurones terminate at various levels within the gut wall, including muscle, mucosal epithelia, and enteric ganglia. Other endings terminate in the serosa and mesenteric attachments and form a dense network around mesenteric blood vessels and their tributaries in the gut wall. These various endings maintain a steady flow of afferent traffic to the central nervous system (CNS), relating information on activity both within and outside the gut wall. The sensory information is conveyed to the CNS by separate vagal and spinal pathways to the brain stem and spinal cord, where information is processed and projected to higher brain areas. Because these afferent nerves run in bundles that contain the autonomic outflow from the CNS, they are often referred to as parasympathetic and sympathetic afferents, but this is something of a misnomer; these terms refer to motor function. More correctly, the vagal, pelvic, and splanchnic nerves are the routes these afferents follow to the brain stem and spinal cord. Vagal afferents are more prevalent in the proximal gut, and spinal, particularly pelvic, afferents predominate in the distal gut.

Approximately 50,000 vagal afferents are estimated to supply the gastrointestinal tract, outnumbering vagal parasympathetic efferent fibers by about 10 : 1. There may be a similar number of spinal afferents, with about 7% of sensory cell bodies in the dorsal root ganglia (DRG) projecting to the viscera and a proportion of these innervating the gastrointestinal (GI) tract. The nodose and DRG neurones have axons that project into the CNS, where they synapse with second-order neurones in the brain stem and spinal cord, respectively.

Vagal and spinal afferent fibers are generally unmyelinated or thinly myelinated fibers transmitting different aspects of sensory information at low conduction velocity (~1 m/sec). Vagal neurones generally process physiological information (for example, the nature and composition of the luminal contents and the presence and amplitude of ongoing motor activity of the gut). In contrast, spinal neurones also process pathophysiological information (for example, potentially noxious mechanical or chemical stimuli arising through tissue injury, ischemia, and inflammation). However, vagal and spinal pathways are not entirely functionally separate, because there is some overlap in their sensitivity, particularly between vagal and pelvic afferents, and there is also some interplay between these two pathways.

The enteric nervous system also contains the cell bodies of sensory neurones that play an integral role in the organization of local reflexes. Of the many millions of myenteric neurones, an estimated 30% are sensory, on the basis of their morphology, chemical phenotype, and electrophysiology. Thus the density of the extrinsic nerve terminals in the bowel wall is sparse compared to the terminals of enteric sensory neurones. However, because these intrinsic afferents do not project beyond the bowel wall, they do not contribute to visceral sensations except indirectly as a consequence of reflex changes in secretion, blood flow, or motor activity. Other myenteric neurones project out from the bowel wall and are therefore referred to as intestinofugal fibers. These project to the prevertebral ganglia and make synaptic contact with postganglionic sympathetic neurones that project back to the intestinal wall, again without being directly involved in visceral perception.

SENSORY ENDINGS

The majority of gastrointestinal afferents terminate within the gut wall as bare nerve endings. Thus, the differential sensitivity of afferents arises from their location in the gut wall, their relationship with other structures, and the receptors and ion channels that they express.

Vagal Afferents

Vagal neurones terminate predominantly in the mucosa and the muscle. Afferent endings in the mucosa are in close association with the lamina propria adjunct to the mucosal epithelium, but are never exposed directly to the contents of the lumen. Thus, vagal afferents within the mucosa are in a position to monitor the chemical nature of luminal contents either directly following absorption across the mucosal epithelium or

indirectly via other cells in the epithelium that are exposed to luminal content. Mucosal afferents are also exquisitely sensitive to any local mechanical stimulation that deforms the mucosal epithelium.

Vagal afferent endings in the muscle can be classified into two types: intramuscular arrays (IMAs) and intraganglionic laminar endings (IGLEs). IMAs are distributed within the muscle sheets, especially in the longitudinal muscle, parallel to the long axes of muscle fibers. They appear to make direct contact with the muscle fibers, but they also course on, and form appositions with intramuscular interstitial cells of Cajal, which may play a role in mechanotransduction. IGLEs are basketlike structures surrounding myenteric ganglia. Because IGLEs are located between the circular and longitudinal muscle layers, they are exposed to shearing forces generated during muscle stretch or contraction and have been suggested to be a source of mechanosensitivity. Evidence supporting this view has been elaborated recently by mapping the receptor fields of vagal afferent endings in the esophagus and showing morphologically that these "hot spots" correspond to the locations of IGLEs. IGLEs are the primary candidates for conveying mechanosensory information relevant to distension and contraction of the bowel wall. However, another intriguing possibility arising from the close proximity of IGLEs to the myenteric ganglion is that IGLEs are chemosensitive, responding to neurotransmitters and neuromodulators released into the synaptic neuropil. In the absence of clearly defined synapses between IGLEs and myenteric neurons, communication may arise following simple diffusion from the site of release to the afferent nerve terminals. Many cell types (neurons, glial cells, endothelium) and many different kinds of substances (ions, purines, amino acids, monoamines, peptides, gases) released through vesicular or nonvesicular mechanisms could potentially participate in such communication.

Spinal Afferents

Spinal nerve terminals are distributed throughout the gut wall but are also located in the serosa and mesenteric attachments, often associated with mesenteric blood vessels. Spinal afferents are largely unmyelinated and have multiple branching punctate endings that correspond to multiple receptive fields, often extending over several visceral structures. Their location and response characteristics suggest that spinal afferents respond to distortion of the viscera during distension and contraction. However, spinal afferent terminals are also found in the mucosa. These spinal afferents respond to mechanical stimulation and to the chemical environment within the lamina propria and in particular respond to the changing chemical milieu following injury, ischemia, or infection.

Axon Reflexes

Spinal afferents have collateral branches that supply blood vessels and innervate the enteric ganglia. These fibers have a beaded appearance and are described as being varicose, with the varicosities being the site of neurotransmitter storage and release. Activation of an afferent terminal causes an action potential to be propagated centrally, but action potentials can also propagate down axon collaterals and stimulate the release of neurotransmitters in a local axon reflex, which serves to modulate blood flow and enteric reflex pathways. The main transmitters present in spinal afferents are calcitonin gene-related peptide (CGRP) and substance P (SP). Both of these peptides are implicated in neurogenic inflammation and so their release via axon reflexes may be involved in the development of an inflammatory response. In addition, CGRP released via local axon reflexes may play a cytoprotective role by increasing blood flow to the mucosa. A small proportion of vagal afferent terminals also contain CGRP and SP, and collateral axon branches have been described, but there is little functional evidence to suggest that axon reflexes occur in vagal neurones.

ADEQUATE STIMULUS AND SIGNAL TRANSDUCTION

The early literature is filled with anecdotal evidence that gastrointestinal pain is dull, aching, ill-defined, and badly localized. Stimuli such as cutting, crushing, and burning, which cause pain if applied to the skin, are not perceived when applied to patients with open colostomies, for example. One explanation for this is that sensory information from the skin and viscera are processed differently in the CNS. Brain imaging studies in humans have shown that, unlike somatic afferents, which on activation cause the S1 somatosensory cortex to "light up", gastrointestinal stimulation leads to activation of secondary somatosensory areas, including the anterior cingulate and prefrontal cortex. This indicates that visceral information activates divergent pathways in the CNS, ascending in the spinal cord via spinothalamic and spinoreticular pathways and also via the dorsal columns. Within the spinal cord, visceral and somatic inputs can converge onto the same second-order neurone.

This convergence gives rise to the phenomenon of referred pain, whereby pain from visceral organs is felt at a remote area of the body, the classic example being angina pain referred to the left shoulder.

Another important consideration when comparing visceral and somatic afferent sensitivity is that of adequate stimulus. What is adequate for one set of sensory endings in the gut may be inappropriate for another in the skin. A case in point is that of sensitivity to noxious heat. A heat-sensitive ion channel with a threshold of about 42° C is present on a subset of somatic nociceptive neurones. This channel can also be opened by capsaicin, the pungent ingredient of hot peppers, and is therefore referred to as vanilloid receptor (VR_1), a member of the transient receptor potential (TRP) family of proteins. VR_1 is expressed by most gastrointestinal afferents, but the gut is unlikely to encounter temperature sufficient to activate this channel, suggesting a role other than detecting body temperature. In this respect, protons, at a pH <6.8, are also known to activate VR_1 and augment thermosensitivity. Low extracellular pH occurs during tissue injury and ischemia, but in the GI tract, most obviously in the stomach, a low pH is the normal luminal environment and proton sensitivity may have particular significance here. However, other ion channels are sensitive to protons—for example, members of the epithelial Na^+ channel (ENaC) family such as acid-sensing ion channels (ASICs) and some members of the P2X receptor family that respond to extracellular ATP.

Gut stimuli that readily cause perception include bowel distension and powerful contraction. These then are the adequate stimuli for gastrointestinal mechanosensitive afferents. The locations of sensory endings in the muscle and in the serosal and mesenteric attachments are consistent with this pattern of sensitivity because distension and contraction will generate tension in the muscle layers and concomitant distortion of the serosa and mesenteric attachments. The sensitivity to both stretch and contractions has led to the term "in-series tension receptor," implying that, by analogy with Golgi tendon organs in skeletal muscle, the sensory endings are linked to gut wall connective tissue elements that transmit tension during contraction or when stretched. However, as previously discussed, the IGLEs that lie in parallel with the muscle appear to be the morphological substrate for tension receptors, and thus "in-series" sensitivity may reside in an "in-parallel" location, probably because of the shear forces that are generated within the tissue. The same may be true for serosal and mesenteric afferents, which are clearly not in series with the muscle and in many cases are actually outside the bowel wall. Thus, sensitivity to distension and contraction arises in these endings as a consequence of the distortion of these structures as the bowel wall moves.

Such mechanical deformations are the basis of mechanosensitivity, which is clearly important both for the reflex mechanisms that control gastrointestinal function and for visceral pain processing. Mechanosensitivity can arise indirectly as a consequence of mechanical forces, causing the release of a chemical mediator that in turn acts on a receptor present on the afferent nerve terminal. ATP, potentially one such substance that is released by mechanical distortion, can act on P2X receptors on the nerve terminal. The P2X receptor is an ion channel that, on activation, leads to depolarization and the generation of action potentials. In contrast, direct mechanosensitivity arises because of the presence of mechanically sensitive ion channels in the nerve terminal membrane. Mechanical deformation of the nerve ending leads to the opening or closing of ion channels, allowing charged molecules to pass in or out of the cell, which in turn leads to an alteration in the excitability of the nerve terminal. A variety of mechanosensitive ion channels have been identified, including a class of receptors exemplified by the ENaC/degenerin family. However, at present, the ion channels and receptors that underlie mechanosensitivity in the gastrointestinal tract remain unknown.

STIMULUS–RESPONSE FUNCTION

The relationship between the intensity of stimulation and the degree of activation of sensory afferent is known as the stimulus–response function. Afferent information conveyed by spinal and vagal mechanosensitive afferents is somewhat different in its sensory–response function, as revealed by direct electrophysiological recordings of afferent traffic en route to the CNS. When the bowel is distended, vagal muscle mechanoreceptors have low thresholds of activation, responding to pressure rises of just a few millimeters of Hg, and reach maximal responses within physiological levels of distension. These are referred to as low-threshold mechanoreceptors and convey information relating to normal physiological events in the bowel. These endings also respond during contraction and therefore signal to the CNS information relevant to each and every contraction that occurs anywhere along the length of the GI tract. Many spinal afferents also have low thresholds for activation but continue to respond beyond the physiological range and thus encode both physiological and noxious levels of stimulation; these are called wide-dynamic-range fibers. These spinal endings can contribute to signaling visceral pain through some intensity code that recognizes

extreme levels of distension or contraction. Other spinal afferents, particularly those with endings in the serosa and mesenteric attachments, respond only to noxious levels of distension (high-threshold mechanoreceptors) and are referred to as nociceptors. These only respond when the bowel is overdistended or during powerful contractions that distort the bowel.

The factors that determine the sensitivity of low, wide-dynamic-range, and high-threshold mechanosensitivity may include the location of the endings within the bowel wall, the latter being influenced by the extent to which mechanical forces are distributed and dissipated by nonneural structures in the bowel wall. Different mechanosensitive channels may also contribute to their different stimulus–response functions. An extreme example of mechanosensitivity has been described in sensory endings that fail to respond even to severe levels of distension. These are the "sleeping," or silent, nociceptors, which can be awakened to become mechanically sensitive under conditions of injury or inflammation (see later). The latter illustrate the fact that mechanosensitivity is not fixed either in terms of threshold for activation or in gain in the stimulus–response relationship, and as such the threshold can be reduced and the gain increased after injury or during inflammation. Under these conditions, the afferents are sensitized and this process is believed to underlying hypersensitivity observed in some clinical conditions.

CHEMOSENSITIVITY

An enormous range of chemical mediators can influence the sensitivity of visceral afferents. Luminal nutrients, for example, are transferred across the epithelium to reach the afferent nerve terminals that lie in the lamina propria. Other chemicals are released from within the epithelia following detection of certain contents of the intestinal lumen and are important for initiating reflex mechanisms that optimize digestion and absorption or trigger expulsion via vomiting or diarrhea. Yet other chemicals are released from the many and varied cell types that are present in the gut wall and that are part of the process of immune surveillance. Electrophysiological, immunocytochemical, and molecular biological techniques have revealed the functional expression of receptors to various mediators on the cell bodies of visceral sensory neurons in the dorsal root or nodose ganglia or on their processes in the gut wall. These diverse mediators, which include amines, purines, prostanoids, proteases, and cytokines (Fig. 1), produce their effects on visceral afferent nerves by three distinct processes: (1) by direct activation, which ultimately involves the opening of ion channels present on the nerve terminals, (2) by sensitization, which may occur in the absence of a direct stimulation, but which usually results in afferent hyperexcitability to both chemical and mechanical modalities (this may arise following G-protein-coupled alterations in second messenger systems, which often lead to phosphorylation of membrane receptors and ion channels that control excitability), and (3) by altering the phenotype of the afferent nerve—for example, through alterations in the expression of mediators, channels, and receptors or modulating the activity of these by changing the ligand-binding characteristics or coupling efficiency of other receptors. Neurotropins, in particular nerve growth factor (NGF) and glial cell-derived neurotropic factor (GDNF), influence different populations of visceral afferents and play an important role in adaptive responses to nerve injury and inflammation.

Chemotransduction

Mediators that produce a direct stimulation of visceral sensory nerve endings may do so as part of a discrete sensory signaling pathway. In this case, the afferent neurone does not respond directly to a particular stimulus, but does so following the release of a mediator from another cell that functions as the sensory detector. One example of such primary sensory cells is the gustatory receptors on the tongue. Cells subserving a similar function in the gut, and often referred to as intestinal "taste" cells, include enterochromaffin (EC) cells, which release 5-hydroxytryptamine (5-HT), and enteroendocrine cells, which release many peptides, including cholecystokinin (CCK), peptide YY (PYY), secretin, and melatonin. The apical tuft of microvilli on these cells is exposed to the intestinal lumen and is proposed to monitor luminal contents; in response to an appropriate stimulus, it releases the contents of storage granules across the basolateral membrane to stimulate afferent terminals in close proximity within the lamina propria. Evidence suggests that these different mediators may act on distinct subpopulations of vagal mucosal afferent nerves. As such, this sensitivity represents an example of a high-fidelity, modality-specific signal transduction pathway. This mechanism, more than likely, functions in the detection of moment-to-moment changes in luminal composition and operates, in the main, below the level of consciousness. However, 5-HT derived from the intestinal mucosa, acting on vagal afferent fibers, is implicated in protection from ingested toxins and is a powerful trigger for vomiting and the associated nausea. In contrast, CCK is a satiety hormone that via vagal activation plays a pivotal role in the control of food intake. In

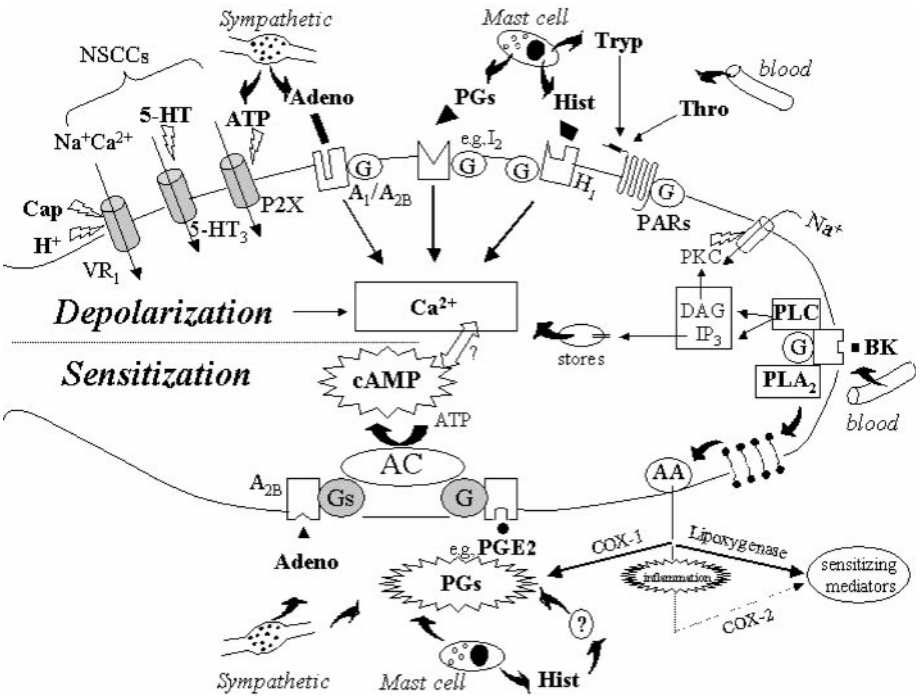


FIGURE 1 Some of the potential receptor mechanisms underlying activation and sensitization of gastrointestinal sensory afferents. Mediators such as serotonin (5-hydroxytryptamine; 5-HT) cause activation, whereas others, such as prostaglandin E2 (PGE2), sensitize visceral afferent responses to other stimuli. Others still, e.g., adenosine (Adeno), cause both stimulation and sensitization, possibly through distinct receptor mechanisms. Bradykinin has a self-sensitizing action, stimulating discharge through activation of phospholipase C (PLC) and enhancing excitability via prostaglandins (PGs) following activation of phospholipase A_2 (PLA_2). Inflammatory mediators can be released from different cell types (e.g., sympathetic nerves, mast cells, and blood vessels) present in or around the afferent nerve terminal. 5-HT, adenosine triphosphate (ATP), and capsaicin (Cap) can directly activate nonselective cation channels (NSCCs) whereas adenosine, histamine, prostaglandins (not PGE2), and proteases such as mast cell tryptase (Tryp) and thrombin (Thro) act on G-protein-coupled receptors, leading to a Ca^{2+}-dependent modulation of ion channel activity. Sensitization, however, may be mediated by elevated levels of intracellular cyclic adenosine 3′,5′-monophosphate (cAMP). Adenosine and PGE2 can generate cAMP directly through G-protein-coupled stimulation of adenylate cyclase (AC). In contrast, histamine may act indirectly through the generation of prostaglandins. The actions of cAMP downstream may involve modulation of ion channels, interaction with other second messengers (e.g., Ca^{2+}), or even changes in receptor expression. Other abbreviations: VR_1, vanilloid receptor; PARs, protease-activated receptors; PKC, protein kinase C; BK, bradykinin; COX-1, COX-2, cyclooxygenase-1 and -2; AA, arachidonic acid; DAG, diacylglycerol; IP_3, inositol 1,4,5-trisphosphate. Modified from Kirkup *et al.* (2001), with permission from the American Physiological Society.

this respect, recent data suggest that other satiety factors, e.g., leptins, orexins, and ghrelin, may interact with CCK signals at the level of the vagal afferent nerve terminal.

Also implicated in sensory signal transduction are the epithelial "brush cells." These differ from EC cells in that they do not have storage granules within the basolateral aspect of the cell. They are morphologically similar to receptor cells within lingual taste buds and express similar G-protein-coupled receptors (for example, α-gustducin), suggesting they play a role in chemosensitivity. However, the way in which brush cells transfer these signals to sensory afferents is currently unknown.

Promiscuous Chemosensitivity

In contrast to the specific signaling pathways that exist in vagal mucosal afferents, it is apparent that a battery of mediators can influence the sensitivity of

spinal afferents in a more promiscuous manner. Such substances are usually released under conditions of inflammation, injury, or ischemia from a plethora of cell types, e.g., platelets, leukocytes, lymphocytes, macrophages, mast cells, glia, fibroblasts, blood vessel cells, muscle cells, and neurons. Each of these specific cells (e.g., mast cells) may release several of these modulating agents, some of which may act directly on the sensory nerve terminal and some of which may act indirectly, following release of other agents from other cells in a series of cascades.

The net effect of this promiscuous chemosensitivity is that the properties of sensory neurones can change (often referred to as plasticity). Of clinical relevance is the increased sensitivity to both mechanical and chemical stimulation that may contribute to chronic pain states. Moreover, because these afferents also serve to trigger reflex mechanisms that control and coordinate gut function, their sensitization may also cause hyper- or dysreflexia. Sensory neuronal plasticity may have a rapid onset, and this is described as peripheral sensitization because the changes take place at the level of the sensory nerve terminal following release of a great many algesic chemicals. Some of the key mediators, their cellular source, and their action on visceral afferents are illustrated in Fig. 1. Following sensitization, there is a leftward shift in the stimulus–response function, which means that for a given level of stimulation, there is a greater afferent barrage generated. This can give rise to altered perception such that stimuli that are normally innocuous can cause pain and the response to a painful stimulus becomes exaggerated (hyperalgesia). Peripheral sensitization normally develops rapidly and is relatively short-lived. However, in the presence of maintained injury or inflammation, the sensitization can be prolonged, and this depends on changes in gene expression. Genes influenced in this way include those that determine the amount and pattern of neurotransmitters released from the central nerve terminals in the brain and spinal cord, thus altering the way sensory signals are relayed within the CNS. This is the basis of central sensitization. Other genes influence signal transduction and sensory neuronal excitability. Voltage-gated sodium channels are one example of a family of ion channel, some of which play an important role in regulating pain thresholds.

CONCLUSION

Afferent fibers convey a vast amount of sensory information to the brain stem and spinal cord, but the nature of this information is different for vagal and spinal pathways. Vagal afferents convey predominantly physiological information, whereas spinal afferents are also able to encode noxious events. These spinal nociceptors are influenced by peripherally acting chemicals, released during inflammation and injury, which are thought to trigger the processes leading to sensitization and increased nociceptive activity. Other chemicals act in a more selective way to activate vagal afferents and are implicated in nutrient signaling from the GI tract.

See Also the Following Articles

Autonomic Innervation • Brain–Gut Axis • Calcitonin Gene-Related Peptide (CGRP) • Cholecystokinin (CCK) • Enteric Nervous System • Interstitial Cells of Cajal • Parasympathetic Innervation • Substance P • Vagus Nerve

Further Reading

Boucher, T. J., and McMahon, S. B. (2001). Neurotrophic factors and neuropathic pain. *Curr. Opin. Pharmacol.* **1**(1), 66–72.

Buchan, A. M. (2000). Nutrient tasting and signaling mechanisms in the gut III. Endocrine cell recognition of luminal nutrients. *Am. J. Physiol. Gastrointest. Liver Physiol.* **277**, G1103–G1107.

Cervero, F., and Laird, J. M. (1999). Visceral pain. *Lancet* **353**, 2145–2148.

Gebhart, G. F. (2000). Pathobiology of visceral pain: Molecular mechanisms and therapeutic implications IV. Visceral afferent contributions the pathobiology of visceral pain. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G834–G838.

Holzer, P. (2001). Gastrointestinal afferents as targets of novel drugs for the treatment of functional bowel disorders and visceral pain. *Eur. J. Pharmacol.* **429**, 177–193.

Kirkup, A. J., Brunsden, A. M., and Grundy, D. (2001). Receptors and transmission in the brain–gut axis: Potential for novel therapies. I. Receptors on visceral afferents. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G787–G794.

Raybould, H. E. (1998). Does your gut taste? Sensory transduction in the gastrointestinal tract. *News Physiol. Sci.* **13**, 275–280.

Wood, J. N. (2000). Pathobiology of visceral pain: Molecular mechanisms and therapeutic implications II. Genetic approaches to pain therapy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **278**, G507–G512.

Zagorodnyuk, V. P., and Brookes, S. J. (2000). Transduction sites of vagal mechanoreceptors in the guinea pig esophagus. *J. Neurosci.* **20**, 6249–6255.

Serotonin

JEROME M. FELDMAN
Duke University Medical Center

APUDoma Amine precursor uptake and decarboxylation tumor; another name for neuroendocrine tumor.

argentaffin cells Containing a potent reducing agent, usually serotonin, that reduces added silver salts to a black pigment in a manner analogous to development of photographic film.

argyrophil cells Neuroendocrine in nature, but do not store serotonin; after exposure to silver salts, an external reducing agent has to be added to reduce the silver salts to a black pigment.

body mass index Measure of body fat, derived as (weight in kilograms)/(height meters)^2 or (weight in pounds)/(height inches)$^2 \times 703.1$. A body mass index of 30.0–34.9 is obese; >40 is extremely obese.

carcinoid tumors Arise from the neuroendocrine cells found throughout the body; the most frequent origin is from the gastrointestinal tract or bronchus. Many, but not all, release neurohormonal agents such as serotonin.

diarrhea Increase in frequency, weight, or consistency of stools; patients with stool weights of greater than 200 g/day are usually considered to have diarrhea.

enterochromaffin cells Diffuse system of cells in the submucosa of the intestinal tract; contain potent reducing agents, such as serotonin or epinephrine, which can reduce chromium salts that are added during tissue fixation. This forms a brown pigment that is visible in the cells when they are examined with a microscope.

irritable bowel syndrome Part of the group of functional gastrointestinal disorders seen frequently by gastroenterologists; patients have chronic and recurrent gastrointestinal symptoms such as abdominal pain, bloating, diarrhea, and constipation. Because there is no known structural basis, the syndrome is thought to be due to altered sensory or motor regulation of the gastrointestinal tract.

neuroendocrine system Group of cells with both neural and endocrine characteristics; thought to arise from cells that migrate from the dorsal neural crest area of the embryo during embryological development.

neuroendocrine tumors Can arise from the cells of the neuroendocrine system; include carcinoid tumors, pheochromocytomas, pancreatic islet cell tumors, and medullary carcinoma of the thyroid.

serotonin Amine (5-hydroxytryptamine) synthesized from the amino acid L-tryptophan; has potent biological effects.

Serotonin, a neurotransmitter that is synthesized from the amino acid L-tryptophan, exerts its biological effects after it attaches to a variety of serotonin receptors that are present on cell membranes. Serotonin has both physiological and pathological effects. Serotonin is the predominant neurohormone secreted from carcinoid tumors and is responsible for most of the symptoms. Serotonin agonists can increase intestinal motility and serotonin antagonists can ameliorate some forms of irritable bowel syndrome. Selective serotonin reuptake inhibitors are used to treat patients with affective disorders and have been used in the treatment of severe obesity, although they can have negative side effects on the cardiovascular system. Serotonin antagonists markedly decrease the severe nausea and vomiting experienced by patients who receive antineoplastic chemotherapy.

INTRODUCTION

The gastrointestinal and circulatory systems have played important roles in the discovery of serotonin. In 1868, it was noted that there was a factor in defibrinated blood that increased vascular resistance in perfused muscle. In 1939, Dr. Vitorio Erspamer in Parma, Italy, discovered enteramine, a substance released from the enterochromaffin system of the gastrointestinal tract that increased intestinal peristalsis. In 1948, Drs. Page, Rappaport, and Green in Cleveland, Ohio, isolated and identified the vasoconstricting substance in defibrinated blood as serotonin. In 1948, Erspamer determined that enteramine was also serotonin. In the ensuing 50 years, studies have indicated that serotonin probably existed in plants even before it existed in animals. There have been numerous studies on the role of serotonin in health and disease. The focus in this article is on the important role of serotonin in medical problems encountered by gastroenterologists.

SYNTHESIS AND METABOLISM

Serotonin is synthesized from the essential amino acid L-tryptophan. L-Tryptophan is first converted to

 Copyright 2004, Elsevier (USA). All rights reserved.

5-hydroxytryptophan by the enzyme tryptophan hydroxylase. 5-Hyroxytryptophan is then converted to serotonin by the enzyme aromatic amino acid (L-dihydroxyphenylalanine; L-DOPA) decarboxylase. Serotonin(5-hydroxytryptamine; 5-HT) is the biologically vasoactive substance. Because tryptophan hydroxylase is predominantly located in the GI tract and the brain, these areas are the major sites of serotonin production and storage. Some of the 5-HT is secreted into the vascular compartment. Although blood platelets cannot synthesize 5-HT, a major portion of the secreted 5-HT is taken up and stored in the dense granules of platelets. A small but biologically important portion remains free in the plasma.

Three different enzymes inactivate serotonin. Monoamine oxidase oxidatively deaminates serotonin to the intermediate 5-hydroxyindoleacetaldehyde. This aldehyde is then converted to 5-hydroxyindoleacetic acid (5-HIAA) by the enzyme aldehyde dehydrogenase or to 5-hydroxytryptophol by the enzyme alcohol dehydrogenase. In humans, the major pathway is to 5-HIAA. To assess serotonin production, serotonin can be measured in serum, platelets, blood, or plasma. The 24-hour urinary excretion of 5-HIAA or, if the methodology is available, the 24-hour urinary excretion of serotonin can also be measured.

PHYSIOLOGICAL EFFECTS

Circulating serotonin attaches to serotonin receptors that are located on the cell membrane. There are seven different serotonin receptor classes designated 5-HT1 through 5-HT7. Within each class, subtypes are further designated by an additional letter (e.g., 5-HT1A or 5-HT1B). These serotonin receptors are coupled to a so-called G protein that delivers serotonin-stimulated intracellular messages. However, the exact mechanism of stimulation is not known for all of the serotonin receptor subtypes. The 5-HT4 subtype is positively coupled to adenylyl cyclase. Various medications are agonists (stimulators) of specific serotonin subtypes, and other medications are antagonists (blockers) of specific serotonin subtypes. The serotonin subtypes mediate a variety of physiological responses; in the GI tract, responses include activation of secretory cells, afferent and efferent neuron activation, and direct effects on gut smooth muscle resulting in either smooth muscle contraction or smooth muscle relaxation.

One model that has proved useful in understanding serotonin action is the synaptic cleft. This is the space between a serotonin-secreting neuron and a serotonin-responding neuron that contains the serotonin receptors. The serotonin-secreting neuron not only secretes serotonin, but it also takes up the secreted serotonin that is in the space between the two neurons. The longer the serotonin remains in the synaptic cleft, the greater effect it has on the serotonin-responding neuron. A number of drugs decrease the rate of reuptake of serotonin, thus prolonging the action of the secreted serotonin.

PHARMACOLOGICAL REGULATION

Many medications can lower the synthesis or release of serotonin through a variety of different mechanisms. Parachlorophenylalanine is a potent inhibitor of tryptophan hydroxylase, the rate-limiting enzyme in the synthesis of serotonin. Administration of this drug dramatically lowers tissue levels of serotonin, and although it is no longer in clinical use, this drug has contributed much to the knowledge of serotonin action. Octreotide acetate, a synthetic analogue of the natural hormone somatostatin, decreases secretion of serotonin from normal tissues and from carcinoid tumors. A large variety of medications are antagonists of various classes and subtypes of serotonin receptors. These medications, including methysergide maleate, cyproheptadine, buspirone, and clozapine, decrease the physiological effect of biologically available serotonin.

Some medications potentiate the effect of biologically available serotonin by blocking its reuptake from the synaptic cleft. They include first-generation antidepressants such as amitriptyline, second-generation antidepressants such as trazodone, and third-generation antidepressants such as fluoxetine. Although first- and second-generation antidepressants also block the reuptake of norepinephrine, third-generation antidepressants block only the reuptake of serotonin. They are thus called selective serotonin reuptake inhibitors (SSRIs). A second mechanism of increasing the effect of serotonin is to use serotonin agonists that bind to specific subtypes of serotonin receptors. Two examples of this class of drugs are the 5-HT4 agonist cisapride, used in treating GI disease, and the 5-HT1D agonist sumatriptan succinate, used in treating migraine and cluster headaches. Finally, monoamine oxidase (MAO) inhibitors such as tranylcypromine and phenelzine inhibit MAO in the liver and other tissues. Because of the reduced inactivation of serotonin, serotonin levels can increase in the body. Clinicians must be careful not to prescribe medications such as an MAO inhibitor and a specific serotonin reuptake inhibitor for simultaneous administration because of the potential for patients to develop dangerously high concentrations of serotonin.

IRRITABLE BOWEL SYNDROME

The GI tract has played a unique role in the investigation of serotonin and its actions. Because of the enterochromaffin cells in the submucosa of the small intestine, the small intestine has the largest concentration as well as the greatest total quantity of serotonin in the human body. The first evidence that there were distinct types of serotonin receptors came from the work of Gaddum and Picarelli on isolated sections of ileum in 1957. These investigators showed that activation of cholinergic nerves by serotonin and a separate direct action of serotonin on smooth muscle were due, respectively, to an M receptor (now called 5-HT3) and a D receptor (now called 5-HT2).

Cisapride, a 5-HT4 receptor agonist, enhances the release of acetylcholine in the myenteric plexus. Cisapride has been used to stimulate intestinal motility, to increase lower esophageal sphincter pressure for the treatment of nocturnal symptoms of gastroesophageal reflux disease (GERD), and in the treatment of diabetic enteropathy. The greatest success in modifying serotonin action has been in the treatment of a type of irritable bowel syndrome in women in which diarrhea is the predominant symptom. Alosetron (Lotronex) is a highly selective and long-acting 5-HT3 receptor antagonist. This drug is effective in decreasing the troublesome symptoms of this group of patients.

DEPRESSION

Serotonin has played a major role in our understanding of psychiatric disease. Although dopamine has been the principal neurotransmitter thought to be involved in schizophrenia, there is evidence that serotonin may also be relevant to this problem. A number of effective medications used in treating schizophrenia, such as clozapine, sertindole, and resperidone, are 5-HT2 receptor antagonists. Serotonin may also play a role in anxiety attacks, based on evidence that buspirone, a 5-HT1A and 5-HT2 antagonist, is effective in treating patients with this problem.

Serotonin plays a major role in the treatment of affective disorders such as depression and mania. It is thought that decreased serotonin activity in certain areas of the brain is associated with depression. The administration of first-, second-, and third-generation antidepressants increases serotonin action. Third-generation antidepressants such as fluoxetine, which have a specific action on serotonin and little effect on norepinephrine, have markedly improved the ability of all clinicians to treat patients with depression effectively.

APPETITE AND OBESITY

Obesity, which is defined as a body mass index of over 30, has become a major problem in the United States. Of even greater concern are extremely obese patients, who are defined as having a body mass index of over 40. This degree of obesity exacerbates diabetes mellitus, coronary artery disease, and hypertension. The ventromedial hypothalamus plays an important role in the regulation of food intake. It contains both an appetite center and a satiety center. This area of the brain has a high concentration of serotonin and norepinephrine. There is evidence that 5-HT2C receptors in this region play a key role in appetite regulation.

Although behavior modification of exercise and diet are the cornerstones of the treatment of obesity, attempts have been made to reduce food intake with medications such as amphetamine and dextroamphetamine. Unfortunately these medications are not suitable for clinical use because they cause a generalized stimulation of the brain, they soon lose their effect on decreasing appetite, and they can be habit forming. Other medications that have been used to decrease appetite include fenfluramine and dexfenfluramine. These medications both inhibit the reuptake of serotonin and cause an increase in the secretion of serotonin from the brain. Many patients have received both fenfluramine and phentermine, a combination that is popularly known as Fen-Phen. Although these medications have helped many obese patients to lose weight, a significant number of patients receiving them developed pulmonary hypertension and others were reported to develop fibrosis of the mitral and aortic valves. Because of this, fenfluramine and dexfenfluramine have been withdrawn from the market. There remains controversy about the long-term significance of the valvular problem.

Sibutramine is the only medication that alters serotonin action that is presently on the market in the United States. Like fenfluramine, sibutramine is a serotonin reuptake inhibitor. However, in contrast to fenfluramine, sibutramine does not stimulate release of serotonin from the brain. Because of reports linking sibutramine administration to increases in blood pressure and heart rate, the Italian Health Ministry has just suspended sales of this drug in Italy. The Food and Drug Administration continues to approve its use in the United States.

NAUSEA AND VOMITING

Nausea is the subjective sensation of an impending urge to vomit. Vomiting is the forceful ejection of gastroesophageal contents from the mouth. Nausea and

vomiting are nonspecific responses to a variety of agents. However, the nausea and vomiting that has been most resistant to therapy is provoked by cancer chemotherapy agents such as cisplatinum, doxirubicin, and cyclophosphamide. Although many antiemetic agents have been used to treat chemotherapy-induced nausea and vomiting in the past, the first truly effective agent was metoclopramide. Because metoclopramide is a dopamine receptor antagonist, its antiemetic activity was originally attributed to this mechanism. It subsequently became clear that in the high doses in which it is administered, metoclopramide also blocks 5-HT3 receptors.

When the potent and specific 5-HT3 receptor antagonist ondansetron was developed, it was evaluated in many cancer chemotherapy centers. Ondansetron is extremely effective in blocking chemotherapy-induced nausea and vomiting. It is as effective after oral administration as it is after parenteral administration. Although there are 5-HT3 receptors in the brain and in the small intestine, it is thought that ondansetron works by blocking the 5-HT3 receptors in the intestine. Two additional 5-HT3 antagonists, granisetron and dolasetron, are also now available. All three 5-HT3 antagonists are equivalent in their clinical efficacy in cancer chemotherapy patients and are very valuable in allowing patients to avoid the severe nausea and vomiting that accompany treatment. However, the three antiemetic agents do not completely prevent the delayed nausea and vomiting that may occur 24 hours after chemotherapy. Metoclopramide and high-dose dexamethasone may be helpful in ameliorating this delayed response. Because all of these antiemetics are expensive, there has not yet been much clinical experience in evaluating these agents for nausea and vomiting of other etiologies.

CARCINOID TUMORS AND CARCINOID SYNDROME

Clinical Manifestations

The term *Karzinoide* (carcinoid) was used by S. Oberdorfer in Germany in 1907 to describe small-intestinal tumors that histologically resembled, but did not behave in the aggressive manner of, adenocarcinomas. In 1952, some patients harboring carcinoid tumors were reported to develop carcinoid syndrome, consisting of diarrhea, facial flushing, and heart valve lesions. In 1953, the chemical released from the carcinoid tumor was identified as serotonin.

Carcinoid tumors are the most common neuroendocrine tumors in clinical practice. Data from evaluation of 840 patients with carcinoid tumors over a 30-year period in one clinical practice are detailed in Table I, showing the site of tumor origin. Two-thirds of the tumors originated from the ileum, the bronchus, or an unknown site; the remaining one-third of the tumors arose from 17 other diverse sites.

Patients with carcinoid tumors of small-intestinal origin have presenting problems of chronic abdominal pain, bowel obstruction, diarrhea, and facial flushing. During the course of their illness, many develop fibrosis of the pulmonic and tricuspid heart valves, leading to carcinoid heart disease. Some patients require cardiac surgery with replacement of the damaged heart valves.

Although patients with carcinoid tumors can live for a long time, ultimately the illness results in the death of many of the patients. The survival time of patients, from the date of diagnosis to date of death, ranges from 1 day to 260 months, with a mean survival time of 36 months. At the time of death, 10% of the patients were younger than 40 years old and 50% of the patients were younger than 60 years old, an indication of the deleterious effect of carcinoid tumors.

The stage of the disease at the time of diagnosis is the most important factor in determining prognosis. In descending order of survival are patients with localized disease, regional metastasis, and distant metastases. Site of origin also plays a role in prognosis, in that patients with ileal carcinoid tumors and distant metastases have a better prognosis than do patients with pancreatic carcinoid tumors and distant metastases.

The major features of carcinoid syndrome are diarrhea, facial flushing, and heart valve disease. In

TABLE I Origin of Carcinoid Tumors of 840 Patients[a]

Origin	Number of patients	Percent of total
Ileum	226	26.9
Unknown	198	23.6
Bronchus	162	19.3
Pancreas	58	6.9
Stomach	40	4.8
Appendix	34	4.0
Rectum	34	4.0
Duodenum	21	2.5
Cecum	20	2.4
Jejunum	15	1.8
Thymus	15	1.8
Ampulla of.vater	8	1.0
Miscellaneous	5	0.6
Meckel's diverticulum	4	0.5

[a] Patients evaluated by the author over a 30-year period in the author's clinical practice.

previously published reports, it has been estimated that only 6% of patients with carcinoid tumors had carcinoid syndrome and it was assumed that only these patients had serotonin overproduction. A policy of measuring serotonin production in all patients with carcinoid tumors regardless of the presence or absence of symptoms, however, leads to quite different conclusions: there is evidence of serotonin overproduction in 53% of the patients with carcinoid tumors (personal experience). Even this represents a minimal estimate, because the presence of carcinoid tumors—particularly tumors originating in the bronchus—is not always suspected before surgery. Thus, by the time serotonin production can be evaluated in some patients, the tumor has already been resected. Indeed, measurement of serotonin production in 44 consecutive patients with definite carcinoid tumors still present in their body has provided evidence of serotonin overproduction in 84% of the patients (personal experience).

Although carcinoid tumors can secrete a variety of neurohormones, serotonin is almost certainly responsible for the symptomatic diarrhea. Serotonin, along with other neurohormones, also plays a role in the facial flushing. When patients with carcinoid tumors and increased serotonin production are questioned about symptoms of facial flushing and diarrhea, alone or in combination, the results are surprising: 44% report both diarrhea and facial flushing, 17% report diarrhea but no facial flushing, 6% report facial flushing but no diarrhea, and 33% report neither diarrhea nor facial flushing. The total absence of classical symptoms of carcinoid syndrome, despite elevated levels of serotonin, is surprising. Possible explanations with respect to patients with carcinoid tumors include a decrease in the number of serotonin receptors in the intestinal tract and the skin, or that the serotonin attaches to the receptors, but some patients have a blunted postreceptor physiological response to the serotonin.

Diagnostic Tests

The overproduction of serotonin is estimated by measuring 5-HIAA excretion in a 24-hour urine collection. Elevated levels of serotonin in the blood are estimated by measuring serum serotonin concentration. It is frequently difficult to visualize the small primary carcinoid tumor in the submucosa of the intestine. Patients frequently have liver metastases at the time of diagnosis. These can be evaluated by computer tomography (CT) or magnetic resonance imaging (MRI) of the abdomen. Because carcinoid tumors concentrate radioactive indium-111 octreotide and iodine-131 metaiodobenzylguanidine ([^{131}I]MIBG), nuclear medicine scans using these radioisotopes are also used to visualize the metastatic carcinoid tumor.

Therapy

If the carcinoid tumor is not metastatic at the time of diagnosis, the primary tumor can be surgically removed. However, in the majority of patients there is tumor involvement of the lymph nodes or liver at the time of diagnosis. Many of the patients have facial flushing and diarrhea. This can be ameliorated with the serotonin receptor antagonists cyproheptadine or methysergide. If this is not effective, the patient can receive injections of either soluble or long-acting octreotide acetate. Octreotide acetate reduces secretion of serotonin from the carcinoid tumor. If the carcinoid tumor shows progressive growth, the patients can receive chemotherapy with streptozotocin and 5-fluorouracil or etoposide and cisplatinum. Some patients have had a beneficial response to moderate doses of interferon α. If the carcinoid tumor concentrates [^{131}I]MIBG, large doses of this radioactive compound can be administered.

SUMMARY

A variety of pharmacological agents can increase or decrease the physiological or pathological effects of serotonin. Serotonin agonists such as cisapride can increase intestinal motility, and serotonin antagonists such as alosetron can ameliorate some forms of irritable bowel syndrome. Selective serotonin reuptake inhibitors have dramatically improved the treatment of patients with affective disorders. Although serotonin reuptake inhibitors have been used in the treatment of severe obesity, because of their troublesome side effects on the cardiovascular system, they have not realized their clinical potential. Serotonin antagonists have markedly decreased the severe nausea and vomiting experienced by patients who receive antineoplastic chemotherapy. Finally, serotonin is the predominant neurohormone secreted from carcinoid tumors. It is responsible for the diarrhea, carcinoid heart disease, and to some extent the facial flushing that constitute the carcinoid syndrome. For reasons not yet identified not all patients with elevated serotonin production develop carcinoid syndrome. Although there is presently no cure for carcinoid tumors once they have spread to lymph nodes and liver, effective therapy has been developed to ameliorate the symptoms of these tumors. Carcinoid tumors, which have been called cancer in slow motion by Dr. Charles Moertel, remain an important tumor model for future investigation.

See Also the Following Articles

Appetite • Emesis • Gastric Acid Secretion • Nausea • Obesity, Treatment of • Small Intestine, Benign and Malignant Neoplasms of the

Further Reading

Azmitia, E. C. (2001). Modern views on an ancient chemical: Serotonin effects on cell proliferation, maturation, and apoptosis. *Brain Res. Bull.* **56**, 413–424.

De Ponti, F., and Tonini, M. (2001). Irritable bowel syndrome: New agents targeting serotonin receptor subtypes. *Drugs* **61**, 317–332.

Feldman, J. M. (1989). Carcinoid tumors and the carcinoid syndrome. *Curr. Prob. Surg.* **26**, 834–885.

Gershon, M. D. (1999). Review article: Roles played by 5-hydroxytryptamine in the physiology of the bowel. *Aliment. Pharmacol. Ther.* **13**(Suppl. 2), 15–30.

Hoyer, D., Clark, D. E, Fozard, J. R., Hartig, P. R., Martin, G. R., Mylechrane, E. J., Sxena, P. R., and Humphrey, P. P. (1994). International union of pharmacology classification of receptors for 5-hydroxytryptamine (serotonin). *Pharmacol. Rev.* **46**, 157–203.

Lindley, C., and Blower, P. (2000). Oral serotonin type 3 receptor antagonists for prevention of chemotherapy induced emesis. *Am. J. Health Syst. Pharm.* **57**, 1685–1697.

Meguid, M. M., Fettissov, Varima, M., Sato, T., Zhang, L., Laviano, A., and Rossi-Fanelli, F. (2000). Hypothalamic dopamine and serotonin in the regulation of food intake. *Nutrition* **16**, 843–847.

Oberg, K. (2002). Carcinoid tumors: Molecular genetics, tumor biology, and update of diagnosis and treatment. *Opin. Oncol.* **14**, 38–45.

Onaitis, M. W., Kirshbom, P. M., Hayward, T. S., Qualye, F. J., Feldman, J. M., Seigler, H. F., and Tyler, D. S. (2000). Gastrointestinal carcinoid: Characterization by site of origin and hormone production. *Ann. Surg.* **232**, 549–556.

Sexually Transmitted Diseases

PETER V. CHIN-HONG AND ROBERT L. OWEN

University of California, San Francisco

darkfield microscopy The most direct method of diagnosing syphilis. The surface of the genital ulcer is first gently abraded with gauze and the serous exudate of the lesion is then expressed onto a glass slide; *Treponema pallidum* has a characteristic corkscrew appearance.

enteritis Infection with small intestinal symptoms such as nausea, bloating, and abdominal pain, but without sigmoidoscopic changes; can follow ingestion of material contaminated with feces.

high-resolution anoscopy Technique similar to cervical colposcopy; uses identical equipment (a powerful light source and binocular lenses) to allow identification and biopsy of lesions that have contributed to abnormal anal cytologic findings. Acetic acid (3%) and Lugol solution (iodine) can be used to identify human papillomavirus-infected tissue in the anal canal as well as in the cervix.

proctitis Inflammation that may be caused by infections of the anorectal mucosa; associated with sigmoidoscopic findings limited to the distal 15 cm of the rectum. A range of cells can be affected, from keratinized, stratified squamous epithelium in the perianal area to columnar epithelium in the rectum.

proctocolitis Inflammation that may be caused by infections of the rectum and colon; sigmoidoscopic findings extend proximally above 15 cm of the rectum.

Sexually transmitted diseases in the intestinal tract include those caused by protozoa, helminths, bacteria, and viruses. Many of these diseases present acutely; others persist and may have malignant potential. Infection is commonly a direct consequence of oral genital contact, oral–anal contact, or anal receptive intercourse. Sexually transmitted enteric infections also reflect a high prevalence of carriage of intestinal pathogens, particularly in certain subpopulations, such as men who have sex with men. Preventative messages of safe sex, screening of at-risk individuals, and vaccines are important medical interventions.

INTRODUCTION

Sexual behavior is a major force in the transmission of bacterial, protozoan, and viral enteric diseases in adults.

Copyright 2004, Elsevier (USA). All rights reserved.

See Also the Following Articles

Appetite • Emesis • Gastric Acid Secretion • Nausea • Obesity, Treatment of • Small Intestine, Benign and Malignant Neoplasms of the

Further Reading

Azmitia, E. C. (2001). Modern views on an ancient chemical: Serotonin effects on cell proliferation, maturation, and apoptosis. *Brain Res. Bull.* **56**, 413–424.

De Ponti, F., and Tonini, M. (2001). Irritable bowel syndrome: New agents targeting serotonin receptor subtypes. *Drugs* **61**, 317–332.

Feldman, J. M. (1989). Carcinoid tumors and the carcinoid syndrome. *Curr. Prob. Surg.* **26**, 834–885.

Gershon, M. D. (1999). Review article: Roles played by 5-hydroxytryptamine in the physiology of the bowel. *Aliment. Pharmacol. Ther.* **13**(Suppl. 2), 15–30.

Hoyer, D., Clark, D. E, Fozard, J. R., Hartig, P. R., Martin, G. R., Mylechrane, E. J., Sxena, P. R., and Humphrey, P. P. (1994). International union of pharmacology classification of receptors for 5-hydroxytryptamine (serotonin). *Pharmacol. Rev.* **46**, 157–203.

Lindley, C., and Blower, P. (2000). Oral serotonin type 3 receptor antagonists for prevention of chemotherapy induced emesis. *Am. J. Health Syst. Pharm.* **57**, 1685–1697.

Meguid, M. M., Fettissov, Varima, M., Sato, T., Zhang, L., Laviano, A., and Rossi-Fanelli, F. (2000). Hypothalamic dopamine and serotonin in the regulation of food intake. *Nutrition* **16**, 843–847.

Oberg, K. (2002). Carcinoid tumors: Molecular genetics, tumor biology, and update of diagnosis and treatment. *Opin. Oncol.* **14**, 38–45.

Onaitis, M. W., Kirshbom, P. M., Hayward, T. S., Qualye, F. J., Feldman, J. M., Seigler, H. F., and Tyler, D. S. (2000). Gastrointestinal carcinoid: Characterization by site of origin and hormone production. *Ann. Surg.* **232**, 549–556.

Sexually Transmitted Diseases

PETER V. CHIN-HONG AND ROBERT L. OWEN

University of California, San Francisco

darkfield microscopy The most direct method of diagnosing syphilis. The surface of the genital ulcer is first gently abraded with gauze and the serous exudate of the lesion is then expressed onto a glass slide; *Treponema pallidum* has a characteristic corkscrew appearance.

enteritis Infection with small intestinal symptoms such as nausea, bloating, and abdominal pain, but without sigmoidoscopic changes; can follow ingestion of material contaminated with feces.

high-resolution anoscopy Technique similar to cervical colposcopy; uses identical equipment (a powerful light source and binocular lenses) to allow identification and biopsy of lesions that have contributed to abnormal anal cytologic findings. Acetic acid (3%) and Lugol solution (iodine) can be used to identify human papillomavirus-infected tissue in the anal canal as well as in the cervix.

proctitis Inflammation that may be caused by infections of the anorectal mucosa; associated with sigmoidoscopic findings limited to the distal 15 cm of the rectum. A range of cells can be affected, from keratinized, stratified squamous epithelium in the perianal area to columnar epithelium in the rectum.

proctocolitis Inflammation that may be caused by infections of the rectum and colon; sigmoidoscopic findings extend proximally above 15 cm of the rectum.

Sexually transmitted diseases in the intestinal tract include those caused by protozoa, helminths, bacteria, and viruses. Many of these diseases present acutely; others persist and may have malignant potential. Infection is commonly a direct consequence of oral genital contact, oral–anal contact, or anal receptive intercourse. Sexually transmitted enteric infections also reflect a high prevalence of carriage of intestinal pathogens, particularly in certain subpopulations, such as men who have sex with men. Preventative messages of safe sex, screening of at-risk individuals, and vaccines are important medical interventions.

INTRODUCTION

Sexual behavior is a major force in the transmission of bacterial, protozoan, and viral enteric diseases in adults.

Copyright 2004, Elsevier (USA). All rights reserved.

With the advent of the acquired immune deficiency syndrome, a sexually transmitted disease (STD) with an important enteric mechanism of infection, emphasis on safe sex temporarily reduced the incidence of sexually transmitted enteric diseases, but has not dramatically altered the importance of sexual transmission of the specific organisms. Indeed, there is a disturbing recent trend of increasing enteric STDs, particularly among men who have sex with men (MSM) in major metropolitan areas. Furthermore, unprotected anal sex is common in other populations, such as injection drug users and heterosexual adolescents and adults of both sexes, but they are not often targeted for safe anal sex messages.

APPROACH TO THE PATIENT

Any patient who presents with intestinal complaints and a history of high-risk sexual activity must first be considered for *human immunodeficiency virus* (HIV) testing. Concomitant with this, an approach for the diagnosis of intestinal and anorectal symptoms focuses on symptom complexes of enteritis, proctocolitis, and proctitis (Table I).

PROTOZOAL AND HELMINTHIC INFECTIONS

Amebiasis

The protozoa *Entamoeba histolytica* is the causative agent in intestinal amebiasis. Although most infection is asymptomatic, amebic dysentery and other extraintestinal manifestations of infection, such as liver abscess, can occur. Sexual transmission is not the predominant means of acquisition—exposure to contaminated food and water is—but individuals who practice oral–anal sex are at higher risk. Cysts are the infectious form and as little as one cyst can cause infection after ingestion. Patients usually present subacutely over 1 to 3 weeks with diarrhea, abdominal pain, and bloody stools. Fever can occur in a minority of patients. There is also a chronic form of the disease that may mimic inflammatory bowel disease. A diagnosis may be made by stool examination, but this does not distinguish between *E. histolytica* and the less pathogenic *Entamoeba dispar*. Antigen detection assays have the highest sensitivity and provide the most information but are not available in all centers. First-line treatment is metronidazole (750 mg, orally for 10 days), which acts by treating both the invading trophozoites and the intraluminal cysts. A second luminal cysticidal agent, such as paromomycin (30 mg/kg per day, orally in three divided doses for 5 to 10 days) or diiodohydroxyquin (iodoquinol) (650 mg, orally three times daily for 20 days), is usually recommended.

Giardiasis

Giardia lamblia, a flagellated protozoan, is an important cause of diarrhea in the United States. Direct person-to-person transfer can occur during sex. MSM have a higher prevalence of giardiasis, with anal intercourse as a risk factor. Infection may also be acquired by contaminated food and water. Presentation is highly variable and many cases are asymptomatic. Less than 50% of those infected have acute giardiasis, which is marked by the sudden onset of watery diarrhea that is foul smelling and is associated with abdominal

TABLE I Common Sexually Transmitted Gastrointestinal Syndromes

Variable	Characteristics of syndrome		
	Proctitis	Protocolitis	Enteritis
Symptoms	Rectal pain, discharge, tenesmus	Proctitis symptoms plus cramps, diarrhea	Diarrhea, cramps, bloating, nausea
Pathogen(s)	*Neisseria gonorrhoeae, Chlamydia trachomatis, Treponema pallidum*, herpes simplex virus	*Entamoeba histolytica, Campylobacter jejuni, Shigella flexneri, C. trachomatis* (LGV[a])	*Giardia lamblia*
Mode of acquisition	Receptive anal intercourse	Direct or indirect fecal–oral contact	Direct or indirect fecal–oral contact
Anoscopic findings	Rectal exudate ± friability	Rectal exudate, friability that may extend into the sigmoid colon	Normal

[a] Lymphogranuloma venereum strains.

From Rompalo, A. M. (1999). Diagnosis and treatment of sexually acquired proctitis and proctocolitis: An update. *Clin. Infect. Dis.* **28**(Suppl. 1), S84–S90, with permission from The University of Chicago Press.

cramps. Fever occurs in only 10% of patients. Diagnosis is usually via stool microscopy; cysts or trophozoites may be detected in 90% of cases when three stool specimens are submitted. *Giardia* antigen enzyme-linked immunoassay (ELISA) may provide a higher yield. Patients with symptomatic disease should be treated. Metronidazole (250 mg, orally three times a day for 5 days) is the preferred treatment. Albendazole is an alternative agent.

HELMINTHS

Only helminths that do not require a period of maturation outside the host are susceptible to direct person-to-person transmission. *Enterobius* (pinworm) and *Strongyloides* species have been associated with sexual transmission. *Taenia solium* (pork tapeworm), *Taenia saginata* (beef tapeworm), and *Hymenolepsis nana* may be transmitted through oral–anal contact as well.

BACTERIAL INFECTIONS

Gonorrhea

Gonorrhea is a common sexually transmitted disease in the United States and an etiologic agent of pharyngitis and proctitis. It is also an important cause of urethritis in men and cervicitis in women, with sequelae of pelvic inflammatory disease and infertility. Subphrenic gonococcal infection (Fitz–Hugh–Curtis syndrome) may arise by contiguous spread from infected fallopian tubes through the peritoneal cavity. Although the early behavioral response to the HIV epidemic in MSM resulted in lower rates of gonorrhea, there has been a recent trend in increasing rates of anorectal gonorrhea in some communities. Anorectal gonorrhea, which is often asymptomatic, occurs in men and women who practice anal intercourse. Patients may present with purulent rectal discharge, constipation, tenesmus, and pain. Gonorrhea-associated pharyngitis is frequently asymptomatic, with pharyngeal exudates as the only evidence of infection. Diagnosis is usually via culture on Thayer–Martin agar. Unlike urethritis, diagnosis by gram stain is less reliable in extragenital infection. DNA amplification methods such as the ligase chain reaction (LCR) initially were approved for diagnosis of urethral gonorrhea but have promise for rectal and pharyngeal gonorrhea as well. Current treatment recommendations include third-generation cephalosporins (cefixime, 400 mg, orally, or ceftriaxone, 125 mg, intramuscularly). Fluoroquinolones are no longer recommended in some parts of the United States and other countries due to increasing microbial resistance. Coinfection with *Chlamydia* is assumed and the Centers for Disease Control and Prevention (CDC) recommend concomitant treatment. All sexual partners who have had contact with the patient in the last 60 days should be offered treatment.

Syphilis

Syphilis is a chronic disease caused by *Treponema pallidum*. Except for perinatal transmission, virtually all cases of syphilis are acquired from sexual transmission. During sexual activity, microscopic tears enable the treponeme to invade the subcutaneous tissue, where an initial chancre develops. Secondary syphilis corresponds to the subsequent treponemia that occurs weeks to months later, despite a host immune response. The patient may then undergo an asymptomatic period called latent disease. Finally, late or tertiary disease may occur in untreated patients. The manifestations of disease are famously protean. They range from a painless chancre in primary syphilis (early disease) to central nervous system involvement and aortitis in late disease. Gastrointestinal involvement can occur at any stage of the disease, from oral and anorectal chancres in primary disease, mucocutaneous patches (oral, gastric, and rectal) and hepatitis in secondary disease, and gummatous syphilis, which can affect any organ in late disease. Diagnosis of disease is mainly by serologic testing, though the organism can also be visualized using dark-field microscopy in the early (primary and secondary) stages of disease. There are two types of serology testing. Nontreponemal tests, such as the Venereal Disease Research Laboratory (VDRL) test and the Rapid Plasma Reagin (RPR) test, are typically used for screening and the reported titer can be followed. Treponemal tests, such as the fluorescent treponemal antibody absorption (FTA-ABS) test and the microhemagglutination test for antibodies to *Treponema pallidum* (MHA-TP), are used as confirmatory tests when the screening tests are positive. Penicillin remains the drug of choice for treatment, but the choice of formulation and the duration of therapy depend on the stage of disease diagnosed. All forms of early syphilis can be treated by a single dose of benzathine penicillin G (2.4 million units, intramuscularly). This dose of benzathine penicillin G weekly for 3 weeks is the recommended treatment for syphilis of unknown stage and for late syphilis, with the exception of neurosyphilis. For neurosyphilis, intravenous penicillin G (3–4 million

units, intravenously every 4 hours for 10 to 14 days) is the regimen of choice.

Chlamydia

Chlamydia trachomatis is the most common bacterial sexually transmitted infection in men and women. It is a rare cause of proctitis and perihepatitis. The *Chlamydia*-associated genital ulcer disease, lymphogranuloma venereum (LGV), is predominantly a disease of tropical and subtropical areas of the world and can progress to a particularly severe form of proctitis. This is associated with certain serovars in individuals who have had anal intercourse. Anorectal fibrosis, strictures, perirectal abscesses, and fistulas can be late complications in untreated disease. Outside the gastrointestinal tract, urethritis (in men and women) and cervicitis and pelvic inflammatory disease (in women) are diagnosed more frequently. Because chlamydial infections are so prevalent, serologic tests such as complement fixation or microimmunofluorescence are of limited use in confirming a causal role for *Chlamydia* in a specific symptom complex. However, an appropriate clinical presentation combined with a positive serologic test is usually adequate for a presumptive diagnosis of LGV. Serial titers demonstrating a titer rise are confirmatory. Chlamydial urethritis and cervicitis can be diagnosed by nucleic acid amplification tests (LCR or polymerase chain reaction). Culture, antigen detection tests, and genetic probe methods are alternative methods of diagnosis. Azithromycin (1 g, orally as a single dose; may not be effective in LGV) or doxycycline (100 mg, orally twice a day for 7 days; 21 days in LGV) is the treatment regimen of choice.

Shigella

Shigella species are an important cause of diarrhea worldwide. Shigellosis is caused by ingestion of contaminated food and water or by direct person-to-person spread, including sexual transmission, perhaps because only 10 to 100 organisms are needed before infection occurs. After ingestion, organisms travel through the stomach to the small intestines and the colon. Symptoms may include the abrupt onset of fever, nausea, and crampy diarrhea, which may be watery or contain blood, mucus, and pus. Diagnosis is by culture. Ciprofloxacin (500 mg, orally twice a day for 3 days) is the treatment of choice.

Salmonella

Salmonella species are gram-negative bacilli that cause a spectrum of disease, ranging from diarrhea and enteric fever to bacteremia, osteomyelitis, and abscesses. Most transmission is via contaminated food and water; *Salmonella* species are very rarely transmitted by sexual contact. This is possibly because as many as 10^4 to 10^6 organisms are required before infection. However, *Salmonella* has been implicated as a cause of enteritis in MSM.

Campylobacter

Campylobacter species are curved, motile gram-negative rods that are one of the most common causes of acute diarrhea in the United States. Infection is typically through contaminated food and water, commonly chicken or dairy products. These organisms have also been identified as one etiology of proctocolitis in MSM.

VIRAL INFECTIONS

HIV

Sexual transmission accounts for most HIV acquisition worldwide. Anal intercourse is the predominant means of transmission among MSM and an important means of acquiring infection by heterosexual men and women. Patients infected with HIV often have gastrointestinal symptoms, including thrush and diarrhea. Dysphagia, odynophagia, abdominal pain, hepatobiliary disease, and anorectal disease can be common in patients with advanced disease. However, with the advent of complex combined antiretroviral and other therapies, adverse effects of antiretroviral therapy are often an important explanation for these symptoms.

Herpes Simplex Virus

Sexual transmission of herpes simplex virus (HSV) occurs by direct skin-to-skin contact. HSV-1 typically affects the oral cavity, skin, eyes, central nervous system, and liver, whereas HSV-2 is mainly implicated in anogenital infections, though either serotype can cause infection in any location. Disease is either primary or recurrent because the virus goes into latency after initial infection. Most HSV-1 infection is asymptomatic. When diagnosed, presentation is usually sudden, with groups of painful vesicles with an erythematous base. Oral infection can present as an exudative pharyngitis in adults. Lesions may occur anywhere in the oral mucosa. In most patients, the disease is self-limited and resolves after 10–14 days. HSV-2 commonly causes anorectal disease. Initial presentation is usually more severe than recurrent disease, with fever, severe pain, and constipation. HIV-positive

patients may have severe acute, recurrent, or chronic disease, with involvement anywhere along the gastrointestinal tract, presenting as odynophagia, gastrointestinal bleeding, anorectal pain, and occasionally colitis. HSV hepatitis is a rare disease that can be caused by either serotype. Patients with HSV hepatitis are generally immunocompromised, such as transplant patients on antirejection medications and HIV-positive patients. Presentation is usually fulminant and with a high fatality rate. Diagnosis is often clinical, with confirmation by viral culture or polymerase chain reaction (PCR), which is sensitive but has limited availability. Treatment depends on whether disease is primary or recurrent, the location of infection, and the immune status of the host. Acyclovir, famciclovir, or valacyclovir is typically used to shorten the duration of symptoms, to reduce pain, and for prophylaxis, but does not reduce asymptomatic shedding and transmission. Vaccines have the potential to stem transmission of disease and reduce the frequency and severity of recurrent disease but are not yet clinically available.

Human Herpesvirus 8

Human herpesvirus-8 (HHV-8), or Kaposi's sarcoma-associated herpesvirus (KSHV), is a novel herpesvirus identified in 1994. It is thought to be transmitted primarily through sexual contact and to be the etiologic agent of Kaposi's sarcoma (KS), body cavity-based lymphoma (a variant of non-Hodgkin's lymphoma), and perhaps multicentric Castleman's disease. The precise method of transmission is unknown but virus has been found in saliva and to a lesser extent in semen, the female genitourinary tract, the gastrointestinal tract, and the prostate. There are also reports of the virus being transmitted by organ transplantation. Diagnosis has not yet been standardized but PCR and serologic methods have been developed. Although the best treatment options for HHV-8 are still unknown, it has been observed that highly active antiretroviral therapy (HAART) is associated with regression of KS lesions.

Human Papillomavirus

Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States. It is the etiologic agent of anogenital, oral, respiratory, and skin condylomata; anal and cervical intraepithelial neoplasia; and anogenital malignancy, including the anus and the cervix. Diagnosis is frequently made by clinical appearance of exophytic warts. However, for most cervical and anal intraepithelial neoplasia, cytology is usually the first diagnostic step, followed by colposcopy or high-resolution anoscopy-directed biopsy for histopathologic confirmation. The application of acetic acid and iodine increases the sensitivity of colposcopy to detect lesions. PCR and hybrid capture assays can determine the HPV type. These tests are not used routinely in clinical practice but are currently being investigated. Treatment of warts is usually for cosmetic reasons, though occasionally they may cause symptoms and warrant removal in the anogenital, oral, and respiratory areas. High-grade anal and cervical lesions are ablated in a variety of ways to prevent invasive cancer. Low-grade disease is followed closely. Both therapeutic and preventative vaccines are currently under development.

Hepatitis A

Hepatitis A virus (HAV) is an RNA virus transmitted via the oral–fecal route, either by ingesting contaminated food and water or by sexual transmission. Infection usually results in an acute, self-limited disease. Prodromal symptoms of malaise, nausea, vomiting, and fever lead to marked jaundice and significant elevations in aminotransferases (to over 1000 μ/liter). Only a small proportion of cases progress to fulminant disease, with an increased risk in patients with chronic hepatitis C or other underlying liver diseases. Diagnosis is by serology, with a positive serum immunoglobulin M (IgM) anti-HAV in acute infection. Treatment is mainly supportive but transplantation may be necessary in fulminant disease. For susceptible adults, preventative hepatitis A vaccine is important in at-risk populations, especially in MSM.

Hepatitis B

Sexual transmission remains the primary mode of infection by hepatitis B virus (HBV)in the developed world. Perinatal transmission, infection during childhood, injection drug use, and transfusions are other methods of acquiring infection. Most infection is subclinical. Individuals who develop symptoms may present with malaise, fatigue, and right upper quadrant tenderness, followed by jaundice with aminotransferases values over 1000 μ/liter. Less than 1% of patients will develop fulminant disease. The rate of progression to chronic hepatitis is related to the age of infection; in adults, this is less than 5%. The diagnosis of HBV infection can also be made by serology with the detection of hepatitis B surface antigen (HBsAg). HBV DNA by PCR is often used to assess response to therapy. Treatment of acute disease is generally supportive. Selected patients

with chronic disease may benefit from lamivudine and the newer antivirals that are being developed. Preventative hepatitis B vaccines remain the best strategy for control of disease and are widely available.

Hepatitis C

The risk of sexual transmission of hepatitis C virus (HCV) is thought to be low, but partners of patients infected with hepatitis C, MSM, and individuals with multiple sexual partners have an increased risk of infection. Most individuals have acquired infection parentally through injection drug use or via contaminated blood transfusions. However, more than 40% of newly infected patients have no identifiable risk factor. Most patients with acute infection are asymptomatic and most acute infections will become chronic. Diagnosis is by serology for detection of anti-HCV antibodies. Detection of HCV RNA by PCR is used increasingly to confirm the diagnosis and to assess response to therapy of hepatitis C. Treatment options in selected patients with chronic disease include combination therapy with pegylated interferon and ribavirin.

See Also the Following Articles

AIDS, Biliary Manifestations of • AIDS, Gastrointestinal Manifestations of • AIDS, Hepatic Manifestations of • *Campylobacter* • Giardiasis • Helminth Infections • Hepatitis A • Hepatitis B • Hepatitis C • Proctitis and Proctopathy • *Salmonella* • *Shigella*

Further Reading

Centers for Disease Control (2003). Web site: http://www.cdc.gov.

Chin-Hong, P., and Palefsky, J. (2002). Natural history and clinical management of anal HPV disease in men and women infected with human immunodeficiency virus. *Clin. Infect. Dis.* **35**, 1127–1134.

Martin, J. N., Ganem, D. E., Osmond, D. H., Page-Shafer, K. A., Macrae, D., and Kedes, D. H. (1998). Sexual transmission and the natural history of human herpesvirus 8 infection. *N. Engl. J. Med.* **338**, 948–954.

Owen, R. L. (1998). Proctitis and sexually transmissible intestinal diseases. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management" (M. Feldman, B. Scharschmidt, and M. Sleisenger, eds.), 6th Ed., pp. 1943–1960. W. B. Saunders, Philadelphia.

Quinn, T. C. (1986). Clinical approach to intestinal infections in homosexual men. Med. *Clin. North Am.* **70**, 611–634.

Quinn, T. C., Stamm, W. E., Goodell, S. E., Mkrtichian, E., Benedetti, J., Corey, L., Schuffler, M. D., and Holmes, K. K. (1983). The polymicrobial origin of intestinal infections in homosexual men. *N. Engl. J. Med.* **309**, 576–582.

Rompalo, A. M. (1999). Diagnosis and treatment of sexually acquired proctitis and proctocolitis: An update. *Clin. Infect. Dis.* **28**(Suppl. 1), S84–S90.

Royce, R. A., Sena, A., Cates, W., Jr., and Cohen, M. S. (1997). Sexual transmission of HIV. *N. Engl. J. Med.* **336**, 1072–1078.

Shigella

JÉRÔME VIALA AND DANA J. PHILPOTT
Institut Pasteur, Paris

bacillary dysentery Diarrheal illness caused by bacteria belonging to the genus *Shigella*.
pathogenicity island Mobile genetic element, often encoding virulence genes; found in many pathogenic bacteria and propagated by horizontal gene transfer.
type III secretion apparatus Macromolecular structure on the surface of some gram-negative pathogenic bacteria; required for the direct translocation of bacterial virulence proteins into the cytosol of the host cell.

Bacillary dysentery is a major cause of morbidity and mortality throughout the world, especially in infants and young children in developing countries. *Shigella* spp. are the causative agents of this disease. Bacillary dysentery is a major public health problem in light of the highly contagious nature of this infection, the emergence of multiple antibiotic resistance, and the lack of an effective vaccine.

MICROBIOLOGY

Shigella spp. belong to the family Enterobacteriaceae and are closely related to *Escherichia coli*. They are gram-negative, nonmotile, and nonencapsulated bacilli. The four species, *Shigella dysenteriae, Shigella flexneri, Shigella boydii*, and *Shigella sonnei*, are differentiated by lipopolysaccharide (LPS) antigens (A, B, C, and D, respectively), biochemical properties, and phage or colicin susceptibility. Different serotypes subdivide each species.

Virulence Factors

The infectivity of *Shigella* spp. is dependent on bacterial entry into host cells. The invasive phenotype depends on the presence of a large plasmid; strains that lack this plasmid are no longer invasive or virulent. Proteins encoded by genes contained within a pathogenicity island in the plasmid induce bacterial entry into host cells. These proteins form a type III secretion apparatus that transfers a number of virulence proteins directly from the bacteria into the host cytosol, to induce a form of macropinocytosis that effectuates bacterial invasion. Once inside the cell, *Shigella* moves within the cytosol and from cell to cell using actin-based motility (see Fig. 1).

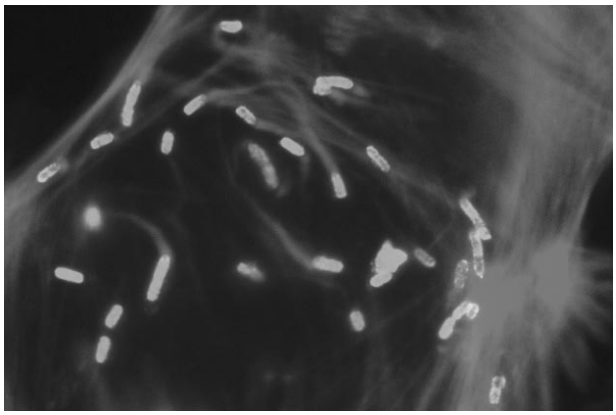


FIGURE 1 *Shigella* is able to recruit host cell actin to form an actin comet tail, necessary for intra- and intercellular propulsion. Here, primary cultures of mouse enterocytes are infected by *S. flexneri*, which express green fluorescent protein (GFP); the bacterial-associated actin comet tails are stained with phalloidin/tetraethylrhodamine isothiocyanate. Micrograph courtesy of Drs. Rafika Athman and Sylvie Robine, Institut Curie, Paris, France.

The virulence of *S. dysenteriae* type 1 is compounded by the expression of a potent cytotoxin, called shiga toxin. This toxin is an AB subunit toxin that mediates cell death through inhibition of protein synthesis. Shiga toxin effectively targets those cells that express globotriaosyl ceramide (Gb_3), which is the receptor for the toxin. Shiga toxin is important for the development of hemolytic uremic syndrome (HUS; see later).

Immune Response

Shigella infection induces local innate immune defense systems, including the recruitment of bactericidal neutrophils to the infected site. In terms of adaptive immunity, the production of secretory immunoglobulin (IgA) directed against the O antigen of

 Copyright 2004, Elsevier (USA). All rights reserved.

lipopolysaccharide has been shown to be protective in animal models of shigellosis. However, this immunity is relatively short-lived and serotype specific, which, together, greatly hampers the successful design of effective vaccine candidates.

EPIDEMIOLOGY

Shigella is exclusively a human pathogen and is transmitted by the fecal–oral route through close personal contact or by way of infected food or water. In contrast to the other enteropathogens, *Shigella* is highly contagious; as few as 200 *S. flexneri* organisms are sufficient to induce diarrhea and fever. For comparison, similar symptoms and a similar attack rate due to *Salmonella* require an inoculum 100 times greater.

In developed countries, shigellosis causes disease primarily in custodial institutions, nursing homes, or day-care centers. Each year, 15,000 cases of shigellosis, essentially due to *S. sonnei*, are reported in the United States. Shigellosis is common in developing countries where poverty, overcrowding, poor hygiene, and malnutrition prevail. The World Health Organization (WHO) estimates that shigellosis causes 160 million cases of diarrhea each year and 1.1 million deaths worldwide. *Shigella flexneri*, and, to a lesser extent, *S. sonnei*, are responsible for endemic disease whereas epidemic outbreaks are due to *S. dysenteriae* type 1. *Shigella* infection is a disease essentially of children less than 5 years of age.

DIAGNOSIS

Clinical Diagnosis

After an asymptomatic incubation period of 1–7 days, shigellosis begins suddenly with abdominal pain, vomiting, anorexia, and fever. Soon after, watery diarrhea develops; this becomes bloody in 50% of patients and presents with tenesmus and abdominal cramps. These intestinal contractions may induce rectal prolapse. In healthy volunteers without antibiotic treatment, diarrhea abates by the seventh day, although stool cultures remain positive for an average of 27 days.

Paraclinical Exams

Leukocytosis is often mild but can become leukemoid (more than 50 g/liter). The examination of fresh stool stained by methylene blue can reveal significant leukocyte counts, consistent with a nonspecific invasive bacterial pathogen. Stool culture is the most informative exam.

Complications

Complications occur more frequently among very young or malnourished children in developing countries. During the acute phase, metabolic disorders such as hypoglycemia and hyponatremia can be observed. Dehydration and septicemia are rare during shigellosis. Among digestive complications, toxic megacolon and intestinal perforation are more frequently observed with *S. dysenteriae* type 1, which can result in a high mortality rate. Neurological complications can also occur. For example, seizures may occur in children before fever appears and *S. flexneri* has been documented to cause encephalopathy without hypoglycemia and bacterial meningitis.

Shigella infection can induce protein-losing enteropathy, which may be responsible for growth retardation in children. Immune deregulation may cause Reiter's syndrome in HLA-B27 patients in association with arthritis, urethritis, and conjunctivitis. HUS can be a complication following infection with *S. dysenteriae* type 1, due to shiga toxin. HUS is characterized by acute hemolytic anemia, thrombocytopenia, and renal failure. Both hemolytic uremic and Reiter's syndromes can occur during the convalescence phase after diarrhea has subsided.

TREATMENT

Curative

Except for symptomatic treatment to control hypoglycemia, hyponatremia, or seizures, the main therapy of shigellosis is antibiotics. In developed countries, a majority of *Shigella* infections are mild and self-limited and often do not require antibiotic treatment. However, in severe shigellosis, several antibiotics have been demonstrated to reduce the duration of the disease and eliminate *Shigella* from the stool. However, *Shigella* spp. have acquired resistance to a number of antibiotics, including tetracycline, ampicillin, trimethoprim-sulfamethoxazole, and fluoroquinolone. The choice of antibiotic should be adapted to the local epidemiology of resistance and to the results of the stool culture, if available.

Preventive

Person-to-person transmission of *Shigella* can be reduced by hand washing with soap, segregating the ill persons, and separating eating areas from care zones. Increased sanitary conditions in developing countries will certainly decrease the incidence of this

infection. Efficient vaccines are not currently available; however, the development of a multivalent vaccine effective against the most prevalent species of *Shigella* is a major priority for the WHO.

See Also the Following Articles

Bacterial Toxins • Foodborne Diseases • Food Poisoning • Food Safety • *Salmonella* • *Yersinia*

Further Reading

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., and Sansonetti, P. J. (1999). Global burden of *Shigella* infections: Implications for vaccine development and implementation of control strategies. *Bull. World Health Organ* **77**(8), 651–666.

Philpott, D. J., Edgeworth, J. D., and Sansonetti, P. J. (2000). The pathogenesis of *Shigella flexneri* infection: Lessons from *in vitro* and *in vivo* studies. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **355**(1397), 575–586.

Short Bowel Syndrome

THERESA BYRNE*,†, SUZANNE COX*, MARIA KARIMBAKAS*, AND DOUGLAS WILMORE†

**Nutritional Restart Center, Wellesley, Massachusetts, and †Harvard Medical School, Boston*

cholerrheic diarrhea Resulting from the presence of malabsorbed bile acids.

dual-energy X-ray absorptiometry Noninvasive technique used to assess body composition, specifically bone and soft tissue.

hyperosmolar fluid Highly concentrated solution that contains several osmotically active particles, such as glucose or sodium.

isotonic fluid Solution that contains electrolytes, nonelectrolytes, or a combination of both, having the same concentration as the solution to which it is being compared (e.g., blood).

polymeric Containing complete proteins, carbohydrates, and fat rather than predigested substances.

probiotics Live microbial supplements; used to reestablish normal intestinal flora.

secretory diarrhea Resulting from the excessive secretion of water and electrolytes into the lumen of the bowel; persists or slows only partially after 24–48 hours of fasting.

Short bowel syndrome is a relatively rare, but devastating, clinical problem characterized by severe diarrhea, malabsorption, fluid and electrolyte abnormalities, and progressive malnourishment resulting from the loss of functional small bowel absorptive surface area. Although the care of patients has often focused on minimizing symptoms and the appropriate replacement of fluid and nutrient losses, therapeutic advancements and comprehensive treatment programs offer new options.

ANATOMIC AND PHYSIOLOGIC FACTORS

Estimates of adult small intestinal length vary from 12 to 20 feet (or 365 to 600 cm). Length has been found to vary with the height and sex of the individual, being slightly longer in men. The duodenum is the first segment of the small bowel and measures approximately 10 inches (25 cm). The remainder of the small bowel is composed of the jejunum and ileum; with the more proximal jejunum comprising approximately two-fifths of this length and the ileum comprising three-fifths. Throughout the small bowel, the mucosa is composed of convoluted folds containing fingerlike projections (villi) that protrude into the lumen. These villi are further augmented by the presence of microvilli (approximately 2×10^8 per square centimeter) on the outer, or brush border, region of the epithelial cells, contributing to the enormous total absorptive surface area of the adult small intestine.

Many substances are absorbed throughout the length of the small intestine, but certain nutrients

 Copyright 2004, Elsevier (USA). All rights reserved.

infection. Efficient vaccines are not currently available; however, the development of a multivalent vaccine effective against the most prevalent species of *Shigella* is a major priority for the WHO.

See Also the Following Articles

Bacterial Toxins • Foodborne Diseases • Food Poisoning • Food Safety • *Salmonella* • *Yersinia*

Further Reading

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., and Sansonetti, P. J. (1999). Global burden of *Shigella* infections: Implications for vaccine development and implementation of control strategies. *Bull. World Health Organ* **77**(8), 651–666.

Philpott, D. J., Edgeworth, J. D., and Sansonetti, P. J. (2000). The pathogenesis of *Shigella flexneri* infection: Lessons from *in vitro* and *in vivo* studies. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **355**(1397), 575–586.

Short Bowel Syndrome

THERESA BYRNE*,†, SUZANNE COX*, MARIA KARIMBAKAS*, AND DOUGLAS WILMORE†

**Nutritional Restart Center, Wellesley, Massachusetts, and †Harvard Medical School, Boston*

cholerrheic diarrhea Resulting from the presence of malabsorbed bile acids.

dual-energy X-ray absorptiometry Noninvasive technique used to assess body composition, specifically bone and soft tissue.

hyperosmolar fluid Highly concentrated solution that contains several osmotically active particles, such as glucose or sodium.

isotonic fluid Solution that contains electrolytes, nonelectrolytes, or a combination of both, having the same concentration as the solution to which it is being compared (e.g., blood).

polymeric Containing complete proteins, carbohydrates, and fat rather than predigested substances.

probiotics Live microbial supplements; used to reestablish normal intestinal flora.

secretory diarrhea Resulting from the excessive secretion of water and electrolytes into the lumen of the bowel; persists or slows only partially after 24–48 hours of fasting.

Short bowel syndrome is a relatively rare, but devastating, clinical problem characterized by severe diarrhea, malabsorption, fluid and electrolyte abnormalities, and progressive malnourishment resulting from the loss of functional small bowel absorptive surface area. Although the care of patients has often focused on minimizing symptoms and the appropriate replacement of fluid and nutrient losses, therapeutic advancements and comprehensive treatment programs offer new options.

ANATOMIC AND PHYSIOLOGIC FACTORS

Estimates of adult small intestinal length vary from 12 to 20 feet (or 365 to 600 cm). Length has been found to vary with the height and sex of the individual, being slightly longer in men. The duodenum is the first segment of the small bowel and measures approximately 10 inches (25 cm). The remainder of the small bowel is composed of the jejunum and ileum; with the more proximal jejunum comprising approximately two-fifths of this length and the ileum comprising three-fifths. Throughout the small bowel, the mucosa is composed of convoluted folds containing fingerlike projections (villi) that protrude into the lumen. These villi are further augmented by the presence of microvilli (approximately 2×10^8 per square centimeter) on the outer, or brush border, region of the epithelial cells, contributing to the enormous total absorptive surface area of the adult small intestine.

Many substances are absorbed throughout the length of the small intestine, but certain nutrients

 Copyright 2004, Elsevier (USA). All rights reserved.

tend to be absorbed more in one region than another. The proximal intestine is the major site for the absorption of iron, calcium, water-soluble vitamins, monoglycerides, and simple fatty acids. Sugars are absorbed in the proximal and midintestine. Amino acids appear to be absorbed primarily in the midintestine or the jejunum, although some absorption also occurs in the proximal and distal segments. Intestinal intubation studies have shown that the absorption of carbohydrate, protein, and simple fatty acids is primarily complete within the first 100 cm of the jejunum. The distal small bowel (ileum) appears to be the major absorptive area for vitamin B_{12} and for bile salts. The bile salts play an important role in the absorption of fat (or triglycerides) and fat-soluble vitamins. Although the colon is an important site for the absorption of water and electrolytes, the small bowel also plays an important role. Under normal circumstances, the intestine is presented with ~2000 ml of ingested fluids and ~7000 ml of secretions from the gastrointestinal tract and associated glands. Approximately 98% of this fluid is reabsorbed (with estimates of ~5500 ml in the jejunum, ~2000 ml in the ileum, and ~1300 ml in the colon), with a daily fluid loss of ~200 ml in the stool.

ETIOLOGY

Significant loss or dysfunction of the intestinal absorptive surface area results in the short bowel syndrome. This syndrome is most often the result of extensive resection of the small bowel due to infarction of the mesenteric vessels, intestinal volvulus, trauma, malignancy, congenital abnormalities, or complications of Crohn's disease. Less often the defect is functional, rather than anatomical, such as in the case of radiation enteritis or severe inflammatory bowel disease.

CLINICAL DESCRIPTION

Short bowel syndrome is a complex of symptoms consisting of severe diarrhea and macro- and micronutrient malabsorption resulting in suboptimal hydration, electrolyte disturbances, progressive weight loss, and nutrient deficiencies. These problems can be physically debilitating and socially incapacitating and can require aggressive interventions that can contribute to a myriad of complications. The severity of the short bowel syndrome is determined by a number of factors, including (1) the degree to which the bowel has adapted following resection, (2) the length, location, and health of the remaining small bowel, and (3) the presence or absence of the colon.

FACTORS INFLUENCING SEVERITY

Bowel Adaptation

The process of bowel adaptation is characterized by an elongation and dilation of the remnant bowel, and in animal models an increase in villus height, crypt depth, cell proliferation, and enzyme activity. These alterations in bowel morphology and function are thought to be mediated in part by factors extrinsic to the gastrointestinal tract (hormones, growth factors, prostaglandins, etc.), by local factors brought into play by the provision of oral or enteral feedings (e.g., enteric hormones, pancreatic–biliary secretions), and by exposure of the mucosa to specific nutrients or nonnutrient components of the diet (e.g., short chain fatty acids, glutamine, fiber). In animals, the absence of luminal nutrients inhibits intestinal hyperplasia even when necessary amounts of calories are provided parenterally. Clinically, bowel adaptation is marked by gradual improvements in symptoms—a decrease in diarrhea and an improved tolerance to and absorption of enteral nutrients. Although the process of bowel adaptation begins almost immediately following extensive resection, the process may not be maximized for 1–2+ years.

Length, Location, and Health of Remnant Bowel

Because of the tremendous length of the small intestine and its ability to adapt and compensate for the loss of absorptive surface area, relatively normal intestinal function could be expected after resection of approximately one-third of the bowel. However, resections necessitating the removal of greater than 50% of the small bowel are associated with metabolic complications and often require more aggressive interventions.

In addition to remnant length, the site of the resection influences the clinical sequelae. Loss of the distal small intestine is often more devastating than loss of the more proximal bowel. If the jejunum is removed, the ileum often adapts and assumes its absorptive functions. However, because of the unique functions of the ileum, particularly bile salt absorption, loss of even 100 cm or less of ileum can result in watery, cholerrheic diarrhea. When more than 100 cm of ileum is resected, bile salt loss in the stool can be considerable. Consequently, fewer bile salts are available, limiting the absorption of fat and fat-soluble vitamins, with resultant steatorrhea. The unabsorbed free fatty acids bind with calcium, magnesium, and zinc, forming insoluble intraluminal soaps. The prolonged malabsorption of these substances necessitates supplementation to avoid deficiency states and related metabolic complications. The formation of

unabsorbable calcium soaps prevents intraluminal calcium from binding to dietary oxalates. If colon is present, the unbound oxalates pass to the colon, where they are reabsorbed and subsequently excreted in the urine. A prolonged state of hyperoxaluria renders the patient prone to the development of calcium oxalate nephrolithiasis. An additional factor influencing calcium absorption is a reduction in serum 25-hydroxy vitamin D levels, which can be related to the loss of ileal surface area and the associated fat and fat-soluble vitamin malabsorption. Prolonged suboptimal levels of calcium and vitamin D are thought to contribute to the bone disease that can accompany the short bowel syndrome.

In addition to the absorption of fat and fat-soluble vitamins, the ileum and the colon have a greater ability than the jejunum to conserve salt and water. In addition, the ileum and colon have a marked effect on slowing intestinal transit. Consequently, resections involving only jejunum often result in very little diarrhea, because the remaining ileum and colon can accommodate the fluid and electrolyte load. In contrast, following ileal resections, the colon receives a large, relatively isotonic fluid load. However, the presence of unabsorbed bile salts and free fatty acids can alter the tonicity of the luminal contents and produce a secretory state within the colon. Although the colon can handle $\sim$5 liters of fluid per day, volumes in excess of this can result in diarrhea and excessive fluid and electrolyte losses. For patients with jejunostomies (no functional ileum or colon), the remaining jejunum is often unable to concentrate the luminal contents, and water and sodium loss is often severe.

The health of the remnant bowel influences the severity of the symptoms confronting the patient with short bowel syndrome. Disease (e.g., Crohn's) or damage (e.g., radiation injury) can impair the functioning capacity of the remaining bowel, rendering these patients to more pronounced symptoms and long-term complications.

Presence or Absence of Colon

In addition to the colon's important role in the absorption of fluid and electrolytes, it plays a unique role for patients with short bowel syndrome. Bacteria within the colon ferment malabsorbed carbohydrate (and to a lesser extent protein) into short-chain fatty acids, which can then be utilized for energy. Thus, for patients with short bowel, the presence of the colon is typically a good predictor of a more favorable outcome and less dependency on parenteral support. The minimal amounts of small bowel required to sustain a patient without parenteral support have been estimated to be approximately 50–70 cm, when the segment is anastomosed to functional colon, but 110–150 cm if no colon is present. These estimates assume that the remnant small bowel and colon are healthy, adequate adaptation has occurred, and the patient has received appropriate medical and nutritional care.

MEDICAL AND NUTRITIONAL MANAGEMENT

The primary short- and long-term objectives of standard medical and nutritional management of patients with short bowel syndrome should be (1) to enhance bowel adaptation and compensation, (2) to improve absorption and reduce diarrhea, (3) to replace nutrient and fluid losses appropriately, and (4) to minimize and/or avoid long-term complications.

Acute Postoperative Period

The immediate postoperative phase following extensive intestinal resection is characterized by massive diarrhea and fluid and electrolyte disturbances. During this time, aggressive replacement of fluid and electrolytes is required. Calorie and nutrient requirements are met via parenteral nutrition (PN) and should be prescribed according to previously published guidelines. Very small amounts of luminal nutrition should be initiated as early as possible for the purpose of encouraging bowel adaptation. The composition of the diet should be based on the presence or absence of colon, with the quantity of fat, carbohydrate, and protein evenly distributed throughout the day. Simple sugars (particularly sucrose, fructose, and lactose) and hyperosmolar beverages should be avoided. Oral rehydration solutions can be trialed. If the patient is unable to eat, a polymeric, isotonic liquid formula should be utilized. Initially, the quantity of the feeding should be restricted (e.g., $\leq$ 500 ml per 24-hour period) to avoid further exacerbation of diarrhea. An antimotility agent should be initiated, and antinausea medication utilized, if indicated. Due to acid hypersecretion following resection, antacid therapy should be started. If either secretory or cholerrheic diarrhea is documented, additional antisecretory medication can be considered (Table I). This initial postoperative phase may last for several weeks or even months, depending on the extent of the resection.

Long-Term Management

The second phase following extensive resection is marked by a decrease in diarrhea. The increased

TABLE I Frequently Utilized Medications

Indication	Medication	Dose[a]
Rapid transit	Loperamide	2 mg po qid
	Atropine sulfate	2.5 mg po qid
	Opium tincture	0.25–1.0 ml po qid
Secretory diarrhea	Octreotide	50–150 μg sc tid
Cholerrheic diarrhea	Cholestyramine	4 g po bid–qid
Acid hypersecretion		
H2 receptor antagonists	Ranitidine	150 mg po bid or 150 mg iv qd
	Famotidine	20 mg po bid or 40 mg iv qd
Proton pump inhibitors	Omeprazole	20 mg po bid
	Lansoprazole	30 mg po bid
Antinausea	Compazine	5–10 mg po tid–qid
	Zofran	10 mg iv qd–tid
Bacterial overgrowth	Metronidazole	250 mg po tid[b]
	Tetracycline	250 mg po qid[b]
	Ciprofloxin	500 mg po qd–bid[b]

[a] For adult patients. Abbreviations: po, periorbital; qid, quater in die (four times daily); sc, subcutaneous; tid, ter in die (three times daily); bid, bis in die (twice daily); iv, intravenous; qd, quaque die (once daily).
[b] Typically administered for 10–14 days.

adaptation is accompanied by improvements in nutrient absorption. Although the composition of the oral diet continues to be based on the presence or absence of colon, the volume of food and fluid is liberalized. Parenteral support is gradually reduced and medications are adjusted, as indicated. Following extensive resection, it can take up to 1–2+ years before maximal adaptation is achieved, and during this time the overall treatment plan may need to be adjusted multiple times. Some patients with short bowel syndrome progress to a third phase, one of full adaptation, which is marked by the achievement of positive nutritional balance via oral or enteral nutrition alone. For these patients, as well as those who remain dependent on parenteral support, it is critical that they be monitored periodically to minimize the risk of potential complications associated with short bowel syndrome and/or their long-term need for PN. Recommendations for routine monitoring are provided in Table II.

POTENTIAL LONG-TERM PROBLEMS

Catheter-Related Complications

For those patients with short bowel syndrome that do require long-term PN, catheter-related complications, including catheter occlusion (due to thrombosis) and catheter-related infections, are the most common problem. Prevention of catheter occlusion remains a clinical challenge, because the cause can be

TABLE II Guidelines for Long-Term Monitoring

Parameter	Baseline	Monthly	Semiannual	Annual
Weight	X	X	X	X
Skeletal muscle mass[a]	X		X	X
Electrolytes	X	X	X	X
Vitamin and trace elements[b]	X		X	X
Essential fatty acid profile	X		X	X
Liver function panel[c]	X	X	X	X
Kidney function[d]	X		X	X
Bone health[e]	X			X

[a] Assessed by creatinine height index.
[b] Vitamin and mineral profile to include vitamins A, C, D (25-hydroxy), E, and B_{12}, and folic acid, zinc, selenium, ferritin, and prothrombin.
[c] Liver function panel should include total bilirubin, alkaline phosphatase, serum glutamic oxaloacetic transaminase, and serum glutamic pyruvic transaminase.
[d] Assessed by 24-hour creatinine clearance.
[e] Assessed by dual-energy X-ray absorptiometry.

multifactorial. Efforts to reduce the incidence of septic complications focus on the use of aseptic techniques in catheter placement and maintenance and in solution preparation and administration.

Liver Dysfunction

Although the use of PN may result in altered liver function tests as soon as 1–2 weeks after therapy has been initiated, liver dysfunction is more frequently thought to be a long-term complication. Patients with the shortest residual intestine, particularly those without functioning remnant colon, are at greatest risk of developing eventual liver failure, suggesting that the degree of malabsorption and/or the level of dependence on parenteral nutrition are the likely causes of the dysfunction. Although nutrient deficiencies (e.g., choline, carnitine, glutamine, vitamin E, taurine) and toxicities (e.g., manganese) have been suggested as potential causes of liver dysfunction, dextrose overfeeding, lipid overload, and bacterial overgrowth are more typically associated with the abnormalities. To prevent PN-associated liver disease, carbohydrate overfeeding and the excessive use of lipid emulsion should be avoided, PN should be cycled, oral intake encouraged, and episodes of bacterial overgrowth appropriately treated and managed.

Metabolic Bone Disease

Patients with short bowel syndrome are at risk of developing metabolic bone disease, with the two most common forms being osteoporosis and osteomalacia. In addition to chronic abnormalities in calcium and vitamin D homeostasis due to impaired nutrient absorption, factors such as chronic metabolic acidosis, the need for long-term PN, prolonged periods of inactivity, and limited sun exposure may further contribute to the problem. Efforts to minimize risk of progressive bone loss should include appropriate vitamin D and calcium supplementation and, if appropriate, sunlight exposure and an exercise regimen. Chronic acidosis should be corrected and the use of PN minimized, if possible. Dual-energy X-ray absorptiometry (DEXA) provides a noninvasive method to assess bone mineral density and should be used as a screening device for patients with short bowel syndrome.

Nutrient Deficiencies

Vitamin, mineral, and essential fatty acid deficiencies are common in patients with short bowel syndrome. As previously mentioned, the location and the extent of intestinal resection impacts on the incidence and severity. Deficiencies can occur despite the use of routine supplementation, thus periodic monitoring is recommended (Table II). When deficiencies are identified, they should be repleted (Table III) to avoid potentially debilitating complications.

Other Problems

Bacterial overgrowth and D-lactic acidosis are additional problems that may occur in patients with short bowel syndrome. Bacterial overgrowth is an abnormal proliferation of bacteria; changes in intestinal flora can be induced by alterations in intestinal motility and in diet (e.g., increased intake of simple carbohydrates) and/or the need for antibiotic therapy. Overgrowth of

TABLE III Guidelines for Nutrient Repletion

Nutrient deficiency	Recommended repletion dosages[a]
Vitamin A[b]	po, 10,000–50,000 IU for 1 month; iv, 50,000–100,000 IU for 1–3 days
Vitamin B_{12}	im, 100–1000 μg daily for 1–2 weeks to replace body stores
Vitamin C	po, 250–500 mg qd–bid
Vitamin D[c]	po, 50,000 IU daily for 1 month; im, 500,000 IU in 1–2 injections
Vitamin E	po, 400 IU qd–tid
Vitamin K[d]	po, 5–20 mg qd; iv or sc, 2.5–10 mg and monitor levels
Iron	po, 27–38 mg of elemental iron tid
Zinc	po, 50 mg of elemental zinc qd–bid
Selenium	po, 50–100 μg qd; iv, 20–60 μg
Magnesium[e]	po, 200–600 mg qd; iv or im, 1–2 g if levels are less than 1.0 mEq/liter
Essential fatty acids	po, 1–3 tablespoons of safflower and/or flaxseed oil; iv, 250 ml of 20% lipid in 12 doses

[a] For adult patients. Abbreviations: po, periorbital; qd, quaque die (once daily); sc, subcutaneous; tid, ter in die (three times daily); bid, bis in die (twice daily); iv, intravenous.

[b] As retinol.

[c] As ergocalciferol.

[d] As phytonadione.

[e] As magnesium glycinate.

pathogenic bacteria has the potential to increase diarrhea and to compete with vitamins (e.g., B_{12}) and other nutrients. Treatment usually consists of a 10- to 14-day course of oral antibiotics (Table I) and a probiotic to repopulate the intestinal flora.

Bacterial fermentation of unabsorbed sugars can cause increased D-lactate production in the colon in some patients. These fermenting organisms can lead to D-lactic acidosis, with related neurologic impairment, including confusion, somnolence, unsteady gait, and lethargy. A specific D-lactate level is used to confirm the diagnosis. Treatment includes intravenous fluids and administration of antibiotics to reduce colonic bacterial mass. Patients who may be susceptible to D-lactic acidosis should be strongly encouraged to avoid refined sugars and to decrease their total carbohydrate intake.

Depression and narcotic dependency are additional problems that may confront the patient with short bowel syndrome. These problems should be addressed and appropriate treatment and support provided.

ADDITIONAL TREATMENT OPTIONS

In addition to the use of approved medications and appropriate oral/enteral and parenteral nutrition regimens, other treatment options are being explored and/or utilized. These options include the use of growth factors, nontransplant surgical procedures, and intestinal transplantation.

Growth Factors

Specific growth factors (e.g., growth hormone, glucagon-like peptide-2) have been proposed as potential adjuncts to the standard treatment of patients with short bowel syndrome. Their ability to augment bowel function and/or morphology and thereby reduce PN requirements is the focus of much research. Preliminary studies have produced conflicting results, but results from larger trials are anticipated to help clarify some of the controversy over this therapeutic approach.

Nontransplant Surgical Procedures

In an attempt to increase the absorptive surface area and to minimize problems of bacterial overgrowth, intestinal lengthening procedures have been proposed for some patients with short bowel syndrome. This technically challenging procedure has been reported to yield favorable results in some patients, particularly those with dilated remnants. The procedure has primarily been utilized in children and clinical experience is limited to a small series in a few centers. For patients with longer remnants, reversal of intestinal segments is intended to slow intestinal transit by the interruption of normal peristalsis and the introduction of retrograde motility. Prolongation of intestinal transit and the associated improvement in nutrient absorption have resulted in some patients regaining enteral autonomy or experiencing a reduction in PN requirements. Although some authors report lasting benefit, others report only short-term success. As with all procedures, the potential benefits need to be weighed against the potential risks.

Intestinal Transplantation

Intestinal transplantation has become a treatment option for those patients who develop life-threatening complications associated with long-term dependence on PN. An international registry has tracked the world experience on this procedure and reports indicate that the 1-year graft/patient survival rate for transplants performed after 1995 is 55/69% for intestinal grafts, 63/66% for those who undergo a transplant of both the small bowel and liver, and 63/63% for those who undergo multivisceral grafts. Despite such aggressive intervention, approximately one-quarter of the survivors still required some parenteral support. This procedure is performed more commonly in children and teenagers than in adult patients.

INTESTINAL REHABILITATION

Comprehensive treatment programs for patients with short bowel syndrome have been developed over the past decade in specialized centers that have the opportunity to care for large numbers of patients. Well-defined protocols and the experienced clinical teams allow standard therapeutic interventions to be optimized. Interventions often include daily monitoring and adjustments in nutritional and medication regimens, comprehensive education and behavior modification programs, and long-term monitoring. If indicated, other therapeutic approaches (e.g., the use of growth factors, and nontransplant surgical interventions) may be utilized. The intent of intestinal rehabilitation is to improve the functioning capacity of the remnant bowel with the goal of avoiding significant complications or the need for intestinal transplantation.

See Also the Following Articles

Bacterial Overgrowth • Colonic Absorption and Secretion • Diarrhea • Malabsorption • Parenteral Nutrition • Short

Bowel Syndrome and Intestinal Transplantation, Pediatric • Small Bowel Transplantation • Small Intestine, Absorption and Secretion

Further Reading

American Society for Parenteral and Enteral Nutrition (ASPEN). (2002). Board of Directors and the Clinical Guidelines Task Force. Guidelines for the use of parenteral and enteral nutrition in adult and pediatric patients. *J. Parenter. Enteral Nutr.* **26**, 1SA–138SA.

Buchman, A. (2002). Total parenteral nutrition-associated liver disease. *J. Parenter. Enteral Nutr.* **26**, S43–S48.

Byrne, T., Veglia, L., Camelio, M., Cox, S., Anderson, S., Wilson, J., and Bennett, H. (2000). Beyond the prescription: Optimizing the diet of patients with short bowel syndrome. *Nutr. Clin. Practice* **15**, 306–311.

Grant, D. (1996). Current results of intestinal transplantation. The International Intestinal Transplant Registry. *Lancet* **347**, 1801–1803.

Grant, J. (2002). Recognition, prevention, and treatment of home total parenteral nutrition central venous access complications. *J. Parenter. Enteral Nutr.* **26**, S21–S28.

Seidner, D. (2002). Parenteral nutrition-associated metabolic bone disease. *J. Parenter. Enteral Nutr.* **26**, S37–S42.

Tavakkolizadeh, A., and Whang, E. (2002). Understanding and augmenting human intestinal adaptation: A call for more clinical research. *J. Parenter. Enteral Nutr.* **26**, 251–255.

Thompson, J., and Langnas, A. (1999). Surgical approaches to improving intestinal function in the short-bowel syndrome. *Arch. Surg.* **134**, 706–709.

Short Bowel Syndrome and Intestinal Transplantation, Pediatric

JON A. VANDERHOOF AND ROSEMARY J. YOUNG
University of Nebraska Medical Center, Omaha

adaptation Enhancement of the intestinal structure and function to compensate for loss of small bowel length.

elemental formulas Solutions composed of individual amino acids, containing no whole proteins; used primarily for rapid absorption and reduction of allergic potential.

intestinal lengthening Surgical procedure in which the dilated small intestinal diameter is halved and joined end to end.

rejection Immune reaction of a transplant recipient to foreign tissues (antigens) after allograft transplantation, with production of antibodies and ultimate destruction of the transplanted organ.

short bowel syndrome Clinical state following bowel resection leading to malabsorption of nutrients, fluids, and/or electrolytes.

small bowel bacterial overgrowth Excess bacterial counts in any area of the small bowel, usually greater than 10^{10} colony-forming units/ml.

total parenteral nutrition Provision of complete nutrition via the intravenous route.

transplantation Transfer of living organs from one individual to another.

tropic hormones Play a role in stimulating bowel adaptation by encouraging intestinal villous growth and in regulating intestinal motility.

Quantification of an anatomical definition of short bowel syndrome after surgical resection has been attempted. No consistent definition can be established because children with as little as a few centimeters of small intestine have been successfully weaned from parenteral nutrition whereas some children with longer segments of bowel have succumbed. In adults, it is generally agreed that less than 100 cm of remaining small bowel constitutes short bowel syndrome. A more practical definition focuses on the functional state of the remaining gastrointestinal tract, whereby short bowel syndrome exists when malabsorption of nutrients, fluids, and/or electrolytes occurs in the presence of any intestinal loss.

INTRODUCTION

Short bowel syndrome, although not common in pediatrics, remains a continuous challenge to those involved

 Copyright 2004, Elsevier (USA). All rights reserved.

Bowel Syndrome and Intestinal Transplantation, Pediatric • Small Bowel Transplantation • Small Intestine, Absorption and Secretion

Further Reading

American Society for Parenteral and Enteral Nutrition (ASPEN). (2002). Board of Directors and the Clinical Guidelines Task Force. Guidelines for the use of parenteral and enteral nutrition in adult and pediatric patients. *J. Parenter. Enteral Nutr.* **26**, 1SA–138SA.

Buchman, A. (2002). Total parenteral nutrition-associated liver disease. *J. Parenter. Enteral Nutr.* **26**, S43–S48.

Byrne, T., Veglia, L., Camelio, M., Cox, S., Anderson, S., Wilson, J., and Bennett, H. (2000). Beyond the prescription: Optimizing the diet of patients with short bowel syndrome. *Nutr. Clin. Practice* **15**, 306–311.

Grant, D. (1996). Current results of intestinal transplantation. The International Intestinal Transplant Registry. *Lancet* **347**, 1801–1803.

Grant, J. (2002). Recognition, prevention, and treatment of home total parenteral nutrition central venous access complications. *J. Parenter. Enteral Nutr.* **26**, S21–S28.

Seidner, D. (2002). Parenteral nutrition-associated metabolic bone disease. *J. Parenter. Enteral Nutr.* **26**, S37–S42.

Tavakkolizadeh, A., and Whang, E. (2002). Understanding and augmenting human intestinal adaptation: A call for more clinical research. *J. Parenter. Enteral Nutr.* **26**, 251–255.

Thompson, J., and Langnas, A. (1999). Surgical approaches to improving intestinal function in the short-bowel syndrome. *Arch. Surg.* **134**, 706–709.

Short Bowel Syndrome and Intestinal Transplantation, Pediatric

JON A. VANDERHOOF AND ROSEMARY J. YOUNG
University of Nebraska Medical Center, Omaha

adaptation Enhancement of the intestinal structure and function to compensate for loss of small bowel length.

elemental formulas Solutions composed of individual amino acids, containing no whole proteins; used primarily for rapid absorption and reduction of allergic potential.

intestinal lengthening Surgical procedure in which the dilated small intestinal diameter is halved and joined end to end.

rejection Immune reaction of a transplant recipient to foreign tissues (antigens) after allograft transplantation, with production of antibodies and ultimate destruction of the transplanted organ.

short bowel syndrome Clinical state following bowel resection leading to malabsorption of nutrients, fluids, and/or electrolytes.

small bowel bacterial overgrowth Excess bacterial counts in any area of the small bowel, usually greater than 10^{10} colony-forming units/ml.

total parenteral nutrition Provision of complete nutrition via the intravenous route.

transplantation Transfer of living organs from one individual to another.

tropic hormones Play a role in stimulating bowel adaptation by encouraging intestinal villous growth and in regulating intestinal motility.

Quantification of an anatomical definition of short bowel syndrome after surgical resection has been attempted. No consistent definition can be established because children with as little as a few centimeters of small intestine have been successfully weaned from parenteral nutrition whereas some children with longer segments of bowel have succumbed. In adults, it is generally agreed that less than 100 cm of remaining small bowel constitutes short bowel syndrome. A more practical definition focuses on the functional state of the remaining gastrointestinal tract, whereby short bowel syndrome exists when malabsorption of nutrients, fluids, and/or electrolytes occurs in the presence of any intestinal loss.

INTRODUCTION

Short bowel syndrome, although not common in pediatrics, remains a continuous challenge to those involved

 Copyright 2004, Elsevier (USA). All rights reserved.

in the care of children with this syndrome. Advances in neonatology, pediatric surgery, and intestinal feeding practices have facilitated survival of these children but enhancement of growth and development along with the avoidance of therapy-related complications present ongoing challenges.

Enteral nutrition is now the mainstay of therapy for children with extensive resections resulting in short bowel syndrome. Parenteral nutrition is required initially while the functional capacity of the small intestine is maximized. The practice of parenteral nutrition has become very sophisticated, with minimal complications occurring in patients managed by teams experienced in such therapy. Death from complications of short bowel syndrome therapy is infrequent because intestinal and liver transplantation procedures have become increasingly more reliable.

ETIOLOGY

Although congenital short bowel syndrome does occur, most cases of short bowel syndrome in children occur following intestinal resection in previously anatomically normal infants. Typically, small bowel resection is performed to treat events occurring in the ileum and colon; however, varying degrees of jejunal resection may also be required. Table I lists the common etiologies of short bowel syndrome in both infants and older children. Congenital anomalies resulting in short bowel syndrome include intestinal atresias, which may occur anywhere in the intestine and can be either isolated or multiple. Most infants with short bowel syndrome have undergone resection due to necrotizing enterocolitis.

INTESTINAL PHYSIOLOGY

The embryonic midgut is the origin of the small intestine, which begins rapid growth at the end of the fifth week of gestation. Intestinal lengthening exceeds the rate of growth of the embryonic body and, therefore, it must grow outside the abdomen until the tenth week. The average intestinal length at 40 weeks of gestation is 200–250 cm. Mean adult small intestinal length is approximately 500–600 cm, with primary increases in length occurring in early childhood.

TABLE I Conditions Leading to Short Bowel Syndrome

Infants	Children
Congenital short bowel	Malignancy/tumors
Intestinal atresias	Radiation enteritis
Gastroschisis	Crohn's disease
Apple peel/Christmas tree deformity	Mesenteric vascular occlusion
Hirschsprung's disease involving ileum and colon	Trauma
Necrotizing enterocolitis	Chronic pseudo-obstruction
Volvulus	

The most obvious change in short bowel syndrome is the loss of overall absorptive surface area. The inherent characteristics of the remaining intestine are, however, critical in determining the overall prognosis. Most commonly, there is retention of the jejunum with part of the colon. The jejunal epithelium differs significantly from that of the ileum in that the latter contains the specific transporters for bile acids and vitamin B_{12}. The intestinal villi create a large surface area for nutrient absorption, with most carbohydrate, protein, and water-soluble vitamins being absorbed in the proximal small intestine. Fat absorption occurs over a much larger proportion of the whole small intestine. The jejunum also has a higher concentration of enzymes and other carrier proteins, making it capable of enhanced absorptive function after distal intestinal resection.

Normally, a rapid infusion of hyperosmolar substances into the upper jejunum will result in an influx of fluid from the plasma to the lumen to equalize the osmotic differences. Reabsorption will then occur distally if the load can be absorbed. Otherwise, diarrhea ensues. In short bowel syndrome, this compensating mechanism is altered. The presence of the colon after small bowel resection provides for sodium and short-chain fatty acid absorption. Loss of the colon generally worsens the prognosis of short bowel syndrome. However, the presence of malabsorbed long-chain fats and bile acids in the colon may worsen watery diarrhea and increase oxalate absorption, with subsequent development of oxalate renal stones.

The ileum generally provides less absorptive function than the jejunum but is rapidly recruited to enhance nutrient absorption if digestive loads are large. Resection of a large portion of the ileum will result in bile salt malabsorption (with subsequent fat-soluble vitamin malabsorption) and vitamin B_{12} deficiency because the jejunum cannot compensate for the loss of the appropriate receptors. The ileocecal valve plays an important role in preventing bacterial flux from the large bowel into the small bowel. The valve also is important for regulating intestinal transit. Its presence was once thought to be key to establishing enteral nutrition and weaning the pediatric patient from parenteral nutrition. In practice, however, other factors such as the length of resection and the hormonal and neural responses to enteral nutrients are also important. In fact, any factors

that result in slower transit may enhance the risk of bacterial overgrowth.

Intestinal adaptation begins rapidly after the initial intestinal resection. In addition, the small intestine appears to undergo enhanced hyperplasia with greater nutrient exposure, although data studies in human are conflicting. Villus hypertrophy occurs by several mechanisms, including neural influences, direct nutrient contact with the enterocytes, and release of tropic substances from the stomach, liver, pancreas, and the small bowel. Hormones shown in animals to be tropic include enteroglucagon, neurotensin, epidermal growth factor, insulin-like growth factor-1 (IGF-1), glucagon-like peptide-2 (GLP-2), and gastrin. High levels of peptide YY have a role in slowing gastric emptying rate and enhancing the colonic brake phenomenon. Ghrelin, a recently identified hormone, has been investigated for potential satiety effects. Peptides such as peptide YY and enteroglucagon also slow intestinal motility. Transforming growth factor-β has been identified as having an inhibitory effect on bowel adaptation after loss of absorptive surface area.

STAGES OF THERAPY

Initially, fluid and electrolyte replacement in the early postoperative period comprises the major focus of treatment. Replacement of nutrient losses and treatment of diarrhea quickly become additional challenges.

Parenteral Nutrition

The practice of parenteral nutrition has become quite refined, involving development of age-related and disease-specific amino acid solutions and the recognition that there are age-dependent micronutrient requirements. The immediate postoperative period requires aggressive monitoring of fluid and electrolytes. A logical approach is to provide a standard parenteral nutrition solution, dependent on age of the patient, in combination with a more patient-specific fluid replacement protocol. Replacement fluids can be prescribed by monitoring electrolyte losses and overall volume loss in the previous 2–4 hours. Replacement of electrolytes lost in intestinal fluids may require very large quantities of electrolytes. Utilizing such protocols for replacement fluids rather than frequent altering of expensive total parenteral nutrition (TPN) solutions can be done quickly and easily at the bedside. Continued replacement of electrolyte and fluid losses can be difficult in the home environment; therefore, this constitutes one of the exclusions for home therapy.

TABLE II Average Caloric Requirements by Age Group

Age	Kilocalories[a]
Premature infants	120–150
Newborns	100–120
1–12 months	100
1–6 years	75–90
7–12 years	60–75
13–18 years	30–60
Adult	25–35

[a] Per kilogram body weight per day.

Initially, total parenteral nutrition solution should replace all basic nutritional needs. Table II identifies the average caloric requirements (kilocalories/kilogram/day) for different age groups. Table III lists the components of a standard pediatric total parenteral nutrition solution. Additional parenteral energy supply may be necessary during the early postoperative period or during times of stress, such as that related to fever or severe infections. One to two times the basal energy requirements may be required. Careful monitoring of parenteral nutrition via both clinical and laboratory parameters is required.

Enteral Nutrition

When fluid and electrolyte status has been stabilized with the parenteral regimen, enteral feeding can be considered and should begin as soon as possible. In infants, elemental diets are delivered via continuous enteral infusion. Elemental formulas are well tolerated and avoid the risk of allergy to proteins in more complex feedings. Enteral feeding is typically started very slowly with a dilute concentration (5 cal/ounce), which is slowly increased to 20 cal/ounce for patients less than 1 year of age and 30 cal/ounce for older patients. When final concentration is reached, the volume is slowly advanced. The technique of reaching final concentration prior to increasing volume avoids fluid overloading for the patient also receiving parenteral nutrition.

TABLE III Standard Pediatric Total Parenteral Solution

Dextrose, 20%
Amino acids, 2.5%
Sodium chloride, 15 mEq/liter
Sodium acetate, 15 mEq/liter
Potassium phosphate, 20 mEq/liter
Calcium gluconate, 10 mEq/liter
Magnesium sulfate, 5 mEq/liter
Pediatric trace elements
Pediatric multivitamins

The continuous and aggressive use of enteral nutrition should be encouraged unless significant dehydrating diarrhea ensues, in which case the infusion should be adjusted so that overall fluid balance improves. Continuous enteral infusion is inconvenient and is thought to decrease the normal developmental processes of eating. However, this can be managed later. Newer, small enteral infusion pumps along with backpack devices have been developed to allow the patient greater mobility. When long-term enteral nutrition is anticipated, i.e., greater than 3 months, gastrostomy tube placement facilitates continuous enteral feeding. The presence of a gastrostomy or nasogastric tube does not contraindicate feeding. However, continuous feeding does alter hunger mechanisms, and rejection of oral feeds is common. Use of continuous enteral feeding does decrease the likelihood of gastroesophageal reflux.

As the child advances in age, a more complex formula, such as a protein hydrolysate, is usually well tolerated. For patients older than 1–2 years of age, whole protein formulas stimulate further adaptation by increasing the workload of the surface epithelium. Carbohydrates in enteral formulas are present in the form of one or more sources, including extensively hydrolyzed starch and disaccharidases such as sucrose. Medium-chain fats, although well absorbed, are not as beneficial as long-chain fats in enhancing adaptation. Therefore, a mixture of both types of fat in the formula is most beneficial. Carbohydrate type is probably the least important type of required nutrient for patients with short bowel syndrome. However, lactose may be more slowly hydrolyzed than glucose polymers. Table IV lists some commonly used formulas for infants and children with short bowel syndrome. Theoretically, a formula with enhanced proportions of fat, even to 50% of the total daily energy intake, may be beneficial not only for delivering more calories in less volume, but also because high-fat formulas may slow gastrointestinal motility to enhance absorption.

Tolerance of continuous enteral infusion is based in part on stool losses; losses of greater than 40–50 ml/kg/day, especially if accompanied by the presence of positive reducing substances, suggest that enteral feedings should be reduced or halted. Infants should be given small volumes of nipple feedings to facilitate developmentally normal stages of eating. At the appropriate ages, introduction of solid foods should also be attempted. Caution should be given to types of solids initially utilized. Avoidance of foods containing high carbohydrate levels reduces osmotic losses. Meats are often well tolerated. Nutrient delivery by the oral route may not be significant due to malabsorption but is key in later stages of therapy. In older infants and toddlers, when the colon is intact, complex diets may be beneficial in enhancing colonic salvage of short-chain fatty acids.

TABLE IV Commonly Used Formulas

Elemental	Semielemental	Whole protein
Neocate	Alimentum	Peptamen
Peptide 1+	Pregestimil	Pediasure
Pediatric Vivonex	Nutramigen	Vital
Elecare		

Chronic Therapy

Parenteral nutrition is weaned as enteral nutrition is advanced on a calorie-for-calorie basis. While continuous enteral nutrition is utilized, parenteral nutrition is suspended for a few hours each day, with the time off parenteral nutrition gradually increasing, decreasing total parenteral nutrition volume and calories. Eventually, delivery of TPN on an every-other-day basis is possible. This is most easily accomplished by continuous enteral delivery until parenteral nutrition is significantly reduced and the child remains stable. At that time, weaning of enteral nutrition can slowly be attempted as solid food intake increases. As the child ages, monitoring of weight gain becomes vitally important. Because caloric requirements do decrease slightly with advancing age, and if adequate intestinal adaptation occurs, many children whose weight seems to plateau over time can still be weaned successfully from parenteral nutrition.

Once parenteral nutrition is weaned, frequent monitoring of weight, height, and macro- and micronutrient intake becomes increasingly important. Deficiency states are incurred based on the area of intestine resected. It is not uncommon for children to require replacement of vitamin B_{12}, magnesium, and fat-soluble vitamins. If enteral nutrition advancement appears to stall, complications such as small bowel bacterial overgrowth or micronutrient deficiency should be considered.

Complications

The most challenging complication of nutritional therapy in short bowel syndrome is diarrhea. Management should be based on etiology. Average stool losses should be established for each patient and alterations from baseline should be investigated. Possible changes in dietary intake and bacterial overgrowth are the most common causes of new-onset diarrhea. Utilization of bolus rather than continuous enteral feeding may result in an increased osmotic load and may not be well

tolerated. As the child grows, small, frequent meals, for example, may be more beneficial than three daily standard meals. In cases of bile acid malabsorption causing diarrhea, use of cholestyramine may be beneficial. However, it should be used with caution as the attendant sequestration of luminal bile salts may exacerbate fat and fat-soluble vitamin malabsorption. Antisecretory drugs are sometimes utilized. Somatostatin has been used to reduce diarrhea with modest efficacy short term, but long-term administration has not proved to be of benefit. Newer drugs such as racecadotril are currently under investigation. Antimotility drugs such as Imodium are often prescribed for patients with short small bowel and may be of benefit particularly if the ileocecal sphincter is intact. Antimotility drugs should be used with caution in children because they may enhance problems with small bowel bacterial overgrowth.

One of the most difficult complications to deal with in the therapy of short bowel syndrome is small bowel bacterial overgrowth. This condition occurs in virtually all patients; however, it does not become problematic in all. Diagnosis is difficult and no gold standard exists. Culture of jejunal aspirates, measurement of excess hydrogen production after a glucose load, serum D-lactate, and urine indican levels have all been attempted as possible correlates to pathological bacterial overgrowth. Symptoms include an increase in diarrhea, abdominal distension, and flatulence. At times, even bloody diarrhea may occur. Bacteria normally present the bowel have a role in deconjugating bile salts and produce micronutrients such as B_{12}. However, if bacterial numbers become excessive ($>10^5$ jejunal and $>10^8$ ileal), mucosal inflammation may result. Another significant complication of small bowel bacterial overgrowth is D-lactic acidosis. Particularly in infants and small children, this can lead to lethargy and even coma.

Empiric treatment of small bowel bacterial overgrowth with antibiotics is often helpful in establishing the diagnosis. Broad-spectrum antibiotics such as oral gentamicin, trimethoprim sulfate, and/or metronidazole are often utilized initially. Low doses may result in significant improvement in symptoms. Endoscopy is rarely helpful, except in evaluating the presence of enteritis and colitis. If present, these situations may also respond to antiinflammatory therapy with mesalamine or even corticosteroids. If beneficial, antibiotic therapy is often not required continuously, but may be used on a rotating schedule. Eventually, the development of resistant organisms may require the alteration of antibiotic type. Use of probiotics may also be beneficial in preventing or modulating the response of more harmful or excessive bacterial species.

TABLE V Commonly Acquired Nutrient Deficiencies

Nutrient	Signs and symptoms
Vitamin A	Night blindness
Vitamin D	Bony demineralization
Vitamin B_{12}	Megaloblastic anemia
Magnesium	Lethargy and tetany

As food-based intake is advanced, specific nutrient deficiency states may occur when supplemental enteral formula is no longer required. Table V lists commonly acquired nutrient deficiencies. Nutrient supplements are readily available and usually dosing can be adjusted based on blood levels of the nutrients. Magnesium presents a particular challenge because supplementation with almost any magnesium supplement can enhance diarrhea. Careful administration and followup monitoring of supplements are required, especially when levels become critically low. The development of acid peptic disease commonly occurs in children with short bowel syndrome. Many pediatric patients with short bowel syndrome eventually become symptomatic with acid reflux and require neutralizing or acid suppression medications on a long-term basis.

The most life-threatening complication of long-term nutritional therapy in infants with short bowel syndrome is parenteral nutrition-related liver disease. The etiology of this entity is not well understood. Toxicity of amino acid solutions, excess administration of lipid solutions, production of toxins by bacteria in the bowel, excessive nutrient administration, toxicity of unknown substances in the parenteral nutrition solution, and absence of stimulation of gastrointestinal hormones have all been implicated as possible contributing factors. It is well understood that even minimal enteral nutrition is important in lessening the impact of this complication. Ursodeoxycholic acid has been advocated to prevent the development of parenteral nutrition liver disease. Recurrent central venous catheter sepsis has been more recently identified as a possible factor in the development of TPN-associated liver disease. Efforts should be made to prevent as many of these complications as possible. In extreme cases of intestinal failure with mild TPN-associated liver disease, reversal of liver disease has been noted following successful intestinal transplantation.

SURGICAL THERAPY AND TRANSPLANTATION

Patients with short bowel syndrome who have undergone intestinal resection frequently experience

problems with anastomotic strictures and ulcerations. These can occur at any time, but typically do so within a few years after the initial resection. Symptoms include recurrent or sudden anemia, blood in the stools, or progressive worsening of vomiting and abdominal bloating. The adaptive response, although beneficial in enhancing the absorptive surface area, may be become extreme and result in dilatation of the bowel, with poor motility and significant small bowel bacterial overgrowth. Such bowel dilatation can be identified on radiographic imaging studies and may be amenable to surgical tapering with or without concomitant bowel lengthening (Bianchi procedure). These procedures, if done at centers with expertise in performing them, may be beneficial in alleviating symptoms without sacrificing intestinal absorptive surface area. Surgical procedures designed to slow intestinal transit via creation of intestinal valvelike areas or reversed intestinal segments are not recommended. Fundoplication is sometimes necessary in patients who experience significantly altered upper gastrointestinal motility.

Intestinal or combined liver/intestinal transplantation has become an alternative, although not an early interventional strategy, for patients requiring chronic parenteral nutritional support due to short bowel syndrome. Transplantation prior to irreversible liver or bowel failure is often met with great success. At the present time, it is not possible to predict which patients will fail intestinal rehabilitation. Attempts to predict ultimate failure of parenteral nutrition have centered on criteria including the presence of less than 30 cm of small bowel, absence of the ileocecal valve, some colon resection, only minimal tolerance of feedings 2 months after resection, patients without a jejunum or ileum, patients with poor quality of residual bowel, and patients with repeated sepsis episodes. Advances in surgical technique and immunosuppression have enhanced survival; however, graft rejection and complications of immunosuppression may ultimately occur because these patients survive longer. New medications for immunosuppression, such as interleukin-2 antagonist, bone marrow augmentation, and tolerance induction, have all enhanced the success of intestinal transplantation. Results vary significantly among centers performing intestinal transplants, with an overall 1- to 2-year survival at 75% decreasing to about 50% at 3–5 years. The shortage of donors remains a significant problem. Although intestinal rehabilitation may offer the best chance for a meaningful quality of life and the best chance for longevity, transplantation as a newer form of therapy may be an option for patients failing on parenteral nutrition.

PROGNOSIS

Parenteral and enteral nutrition programs remain the mainstay of therapy for infants and children with short bowel syndrome. The majority of these patients, if adequately managed, will require parenteral nutrition only as an interim phase of rehabilitation. Some patients will eventually develop life-threatening complications, including liver disease. Transplantation of either the bowel or the bowel and liver have become possible options for selected patients unable to succeed with parenteral nutrition. Although home-based parenteral nutrition does permit optimal psychosocial development, it requires the utilization of an aggressive nutrition team approach by nurses, physicians, and nutritionists. Augmentative surgical procedures are sometimes helpful in restoring bowel function.

See Also the Following Articles

Bacterial Overgrowth • Enteral Nutrition • Parenteral Nutrition • Transplantation Immunology

Further Reading

Andorsky, D. J., Lund, D. P., Lillehei, C. W., *et al.* (2001). Nutritional and other postoperative management of neonates with short bowel syndrome correlates with clinical outcomes. *J. Pediatr.* **139**(1), 27–33.

Georgeson, K., Halpin, D., Figueroa, R., *et al.* (1994). Sequential intestinal lengthening procedure for refractory short bowel syndrome. *J. Pediatr. Surg.* **29**(2), 316–320.

Hart, M. H. (1992). Growth factors in intestinal resection. *J. Pediatr. Gastroenterol. Nutr.* **14**, 2.

Jeppson, P. B., Hartmann, B., Thulsesen, J., *et al.* (2001). Glucagon-like peptide 2 improves nutrient absorption and nutritional status in short bowel patients with no colon. *Gastroenterology* **120**(4), 806–815.

Kaufmann, S. S., Atkinson, J. B., Bianchi, A., *et al.* (2001). Indications for pediatric intestinal transplantation: A position paper of the American Society of Transplantation. *Pediatr. Transplant.* **5**(2), 80–87.

Michail, S., and Vanderhoof, J. A. (1996). Parenteral nutrition in clinical practice. *Ann. Nestle* **54**, 53–60.

Nightingale, J. M., Kamm, M. A., van-der-Sijp, J. R., *et al.* (1993). Disturbed gastric emptying in the short bowel syndrome. Evidence for a "colonic brake." *Gut* **34**(9), 1171–1176.

Nightingale, J. M. D., Kamm, M. A., van der-Sijp, J. R. M., Ghatei, M. A., Bloom, S. R., and Lennard-Jones, J. E. (1996). Gastrointestinal hormones in short bowel syndrome. Peptide YY may be the "colonic brake" to gastric emptying. *Gut* **39**, 267–272.

Nucci, A. M., Barksdale, E. M. Jr., Beserock, N., *et al.* (2002). Long-term nutritional outcome after pediatric intestinal transplantation. *J. Pediatr. Surg.* **37**(3), 460–463.

Vanderhoof, J. A. (1996). Short bowel syndrome. *In* "Pediatric Gastrointestinal Disease" (W. A. Walker, P. R. Durie, J. R. Hamilton, J. A. Walker-Smith, and J. B. Watkins, eds.), 2nd Ed., pp. 830–840. Mosby, St. Louis, Missouri.

Vanderhoof, J. A. (1999). Short bowel syndrome: Parenteral nutrition versus intestinal transplantation. *Pediatr. Nutr. ISPEN* **1**(1), 8–15.
Vanderhoof, J. A., Young, R. J., Murray, N., and Kaufman, S. S. (1998). Treatment strategies for small bowel bacterial overgrowth in short bowel syndrome. *J. Pediatr. Gastroenterol. Nutr.* **27**, 155–160.
Vargas, J. H., Ament, M. E., and Berquist, W. E. (1987). Long term home parenteral nutrition in pediatrics: 10 years of experience in 102 patients. *J. Pediatr. Gastroenterol. Nutr.* **6**(1), 24–32.
Warner, B. W., Vanderhoof, J. A., and Reyes, J. D. (2000). What's new in the management of short gut syndrome in children. *J. Am. Coll. Surg.* **190**(6), 725–736.

Sigmoidoscopy

NATHAN SCHMULEWITZ AND DAVID A. TENDLER
Duke University Medical Center

cathartic An agent that results in the purging of bowel contents.
colitis Inflammation of the colonic mucosa; may be acute or chronic.
fissure A painful split in the mucous membrane of the anus.
hemorrhoids Varicosities of the external hemorrhoidal veins that may result in a painful swelling at the anus and have a propensity for bleeding.
mini-perforation A small rent in the bowel, which typically is contained by the overlying omentum and heals without the need for surgical repair.
perineum The external region between the anus and the genitalia.
pneumatosis coli The presence of gas-filled cysts within the intestinal mucosa, which may occur after colonic insufflation or with colonic ischemia.
postpolypectomy syndrome An acute syndrome composed of pain, fever, and focal colitis stemming from transmural thermal injury following the removal of a polyp with cauterization.
proctitis Inflammation of the rectal mucosa; may be acute or chronic.
sigmoidoscopy A technique for obtaining intralumenal visualization of the rectum, sigmoid, and, usually, left colon, utilizing a fiber-optic endoscope.
splenic flexure The section of colon representing the transition from the transverse colon to the left colon, typically located in the left upper quadrant of the abdomen, inferior to the spleen.
tenesmus A painful cramping sensation of incomplete defecation, accompanied by ineffectual straining to further evacuate the bowel.
valves of Houston The three or four crescentic transverse folds of the rectum.

Sigmoidoscopy, by definition, is a manual technique whereby intralumenal visualization of the rectum and sigmoid colon is accomplished using a fiber-optic endoscope. However, the procedure generally includes examination of the left colon as well. With the advent of flexible endoscopy in the 1960s, it became possible to survey the entire large intestine. Yet, even as full colonoscopy has become an important tool in the workup and surveillance of various conditions, flexible sigmoidoscopy continues to play an important role in gastrointestinal diagnostics. That role is based on several key facts: (1) flexible sigmoidoscopy is safe and, most often, does not require sedation; (2) the procedure can be performed relatively quickly, in an office setting, and by a variety of appropriately trained health care providers; and (3) findings in the distal colon often answer important clinical questions regarding disease states such as acute colitis, chronic colitis, and rectal bleeding. Although evaluation of the sigmoid colon can also be accomplished with a rigid sigmoidoscope, this technique has largely been supplanted by flexible sigmoidoscopy, which is better tolerated by patients and typically allows for a more extensive examination. Therefore, the focus of this article will be on the use of flexible sigmoidoscopy.

PATIENT PREPARATION

Preparation for sigmoidoscopy is far easier for the patient than that necessary for colonoscopy. As only the distal portion of the colon needs proper cleansing, enemas prior to procedure often suffice. A common

Copyright 2004, Elsevier (USA). All rights reserved.

Vanderhoof, J. A. (1999). Short bowel syndrome: Parenteral nutrition versus intestinal transplantation. *Pediatr. Nutr. ISPEN* **1**(1), 8–15.
Vanderhoof, J. A., Young, R. J., Murray, N., and Kaufman, S. S. (1998). Treatment strategies for small bowel bacterial overgrowth in short bowel syndrome. *J. Pediatr. Gastroenterol. Nutr.* **27**, 155–160.
Vargas, J. H., Ament, M. E., and Berquist, W. E. (1987). Long term home parenteral nutrition in pediatrics: 10 years of experience in 102 patients. *J. Pediatr. Gastroenterol. Nutr.* **6**(1), 24–32.
Warner, B. W., Vanderhoof, J. A., and Reyes, J. D. (2000). What's new in the management of short gut syndrome in children. *J. Am. Coll. Surg.* **190**(6), 725–736.

Sigmoidoscopy

NATHAN SCHMULEWITZ AND DAVID A. TENDLER
Duke University Medical Center

cathartic An agent that results in the purging of bowel contents.
colitis Inflammation of the colonic mucosa; may be acute or chronic.
fissure A painful split in the mucous membrane of the anus.
hemorrhoids Varicosities of the external hemorrhoidal veins that may result in a painful swelling at the anus and have a propensity for bleeding.
mini-perforation A small rent in the bowel, which typically is contained by the overlying omentum and heals without the need for surgical repair.
perineum The external region between the anus and the genitalia.
pneumatosis coli The presence of gas-filled cysts within the intestinal mucosa, which may occur after colonic insufflation or with colonic ischemia.
postpolypectomy syndrome An acute syndrome composed of pain, fever, and focal colitis stemming from transmural thermal injury following the removal of a polyp with cauterization.
proctitis Inflammation of the rectal mucosa; may be acute or chronic.
sigmoidoscopy A technique for obtaining intralumenal visualization of the rectum, sigmoid, and, usually, left colon, utilizing a fiber-optic endoscope.
splenic flexure The section of colon representing the transition from the transverse colon to the left colon, typically located in the left upper quadrant of the abdomen, inferior to the spleen.
tenesmus A painful cramping sensation of incomplete defecation, accompanied by ineffectual straining to further evacuate the bowel.
valves of Houston The three or four crescentic transverse folds of the rectum.

Sigmoidoscopy, by definition, is a manual technique whereby intralumenal visualization of the rectum and sigmoid colon is accomplished using a fiber-optic endoscope. However, the procedure generally includes examination of the left colon as well. With the advent of flexible endoscopy in the 1960s, it became possible to survey the entire large intestine. Yet, even as full colonoscopy has become an important tool in the workup and surveillance of various conditions, flexible sigmoidoscopy continues to play an important role in gastrointestinal diagnostics. That role is based on several key facts: (1) flexible sigmoidoscopy is safe and, most often, does not require sedation; (2) the procedure can be performed relatively quickly, in an office setting, and by a variety of appropriately trained health care providers; and (3) findings in the distal colon often answer important clinical questions regarding disease states such as acute colitis, chronic colitis, and rectal bleeding. Although evaluation of the sigmoid colon can also be accomplished with a rigid sigmoidoscope, this technique has largely been supplanted by flexible sigmoidoscopy, which is better tolerated by patients and typically allows for a more extensive examination. Therefore, the focus of this article will be on the use of flexible sigmoidoscopy.

PATIENT PREPARATION

Preparation for sigmoidoscopy is far easier for the patient than that necessary for colonoscopy. As only the distal portion of the colon needs proper cleansing, enemas prior to procedure often suffice. A common

Copyright 2004, Elsevier (USA). All rights reserved.

prescription is one or two hypertonic phosphate enemas taken an hour prior to examination. Delay of over an hour from the last enema may allow more proximal fecal debris to migrate to the left colon. Many different bowel preparations for sigmoidoscopy have been studied and recommendations vary. Some investigators advocate using oral cathartics, such as magnesium citrate, the night before the examination, followed by two hypertonic phosphate enemas on the day of the exam. Others suggest using oral laxatives, such as magnesium citrate or oral hypertonic phosphate solution, alone or in combination with a stimulant laxative, such as bisacodyl. Although hypertonic phosphate enema is the standard regimen, alternative preparations can be used based on the preferences of the examiner and medical history of the patient. Adding oral cathartics the night before a procedure will usually lead to a better preparation and might be useful in the patient in whom enemas alone are suspected to give marginal results. Such a patient might be someone with severe diverticular disease, one with poor anal sphincter tone, who has difficulty retaining an enema, or someone with a poor preparation previously. However, patient tolerance and satisfaction may be compromised, in light of the common side effects of oral laxatives, such as abdominal cramping, nausea, and vomiting. The goal of the bowel preparation is threefold: safety, efficacy, and tolerance. A properly cleansed bowel allows the endoscopist to advance the instrument with a clear view of the lumen and hence, achieve a thorough and safe examination. A poor preparation with residual stool may result in an inadequate assessment of the colonic mucosa, as well as an increased risk of perforation due to suboptimal visualization of the colonic lumen. Pathology may be missed with stool in the way, even with honest efforts to irrigate and aspirate. Therefore, the patient is best served by having the examination rescheduled. Finally, a good colonic preparation usually yields a shorter, more comfortable examination for the patient. Minimizing discomfort is paramount for ensuring compliance with potential future examinations.

When choosing a preparation, it is important to account for the medical condition of the patient. Oral hypertonic sodium phosphate solutions are generally contraindicated in patients with renal impairment, due to the potential for clinically significant hyperphosphatemia and/or hypocalcemia. Similarly, patients with renal impairment should avoid magnesium citrate preparations, due to the risk of developing hypermagnesemia. Sodium phosphate solutions may also be hazardous in those with congestive heart failure or cirrhosis, due to the large sodium load that occurs with these preparations. Other conditions that may carry an increased risk for sodium phosphate-induced hyperphosphatemia include severe ulcerative colitis and pregnancy (the placenta actively transports phosphates). Phosphate preparations can also cause epithelial sloughing and/or aphthous ulcerations, which may cause confusion in cases of suspected colitis. Mucosal biopsies will usually distinguish preparation-induced changes from a chronic, idiopathic colitis. In moderate to severe colitis, with significant diarrhea, bowel preparation may not be needed at all, as any retained stool will be liquid and easily aspirated. Alternatively, tap water enemas may be used.

Some patients with anorectal disease, particularly those with fissures, or patients with low thresholds for discomfort may need topical anesthetics, such as viscous lidocaine, or even conscious sedation for the examination. If conscious sedation is anticipated, then overnight fasting is necessary to reduce the chance of nausea, vomiting, and pulmonary aspiration. In this circumstance, cardiopulmonary monitoring is also required, as well as the presence of providers that are trained and certified in the administration of conscious sedation.

SETTING UP

It is common for patients to experience varying degrees of anxiety prior to the procedure. Therefore, the procedure should be reviewed in detail. The patient should be reminded of why the test is being performed, abnormal findings that might reasonably be anticipated, and the fact that mild discomfort may be experienced. The patient should know that discomfort can be ameliorated by removing air or partially withdrawing the scope, to avoid the notion that minor discomfort should necessarily lead to aborting the procedure. Risks ought to be recounted, including discomfort, bleeding (especially if biopsies are taken), and perforation. In the patient receiving conscious sedation for the test, the potential for adverse reaction to a sedative must be discussed and a review of comorbid conditions, medications, and allergies is necessary. It is important to provide ample time for patients to have any questions answered and for informed written consent to be obtained from the patient or designated surrogate.

Once the patient is ready, he or she should lie on the examination table with the left side down and facing away from the examiner. Most endoscopists will use video endoscopes, as opposed to fiber-optic scopes that utilize an eyepiece. The endoscopic power/light source will be alongside the endoscopist to the right, toward the patient's feet, with the video monitor on the

other side of the table, facing the patient and the endoscopist. If a nurse is assisting the procedure, he or she should stand at the patient's feet, on the other side of the light source, so as to most easily hand endoscope accessories to the endoscopist. The light source should be turned on and checked to ensure that air insufflation, suction, and lens-washing functions all operate properly. Discovering technical malfunctions during the procedure can compromise thorough examination as well as patient confidence. At the ready, the endoscopist should have lubricating jelly, a washcloth or gauze for use if needed in handling a lubricated instrument, and a water basin with a prefilled 60 cc syringe for irrigating stool through the endoscope channel.

Next, visual inspection and a digital rectal exam are performed. Rectal examination allows the examiner to lubricate the anal canal (a generous amount of jelly is advised) and to palpate the anal canal and lower rectum. This step is critical for finding pathology of the perineum and anal canal. Conditions such as external hemorrhoids and anal fissures may be best appreciated at this time. The patient should be informed that the maneuver is to occur and it should be performed slowly. Unexpectedly cold lubricant, rapid insertion, or a rough examination may heighten anxiety and possibly lower the threshold for perceived discomfort during the endoscopy. A tender fissure or inflamed hemorrhoids may warrant the use of anesthetic jelly during examination. Circumferential palpation of the canal and lower rectum is important, as this area is the most difficult to examine endoscopically, even with retroflexing the scope. Furthermore, on occasion, lesions that are not easily distinguished visually, such as squamous cell cancers of the anal canal or submucosal lesions, might be palpated.

THE SIGMOIDOSCOPE

Most commonly, flexible sigmoidoscopy is performed with a 60 cm endoscope with a 12 mm outer diameter. All standard endoscopes are right-handed, so that the shaft is grasped with the right hand and the control head with the left. Within the shaft, either fiber-optic bundles (in the older models) or an electronic chip carries the visual image either to a focusing lens beneath an eyepiece or to a video processor, respectively. The shaft also carries the cables that allow the directional dials to control the scope tip, as well as an open channel through which instruments can be passed (forceps, snare, cautery probe, injection needle, etc.). At the control head are two main buttons, two dials, and an entry port for the internal channel. The upper button is depressed for suction, with the degree of depression corresponding to the suction pressure. The lower button has a central hole, which is covered to activate air insufflation at the endoscope tip. Tapping at the button allows for fine control of delivered air volume. Depressing the lower button fully causes the water irrigation system to clean the lens at the scope tip by washing across it. Some endoscopes have other buttons for freezing or capturing images to the video processor. The inner dial controls up and down tip deflection and the smaller outer dial controls left and right tip deflection. On most flexible sigmoidoscopes, the tip can deflect 180° in the up or down direction and 160° left or right. Brakes are usually mounted on these dials to lock them if one wishes to maintain a given tip deflection or neutral position. An umbilical cord inserts into the control head and connects it to the power source that controls the light, air, and irrigation systems. The scope head should be held in the left hand so that the thumb can reach both directional dials and the forefinger comes around to the front of the head to manage the buttons.

Occasionally, the standard flexible sigmoidoscope shaft may be too stiff for some patients. The advantage of a relatively stiff-shafted instrument in the colon is that it reduces scope looping in the often-windy left colon. However, when the sigmoid colon is more tortuous or has restricted mobility, owing to previous surgeries or inflammatory (e.g., diverticular) processes, a standard sigmoidoscope may cause painful looping, as it stretches the bowel upon its mesentery. In such cases, a thinner, more flexible instrument may be helpful. For these occasions, a standard gastroscope is thinner (0.96 cm) and longer and has greater tip deflection. Such an instrument can more easily negotiate sharper turns in a tortuous colon with less pressure applied and should allow the endoscopist to reach the splenic flexure with less patient discomfort.

SIGMOIDOSCOPIC TECHNIQUE

In general, the sigmoidoscope should be advanced until the splenic flexure is reached, solid stool is encountered, or the patient experiences appreciable discomfort. Thus, a proper exam requires some skill on the part of the endoscopist, in order to reach the flexure without running out of scope or convincing the patient never to return. A complete exam usually takes 5 to 10 min.

After the rectal examination, the patient should be told that the endoscopy is to begin and additional lubricant is applied to the distal portion of the endoscope tip. The scope is then inserted. Usually the scope will meet mild resistance at the anal sphincter muscles and then slip through the anal canal. Placement of the right forefinger on the tip of the scope might help guide it into the anal canal.

Once the scope has entered the rectum, it should be withdrawn a few centimeters, air should be insufflated into the rectum, and the tip should be deflected upward a bit. If not pulled back, the scope will likely be pressed against the anterior wall of the rectum with a pink image on the monitor. After the rectum is partially inflated and the lumen is visible, the endoscope may be advanced, with "dark" areas signifying the lumen ahead. In advancing the scope, most endoscopists prefer to guide the instrument with use of the up/down dials and right-handed torque of the shaft. Left and right tip deflection is rarely a prominent feature of scope advancement through the colon, as it is usually simpler to "feel" the turns by gently rolling the scope between the right thumb and forefinger (with or without the middle finger).

Forward progress and risk reduction require that the lumen be visualized at all times. This is not always easy as the bowel has continuous underlying peristalsis and some turns may be sharp. Yet, the most important rule of intestinal endoscopy is to always pull back when the lumen is not clearly visible. Likewise, if blanching of blood vessels is observed, excessive pressure is being exerted on the mucosa and the scope should be promptly withdrawn until a view of the lumen is restored. One should never worry about losing ground in falling back past a turn. Sometimes, sharp turns require pulling back, deciding which angle to take, and then proceeding anew. Once a turn is made with the scope tip, pulling the shaft back slightly, with the scope tip still deflected, usually allows for the lumen to come back into view. Care must always be taken to advance the scope only when the path of the lumen is certain. "Blind pushing" may lead to traumatic complications and must be avoided. Greatest care must accompany the examination of the colon with dense diverticular disease. Large, thin-walled diverticula may mimic the lumen and thus every maneuver with the scope tip must be slow and cautious. When a sharp turn is present and the direction is known, the endoscopist is often tempted to "slide by," that is, gently try to advance the directed scope tip, hoping the lumen will soon come into view. Slide-by maneuvers do carry some risk and should generally be avoided. Sometimes it can be avoided by either hooking a turn and pulling back on the scope shaft, thus bringing the lumen into view, or by pulling back, desufflating the segment of bowel, which may decrease the acuity of the angulation, and slowly advancing through a "memorized" trail with only a small amount of dark lumen leading the way. The circular muscles in the wall of the colon are helpful, as the lumen will follow in the direction of the concave contours, which usually are readily apparent through the mucosa.

Inevitably, when pushing an endoscope into the colon, loops will occur. In looping, forward pressure on the back end of the scope forces the shaft into the wall of the bowel behind the scope tip. The only way for looping not to happen is for the bowel to resemble a straight pipe, which, unfortunately, does not commonly occur. In other words, looping always occurs and forward progress requires that the loop be "reduced." A loop can be removed by pulling back on the scope and usually by trying to torque the scope in the direction opposite to the dominant twist of the bowel, a maneuver that often calls upon trial and error. Loops in the sigmoid colon are most often "alpha" loops, which are usually reduced by torquing the shaft clockwise and pulling the scope backward. Remember that with every turn made by the scope tip, some looping becomes inherent in the shaft. For this reason, the endoscopist should try to withdraw somewhat after every turn. In doing so, the bowel can be straightened along the shaft of the scope, allowing for further forward progress. In unsedated sigmoidoscopy, no significant looping will escape the endoscopist's consciousness, as the patient will become uncomfortable when the bowel is stretched on its mesentery. Passing the scope more proximally will, usually, occur only after reducing the loop and restoring the patient's comfort. In the patient with known or suspected acute colitis, looping should be avoided at all costs, as the less sturdy bowel wall may more easily succumb to traumatic wall pressures applied by the scope.

As mentioned previously, the anatomic goal of sigmoidoscopy is to reach the splenic flexure, the proximal end of the "left colon." Although there are no consistent landmarks in the large intestine between the anal canal and the appendix and ileocecal valve, there are a few ways to help decide where the scope tip might be. One endoscopic marker to look for is the spleen itself. The spleen and liver both are clearly seen through the thin-walled bowel. Unfortunately, a splenic or hepatic impression sometimes occurs in more than one segment of colon. Light pressure on the anterior aspect of the abdominal skin may yield a gross idea of where the scope tip is in the abdomen, although knowing that the tip is somewhere in the left upper quadrant or mid abdomen is not too helpful. The most helpful markers are the appearance of the bowel and the centimeter markings on the shaft of the scope. (Note that the depth of scope insertion has informational value only if the endoscope is fully reduced, as looping will always give the misimpression of the tip being more proximal in the bowel.) When the scope is straight, the splenic flexure is usually approximately 40–45 cm from the anus. Nevertheless, twice that length of scope can easily be pushed

into the bowel before reaching the splenic flexure, if looping is allowed. Of course the standard sigmoidoscope is only 60 cm, meaning that loops must be reduced systematically.

The appearance of the bowel may also be useful. The capacious rectum is approximately 10–15 cm long and is partially divided proximally by the valves of Houston, which causes the entering endoscope to slalom a bit before reaching the rectosigmoid junction, at which time the lumenal diameter narrows appreciably. The sigmoid colon runs approximately 15–20 cm, but is typically quite serpiginous and its length varies widely among patients. The junction of the sigmoid and descending colons is usually not evident endoscopically, as it is on barium enema. One may hit a long "straight-away," which usually means that the scope tip is in the descending colon and that the next major turn will be the splenic flexure. Yet, some patients may have a relatively short descending colon or one that is as tortuous as the sigmoid due to diverticular or other inflammatory disease, previous surgery, or laxity in the mesentery, which may come with age. The transverse colon usually has a typical triangular shape to the colonic haustra and a relatively straight course. Seeing these identifying marks allows the endoscopist to comfortably presume that the splenic flexure has been traversed.

In truth, the splenic flexure is not reached as commonly as it should be during sigmoidoscopies. One study using magnetic imaging to determine scope tip location discovered that a quarter of examinations failed to reach the sigmoid-descending junction and fewer than 10% reached the splenic flexure. The same study used colonoscopes to demonstrate that, on average, 75 cm of scope was needed to reach the splenic flexure. Such information confirms the importance of always trying to continually remove loops, especially when using a 60 cm scope. Keeping the endoscope straight and reproducibly reaching at least the descending colon, however, takes practice and skill. In that vein, effective sigmoidoscopy (particularly for screening purposes, where depth of insertion may determine findings) requires experience and continued repetition.

Once the splenic flexure is reached, solid stool is found, or the patient is overly uncomfortable, the process of withdrawing the scope begins. It is during withdrawal that more careful endoscopic evaluation of the bowel occurs. During insertion it is generally wise to avoid full insufflation, which makes the bowel longer, the turns sharper, and the patient more distended. Yet, on withdrawing the scope, care should be taken to see the entire mucosal surface, which necessitates some insufflation. Much of this air can be suctioned back out as the examination progresses distally. Many endoscopists reserve mucosal biopsies or polypectomies for the withdrawal stage of the test. (A caveat to this option is that polyps that are small or in awkward spots, such as behind folds, should be removed on insertion lest they not be found on withdrawal.) Withdrawing is technically easier, as turns can be more readily anticipated and scope loops, and hopefully air, are being removed, allowing the patient more comfort. Here, it is important to keep the lumen centered in view with fine tip control. Some advocate locking the right/left dial to minimize unwanted minor tip deflections and using gentle torque and the up/down dial to maintain proper bowel visualization. Gentle torquing of the shaft allows the endoscope tip to swivel a bit, enhancing visualization behind folds. If the scope slips back too quickly and a segment of colon is not adequately inspected, the endoscopist should advance once again, so as to avoid missing any pathology. Similarly, if the patient passes flatus, which should not be discouraged as it promotes comfort, or bowel spasm occurs, the endoscopist should stop and re-inflate the segment. Liquid stool should be suctioned out and mucosa with adherent stool should be irrigated with water. These cleaning chores are best avoided during scope insertion, unless the lumen cannot be adequately seen, as they lead to excessive air insufflation early on, making the rest of the insertion process more difficult.

After the scope has been pulled back into the rectum and the rectum examined, retroflexion is performed. The tip should be right at the anal verge and all dial brakes should be released so the scope is as flexible as possible. The scope is then slowly advanced as maximal upward tip deflection is applied with concomitant torque in either direction. This maneuver allows the tip to face back toward the anus and the scope shaft itself. Here, internal hemorrhoids are best seen, as is the most distal rectum. Without retroflexion, many low rectal lesions will be missed. Retroflexion can be uncomfortable for the patient, so the patient should be warned of impending rectal pressure, which will signify the end of the procedure.

Although the variety of endoscopic findings is beyond the scope of this article, several findings bear mention. Polyps appear as raised mucosal lesions; most are only slightly raised off the surface, but some may be large and rounded on stalks. Tiny polyps may be removed (usually with biopsy forceps) during the exam to determine whether they are adenomatous, and thus carrying malignant potential, or hyperplastic, without any malignant potential. Hyperplastic polyps generally are rather small, sessile, and smooth textured, whereas most adenomas have a slightly corrugated surface, although appearances may sometimes be misleading. If

larger (greater than 3–4 mm) or multiple polyps are found, concern for likely adenomas should lead to scheduling a colonoscopy, obviating sampling efforts. Similarly, if a tumor is found, colonoscopy will be necessary to "clear" the proximal colon prior to surgical evaluation. Diverticula, vascular malformations, hemorrhoids, or mucosal inflammation should always be noted, as they may explain previous or future bleeding episodes.

INDICATIONS

Average-Risk Cancer Screening

The most common application of flexible sigmoidoscopy is for colon cancer screening in the average-risk adult. Colon cancer is the second most common cause of cancer-related death in the United States. The survival rates for early-stage malignancies are excellent, but rather poor for advanced tumors. Furthermore, it is well established that adenomatous colon polyps are the precursors of adenocarcinomas of the colon and rectum. Therefore, screening people for early cancers and for polyps that may lead to cancers is worthwhile and has been shown to save lives. Screening modalities currently include fecal-occult blood testing, sigmoidoscopy, barium enema, and colonoscopy. Computerized tomographic (CT) colonography is another modality that is being investigated as a screening technique, but is currently considered experimental and has not yet been approved for this indication. Endoscopic screening is the most sensitive way to find neoplasms, and polyps greater than 5 mm in size are rarely missed. For average-risk patients, the American Cancer Society, American College of Gastroenterology, and the Gastrointestinal Consortium all recommend combining yearly three-sample fecal-occult blood tests with sigmoidoscopy every 5 years or performing colonoscopy every 10 years. The Preventive Services Task Force also recommends sigmoidoscopy every 5 years after the age of 50, with or without concomitant yearly fecal occult-blood testing. The obvious shortcoming of sigmoidoscopy is missing proximal lesions and between 20 and 50% of patients with polyps will have only proximal lesions. In fact, there has been a "rightward shift" in colorectal neoplasia, so that polyps and cancers are found in the proximal colon more commonly than a few decades ago. Yet, as there are insufficient resources to perform screening colonoscopies for the entire population, fecal occult-blood testing and sigmoidoscopy remain important tools to this end. In truth, the addition of fecal occult-blood testing to sigmoidoscopy only marginally increases the yield of advanced (10 mm of larger) neoplastic lesions, from 70 to 76%. Yet, fecal occult-blood testing is relatively inexpensive and will yield more colonoscopies, which means finding more cancers. (Note that patients with positive fecal occult-blood testing, known or suspected adenomatous polyps on sigmoidoscopy, family history of a first-degree relative with colon polyps or cancer, and patients otherwise at above-average risk for colorectal cancer should undergo colonoscopy and are not appropriate for "screening.")

Adjunct to Barium Enema in Screening

Barium enema, although less sensitive than colonoscopy for polyps and cancers, is still used for average-risk screening and above-average-risk indications for colonic evaluation (such as positive fecal occult-blood testing, family history of neoplasia). When barium enema is performed, a catheter-tipped balloon is inserted into the rectum and thereafter barium is introduced to the colon in a retrograde fashion. The presence of the balloon prevents complete visualization of the rectum. Thus, whenever barium enema is used to survey the colon, flexible sigmoidoscopy (or at least proctoscopy) is necessary to complete the task.

Workup of Minor Rectal Bleeding in Patients Less Than 40 Years of Age

The passage of red blood from the rectum generally results from sources in the rectum and anal/perianal regions. Hemorrhoids, anal fissures, proctitis, and rectal cancers are the most common explanations. Other precipitators may include distal neoplasms, proctitis, left-sided colitis, diverticula, angiodysplasias, or colonic ulcers. Some of these conditions may involve multiple segments of the colon, but the finding of a right-sided lesion as an explanation for small amounts of visible red blood, with a normal left colon, is unusual. Thus, when a patient describes seeing small amounts of blood with the stool, or just on the toilet tissue, examination of the distal 60 cm of the colon should be sufficient. However, several studies have shown that patients with non-emergent rectal bleeding may have above-average risk for colorectal cancers. Because of an increased risk for advanced neoplasia, if a patient is over 40 years of age and has not had a previous colonoscopy, a complete examination of the colon is indicated for the evaluation of rectal bleeding. In the patient under 40 years of age, without other risk factors for colorectal cancer, flexible sigmoidoscopy is the recommended test. Obviously, the specific clinical history should help guide management as well. For instance, the new onset of blood mixed with stool, unrelated to any straining, tenesmus, or other

symptoms common to anorectal pathology, should heighten concern for a neoplasm. In such a case, colonoscopy may be the most appropriate study. Additionally, a patient with benign findings on sigmoidoscopy who continues to bleed, despite directed conservative measures, deserves colonoscopy to rule out missed or right-sided lesions.

Suspected Acute Colitis

In the patient complaining of recent cramping, altered bowel consistency, and/or bleeding per rectum, a diagnosis of acute colitis must be considered. Causes of acute colitis are varied and include infections, ischemia, mucosal toxicity related to medications, chemotherapy or radiation, flares of ulcerative colitis or Crohn's colitis, and diverticulitis. Usually the clinical scenario lends to likely etiologies. For instance, the hospitalized individual on antibiotics with new fever and diarrhea may have infectious colitis due to infection with *Clostridium difficile*, whereas an elderly person with vascular disease and new-onset cramping and bloody diarrhea likely suffers from ischemic colitis. Often, empiric therapy for acute colitides is reasonable. Yet if the diagnosis or cause is not certain, empiric therapy may be hazardous to the individual, or symptoms are severe or unresponsive to empiric therapy, further investigation is warranted. Flexible sigmoidoscopy, with or without biopsies, will generally suffice, as it is uncommon for an acute colitis to solely involve the proximal colon. In fact, bowel preparation is often not necessary, as these patients rarely have significant amounts of solid stool in the distal colon and the preparation may alter or worsen mucosal findings. Biopsies of the affected mucosa may be quite helpful, but, sometimes, endoscopic views will clinch a diagnosis, such as the typical pseudomembranes of *C. difficile*, the hypervascular markings of radiation proctitis, or the segmental distribution of ischemic colitis. On occasion, acute colitis may warrant full colonoscopy as in suspected Cytomegalovirus-related colitis, which may affect only the proximal colon, or when CT scanning shows a proximal colitis.

Surveillance of Chronic, Left-Sided Disease

Ulcerative colitis, and less often Crohn's colitis, may be limited to the distal colon. In patients with pancolitis, increased risk of colorectal cancer has led to regular surveillance programs for patients with duration of disease over 8 years. In patients with distal ulcerative colitis only, the time to increased cancer risk is unclear, but most physicians start surveillance programs at 12 years from diagnosis. For these patients, surveillance biopsies of the involved bowel segments can be performed with sigmoidoscopy (although some alternate sigmoidoscopic surveillance of the affected bowel with full colonoscopic surveillance). Some patients with familial adenomatous polyposis will have undergone colectomy but have an ileocolonic anastomosis with a retained rectal remnant. These patients need continued surveillance of the rectum given the persistent risk of cancer involving the rectum.

Directed Investigation of Suspected Left Colon Lesions

Not infrequently, colonoscopy or sigmoidoscopy is precipitated by abnormal radiographs. A patient at low risk for colorectal cancer (e.g., a person under 40 with no family history) or who is a poor candidate for colonoscopy (very elderly, postmyocardial infarction, critically ill, etc.) may have a suspicious finding on barium enema or CT scan that seems to involve the distal bowel. This finding may represent a mass, wall thickening, ulceration, suspected colonic invasion by tumor in an adjacent organ, or simply radiographic artifact. In such a scenario, an evaluation of the colon limited to the distal bowel may suffice. Similarly, in a patient with known gynecologic or urologic malignancy there may be suspicion of colonic involvement, based on hematochezia or suggestive radiographs. If colonic extension of tumor is suspected in the pelvis, as is commonly the case, sigmoidoscopy is adequate to answer the question.

Directed Investigation or Treatment of Known Left Colon Lesions

Uncommonly, a distal colonic lesion needs treatment and the more proximal colon has already been cleared with colonoscopy. An endoscopist may wish to revisit a recent colonic polypectomy site, when residual adenomatous tissue is suspected. Another occasion is planned therapy for radiation proctitis with endoscopic fulguration or vaporization. Note that whenever electrical or photochemical therapies are used in the colon, even in unsedated patients with lesions well within the reach of the sigmoidoscope, a full bowel preparation is necessary. If the bacteria that produce hydrocarbon gases are not purged from the colon, the aforementioned energy sources could, theoretically, spark an explosion within the bowel lumen.

Colonic Symptoms without Bleeding

The use of lower endoscopy in the evaluation of chronic abdominal pain, altered bowel habits, or weight

loss is of low yield, although commonplace. Colonic neoplasia will seldom explain constipation, but if present will occur on the left side where stool is more formed and subject to partial obstruction more easily. Low abdominal pain sometimes is explained by diverticular disease (relapsing diverticulitis) or, more commonly, by spastic colon, which often can be appreciated endoscopically. Colonic insufflation may often reproduce a patient's reported pain, and conversely, a proper bowel prep may alleviate chronic pain, suggesting functional bowel symptoms. Finding melanosis coli (hyperpigmented mucosa) will suggest the prolonged laxative use that often accompanies chronic constipation. Symptomatic patients over 50 years might benefit from full colonoscopy, as it is a better test to screen for neoplasia, but any pertinent information regarding their symptoms should be available in the left colon. Note that when lower endoscopy is used to evaluate chronic diarrhea, full colonoscopy with biopsies is needed to rule out microscopic colitis, which sometimes occurs only in the right colon. Similarly, full colonoscopy is needed in the patient infected with human immunodeficiency virus who presents with chronic diarrhea.

PROCEDURAL RISKS

Discomfort

Some discomfort should be anticipated by the patient, but it is important to reiterate this point. Patients need to know that the endoscopist is attuned to their subjective experience, will try to minimize painful stimuli, and will respond to the patients' input during the test. Air insufflation and scope looping are the primary sources of pain and both these sources can be ameliorated. Anal pain due to fissure or hemorrhoids should be managed proactively with anesthetic lubricant if the preprocedural rectal exam suggests the need.

Bleeding

Mucosal scratches due to the endoscope are rare and seldom bleed. However, friable mucosa associated with colitis or uncorrected coagulopathy may bleed slightly with even the mildest scope trauma. The removal of polyps with forceps (removal of larger polyps with other instruments is rarely undertaken during sigmoidoscopy and is discussed elsewhere) almost never leads to clinically significant bleeding, although rectal polypectomies may result in a small amount of blood with the stool over the subsequent few hours. In colonoscopy studies, bleeding rates for purely diagnostic examinations are approximately 0.02% and the risk for sigmoidoscopy is likely comparable. The bleeding risk for colonoscopic polypectomy cases is significantly higher, but large polyps are not removed during sigmoidoscopy. The added risk with forceps-mediated polypectomies is not known, but unlikely to be much higher than for purely diagnostic studies, based on common experience.

Perforation

Perforation rates for flexible sigmoidoscopy are roughly 1 in 10,000. The risk for this complication is lower than during colonoscopy for several reasons: less scope is inserted, less air is insufflated, few polypectomies utilizing cautery are performed, the thin-walled cecum will become less inflated than during colonoscopy, and the awake patient will complain of the significant looping or distension that increases risk of perforation. Although event rates are low with sigmoidoscopy, each procedure should be carried out with the utmost caution. Every patient must be warned of the risk for perforation with endoscopy.

Infection

The risk of infection related to sigmoidoscopy may be more theoretical than practical. In the past, much was written about postendoscopic bacteremia caused by bowel-flora translocation via a thin, distended bowel wall. Studies have demonstrated bacteremia occurring after sigmoidoscopy, with the greatest concern being for patients who could not cleanse the portal blood, i.e., cirrhotics and patients with vascular prostheses. Yet, the data have never clearly shown a clinically significant infectious risk with sigmoidoscopy, although rare cases may still occur.

Chemically Induced Colitis

Colitis can occur as a result of incompletely rinsed disinfectants. Hydrogen peroxide and glutaraldehyde are commonly used for disinfecting endoscopes. If these agents are not thoroughly washed off the instrument shaft and channels, they can subsequently come in contact with the next patient in whom the instrument is used. These chemicals will cause a self-limited colitis with patients experiencing postprocedural cramping, tenesmus, and bleeding per rectum. Therapy is supportive, but preemptive avoidance via proper endoscopic cleaning and rinsing techniques is most important.

Vasovagal Reaction

It is not uncommon for anal stimulation with the endoscope and/or colonic distension with air to trigger a vasovagal reaction. The endoscopist should be aware of this and realize that air suction and scope withdrawal can correct the condition. Some conversation between the endoscopist and the patient during the procedure allows the latter to report suggestive symptoms.

Other Risks

There are other potential complications of sigmoidoscopy, such as postpolypectomy syndrome, pneumatosis coli, ischemic colitis, electrolyte abnormalities, and "mini-perforation." These are rare with sigmoidoscopy. Similarly, risks associated with conscious sedation will not be discussed here.

CONTRAINDICATIONS

Absolute Contraindications

Suspected Severe or Toxic Colitis, Including Acute Diverticulitis

When the bowel wall is severely inflammed it may become thin and particularly susceptible to perforation. Sigmoidoscopy should not be performed in this setting, unless it is absolutely necessary for diagnosis and management. In that circumstance, it should be performed only by an experienced endoscopist, with care taken to avoid advancing the sigmoidoscope beyond the level of inflammed mucosa. If there is clinical or radiological evidence of toxic megacolon, sigmoidoscopy should not be performed.

Suspected Perforation

A small tear in the bowel will surely enlarge with air insufflation, potentially leading to more leakage of bowel contents into the peritoneal cavity. Some patients with small perforations may be able to avoid surgery if they "self-seal" with overlying omentum, a process likely to be disrupted with added bowel distension. Similarly, the patient who will need an operation anyway will likely need a lengthier surgery after endoscopy creates a larger hole with more spillage.

Suspected Bowel Obstruction

In an obstruction, the dilated bowel has increased wall tension. The inevitable insufflation that occurs during endoscopy may lead to hyperbaric trauma with perforation and peritoneal soilage. Bowel obstruction should be viewed as a surgical emergency and never as an endoscopic question awaiting an answer. That being said, there are times when a known sigmoid volvulus may be amenable to endoscopic reduction (usually after gastrograffin enema has failed to do so) or persistent pseudo-obstruction may benefit from endoscopic placement of a decompression tube. These are procedures with significant risks for perforation, generally performed to postpone surgical intervention. Such procedures should be performed only by the most experienced therapeutic endoscopist and with surgical backup.

Agitated Patient

If a patient is overly anxious, is belligerent, or otherwise cannot lie still for an examination, the risk of traumatic injury outweighs the benefits of the procedure.

Relative Contraindications

Acute Illness

Sigmoidoscopy is usually an elective procedure and, as such, should await a patient's recovery to reasonable health. Even in skilled hands, sigmoidoscopy can be somewhat uncomfortable and its risks, although usually negligible, will rise in the patient who is already uncomfortable or in any way prone to hemodynamic instability.

Coagulopathy

The risk of bleeding with sigmoidoscopic examination is quite low and an examination usually can be performed safely in the anticoagulated patient, although biopsies might well be avoided in such subjects.

Colonoscopy Is Preferable Test

A patient who ought to have colonoscopy for reasons of symptoms, colorectal cancer screening, or surveillance should not undergo sigmoidoscopy as it would not obviate the need for complete colonic evaluation.

Recent Bowel Surgery

Within 5 to 7 days after bowel surgery, hyperbaric trauma, such as that resulting from endoscopic insufflation, should be avoided to reduce risk of anastomotic breakdown.

Recent Myocardial Infarction

Within 3 weeks of acute myocardial infarction, all endoscopy should be limited to emergent cases. Most of the risk relates to administration of conscious sedation and its potential hemodynamic consequences. Yet, even without sedation, sigmoidoscopy can cause tachycardia or vagal reaction, which could incite ischemia in the

recently injured myocardium. If endoscopy must be performed, insufflation should be minimized and electrocardiographic monitoring is advisable.

Pregnancy

Studies have shown that sigmoidoscopy is safe during each trimester of pregnancy. That being said, there is a theoretical risk of placental abruption, so purely elective procedures, as to work up minor bleeding or altered bowel habits, might best be postponed until postpartum. However, more pressing issues, such as ongoing rectal bleeding or colitis of uncertain cause may be safely pursued endoscopically, without delay.

SIGMOIDOSCOPISTS

The number of sigmoidoscopies performed has increased with the widening appreciation of colon cancer screening. And even so, a nationwide 1997 survey revealed that only 30% of potential candidates in the United States were having screening flexible sigmoidoscopies. Although screening colonoscopy has become accepted by most third-party payers, the number of skilled colonoscopists, mostly trained gastroenterologists, is insufficient to meet the needs of the populace. Therefore, sigmoidoscopy will continue to play a role in screening until a less invasive, inexpensive, colorectal cancer screening test replaces it. Until that test is found and accepted, sigmoidoscopy needs to be performed by a wide cadre of professionals. Studies have demonstrated that internists, general practitioners, physician assistants, and nurse practitioners can properly perform the test. In order to allow more patients to benefit from colorectal cancer screening, flexible sigmoidoscopy training programs for nongastroenterologist health care providers should be promoted.

See Also the Following Articles

Colonoscopy • Colorectal Adenocarcinoma • Colorectal Adenomas • Colorectal Cancer Screening • Endoscopy, Complications of • Lower Gastrointestinal Bleeding and Severe Hematochezia • Virtual Colonoscopy

Further Reading

Cotton, P., and Williams, C. (1996). "Practical Gastrointestinal Endoscopy," 4th Ed. Blackwell Science Inc., Cambridge, MA.

Lieberman, D., and Weiss, D. (2001). One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. *N. Engl. J. Med.* **345**, 555–560.

McCallion, K., Mitchell, R., Wilson, R., *et al.* (2001). Flexible sigmoidoscopy and the changing distribution of colorectal cancer: Implications for screening. *Gut* **48**, 522–525.

Miller, B. (2001). A review of sigmoidoscopy. *In* "An Atlas of Sigmoidoscopy and Cystoscopy" (B. Miller, ed.), The Encyclopedia of Visual Medicine Series, Parthenon, New York.

Painter, J., Saunders, D., Bell, G., *et al.* (1999). Depth of insertion at flexible sigmoidoscopy: Implications for colorectal cancer screening and instrument design. *Endoscopy* **31**, 227–231.

Ransohoff, D., and Sandler, R. (2002). Screening for colorectal cancer. *N. Engl. J. Med.* **346**, 40–44.

Shapiro, J., Seeff, L., and Nadel, M. (2001). Colorectal cancer-screening tests and associated health behaviors. *Am. J. Prev. Med.* **21**, 132–137.

Van Rosendaal, G., Sutherland, L., Verhoef, M., *et al.* (2000). Defining the role of fiberoptic sigmoidoscopy in the investigation of patients presenting with bright red rectal bleeding. *Am. J. Gastroenterol.* **95**, 1184–1187.

Waye, J., Kahn, O., and Auerbach, M. (1996). Complications of colonoscopy and flexible sigmoidoscopy. *Gastrointest. Endosc. Clin. North Am.* **6**, 343–377.

Sinusoidal Obstruction Syndrome (Hepatic Venoocclusive Disease)

LAURIE DELEVE
University of Southern California

conditioning regimen Procedures performed in preparation for stem cell transplantation, which consists of high-dose chemotherapy with or without total body irradiation.
hematopoietic stem cell transplantation A procedure formerly known as bone marrow transplantation.
pyrrolizidine alkaloids Plant-derived alkaloids that may cause sinusoidal obstruction syndrome.
sinusoidal obstruction syndrome A nonthrombotic obstruction of the hepatic circulation with subsequent centrilobular sinusoidal fibrosis and, often, fibrotic obliteration of hepatic venules. Formerly known as hepatic venoocclusive disease.

Sinusoidal obstruction syndrome (SOS) is a disease of the hepatic circulation that leads to parenchymal dysfunction. SOS is caused by exposure to plant toxins or to drugs with or without concurrent total body irradiation. In North America and Western Europe, SOS occurs mainly due to the preparative conditioning regimen for stem cell transplantation. Although 70 to 85% of patients with SOS survive, severe SOS is almost uniformly fatal. There is currently no specific therapy for SOS with good efficacy.

ETIOLOGY

The earliest case reports of sinusoidal obstruction syndrome (SOS) in humans were of individuals in South Africa who ingested teas containing pyrrolizidine alkaloids, so-called "bush tea disease." Bras *et al*. in Jamaica coined the name hepatic venoocclusive disease in response to the most prominent histologic feature, notably the narrowing or obliteration of intrahepatic venules. In many non-Western nations, ingestion of teas or of food sources contaminated by pyrrolizidine alkaloids may still be the most common cause of the disease (see Table I). However, in North America and Europe, the most common cause is high-dose myeloablative chemotherapeutic regimens, alone or in conjunction with total body irradiation, used in preparation for hematopoietic stem cell transplantation. SOS is also seen in renal and liver transplant patients and this is attributed to long-term azathioprine therapy. Chemotherapeutic agents that are associated with SOS at conventional doses include gemtuzumab ozogamicin, actinomycin D, dacarbazine, 6-thioguanine, cytosine arabinoside, mithramycin, and urethane. Given the increased use of herbal remedies, sporadic cases of SOS may be seen due to teas that contain pyrrolizidine alkaloids, notably teas from *Crotalaria*, *Senecio*, *Heliotropium*, *Comfrey*, *Gordolobo yerba*, *Ilex*, and *Mate*. Hepatic irradiation in excess of 30 to 35 Gy in adults causes radiation-induced liver disease, a liver disease that shares some of the features of SOS.

In other intrinsic liver diseases, progressive parenchymal dysfunction may eventually cause portal hypertension. Sinusoidal obstruction syndrome differs in that

TABLE I Causes of Sinusoidal Obstruction Syndrome

Chemotherapy
Myeloablative regimens (stem cell transplantation)
Cyclophosphamide–total body irradiation
Busulfan–cyclophosphamide
BCNU–cyclophosphamide–etoposide[a]
Carboplatin–cyclophosphamide–BCNU
Busulfan–melphalan
Conventional dose chemotherapy
Gemtuzumab ozogamicin
Actinomycin D
Dacarbazine
Cytosine arabinoside
6-Thioguanine
Carmustine
Lomustine
Urethane
Indicine *N*-oxide
Pyrrolizidine alkaloids
Crotalaria
Senecio (Adenostyles)
Heliotropium
Comfrey
Gordolobo yerba
Ilex
Mate
Azathioprine immunosuppression

[a] BCNU, carmustine.

 Copyright 2004, Elsevier (USA). All rights reserved.

it is a disease of the hepatic circulation that may cause parenchymal dysfunction. The circulatory impairment is a nonthrombotic obstruction at the level of the sinusoids and involvement of central veins and venules is more common with more severe disease. The change in name from hepatic venoocclusive disease to sinusoidal obstruction syndrome is based on the recognition that the disease is initiated in the hepatic sinusoids and that the clinical signs and symptoms can occur without involvement of the hepatic venules.

CLINICAL FEATURES OF SOS AFTER STEM CELL TRANSPLANTATION

It would be too cumbersome to try to describe SOS in the different settings. The rest of this article will therefore focus on SOS in patients treated with myeloablative chemotherapeutic regimens. It should be noted that SOS due to long-term ingestion of pyrrolizidine alkaloids has a more chronic course than that due to toxicity from short-term exposure to chemotherapy.

Incidence

The reported incidence of SOS varies from 0 to 50% in patients undergoing stem cell transplantation for malignancies. This wide range is largely due to differences in conditioning regimens, i.e., chemotherapeutic regimens used prior to stem cell transplantation. Other important variables relate to patient selection criteria: the risk of SOS is increased in patients undergoing a second transplant, in patients with malignancy not in remission, in patients with chronic hepatitis C or fibrotic liver disease, or in patients with lower performance status. There has been a decline in recent years in the incidence of SOS with the application of strategies to reduce the risk of this syndrome.

Diagnosis

The hallmark features of SOS are tender hepatomegaly, weight gain due to fluid retention, and hyperbilirubinemia. The diagnosis is usually made based on published clinical criteria (Table II). The differential diagnosis includes (hyper)acute graft-versus-host disease, cholestasis associated with sepsis or cyclosporine therapy, hemolysis, congestive heart failure, and decompensated chronic viral hepatitis. Combinations of illnesses that occur posttransplantation may be particularly difficult to distinguish from SOS, e.g., sepsis complicated by cholestasis and renal insufficiency.

Ultrasound may demonstrate features consistent with SOS and may exclude other causes, but cannot establish the diagnosis. Thus, ultrasound may demonstrate hepatomegaly, ascites, lack of biliary dilation, and absence of tumor invasion in the parenchyma or the hepatic vasculature. Early in SOS there may be attenuation of hepatic venous flow and later in the disease there may be reversal of portal flow, but prospective studies of these features did not show them to be diagnostic.

The most useful additional diagnostic tool is transvenous liver biopsy, which will both provide biopsy material and allow measurement of the hepatic venous pressure gradient. Thrombocytopenia due to the conditioning therapy will restrict the use of percutaneous liver biopsy, but the transvenous approach can be safely performed with platelet counts as low as 30,000/mm^3. A hepatic venous pressure gradient of greater than 10 mm Hg is highly specific for SOS in stem cell transplantation patients.

Prognosis

It is likely that all or most patients who undergo myeloablative stem cell transplantation sustain some degree of liver damage. By definition, patients with mild SOS recover without therapy. Moderate SOS may require diuretics or pain medication and the majority of patients survive, whereas severe SOS is almost universally fatal. Patients with severe SOS most commonly die from multiorgan failure, i.e., renal and cardiopulmonary failure. Death usually occurs 30 to 60 days after conditioning therapy, although the outcome is often evident by day 20.

Published case fatality rates for SOS vary widely. Based on findings from several large studies, the case fatality rate for SOS after cyclophosphamide-containing regimens seems to be approximately 30%, but may be approximately 15% for SOS caused by other alkylating

TABLE II Clinical Diagnostic Criteria for SOS in Stem Cell Transplantation Patients

Seattle criteria	Baltimore criteria
At least 2 of 3 findings within 20 days of stem cell transplantation: Bilirubin >2 mg/dl Hepatomegaly or right upper quadrant pain >2% weight gain due to fluid retention	Elevated bilirubin (>2 mg/dl) plus 2 of 3 clinical findings: Tender hepatomegaly >5% weight gain Ascites

agents. There are published graphs that can help predict the outcome for SOS due to cyclophosphamide-containing regimens. Alanine aminotransferase levels greater than 750 U/liter are associated with a poor prognosis. Other predictors of poor outcome include higher hepatic venous pressure gradient, portal vein thrombosis, doubling of baseline serum creatinine, and declining oxygen saturation.

PREVENTION OF SOS

As with any form of drug-induced liver disease, the best approach to primary prevention is to avoid the therapy in those at highest risk. Those at highest risk include patients with hepatitis C, extensive hepatic fibrosis, or cirrhosis, individuals who have previously received myeloablative regimens, patients with malignancy not in remission, and patients with a previous episode of SOS.

There are several strategies that may reduce the risk of SOS. Nonmyeloablative regimens that do not contain hepatotoxic drugs are one possibility. Myeloablative regimens may be modified to reduce risk. The value of therapeutic monitoring with dosage adjustment of busulfan is controversial, given the inconsistent outcome of several studies. Administration of cyclophosphamide before busulfan may be protective. Decreased doses of total body irradiation are associated with a lower incidence of SOS. It is controversial whether modification of the radiation technique reduces the risk. The source of irradiation may be important, since cobalt sources and linear accelerator differ in the dose rate. Avoidance of cyclophosphamide-containing regimens in patients at risk for SOS is another possible approach. Although some of the approaches listed above may reduce the risk of SOS, this needs to be weighed against the potential for increased risk of graft-versus-host-disease, poor engraftment, or unsuccessful treatment of the underlying malignancy.

Heparin infusion or low-molecular-weight heparin is used routinely in many centers, but there are no randomized studies that demonstrate that this successfully prevents fatalities from SOS. Urodeoxycholic acid, prostaglandin E1, and pentoxifylline have also been tried, but the preponderance of evidence does not support the efficacy of prevention by these drugs.

MANAGEMENT OF SOS

Supportive Care

There is currently no specific and satisfactory therapy for SOS. Conventional supportive care is used to manage fluid and electrolyte balance and ascites. Renal and pulmonary failure may necessitate hemodialysis and mechanical ventilation, but outcome will more likely reflect the severity of the underlying liver disease.

Pharmacologic Therapy

The combination of tissue plasminogen activator and heparin is beneficial in approximately 30% of patients with severe SOS. However, in the largest series to date, patients with renal and pulmonary failure did not benefit. Patients at risk for intracerebral or pulmonary hemorrhage should be excluded from consideration for this approach.

Defibrotide is an experimental drug that has shown promise in uncontrolled trials. Various other pharmacologic approaches have been described in case reports, but have not been confirmed in clinical trials.

Transjugular Intrahepatic Portosystemic Shunt

Transjugular intrahepatic portosystemic shunt (TIPS) has been used in these patients with improvement in ascites. However, outcome of SOS is dependent on the underlying liver disease, which is not altered by TIPS placement.

Liver Transplantation

Liver transplantation is not usually a consideration for patients who undergo stem cell transplantation for malignancy. SOS is rare in patients who undergo transplantation for nonmalignant indications, but liver transplantation for SOS should certainly be considered in patients with a favorable prognosis for their underlying disease.

See Also the Following Articles

Cirrhosis • Hepatic Circulation • Hepatitis C • Liver Transplantation

Further Reading

Bearman, S. I. (1995). The syndrome of hepatic venoocclusive disease after marrow transplantation. *Blood* **85**, 3005–3020.

Bearman, S. I., Anderson, G. L., *et al.* (1993). Venoocclusive disease of the liver: Development of a model for predicting fatal outcome after marrow transplantation. *J. Clin. Oncol.* **11**, 1729–1736.

Brooks, S. E. H., Miller, C. G., *et al.* (1970). Acute venoocclusive disease of the liver. *Arch. Pathol.* **89**, 507–520.

Carreras, E. (2000). Venoocclusive disease of the liver after hematopoietic cell transplantation. *Eur. J. Haematol.* **64**, 281–291.
DeLeve, L. D. (2002). Cancer chemotherapy. *In* "Drug-Induced Liver Disease" (N. Kaplowitz and L. D. DeLeve, eds.), Chap. 24., pp. 593–632. Dekker, New York.
DeLeve, L. D., Shulman, H. M., *et al.* (2002). Toxic injury to hepatic sinusoids: Sinusoidal obstruction syndrome (venoocclusive disease). *Semin. Liver Dis.* **22**, 623–638.
Rollins, B. J. (1986). Hepatic venoocclusive disease. *Am. J. Med.* **81**, 297–306.

Sjögrens Syndrome

STANLEY R. PILLEMER
National Institute of Dental and Craniofacial Research, National Institutes of Health

autoantibodies Antibodies directed against a normal cellular component.
cevimeline Cholinergic drug that stimulates secretion of most exocrine glands, including lacrimal and salivary glands.
keratoconjunctivitis sicca Dry eye and associated symptoms due to the absence of the aqueous component of tears.
lacrimal glands The glands producing tear fluids that bathe the surface of the eye (cornea).
pilocarpine Cholinergic drug that stimulates secretion of most exocrine glands, including lacrimal and salivary glands.
rheumatoid factor An autoantibody directed against immune globulin that is present in rheumatoid arthritis.
T cells Thymus-derived lymphocytes.
xerostomia Dry mouth due to the absence of salivary secretions.

Sjögrens syndrome (SS) is a debilitating, systemic, autoimmune disorder with prominent exocrinopathy that has been described as an "epithelitis." SS may be categorized as primary or secondary. In secondary SS, the disorder coexists with other autoimmune diseases, such as rheumatoid arthritis, systemic lupus erythematosus, scleroderma, polymyositis, and polyarteritis nodosa. Salivary and lacrimal gland involvement is typical of SS and is associated with decreased production of saliva and tears. Other epithelial components of the body are commonly involved, including the skin, as well as the urogenital, respiratory, and gastrointestinal tracts. Systemic autoimmune manifestations include synovitis, neuropathy, vasculitis, and autoantibodies, particularly antinuclear antibodies, anti-SSA, and anti-SSB, as well as rheumatoid factor. Immunoglobulin levels are frequently elevated. SS is associated with an increased risk of lymphoma, especially mucosal-associated lymphoid tissue lymphomas of B-cell lineage.

EPIDEMIOLOGY AND CLASSIFICATION CRITERIA

Primary Sjögrens syndrome (SS) affects 0.3 to 4.8% of the population and appears to increase in frequency with age. The physician-diagnosed incidence is 4 per 100,000 population per year. Females outnumber males by 9 : 1. Although the peak incidence occurs in midlife, SS may occur at any age. Onset is often insidious and diagnosis may be delayed for years.

Various classification criteria have been proposed for SS. These criteria have generally included keratoconjunctivitis sicca, xerostomia, and autoantibodies. American European Consensus criteria are currently the most widely accepted. The recently revised criteria are shown in Table I. The rules for applying the criteria are shown in Table II.

PATHOGENESIS

The pathogenesis of SS remains unknown. Viruses that can cause disease in salivary glands, such as cytomegalovirus, hepatitis C, and the retroviruses human T-cell lymphotropic virus-1 and human immunodeficiency

 Copyright 2004, Elsevier (USA). All rights reserved.

Carreras, E. (2000). Venoocclusive disease of the liver after hematopoietic cell transplantation. *Eur. J. Haematol.* **64**, 281–291.
DeLeve, L. D. (2002). Cancer chemotherapy. *In* "Drug-Induced Liver Disease" (N. Kaplowitz and L. D. DeLeve, eds.), Chap. 24., pp. 593–632. Dekker, New York.
DeLeve, L. D., Shulman, H. M., *et al.* (2002). Toxic injury to hepatic sinusoids: Sinusoidal obstruction syndrome (venoocclusive disease). *Semin. Liver Dis.* **22**, 623–638.
Rollins, B. J. (1986). Hepatic venoocclusive disease. *Am. J. Med.* **81**, 297–306.

Sjögrens Syndrome

STANLEY R. PILLEMER
National Institute of Dental and Craniofacial Research, National Institutes of Health

autoantibodies Antibodies directed against a normal cellular component.
cevimeline Cholinergic drug that stimulates secretion of most exocrine glands, including lacrimal and salivary glands.
keratoconjunctivitis sicca Dry eye and associated symptoms due to the absence of the aqueous component of tears.
lacrimal glands The glands producing tear fluids that bathe the surface of the eye (cornea).
pilocarpine Cholinergic drug that stimulates secretion of most exocrine glands, including lacrimal and salivary glands.
rheumatoid factor An autoantibody directed against immune globulin that is present in rheumatoid arthritis.
T cells Thymus-derived lymphocytes.
xerostomia Dry mouth due to the absence of salivary secretions.

Sjögrens syndrome (SS) is a debilitating, systemic, autoimmune disorder with prominent exocrinopathy that has been described as an "epithelitis." SS may be categorized as primary or secondary. In secondary SS, the disorder coexists with other autoimmune diseases, such as rheumatoid arthritis, systemic lupus erythematosus, scleroderma, polymyositis, and polyarteritis nodosa. Salivary and lacrimal gland involvement is typical of SS and is associated with decreased production of saliva and tears. Other epithelial components of the body are commonly involved, including the skin, as well as the urogenital, respiratory, and gastrointestinal tracts. Systemic autoimmune manifestations include synovitis, neuropathy, vasculitis, and autoantibodies, particularly antinuclear antibodies, anti-SSA, and anti-SSB, as well as rheumatoid factor. Immunoglobulin levels are frequently elevated. SS is associated with an increased risk of lymphoma, especially mucosal-associated lymphoid tissue lymphomas of B-cell lineage.

EPIDEMIOLOGY AND CLASSIFICATION CRITERIA

Primary Sjögrens syndrome (SS) affects 0.3 to 4.8% of the population and appears to increase in frequency with age. The physician-diagnosed incidence is 4 per 100,000 population per year. Females outnumber males by 9 : 1. Although the peak incidence occurs in midlife, SS may occur at any age. Onset is often insidious and diagnosis may be delayed for years.

Various classification criteria have been proposed for SS. These criteria have generally included keratoconjunctivitis sicca, xerostomia, and autoantibodies. American European Consensus criteria are currently the most widely accepted. The recently revised criteria are shown in Table I. The rules for applying the criteria are shown in Table II.

PATHOGENESIS

The pathogenesis of SS remains unknown. Viruses that can cause disease in salivary glands, such as cytomegalovirus, hepatitis C, and the retroviruses human T-cell lymphotropic virus-1 and human immunodeficiency

TABLE I Revised American–European Classification Criteria for Sjögren's Syndrome

I. **Ocular symptoms**: A positive response to one or more the questions below:
   1. Have you had daily, persistent, troublesome dry eyes for more than 3 months?
   2. Do you have a recurrent sensation of sand or gravel in the eyes?
   3. Do you use tear substitutes more than 3 times a day?

II. **Oral symptoms**: A positive response to one or more of the questions below:
   1. Have you had a daily feeling of dry mouth for more than 3 months?
   2. Have you had recurrently or persistently swollen salivary glands as an adult?
   3. Do you frequently drink liquids to aid in swallowing dry food?

III. **Ocular signs**: Objective evidence of eye involvement defined as a positive result for one or both of the two tests below:
   1. Positive Schirmer's I test, performed without anesthesia (≤ 5 mm wetting in 5 min)
   2. Positive Rose Bengal or other ocular dye staining with Van Bijsterveld score ≥ 4.

IV. **Histopathology**: In minor salivary glands (obtained through normal-appearing mucosa) focal lymphocytic sialoadenitis, evaluated by an expert histopathologist, with a focus score ≥ 1, defined as a number of lymphocytic foci that are adjacent to normal-appearing mucous acini and contain more than 50 lymphocytes per 4 mm^2 of glandular tissue.

V. **Salivary gland involvement**: Objective evidence of involvement indicated by a positive result for at one or more of the diagnostic tests below:
   1. Unstimulated whole salivary flow (≤ 1.5 ml in 15 min)
   2. Parotid sialography showing the presence of diffuse sialectasias (punctate, cavitary, or destructive pattern), without evidence of obstruction in the major ducts.
   3. Salivary scintigraphy showing delayed uptake, reduced concentration, and/or delayed excretion of tracer.

VI. **Autoantibodies**: Presence in the serum of the following autoantibodies:
   1. Antibodies to Ro(SSA) or La(SSB) antigens, or both.

virus type 1, have been considered as possible triggers for SS. However, no causative organism has been found.

Hormonal factors, such as relatively low levels of androgens, may explain the preponderance of females with SS. Also, hypofunction of the hypothalamic–pituitary–adrenal axis has been described in SS.

Genetic factors may also have a role in SS. An increased prevalence of SS and autoantibodies, particularly anti-Ro/SSA, occurs in family members, and HLA-DR3, HLA-DQ2, and other genetic markers are associated with SS.

Focal mononuclear cell infiltration of exocrine tissues and the presence of autoantibodies, especially anti-Ro/SSA, anti-La/SSB, and rheumatoid factor, are key features of SS. The periductal infiltrate in the salivary and tear glands consists mainly of T cells with fewer B cells, macrophages, and mast cells. Most of the T cells are $CD4^+$ helper cells with the memory phenotype $CD45RO^+$ and appear to be resistant to apoptosis despite increased expression of Fas. Apoptosis may be blocked by the suppressor proto-oncogene Bcl-2, allowing autoreactive cells to persist in the exocrine tissues.

Salivary and lacrimal gland epithelial cells in SS show increased HLA-DR antigen expression, which allows these cells to present antigens, including autoantigens, to the CD4 T cells. Cellular interactions can result in the production of cytokines. Proinflammatory cytokines interleukin-1β (IL-1β), IL-6, and tumor necrosis factor α tend to be produced by epithelial cells, whereas IL-10 and interferon-γ

TABLE II Revised Classification Rules

Primary SS

In the absence of any potentially associated disease, primary SS may be defined as follows:
   1. Presence of ***any 4 out of the 6 items*** is indicative of primary SS, as long as either item 4 (histopathology) or 6 (serology) is positive.
   2. The presence of ***any 3 of the 4 objective criteria items*** (i.e., items III, IV, V, VI in Table I).
   3. A classification tree procedure may be used as a valid alternative method for classification.

Secondary SS

For patients who have a potentially associated disease (e.g., another well-defined connective tissue disease), the presence of ***item I or item II plus any two from among items III, IV, and V*** from Table I may be considered as indicative of secondary SS.

Exclusions

History of head and neck irradiation, preexisting lymphoma, sarcoidosis, graft-versus-host disease, hepatitis C or human immunodeficiency virus infection, use of anticholinergic drugs (within 4 half-lives of taking the drug).

(IFN-γ) are produced mostly by infiltrating T cells. IFN-γ increases HLA-DR and La/SSB expression by glandular epithelial cells and IL-10 can induce B-cell proliferation. In SS, B-cell activation is typical and may progress toward B-cell lymphoid malignancy.

Activated B cells in SS produce increased amounts of immunoglobulins with autoantibody reactivity for immunoglobulin G (IgG) (rheumatoid factor), Ro/SSA, and La/SSB. Also, B cells produce antibodies targeting the muscarinic M3 receptor. In SS, muscarinic (acetylcholine) receptor blockade in exocrine tissue would inhibit the production of secretions. Since complete destruction of the salivary tissue is rarely seen in salivary gland biopsies of SS patients, it has been suggested that cytokines may interact directly with epithelial cells or autoantibodies, other than those directed against muscarinic M3 receptors. For example, interference with the nerve supply might also decrease secretions in a manner disproportionate to the level of tissue destruction.

CLINICAL FEATURES

Typical manifestations of SS are oral and ocular symptoms and signs of dryness, the presence of autoantibodies, especially anti-Ro/SSA and anti-La/SSB antibodies, and a positive labial salivary gland biopsy.

Eyes

The precorneal tear film has three layers, progressing outward from the corneal surface: mucus, water (aqueous), and oil. The dry eyes of SS are due to aqueous tear deficiency, which, however, may exist in the absence of SS because of other lacrimal gland diseases, lacrimal duct obstruction, and loss of reflex tearing. The aqueous tear deficiency of SS produces symptoms of ocular irritation, particularly a gritty sensation in the eye, and a positive Schirmer's test and the finding of ocular surface damage on the slit lamp examination with increased uptake of ocular dyes are observed. The diagnostic eye tests are shown in Table I.

Mouth

SS patients often experience a sensation of decreased saliva, oral dryness on eating, a need to drink liquids to facilitate swallowing of dry foods, spicy food intolerance, altered taste, and burning mouth. Speech difficulties interfere with social interactions and functioning in the workplace. More than 400 medications are known to be associated with dry mouth symptoms, including drugs commonly used in the treatment of hypertension, insomnia, and depression. Head and neck radiation used to treat tumors produces severe oral dryness. Patients treated with radioactive iodine for thyroid disorders may later present with oral dryness, since the isotope tends to concentrate in the salivary glands. Dental erosions and caries are common, especially on the incisal edges of the teeth and at the gingival margins. A dry sticky mucosa and furrowed tongue are often seen. The mucosa often displays the typical erythema and white patches associated with candidiasis. Decreased or absent saliva pooling is often present. Swallowing difficulties may be evaluated by barium swallows or ultrasound studies.

Major salivary gland swelling and tenderness may occur. The parotid gland swelling displaces the earlobe and extends downward over the angle of the jaw. Medial to the angle of the jaw, submandibular salivary gland swelling may be seen or, more often, palpated. The swelling may transient or chronic and is often recurrent, which distinguishes it from mumps.

Other Clinical Features

Gastrointestinal

In addition to the oral manifestations mentioned above, SS is associated with dysphagia, esophageal dysmotility, and esophageal webs. Lymphoid infiltration may occur in the gastrointestinal tract. Chronic atrophic gastritis may give rise to dyspeptic symptoms, including nausea and epigastric discomfort. *Helicobacter pylori* has been suspected to play a role in the gastrointestinal manifestations of SS, but evidence involving normal population controls is not available. The extent to which small and large bowel involvement occurs in SS is unclear. However, patients frequently complain of bloating and constipation. Patients on muscarinic agonists such as pilocarpine or cevimeline may complain of abdominal cramping and may have an increase in the frequency of bowel movements or diarrhea. Pancreatic involvement may occur, but is usually subclinical. Primary biliary cirrhosis may be associated with SS. Hepatitis C has emerged as an important infection that mimics SS, including the salivary gland lymphocytic infiltration, decreased saliva production, hypergammaglobulinemia, and vasculitis. However, such patients generally have no anti-SSA or anti-SSB antibodies. It is therefore important to rule out hepatitis C in individuals who appear to have SS.

Skin

Dry skin affects approximately half of SS patients. Sweating may be decreased. Dry skin and peripheral

neuropathy may result in pruritis. Scratching repeatedly may produce increased hyperpigmentation, excoriations, and lichenification. Also, some patients may develop palpable or nonpalpable purpura and petechiae, most often on the lower extremities, in showers of lesions lasting several days. The lesions, on microscopic examination, are consistent with either leucocytoclastic vasculitis or mononuclear inflammatory vasculopathy.

Thyroid

Thyroid disorders, most often hypothyroidism, are common in SS patients. This may reflect the prevalence of such thyroid disease in individuals of a similar age in the general population or could be a true association of autoimmune thyroiditis with SS.

Respiratory

Most SS patients with pulmonary involvement do not develop progressive disease. A dry cough is experienced by many SS patients and probably reflects tracheal dryness and decreased mucus production. The cough may also be related to hyperreactive airways in both primary and secondary SS. Also, mild interstitial pulmonary disease and rheumatoid-like pulmonary nodules may occur in SS.

Rheumatologic

Manifestations may include low-grade fever, fatigue, lymphadenopathy, myalgias, arthralgias, and symmetrical, nondeforming polyarthritis, which is responsive to standard antirheumatic therapies.

Neurologic

Peripheral neuropathy is not uncommon in SS, is most notable in the lower limbs, and is most often sensory. Autonomic neuropathy also occurs. Also, the central nervous system involvement has been reported with lesions noted in imaging studies.

Hematologic

Lymphoid malignancy is the most important hematologic complication of SS. Up to a 44-fold increase in the risk of B-cell lymphomas has been reported in SS patients. Non-Hodgkin's lymphomas of mucosa-associated lymphoid tissues are the most common in SS patients, and often salivary glands and cervical lymph nodes are involved.

Reproductive

Women often experience vaginal dryness, which may be the first symptom of SS and is commonly associated with dyspareunia. Anti-SSA (Ro) antibodies occur in SS, raising the possibility of congenital heart block, although the frequency is low.

Renal

Symptoms consistent with irritable bladder are not uncommon. Urinary frequency may result from increased fluid intake to alleviate oral dryness, as well renal abnormalities associated with diminished ability to concentrate urine (hyposthenuria). Renal abnormalities include hyposthenuria in approximately half the cases, distal renal tubular acidosis in approximately 15% of cases, as well as nephrocalcinosis, renal stones, and less often interstitial nephritis and glomerular disease. Urine pH is usually in excess of 5.5 in renal tubular acidosis. The associated systemic acidosis results in mobilization of calcium from bone, promoting osteoporosis and resulting in hypercalciuria. Urinary citrate, which normally complexes a substantial proportion of urine calcium, is decreased. This raises the risk of calcium phosphate stone formation. In approximately half of SS cases, tubular proteinuria may occur. Glomerular disease affects approximately 2% of the patients, tends to be associated with cryoglobulins, and occurs more often in those with longer disease duration.

LABORATORY FEATURES

SS is associated with rheumatoid factor (90%), anti-Ro/SSA or anti-La/SSB (50–90%), and often hypergammaglobulinemia. Antinuclear antibodies occur in approximately 80% of cases. The 52 kDa Ro is more often associated with SS, whereas 60 kDa Ro appears to be more frequent in systemic lupus erythematosus. Anti-Ro/SSA antibodies are associated with systemic manifestations of the disease, including anemia, leukopenia, thrombocytopenia, purpura, cryoglobulinemia, hypocomplementemia, lymphadenopathy, and vasculitis. Other autoantibodies have been recognized in SS, including those directed against carbonic anhydrase, pancreatic antigen, α-fodrin, 97 kDa Golgi complex, mitotic spindle apparatus, M3 muscarinic acetylcholine receptors, and Fcγ receptors. These latter autoantibodies are not currently used to diagnose or monitor the disease.

DIAGNOSIS

The diagnosis of SS may be made by applying the published classification criteria. The differential diagnosis for SS includes the disorders listed as exclusions in Table II.

TREATMENT

Patient Education and Self-Care

The patient will benefit from education about the disease and assistance in developing strategies for self-management as well as for coping with physical, mental, and social challenges associated with their condition.

Sicca Symptoms

Keratoconjunctivitis Sicca

Treatment of dry eyes includes tear replacement and conservation, as well as topical ocular and systemic medications. Artificial tears are instilled as eyedrops to ameliorate symptoms. Preservative-free preparations are best to avoid irritation, ocular surface damage, and allergic reactions. The use of small individual dispensers minimizes the risk of bacterial growth and infection.

Hydroxypropylcellulose pellets inserted under the lower eyelids may be used to prolong the effects of artificial tears. Ointments are used at night, since they are viscous and may interfere with vision. Topical steroids or cyclosporine may be beneficial in the treatment of keratoconjunctivitis sicca.

Systemic secretogogues such as pilocarpine and cevimeline may improve ocular symptoms. Typically, pilocarpine is given in a dosage of 5 mg orally four times a day, and the total daily dose usually does not exceed 30 mg. Adverse effects include increased perspiration, feeling hot and flushed, as well as symptoms associated with increased bowel and bladder motility. Caution must be exercised in the presence of bronchospasm. Some patients who experience adverse effects may benefit from one to three 5 mg doses per day to ameliorate symptoms at the most troublesome time of day. Cevimeline has recently been approved for the treatment of dry mouth in SS at a dosage of 30 mg orally three times daily. Like pilocarpine, it is a muscarinic agonist that increases production of saliva and possibly tears and other secretions. The drug is contraindicated for individuals with uncontrolled asthma, iritis, and narrow angle glaucoma. The role of systemic, immunomodulatory treatment for the ocular manifestations of SS remains unclear.

Xerostomia

Frequent dental care, an appropriate diet, limiting sugar intake, daily topical fluoride use, antimicrobial mouth rinses, and exchanging medications that promote oral dryness or its complications for more appropriate ones may limit the development of caries in patients with reduced salivary flow. Artificial saliva, lubricants, and sugar-free chewing gum or candies may ameliorate oral dryness. Oral moisturizers and lubricants, as well as dietary modifications, may improve dysphagia. Sicca symptoms may be improved by using humidifiers. As in the treatment of ocular dryness, secretagogues such as pilocarpine or cevimeline may increase secretions in patients with sufficient exocrine tissue.

Oral candidiasis commonly complicates the dry mouth of SS and is treated by the use of oral troches or vaginal suppositories of antifungal agents, such as Nystatin or clotrimazole. Angular cheilitis may require topical antifungal agents. Bacterial parotitis should be treated with warm compresses, massage of the parotid gland, and, if necessary, antibiotics.

Systemic Manifestations

Immunomodulatory drugs can be used to treat systemic or exocrine autoimmune and inflammatory manifestations. Hydroxychloroquine is often given for milder systemic manifestations of autoimmune disorders, such as fever, rashes, and arthritis. It remains unclear whether hydroxychloroquine is effective for the exocrine component of the disease, although serological measures improve in SS patients on this drug. Methotrexate, prednisone, azathioprine, and other immunomodulatory drugs have been used in patients with prominent systemic manifestations of SS as is done in patients with systemic lupus erythematosus. However, few randomized, double-blind, clinical trials have been carried out to establish whether these immunomodulatory agents are beneficial in SS.

Consideration should be given to the treatment of other problems associated with SS. For renal tubular acidosis, oral alkaline medications containing sodium and potassium citrate at a dose of 1–2 mEq/kg/day may be necessary to correct acidosis and decrease the risk of kidney stones. Also, urine calcium levels should be monitored. Renal tubular acidosis promotes mobilization of calcium from bone, which requires appropriate monitoring and treatment. Bronchodilators may benefit patients with a chronic cough, since SS is associated with increased airway reactivity. Thyroid disorders in SS patients are managed in the same manner as for other patients of similar age and sex. Dyspepsia and gastroesophageal reflux are not uncommon and require standard therapy. The arthritis occurring in SS patients responds well to the standard therapies used in rheumatoid arthritis. Dry skin may be alleviated by decreased frequency of bathing and application of lubricants. Pruritis may be treated with mentholated lotions. Superficial vasculitis and dermatitis are treated with steroids. Vaginal dryness and dyspareunia may respond to

water-soluble lubricants and vaginal estrogen preparations may be beneficial in postmenopausal females.

The understanding of SS continues to advance. However, only symptomatic therapies have been specifically approved for the treatment of this debilitating disorder and further investigation will be necessary to clarify the role of immunomodulatory agents. Replacement of destroyed salivary gland tissue by artificial salivary glands and the possibilities for gene therapies are under consideration as future treatment options.

See Also the Following Articles

Salivary Glands, Anatomy and Histology • Salivary Glands, Physiology • Xerostomia

Further Reading

Anaya, J. M., and Talal, N. (1999). Sjögren's syndrome comes of age [see comments]. *Semin. Arthritis Rheum.* **28**, 355–359.

Atkinson, J. C., and Baum, B. J. (2001). Salivary enhancement: Current status and future therapies. *J. Dent. Educ.* **65**, 1096–1101.

Atkinson, J. C., and Wu, A. J. (1994). Salivary gland dysfunction: Causes, symptoms, treatment. *J. Am. Dent. Assoc.* **125**, 409–416.

Bell, M., Askari, A., Bookman, A., Frydrych, S., Lamont, J., McComb, J., Muscoplat, C., and Slomovic, A. (1999). Sjögren's syndrome: A critical review of clinical management [published erratum appears in *J. Rheumatol.* (1999) **26**, 2718]. *J. Rheumatol.* **26**, 2051–2061.

Bloch, K. J., Buchanan, W. W., Wohl, M. J., and Bunim, J. J. (1992). Sjögren's syndrome: A clinical, pathological, and serological study of sixty-two cases. *Medicine* **71**, 386–401.

Drosos, A. A., Andonopoulos, A. P., Costopoulos, J. S., Papadimitriou, C. S., and Moutsopoulos, H. M. (1988). Prevalence of primary Sjögren's syndrome in an elderly population. *Br. J. Rheumatol.* **27**, 123–127.

Fox, R. I., and Maruyama, T. (1997). Pathogenesis and treatment of Sjögren's syndrome. *Curr. Opin. Rheumatol.* **9**, 393–399.

Fox, R. I., Tornwall, J., and Michelson, P. (1999). Current issues in the diagnosis and treatment of Sjögren's syndrome. *Curr. Opin. Rheumatol.* **11**, 364–371.

Julkunen, H., Siren, M. K., Kaaja, R., Kurki, P., Friman, C., and Koskimies, S. (1995). Maternal HLA antigens and antibodies to SS-A/Ro and SS-B/La: Comparison with systemic lupus erythematosus and primary Sjögren's syndrome. *Br. J. Rheumatol.* **34**, 901–907.

Pillemer, S. R., Matteson, E. L., Jacobsson, L. T., Martens, P. B., Melton, L. J., III, O'Fallon, W. M., and Fox, P. C. (2001). Incidence of physician-diagnosed primary Sjögren syndrome in residents of Olmsted County, Minnesota. *Mayo Clin. Proc.* **76**, 593–599.

Pflugfelder, S. C., Solomon, A., and Stern, M. E. (2000). The diagnosis and management of dry eye: A twenty-five-year review. *Cornea* **19**, 644–649.

Provost, T. T., and Watson, R. (1992). Cutaneous manifestations of Sjögren's syndrome. *Rheum. Dis. Clin. North Am.* **18**, 609–616.

Sheikh, S. H., and Shaw-Stiffel, T. A. (1995). The gastrointestinal manifestations of Sjögren's syndrome. *Am. J. Gastroenterol.* **90**, 9–14.

Strickland, R. W., Tesar, J. T., Berne, B. H., Hobbs, B. R., Lewis, D. M., and Welton, R. C. (1987). The frequency of sicca syndrome in an elderly female population. *J. Rheumatol.* **14**, 766–771.

Tapinos, N. I., Polihronis, M., Tzioufas, A. G., and Moutsopoulos, H. M. (1999). Sjögren's syndrome: Autoimmune epithelitis. *Adv. Exp. Med. Biol.* **455**, 127–134.

Thomas, E., Hay, E. M., Hajeer, A., and Silman, A. J. (1998). Sjögren's syndrome: A community-based study of prevalence and impact [see comments]. *Br. J. Rheumatol.* **37**, 1069–1076.

Vitali, C., Bombardieri, S., Jonsson, R., Moutsoupoulos, H. M., Alexander, E. L., Carsons, S. E., Daniels, T. E., Fox, P. C., Fox, R. I., Kassan, S. S., Pillemer, S. R., Talal, N., and Weisman, M. H. (2002). Classification criteria for Sjögren's syndrome. *Ann. Rheum. Dis.* **61**, 554–558.

Voulgarelis, M., Dafni, U. G., Isenberg, D. A., and Moutsopoulos, H. M. (1999). Malignant lymphoma in primary Sjögren's syndrome: A multicenter, retrospective, clinical study by the European Concerted Action on Sjögren's Syndrome [see comments]. *Arthritis Rheum.* **42**, 1765–1772.

Small Bowel Transplantation

THOMAS M. FISHBEIN AND STUART KAUFMAN
Mount Sinai School of Medicine, New York

intestinal failure An irreversible state of inability of the native gastrointestinal tract to provide for the nutritional and/or fluid and electrolyte demands of the body.
isolated intestinal transplant Vascularized transplantation of the jejuno-ileum from another person (usually a cadaver).
liver–bowel transplant Combined simultaneous transplantation of a liver and jejuno-ileum, requiring removal of the native liver.
multivisceral transplant Transplantation of the stomach, pancreas, and small intestine, and sometimes other organs, which may include the liver and/or kidney.
parenteral nutrition Administration of intravenous solutions (including carbohydrate, protein, fats, and water) sufficient to provide complete nutritional requirements.
portal or mesenteric drainage Venous effluent from a graft is directed into the superior mesenteric vein or portal vein, so that this blood undergoes a first-pass effect through the liver before entering the systemic circulation.
systemic drainage Venous effluent from a graft is directed to the systemic circulation via the vena cava, so that it does not first filter through the liver.

Intestinal (small bowel) transplantation refers to the removal of the jejuno-ileum from one individual and the transplantation of this organ into another individual. This procedure requires reestablishment of mesenteric arterial inflow and mesenteric venous outflow from the graft and placement of the jejuno-ileal graft in the abdominal cavity. The jejuno-ileum can be transplanted alone, as an isolated small bowel graft, or in combination with a variety of other abdominal organs as part of composite or noncomposite grafts. Intestinal transplantation is generally performed for patients with intestinal failure (i.e., patients whose gastrointestinal tract functions too poorly to provide nutritional autonomy) or, less commonly, for patients with central abdominal tumors that cannot be removed without sacrificing the superior mesenteric vessels and small intestine.

BACKGROUND

Intestinal failure (the inability of the native gastrointestinal tract to provide nutritional autonomy) necessitates lifelong intravenous support to maintain caloric, fluid, or electrolyte homeostasis. Provision of replacement therapy in the form of parenteral nutrition, most often at home, permits survival for the majority of patients. Home parenteral nutrition is therefore the primary long-term therapy for patients with intestinal failure. For patients with intestinal failure, home parenteral nutrition can be considered analogous to renal replacement therapy for patients with end-stage kidney failure.

In some patients, however, the use of parenteral nutrition is limited by its complications, which include parenteral nutrition-associated liver disease, recurrent sepsis associated with line infections or bacterial overgrowth in the native intestinal tract, and loss of venous access for parenteral nutrition due to multiple-line-site thromboses. These complications and others contribute to death in 10–30% of patients with intestinal failure during the first 3–5 years on therapy. Patients with these complications are failing parenteral therapy and are therefore candidates for intestinal transplantation.

More than 50,000 North American patients currently receive parenteral nutrition for treatment of intestinal failure. It has been estimated that 15–20% of these patients are young and otherwise healthy and could be candidates for transplantation. It has also been estimated that 2 new live births per million in Western countries will eventually develop intestinal failure. Thus, although the current population receiving intestinal transplantation is small, with approximately 111 such operations having been performed in North America in 2001, there is a large population for whom this therapy could be useful when success is optimized.

INDICATIONS

Indications for intestinal transplantation can be categorized as being related to loss of intestinal length (short bowel syndrome), loss of intestinal function, or tumors (Fig. 1).

Short bowel syndrome, in which intestinal length and absorptive surface area have been lost due to surgical resection, is the most common cause of intestinal failure leading to intestinal transplantation. The loss of

 Copyright 2004, Elsevier (USA). All rights reserved.

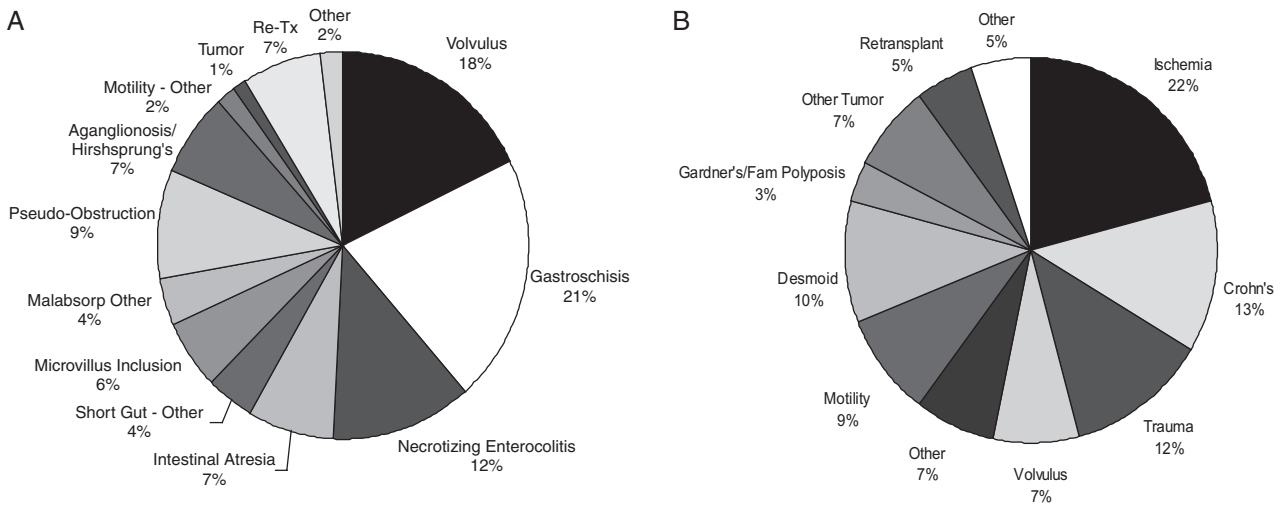


FIGURE 1 Pie charts illustrate the current indications for intestinal transplantation in children (A) and adults (B). Data from the Intestinal Transplant Registry. Available at http://www.intestinaltransplant.org. Accessed January 13, 2003.

mucosal absorptive surface area is associated with malabsorption and rapid transit time through the jejuno-ileum, with resultant malnutrition, recurrent dehydration, and electrolyte abnormalities. Short bowel syndrome may be secondary to a variety of diseases in adults and children, most of which are secondary to vascular or ischemic insults. In children, these include malrotation, volvulus, necrotizing enterocolitis, jejuno-ileal atresias, gastroschisis, omphalocele, and other congenital disorders. Adults frequently suffer short gut syndrome due to trauma, thrombosis or embolism to the mesenteric vessels, inflammatory bowel disease, volvulus, or other causes of infarction.

Functional disorders of the small intestine leading to intestinal failure include disorders of motility and disorders of enterocyte function. Motility disorders can be either myopathic or neuropathic. These include chronic idiopathic intestinal pseudo-obstruction, visceral myopathy, visceral neuropathy, total intestinal aganglionosis, and some forms of mitochondrial respiratory chain disorders that affect gastrointestinal motor function (e.g., mitochondrial neurogastrointestinal encephalomyopathy). Epithelial disorders that lead to secretory diarrhea or failure of absorption in the intestine are more common in children and include microvillus inclusion disease, tufting enteropathy, and autoimmune enteritis.

Tumors involving the base of the jejuno-ileal mesentery are often benign but are locally invasive and therefore lethal. Only complete resection of the tumor and sacrifice of the intestine can provide cure. The most common such lesions are desmoid tumors in patients with familial adenomatous polyposis. This tumor sometimes involves the mesenteric vessels, foreshortens the mesentery, and requires complete exenteration of the small bowel for complete resection. Sometimes these tumors involve other foregut organs that provide portal flow to the liver, such as the pancreas, spleen, stomach, and duodenum. Exenteration of these organs requires concurrent transplantation. Therefore, patients with desmoid tumors sometimes do not have intestinal failure and are not dependent on parenteral nutrition before resection and transplantation, which may be performed concurrently.

TYPES OF TRANSPLANTS

Intestinal transplantation may be performed with an isolated intestinal graft or as a multiorgan transplant procedure. The common element of these procedures is transplantation of the jejuno-ileum, with or without other organs. Under current protocols, an ileostomy is created to allow for easy endoscopic surveillance of the graft for pathology in the early period after transplant. An enteric feeding tube is also placed to allow the early delivery of enteral nutrition, in case oral intake is not complete but the transplanted intestine can provide the full nutritional needs. This is not uncommon among babies who suffer neonatal short gut syndrome and were therefore never fed by mouth during infancy. Even if these children have a fully functional intestinal tract after transplant, they may still require enteral tube feeding while acquiring feeding skills. The different

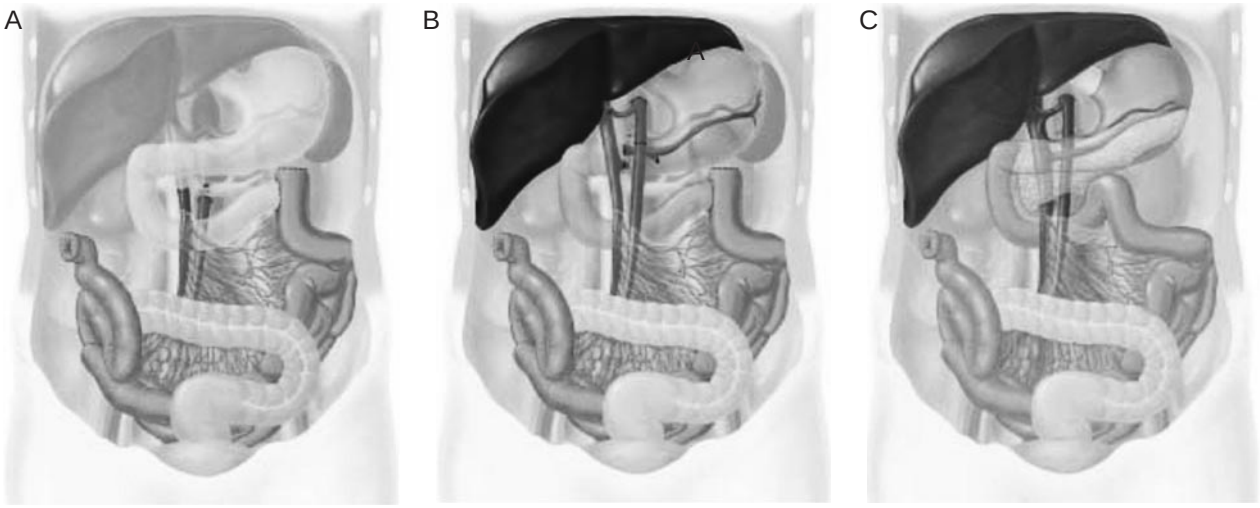


FIGURE 2 Illustrated are the three most common intestinal transplant procedures, which include (A) isolated intestine transplantation with mesenteric venous outflow; (B) liver–intestine transplantation with portacaval shunt draining the native foregut; and (C) multivisceral transplantation. Shaded areas represent native organs; dark areas represent transplanted organs. Reprinted, with permission, from Fishbein, T. M., Gondolesi, G. E., and Kaufman, S. S. (2003). Intestinal transplantation for gut failure (Review). *Gastroenterology*, **124**, 1615–1628.

types of intestinal transplant procedures are illustrated in Fig. 2 and described below.

Isolated Intestinal Transplantation

Isolated intestinal transplantation may be performed with mesenteric/portal drainage or systemic drainage. The drainage refers to the venous outflow of the grafted organ, which may be through the liver (mesenteric/portal) or into the inferior vena cava (systemic), utilizing a variety of surgical techniques. Gastrointestinal continuity is then reestablished, providing the recipient with the ability to take nutrition enterally. Isolated intestinal transplantation is generally indicated when there is intestinal failure without significant underlying liver disease or failure of other organs.

Liver–Intestine Transplantation

Liver–intestine transplantation includes the jejunoileum and liver and generally involves transplantation of these organs with an intact mesenteric–portal circulatory system. Because the native foregut is preserved (native stomach, pancreas, duodenum, and spleen), a portocaval shunt must be performed to provide venous drainage of these organs when the composite liver–bowel graft is transplanted. These two organs may also be transplanted with or without the duodenum and intact biliary system. Transplantation of the organs with an intact duodenum and a portion of the head of the pancreas allows transplantation of the biliary system without disruption and is advantageous in small pediatric patients. This approach is referred to as liver–intestine transplantation with duodenal preservation. Alternatively, the liver and intestine may be transplanted without the duodenum, with a biliary anastomosis created between the donor common bile duct and the transplanted jejunum, with the mesenteric and portal venous system remaining intact. Finally, the patient may receive an orthotopic liver transplant and an isolated intestine transplant, in which the two organs are both transplanted, but not as a composite graft. Combined liver and intestinal transplantation is generally indicated when intestinal failure is accompanied by liver failure, usually due to parenteral nutrition.

Multivisceral Transplantation

Multivisceral transplantation involves transplantation of the stomach, pancreas, and small intestine and sometimes also the liver and/or the kidney. These transplants are always performed as composite organ allografts and involve placement of the new gastrointestinal tract in the native position in the abdomen. Exenteration of the native organs, including the stomach, pancreas, duodenum, spleen, and small bowel, is therefore required. The right colon, receiving its blood supply from the superior mesenteric artery, also must be resected. Since removal of the donor organs necessitates vagal

nerve transection, the pylorus lacks a native vagal relaxation reflex and pyloroplasty is required.

Donor Organ Selection and Procurement

In most cases, donor organ procurement requires a blood-group-identical or -compatible, hemodynamically stable cadaver donor. For isolated intestinal transplantation, the entire jejuno-ileum is usually procured with the superior mesenteric artery and vein and the pancreas and liver may be procured for transplantation elsewhere. When a composite liver and small bowel graft is procured, the shared mesenteric–portal venous drainage of the two organs precludes separate transplantation of the pancreas alone. For multivisceral transplantation, the stomach, pancreas, spleen, small bowel, and liver are procured *en bloc* and the spleen can then be removed in a chilled ice bath. Because the small bowel is a hollow viscus, it is particularly prone to warming and preservation injury, and core cooling with a standard preservation solution (University of Wisconsin solution, Viaspan) is the central tenet of organ preservation. Ischemic time must be kept short and every attempt is made to reperfuse the organ within 12 h after its removal from the donor circulation. Whereas most intestinal transplants have been performed with cadaveric donor organs as described here, only a few transplants have used live-donor partial ileal or jejunal grafts with mixed results. Live donor transplantation is ideal when the patient has an identical twin donor and immunosuppression is therefore not required. Finally, newer reduced-size techniques are being developed to address the shortage of grafts for small children.

MANAGEMENT AFTER TRANSPLANTATION

Management after transplantation requires intensive immunosuppressive therapy to prevent allograft rejection, early recognition of rejection when it occurs, prophylaxis against and recognition and treatment of infectious disorders, and nutritional management.

Surveillance Endoscopy

Surveillance endoscopy is the standard method for diagnosing rejection after transplantation. Currently, there is no noninvasive means to reliably predict or detect rejection of the transplanted jejuno-ileum. Therefore, protocol biopsies are performed at variable intervals, usually weekly or twice weekly, early after transplantation. Rejection cannot be diagnosed solely on the basis of the gross endoscopic appearance of the intestinal mucosa. The process has been shown to be patchy, with areas of normal mucosa interspersed with areas that demonstrate rejection. Therefore, multiple random mucosal pinch biopsies are generally sampled for histologic review.

Review

Rejection is common after intestinal transplantation, with the process directed at the crypt epithelium. Enterocyte apoptosis in the crypt is generally present with rejection early after transplantation (mild rejection), variably associated with an increase in the density of the lamina propria infiltrate, activated lymphocytes, acute cryptitis, and variable degrees of villous blunting (moderate rejection). Such changes are usually, but not always, associated with secretory diarrhea, the clinical hallmark of most posttransplant pathologies. Accompanying changes may include development of mucosal congestion, fever, leukocytosis and sometimes development of ileus. These changes, if not adequately treated, lead to the dissolution of crypts and eventual loss of mucosal architecture, leading to sloughing of the epithelium from the underlying submucosa (severe rejection). Anti-lymphocyte antibody preparations are often required to treat advanced rejection.

Immunosuppression

The gastrointestinal tract harbors approximately 80% of the body's total lymphoid tissue. Due in part to this heightened immunogenicity over other solid organs, the small intestine has been a troublesome organ to transplant. Rejection is more common and more commonly severe than with other solid organ transplants. Immunosuppression after intestinal transplantation is currently based on calcineurin blockade, with other agents variably used in different protocols. Corticosteroid use generally accompanies the calcineurin inhibitor, with global immunosuppression exceeding that required for successful transplantation of the kidney or liver. Prior to the use of tacrolimus in intestinal transplantation, graft survival was poor and marred by high rates of death resulting from rejection and concomitant infections. Recently, various approaches, including the use of monoclonal interleukin-2 antagonists, graft irradiation, and concomitant use of rapamycin, all in combination with tacrolimus, appear to have decreased the rates of acute cellular rejection compared to historical controls. When acute cellular rejection occurs, augmentation of immunosuppression is required. Mild rejection often responds to bolus administration of intravenous corticosteroids,

whereas moderate or severe rejection usually requires administration of depleting anti-lymphocyte antibodies.

Infection Prophylaxis

Prior to procurement, most programs attempt to decontaminate the intestinal allograft of bacterial and fungal organisms by administering enteric antibiotics to the donor. Decontamination using mechanical cleansing is impractical in cadaver donors. Intravenous antibiotics are usually given to the recipient during transplantation, as the donor organ is not sterile. Furthermore, because the level of immunosuppression necessary for successful prophylaxis of rejection puts recipients at risk for viral infection, prophylaxis with hyperimmune globulin and antiviral agents (usually ganciclovir) is generally employed against the most common pathogens (e.g., cytomegalovirus and Epstein-Barr virus). However, newer viral pathogens, including adenovirus and calicivirus, have recently been described after intestinal transplantation.

COMPLICATIONS

Intestinal transplantation is a complex undertaking. The complications seen after this transplant can be categorized as surgical, infectious, or immunological.

Surgical Complications

Because the transplanted intestine has undergone ischemic and reperfusion injuries, some degree of mucosal injury is usually present early after transplantation. Mucosal injury can contribute to poor healing, anastomotic leakage, and loss of barrier function, leading to translocation and peritonitis and, often, to the need for early reoperation. Complex vascular reconstructions are required for this transplant, predisposing to postoperative bleeding. Transection of the mesentery and lymphatic drainage of the intestine may result in the development of chylous ascites after transplant. Chylous or lymphatic ascites may arise from the donor organ or from lymphatics of the native intestine in cases where enterectomy is required at the time of transplantation. Chylous ascites usually presents once a fat-containing enteral diet is begun and it resolves within 6 weeks with administration of a diet low in long-chain triglycerides.

Infectious Complications

Line-related sepsis can occur after transplantation, during the interval before weaning from parenteral therapy. Bacterial and fungal infections that are common before transplant in patients on parenteral therapy may recur with immunosuppression after intestinal transplantation. Peritonitis early after transplant is common, due either to contamination of the field during the transplant procedure or to translocation. Peritonitis usually does not respond to antibiotic therapy and requires reoperation with peritoneal lavage. Later, during the first year after transplantation, viral infections are common; these usually infect the host organ. Cytomegalovirus infection can cause a secretory diarrhea, as can adenovirus infection of the graft. Both viruses may also disseminate, causing a viral syndrome or, less commonly, lethal sepsis. Distinguishing these opportunistic viral infections from rejection can be difficult, as crypt inflammation and diarrhea are the clinical hallmarks of both. Epstein-Barr virus-related B-lymphocyte proliferations also occur after transplantation and have been reported in up to 20% of recipients in some series. This complication, particularly common among naive pediatric transplant recipients, frequently affects the transplanted intestine. It can be seen as expanded lymphoid nodules with activated or atypical lymphocytes on mucosal pinch biopsies of the graft. If untreated, these can progress to lethal monoclonal proliferations. Polymerase chain reaction serum evaluation aids in the early diagnosis and treatment of these viruses.

Rejection

In most clinical reports, more than 80–90% of patients experience at least one episode of acute cellular rejection during the first year after transplantation. Enterocyte apoptosis in the crypt is generally present with rejection early after transplantation (mild rejection) and can be associated with an increase in the density of the lamina propria infiltrate, activated lymphocytes, acute cryptitis, and variable degrees of villous blunting (moderate rejection). If detected early, these changes are usually reversible with bolus steroid administration. If not recognized and treated, however, they may progress to dissolution of crypts, loss of mucosal architecture, and separation of the epithelium from the underlying submucosa (severe rejection). Anti-lymphocyte antibody preparations are required to treat advanced rejection and the process is often not reversible when it has reached this stage.

The rejection process, and the resultant mucosal injury, leads to loss of intestinal epithelial barrier function and is frequently associated with translocation of bacteria and systemic sepsis or peritonitis. Sloughing of the mucosal lining may occur, with bleeding and malabsorption. This combination of events, requiring intensive augmentation of immunosuppression in the face of accompanying infection, accounts for the high

mortality rate associated with intestinal transplantation. If the recipient has received an isolated intestinal allograft, the organ may be sacrificed and removed. Such patients may later undergo successful retransplantation. Referral of patients with sufficient venous access and liver reserve to allow an interval of parenteral therapy is critical to provide this alternative. When the transplanted intestine is part of a multiorgan graft, removal of the donor organs is not feasible. Thus, transplantation of a multiorgan intestinal allograft is an irreversible step, leading either to resolution and repair of the damaged graft or to the ultimate demise of the patient. Despite considerable historical rodent data to support the protective effect of the liver against development of intestinal rejection, the ability to remove the isolated intestinal allograft in part accounts for the improved overall patient survival seen with isolated intestinal transplants in humans.

Chronic rejection, leading to arteriolitis of the graft vasculature, fibrosis of the muscular layers of the small bowel, chronic distortion of the mucosal villous architecture, and allograft dysfunction with altered motility and malabsorption, has also been reported. The clinical diagnosis remains enigmatic, but is usually associated with recurrent diarrhea, worsening nutrition, and motility disturbances.

Recently, various approaches in different centers appear to have decreased the rates of acute cellular rejection compared to historical controls. A number of new agents may have an effect on intestinal allograft survival. Humanized interleukin-2 inhibitors, which have been introduced and studied in other solid organ transplant populations, inhibit the ability of the activated T lymphocyte to up-regulate the immune response through recruitment of other such cells. Sirolimus use in combination with tacrolimus has also significantly decreased the rate of acute cellular rejection and increased the rates of graft survival. Although no randomized studies have evaluated these agents in intestinal transplantation because of the small number of patients involved, experience at several centers indicates a reduction in early rejection rates from nearly 90% historically to under 20% in some series.

Other approaches have included attempts to augment chimerism and improve graft survival rates through the infusion of unmodified donor bone marrow cells. Early reports from Pittsburgh failed to confirm decreased rejection rates among patients treated in this manner. The augmentation of chimerism through infusion of stem cells remains to be studied, however, and may prove more effective. More recently, researchers have attempted to inactivate donor antigen-presenting cells by irradiating donor intestinal allografts prior to implantation.

RESULTS

Success with intestinal transplantation has lagged behind that of other solid organ transplants, owing both to the various factors detailed above and to the advanced state of chronic illness seen in most patients referred for transplantation. The International Intestinal Transplant Registry on Recipient Outcomes last published its results in 1999 and reported a 69% 1-year survival rate for both patient and graft in isolated intestine transplants and 66% patient and 63% graft survival rates at 1 year for liver/bowel and multivisceral transplants. More recently, unpublished registry data demonstrate improved survival for all graft types and better survival at more experienced centers among patients who did not require hospitalization immediately prior to transplantation and among those receiving isolated intestine grafts. Whereas initial reports indicated that pediatric patients had higher rates of survival, more recently no significant survival advantage appears to occur in pediatric recipients.

Overall, series including transplants performed in the early 1990s report low survival rates for all recipients, with 1- and 3-year survival rates of 63 and 55% in one series of 17 patients from the UCLA Medical Center. Recent single-center series have reported higher patient and graft survival rates, including 1-year patient survival rates of up to 89% among isolated intestinal allograft recipients in one report from University of Nebraska and pediatric liver and bowel recipient survival rates of approximately 70% at 2 years in another report from Pittsburgh. Another study reported 1- and 5-year survival rates of 75 and 54% among 155 patients. Another report on 95 transplants from the University of Miami demonstrated 1-year patient and graft survival rates of 84 and 72%, respectively, among recipients of isolated intestine transplants performed since 1998, compared to 1-year liver–survival rates of only 40 and 48%, respectively, in bowel and multivisceral transplant patients. Thus, it appears that overall patient and graft survival rates are improving. Several factors have contributed to improved survival rates in intestinal transplantation. Immunosuppressive management including the use of newer agents has led to decreases in the rate of acute cellular rejection. In one report from The Mount Sinai Medical Center, this correlated with a significant improvement in graft survival rates. Additionally, earlier and more accurate diagnosis and management of viral infectious complications appear to have decreased the mortality rate associated with these diseases.

The majority of survivors (more than 70%) have full graft function with no requirement for parenteral therapy, although most recipients of intestinal

transplants require some antidiarrheal medications. Although linear growth may be attained in the majority of surviving children after successful transplantation, catch-up growth is not necessarily achieved. The majority of indications for intestinal transplantation include congenital anomalies or vascular accidents, leading to surgical resection, and therefore recurrence of the underlying disease is not common. Only one patient with Crohn's disease is known to have suffered recurrence of this disease leading to failure of the allograft.

See Also the Following Articles

Liver Transplantation • Parenteral Nutrition • Short Bowel Syndrome • Transplantation Immunology

Further Reading

Carbonnel, F., Cosnes, J., Chevret, S., Beaugerie, L., Ngo, Y., Malafosse, M., Parc, R., Le Quintrec, Y., and Gendre, J. P. (1996). The role of anatomic factors in nutritional autonomy after extensive small bowel resection. *J. Paren. Enteral. Nutr.* **20**, 275–280.

Fishbein, T. M., Gondolesi, G. E., and Kaufman, S. S. (2003). Intestinal transplantation for gut failure (Review). *Gastroenterology* **124**, 1615–1628.

Fishbein, T. M., Schiano, T., LeLeiko, N., Facciuto, M., Ben-Haim, M., Emre, S., Sheiner, P. A., Schwartz, M. E., and Miller, C. M. (2002). An integrated approach to intestinal failure: Results of a new program with TPN, bowel rehabilitation and transplantation. *J. Gastrointest. Surg.* **6**, 554–562.

Grant, D. (1996). Current results of intestinal transplantation. The International Intestinal Transplant Registry. *Lancet* **347**, 1801–1803.

Intestinal Transplant Registry. Available at http://www.intestinal-transplant.org. Accessed January 13, 2003.

Kaufman, S., Atkinson, J., Bianchi, A., Goulet, O., Grant, D., Langnas, A., *et al.* (2001). Indications for pediatric intestinal transplantation: A position paper of the American Society of Transplantation. *Pediatr. Transplant* **5**, 80–87.

Khan, F. A., and Tzakis, A. G. (1999). Intestinal and multivisceral transplantation. *In* "Transplantation" (L. C. Ginns, A. B. Cosimi, and P. J. Morris, eds.), p. 422. Blackwell Science, Malden, MA.

Messing, B., Crenn, P., Beau, P., *et al.* (1999). Long-term survival and parenteral nutrition dependence in adult patients with short bowel syndrome. *Gastroenterology* **117**, 1043–1050.

Nightingale, J. M. D. (ed.) (2001). "Intestinal Failure." Greenwich Medical Media Ltd., London, UK.

Starzl, T. E., Todo, S., Tzakis, A., *et al.* (1991). The many faces of multivisceral transplantation. *Surg. Gynecol. Obstet.* **172**, 335–344.

Todo, S., and Tzakis, A. G. (1996). Transplantation of the liver with digestive organs. *In* "Transplantation of the Liver" (R. W. Busuttil and G. B. Klintmalm, eds.), p. 503. W. B. Saunders, Philadelphia, PA.

Todo, S., Tzakis, A. G., Abu-Elmagd, K., *et al.* (1992). Intestinal transplantation in composite visceral grafts or alone. *Ann. Surg.* **216**, 223–234.

Wilmore, D. W., Lacey, J. M., Soultanakis, R. P., Bosch, R. L., and Byrne, T. A. (1997). Factors predicting a successful outcome after pharmacologic bowel compensation. *Ann. Surg.* **226**, 288–293.

Small Intestinal Motility

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

migrating motor complex Specific pattern of small intestinal motility that begins when digestion of a meal is complete and ends with intake of the next meal; also called interdigestive motility.
motilin Hormone released by enteroendocrine cells in the gut.
physiologic ileus Normal absence of gastrointestinal contractile activity.
power propulsion Specific pattern of motility that rapidly propels luminal contents over extended distances in the small and large intestine.
segmentation Specific pattern of small intestinal motility that starts with the ingestion of a meal and accomplishes mixing of the contents in the lumen; also called digestive motility.
vagotomy Surgical sectioning of the vagus nerves.

Transit time for meals moving from stomach, to small intestine, to large intestine is measured in hours. In these three compartments, transit occurs most rapidly as ingested food passes through the stomach; the slowest transit time occurs in the large intestine. Three fundamental patterns of motility that influence transit of material through the small intestine are the interdigestive pattern, the digestive pattern, and power propulsion. Each pattern is programmed by the enteric nervous system.

INTERDIGESTIVE MOTILITY PATTERN

The interdigestive pattern of small intestinal motility, called the migrating motor complex (MMC), begins after digestion and absorption of nutrients are complete, about 2–3 hours after a meal. The contractile behavior is detected by placing pressure sensors in the lumen of the intestine or by attaching electrodes to the intestinal surface. Sensors in the stomach show the MMC starting as large-amplitude contractions in the distal stomach. The MMC in the stomach migrates into the duodenum and on through the small intestine to the ileum. As it ends in the terminal ileum, the next cycle starts in the stomach.

At a single recording site in the intestine, the MMC pattern consists of three consecutive phases: (1) a silent period, Phase I, which involves no contractile activity and corresponds to physiologic ileus, (2) phase II, which consists of irregularly occurring contractions, and (3) phase III, which consists of regularly occurring contractions. Phase I returns after Phase III ends and the cycle is repeated after 80–120 minutes in humans. With multiple sensors positioned along the intestine, slow propagation of the Phase II and Phase III activity down the intestine becomes evident.

At any given time, the MMC occupies a limited length of intestine called the "activity front," which has an upper and a lower boundary. The activity front slowly advances (migrates) along the intestine at a rate that progressively slows as the front approaches the ileum. Peristaltic propulsion of luminal contents in the aboral direction occurs between the oral and aboral boundaries of the activity front. The waves of peristalsis start at the oral boundary and propagate to the aboral boundary of the activity front. Successive peristaltic waves start on average slightly further in the aboral direction and propagate on average slightly beyond the boundary where the previous wave stopped. Thus, the entire activity front slowly migrates down the intestine, sweeping the lumen clean as it goes.

Cycling of the MMC continues until it is ended by the ingestion of more food. A sufficient nutrient load terminates the MMC simultaneously at all levels of the intestine. Termination requires the physical presence of a meal in the upper digestive tract. The speed with which the MMC is terminated at all levels of the intestine suggests a neural or hormonal mechanism. Gastrin and cholecystokinin, both of which are released during a meal, terminate the MMC in the stomach and upper small intestine when injected intravenously.

The MMC is organized by the enteric nervous system. It continues in the small intestine after interruption of neural input from the central nervous system, but stops when it reaches a region of the intestine where the enteric nervous system has been damaged. Presumably, command signals to the enteric neural circuits are necessary for initiation of the interdigestive pattern; nevertheless, whether the commands are neural or hormonal or both is unknown. Although levels of the hormone motilin increase in the blood at the onset of

 Copyright 2004, Elsevier (USA). All rights reserved.

the interdigestive state, it is unclear whether motilin triggers the MMC or is released as a consequence of it.

DIGESTIVE MOTILITY PATTERN

Mixing movements in the small intestine characterize the digestive state. Feeding interrupts the interdigestive pattern and converts small intestinal motor behavior to a fed pattern characteristic of the digestive state. The fed pattern is distinguished by peristaltic contractions that propagate for only very short distances. This activity occurs continuously along the length of the intestine. Because each short segmental contraction does not propagate far, it jets the chyme in both directions. These contractions, spaced closely together as they are along the bowel, accomplish mixing of the luminal contents and, over time, net aboral propulsion of the luminal contents.

Commands transmitted from the brain by the vagus nerves sustain the digestive motility pattern in the small intestine. Transmitted vagal command signals are important in the conversion from the fasting to the fed pattern. After vagotomy, a larger quantity of ingested food is necessary for interruption of the interdigestive motor pattern, and interruption of the migrating motor complex is often incomplete. During the fed pattern, blockade of impulse transmission in the vagus nerves results in interruption of the digestive pattern.

POWER PROPULSION

Power propulsion is a defensive response to the presence of harmful agents in the intestinal lumen. Power propulsion involves strong, long-lasting contractions of the circular muscle. These "giant" migrating contractions, which propagate for extended distances along the intestine, are considerably stronger than the peristaltic contractions seen during the migrating motor complex or in the digestive motility pattern. Giant migrating contractions last 18–20 seconds. They reflect a highly efficient propulsive mechanism that rapidly strips the lumen clean as they travel over long lengths of intestine, at a rate of about 1 cm/sec. Small intestinal power propulsion differs from peristaltic propulsion during the migrating motor complex and from the mixing movements of the digestive motility pattern in that circular contractions in the propulsive segment are much stronger and they propagate over much longer reaches of intestine.

Power propulsion occurs in the oral direction during emesis and in the anal direction in response to noxious stimulation in the small intestine. The power propulsion pattern starts in the midjejunum and propagates to the stomach during emesis. When initiated by noxious stimulation, propulsion may start in the midregions of the small intestine and travel all the way to the ileocecal junction. Abdominal cramping sensations and diarrhea are associated with this motility pattern. Application of irritants to the mucosa, the introduction of luminal parasites, enterotoxins from pathogenic bacteria, allergic reactions, and exposure to ionizing radiation all trigger the power propulsion pattern of motility. This suggests that power propulsion is a defensive adaptation for rapid clearance of undesirable contents from the intestinal lumen. Propulsion clears the upper bowel as part of the emetic response and clears the lower small intestine by rapid movement of the material into the colon.

See Also the Following Articles

Emesis • Enteric Nervous System • Migrating Motor Complex • Motilin • Physiologic Ileus • Postprandial Motility • Power Propulsion (590) • Small Intestine, Absorption and Secretion

Further Reading

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Wood, J. (2003). Neurogastroenterology and digestive motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., pp. 449–480, Chap. 26. Lippincott Williams & Wilkins, Baltimore.

Wood, J., and Castel, D. (2000). "Neurogastroenterology and Motility: Upper Gastrointestinal Tract." Milner-Fenwick, Timonium, Maryland.

Small Intestine, Absorption and Secretion

MARSHALL H. MONTROSE
Indiana University, Indianapolis

apical membrane Portion of the intestinal epithelial cell plasma membrane facing the intestinal lumen (the exterior of the body).
basolateral membrane Portion of the intestinal epithelial cell plasma membrane facing the blood (interior of the body).
chyme Slurry of partially digested foods and liquids that is transiently present in the intestines following a meal.
diarrhea Frequent passage of watery stool.
electrolytes Charged ions, such as sodium and chloride, that are part of the ionic composition of body fluids.
epithelial polarity The property of epithelial cell asymmetry. Epithelial cells sit at the interface between two compartments and the cells have different surface and cellular features along an imaginary axis drawn between the two compartments.
osmolyte A molecule that can be dissolved in a solvent (most commonly water) and thereby contribute to the total concentration of particles in the solvent, adding to the osmolarity.
paracellular transport Vectorial transport of material through the extracellular spaces between adjacent epithelial cells. Results in net movement of material between the two compartments faced by epithelial cells.
transcellular transport Vectorial transport of material through the cytosol of epithelial cells. Results in net movement of material between the two compartments faced by epithelial cells.
transepithelial transport Vectorial transport of material across the epithelial layer, using either paracellular or transcellular routes.

The major function of the small intestine is to efficiently absorb a wide variety of nutrients and electrolytes from ingested foods. As a consequence of these absorptive processes, water is also absorbed and the lumenal contents become drier. To keep the intestinal chyme well hydrated (to promote easier mixing, intestinal transit, and solubility of remaining nutrients), the small intestine also secretes fluid into the lumen. For this reason, both absorptive and secretory processes occur simultaneously. During normal intestinal function, fluid absorption exceeds secretion so that virtually all of the lumenal fluid will ultimately be absorbed by the concerted action of both the small and large intestines. This is important to supply water as well as nutrients and electrolytes to the body. However, in diarrheal diseases, secretory activity exceeds absorption. In this case, water is lost to the stool and life-threatening dehydration can result. The intestinal epithelium mediates all vectorial transport of substances between the body and the intestinal lumen. This single layer of epithelial cells sits at the interface between these two compartments and simultaneously acts as a selective barrier to undesirable lumenal contents and as a portal to needed lumenal contents. Each individual epithelial cell has a polarized architecture, with distinct surface features present in the membrane facing the gut lumen (apical membrane) versus the membrane facing the blood (basolateral membrane). It is the specialized properties of the intestinal epithelial cell membranes that underlie the ability to perform the net transport of substances between the body and the gut lumen. The body dynamically regulates the function of proteins that mediate these properties and thereby regulates the processes. It should be recognized that epithelial absorptive functions vary along the length of the small intestine from the duodenum through the ileum, reflecting the need to absorb specific constituents from the lumen at specific locations.

INTRODUCTION

Ingested food must provide the body with necessary salts, nutrients, fluid, and osmolytes to sustain life. Assimilating these molecules must occur with a high level of specificity to avoid absorption of toxins, bacteria, and other undesired substances from the gut lumen. It must also occur in response to a meal that can require activation of these absorptive processes at any time of day or night and in response to a staggeringly diverse set of foodstuffs.

To cope with the massive amount of absorption that is required to sustain the body, the intestine has evolved multiple strategies to increase the surface area for absorption. The small intestine is not a smooth cylindrical tube, but rather has macroscopic folds that provide both enhanced surface area and extra space for distension following entry of chyme. On a smaller scale, fingerlike projections of tissue extend into the lumen several hundred micrometers, with the epithelial cells that mediate

 Copyright 2004, Elsevier (USA). All rights reserved.

absorption lining these intestinal villi. At a microscopic level, the apical membrane of each epithelial cell has abundant specialized membrane structures, termed microvilli, that are 2 to 5 μm long finger-like projections of the cell membrane into the lumen. Combined, these strategies increase the absorptive surface area approximately 600-fold compared to a smooth cylindrical tube of the same macroscopic dimensions.

The small intestine is responsible for absorption of the majority of salts, amino acids, carbohydrates, vitamins, and fats. In most cases, these absorptive processes occur due to the presence of membrane proteins that reside in the plasma membranes of intestinal epithelial cells and specifically bind and transport desired nutrients into the body. There are many such membrane transport proteins residing in the membrane and these products of multiple genes are each selective for one or a small group of substrates whose transport they facilitate. For example, the SGLT1 protein resides in the apical membrane and is responsible for the majority of glucose and galactose absorption. The GLUT5 protein resides in the same membrane, but is responsible for fructose absorption. Dividing the task of absorption among numerous proteins provides a mechanism for regulating the uptake of individual constituents of a meal (through regulation of the relevant transporters) and ensures that a high level of specificity for desirable versus undesirable constituents is sustained.

The small intestine absorbs both water-soluble and water-insoluble compounds. If absorbed molecules are dissolved in the aqueous lumenal fluid, then they act as osmolytes. Therefore, their absorption has the consequence of simultaneously stimulating absorption of water molecules, because of local differences in osmotic pressure between compartments (e.g., between the lumen and the epithelial cell cytosol). It is believed that water transport may be enhanced by the action of aquaporin water channels, specifically AQP8 and AQP4 proteins isoforms. In this manner, the absorption of solutes promotes the absorption of water and thereby tends to dehydrate the lumenal chyme. In contrast, the absorption of water-insoluble fats occurs because these compounds are emulsified by bile salts, compounds that act as solubilizers of fats because of their detergent properties.

Maintaining hydration of the lumenal chyme is essential to promote efficient absorption, a function that requires a well-mixed aqueous environment. To this end, the small intestine also secretes fluid. The secretory process present in all portions of the small intestine is chloride secretion, which promotes an attendant water secretion into the lumen. In addition to hydration of the lumen, the secretion of bicarbonate contributes to the maintenance of optimal pH values in the small intestinal lumen and the secretion of bioactive peptides promotes defense from bacterial pathogens. Coordinated contractions of the intestinal wall are needed to both mix the lumenal contents and propel them down the intestinal tract.

THE INTESTINAL EPITHELIUM

The intestinal epithelium is a single layer of cells that form a lining to separate the intestinal lumen from the body. The lining acts in part as a general barrier to keep the body separate from the outside world represented by the lumen, with specialized junctional structures to fuse neighboring epithelial cells to each other and seal the epithelial layer. Epithelial transport acts to complement this function by providing routes for specific substances to enter (or leave) the body via their transepithelial transport between the lumen and the blood across the epithelial layer. The relevant transport routes are either through the epithelial cells (transcellular transport) or between the epithelial cells for those substances that have a limited permeability through the junctions (paracellular transport). In either case, absorbed substances are delivered directly to the basal side of the epithelial layer where diffusion of water-soluble substances into the portal blood, or diffusion of lipid-soluble materials into the lymphatic lacteals, allows the rapid transit of absorbed materials from the intestinal tissue to other parts of the body.

Transcellular transport requires the sequential flux of a substance across the apical and basolateral membranes. At each step, the involvement of an integral membrane protein is required. There are four classes of these membrane transporters. ATPases drive a substance across the membrane by transducing the energy from the cleavage of ATP. These are called primary active transporters because they are responsible for creating ion gradients that can act as a source of energy for other transporters. Ion-gradient-dependent transporters transduce this ion gradient energy into the transmembrane movement of other ions or solutes and are termed secondary active transporters. Facilitated diffusion transporters bind solutes or ions that are not carried by the ATPases and facilitate their equilibration across the membrane. Channels also dissipate gradients across the membrane, but act more as gated ion pores in the membrane. Transport across the apical membrane is the rate-limiting step for transcellular transport of most substances and is therefore a key site for the regulation of transport rates.

Tight junctions provide the rate-limiting step for paracellular transport. After passing this restrictive

site, fluid and solutes progress through the torturous extracellular spaces between adjacent epithelial cells.

The intestinal epithelium contains a diverse set of cells, some of which mediate absorptive and secretory processes. It is generally agreed that cells on the villus are absorptive and cells in the crypts are secretory. It remains unclear how and where the transition in function occurs along this "vertical axis" and whether crypt cells might also absorb some substances.

ABSORPTION OF WATER-SOLUBLE MOLECULES

The small intestine has a large capacity to absorb water-soluble nutrients. In a normal Western diet, 40–70 g of protein is ingested per day, and even in the presence of an additional 50 g of secreted protein (mostly from the pancreas), 100% of this load is absorbed from the lumen of the small intestine. Similarly, 400 g of ingested carbohydrate is fully absorbed per day. In the latter case, it is known that the maximum capacity for carbohydrate absorption is approximately 1 kg/day. The ability of the intestine to absorb with such high efficiency is dependent on the mechanisms that have evolved to efficiently scavenge lumenal nutrients.

The small intestine uses a common strategy for absorption of many water-soluble molecules. Many lumenal nutrients are presented as polymers that must be digested by lumenal enzymes (either attached to the apical surface of epithelial cells or secreted into the gut lumen) prior to absorption of their simpler constituents. This strategy reduces the number of different molecules that must be recognized by transport proteins as absorbable nutrients. Transcellular transport is then initiated by apical uptake of the simpler constituents via either an ion-gradient-dependent or a facilitated diffusion transporter. Vitamins are absorbed in the original ingested form, since modifying these essential factors prior to absorption would not help the body. The absorbed nutrients within enterocytes then complete their transcellular route by leaving the cell through facilitated diffusion across the basolateral membrane.

Many molecules are absorbed by ion-gradient-dependent transporters in the apical membrane. Most of these transporters use the sodium electrochemical gradient (established by the basolateral Na^+,K^+-ATPase) as the energy source to drive apical cotransport of sodium and a nutrient (sugars, amino acids, vitamins, bile salts) into the enterocyte. The advantage of such cotransporters is that they can cause a high level of intracellular accumulation of the nutrient, leading to efficient absorption even with low lumenal levels of the nutrient. Examples of apical transporters for major dietary nutrients are shown in Table I, followed by a listing of facilitated diffusion carriers that are also present in the apical membrane.

TABLE I Apical Membrane Absorptive Processes and Proteins

Driving ion	Solute class	Substrate specificity	Gene: Common symbol (NCBI symbol)[a]
Sodium	Sugars	Glucose, galactose	SGLT1 (SLC5A1)
Sodium	Amino acids	Neutral	ASCT1 (SLC1A4)
		Anionic	EAAT3 (SLC1A1)
		Imino	Uncloned
Sodium	Water-soluble vitamins	Ascorbate	SVCT1 (SLC23A2)
		Biotin, pantothenate	SMVT (SLC5A6)
Sodium	Bile salts	Bile acids, bile salts	IBAT (SLC10A2)
Proton	Peptides	Dipeptides, tripeptides	PEPT1 (SLC15A1)
—	Sugars	Fructose	GLUT5 (SLC2A5)
—	Amino acids	Cationic	CAT1 (SLC7A2)
			4F2HC (SLC7A7)
		Neutral	RBAT+ATB0 (SLC3A1+SLC7A9?)
—	Water-soluble vitamins	Folate	FOLT1 (SLC19A1)
		Thiamine	THTR1 (SLC19A2)
		Nicotinamide (niacin)	Uncloned
		Pyridoxamine (B_6)	Uncloned
	Salts	Na^+/H^+ exchange	NHE3 (SLC9A3)
			NHE2 (SLC9A2)
		Cl^-/HCO_3^- exchange	DRA (SLC26A3)

[a] NCBI, National Center for Biotechnology Information.

Table I also lists the relevant carriers that absorb the majority of NaCl in the absence of other lumenal nutrients. In this mechanism, apical Na^+/H^+ exchange mediates sodium uptake into the enterocyte. The concomitant alkalinization of the cell causes activation of Cl^-/HCO_3^- exchange in the same membrane, which mediates chloride uptake. Basolateral Na^+,K^+-ATPase promotes Na^+ efflux from the cell, completing the transcellular transit of Na^+. A basolateral K^+/Cl^- cotransporter (KCC1, SLC12A4) mediates K^+ and Cl^- efflux and completes the balance sheet of net NaCl absorption.

Many of the apical membrane proteins are expressed in a few other tissues in addition to the intestinal epithelium. In contrast, basolateral membrane transporters are often broadly distributed in the body. In addition to the function of allowing molecules to complete their transcellular transit, such proteins are also utilized frequently (by many cell types) for normal homeostatic processes. For instance, the ubiquitous Na^+,K^+-ATPase is responsible for cellular Na^+ and K^+ homeostasis. It is expressed in the basolateral membrane of intestinal epithelial cells, where it is also responsible for the final step of transcellular sodium absorption.

The major regulator of nutrient absorption is lumenal availability. Although second messengers may modestly increase or decrease the transport rate via absorptive transporters, it is rare to find a normal physiologic state in which absorption of proteins and carbohydrates has been compromised to the extent that the nutrients are not fully absorbed by the end of the small intestine. This does not indicate that absorption succeeds under all conditions. There are a number of disease states in which absorptive processes are compromised directly or net absorption is blunted by the stimulation of secretory processes.

ABSORPTION OF FAT-SOLUBLE MOLECULES

The major dietary lipids are triacylglycerols (90–95%) with the remainder being free and esterified cholesterol, sterols, and phospholipids. Small amounts (micrograms to milligrams per day) of fat-soluble vitamins (vitamins A, D, E, and K) are also in the diet. Unlike salts and solutes, fats are not ingested in a convenient form for digestion or absorption. Most lipids are water-insoluble, but absorption must occur via an aqueous environment in the lumen. The solution is to transform fats into water-soluble structures through emulsification with amphipathic molecules (dietary proteins, fatty acids, and bile acids secreted into the gut lumen by the liver and gallbladder), followed by digestion via lipases secreted from the pancreas (major source), tongue, and stomach. The finely emulsified fat droplets then spontaneously break down into 4 to 7 nm diameter structures, micelles, with bile salts positioned at the surface and hydrophobic cores that act as a sink for the water-insoluble products of lipolysis (e.g., fatty acids, monoglycerides, and cholesterol) as well as fat-soluble vitamins.

At the mucosal surface, lipids diffuse out of the micelles, so that a saturated aqueous solution of the lipids is maintained in contact with the brush border of the mucosal cells. These lipophilic substances enter the intestinal epithelial cell by diffusion across the plasma membrane. Absorption of lipolytic products is greatest in the upper (proximal) parts of the small intestine, but appreciable amounts are also absorbed in the ileum. Once lipids have diffused from the micelles, bile acids are absorbed in the distal ileum and recycled for another round of fat absorption.

The fate of the fatty acids in enterocytes depends on their size. Fatty acids with less than 10–12 carbon atoms (and some fat-soluble vitamins) transit through the epithelium directly into the portal blood unmodified. Longer-chain FFA and monoglycerides are taken up by the endoplasmic reticulum and metabolically rebuilt into triglycerides. These intracellular lipids are then packaged in 0.5 to 1.0 μm diameter particles termed chylomicrons. The hydrophobic core of chylomicrons contains triglycerides, cholesterol esters, and fat-soluble vitamins. Chylomicrons are carried in secretory vesicles to the basolateral membrane where they are expelled via exocytosis into intercellular spaces at the basal pole of the intestinal epithelial cell. These large particles cannot enter capillaries and instead enter lacteals and are carried in lymphatics to the superior vena cava via the thoracic duct. They are then transported via systemic blood to sites of storage or utilization.

SECRETION

Transcellular secretion of chloride is the main process leading to fluid secretion in the normal small intestine. Chloride enters enterocytes at the basolateral membrane via the electroneutral $Na^+/K^+/2Cl^-$ cotransporter (gene NKCC1, SLC12A2) and then Cl^- exits the enterocyte via chloride channels in the apical membrane. The product of the cystic fibrosis gene, CFTR (ABCC7), encodes one anion channel in the apical membrane, but other channels may also be important. Transcellular chloride secretion moves a net charge across the epithelial layer, so to maintain cellular electroneutrality, potassium channels in the basolateral membrane allow for charge compensation through K^+ efflux. Similarly, it

is believed that Na^+ follows chloride passively by paracellular flux through tight junctions to maintain electroneutrality between lumenal and serosal compartments. Water is secreted in response to the net movement of osmolytes (NaCl) into the lumen.

Chloride secretion can be stimulated by increases in cellular cyclic AMP, cyclic GMP, or calcium. A number of membrane receptors for hormones are expressed in the basolateral membrane of enterocytes, which lead to physiologic changes in these second messengers to meet with the demands of absorbing different nutrients. The target of regulation by cyclic nucleotides appears to be the apical CFTR channel. Increased intracellular calcium predominantly increases the activity of the basolateral K^+ channel, which stimulates secretion through keeping the cell hyperpolarized (promoting greater efflux of the Cl^- anion) and recycling K^+ at the basolateral membrane.

Bicarbonate is secreted by the duodenal epithelium to help neutralize the acidic chyme entering the small intestine from the stomach. Bicarbonate secretion also takes place in the ileum and colon, but its physiologic role in these sites is less clear since pancreatic bicarbonate secretion provides the remaining alkali needed to neutralize lumenal acid. In the duodenum, bicarbonate is secreted by transcellular mechanisms. Carbonic anhydrase acts on intracellular CO_2 (coming from either basolateral uptake or cellular metabolism) to produce intracellular HCO_3^- and a proton. The proton leaves the cell via basolateral Na^+/H^+ exchange and bicarbonate leaves the duodenal cell either by Cl^-/HCO_3^- exchange or via CFTR channels. This vectorial flux of alkali produces an alkaline shift in pH at the lumen or at a minimum serves to neutralize lumenal acid when HCO_3^- and H^+ combine to form CO_2 and water. Similar to regulation of chloride secretion, cyclic nucleotides stimulate bicarbonate secretion through enhancing CFTR activity.

DISEASES OF ABSORPTION AND SECRETION

Osmotic Diarrhea

Water flows in response to transepithelial movement of osmolytes, in direct response to osmotic gradients. If foodstuffs that cannot be absorbed by the intestinal epithelium are ingested, then they act as an osmotic sponge to pull water into the lumen. This is the mechanism of action of laxatives and is the mechanism underlying many diseases of absorption and secretion (see below). It should be noted that such osmotic diarrhea can also be produced in response to ingesting massive amounts of material that exceed the absorptive capacity of the intestine.

Infectious Diarrhea

Diarrhea caused by bacteria and other pathogens is responsible for 5 million deaths per year worldwide. Many microorganisms produce enterotoxins that interfere with the second-messenger pathways in enterocytes. Cholera toxin (from *Vibrio cholera*) binds to GM1 gangliosides on the apical surface of intestinal epithelial cells and simulates a sustained increase in intracellular cyclic AMP. This directly inhibits electroneutral NaCl absorption and stimulates chloride secretion. Traveler's diarrhea is caused by StA toxin (*Escherichia coli*), which binds to guanylin receptors on intestinal epithelial cells and increases cyclic GMP, resulting in the stimulation of chloride secretion.

Disaccharide Intolerance

There is a common genetic variability in levels of lactase, an enzyme in the apical membrane of enterocytes that is responsible for the digestion of carbohydrates. There can also be a genetic sucrase–isomaltase deficiency. Lack of these disaccharidase enzymes will cause osmotic diarrhea when patients ingest carbohydrates. This is because only monosaccharides can be absorbed by enterocytes. The undigested disaccharides thus remain in the lumen.

Celiac Disease

Celiac disease is a genetic disorder causing a sensitivity to gluten: a water-insoluble protein found in certain cereal grains, notably wheat. In response to gluten, the small intestine undergoes loss of villi, damage to remaining epithelial cells, and crypt hyperplasia. Due to the dramatic loss of absorptive cells on the villi and an abundance of secretory cells in the crypt, the disease results in diarrhea and nonspecific nutrient malabsorption. The underlying cause of the reaction to gluten is unknown.

Genetic Defects in Cl^- Secretion

Cystic fibrosis is an autosomal recessive disease causing defective intestinal transport through mutation of the CFTR gene and protein. Human mutations in CFTR cause defective Cl^- secretion either by diminishing the amount of CFTR protein in the membrane or by decreasing Cl^- channel opening. Cystic fibrosis causes intestinal obstruction and meconium ileus in newborns, due to poorly hydrated lumenal contents.

Genetic Defects in Cl^- Absorption

Congenital chloride diarrhea (CLD) is an extremely rare, recessive, autosomal genetic disorder that produces Cl^--rich, acidic diarrhea commencing at birth. The transport defect is limited to the ileum and colon. The gene that causes congenital chloride diarrhea (DRA, SLC26A3) encodes a transmembrane protein that can perform Cl^-/HCO_3^- exchange when the protein is normal, but cannot transport Cl^- when DRA contains mutations found in patients suffering from CLD.

Genetic Defects in Na^+-Dependent Sugar Absorption

Glucose–galactose malabsorption is a rare, autosomal recessive disorder in which Na^+-coupled uptake of glucose and galactose is defective. Sugar ingestion leads to osmotic diarrhea, which can be treated by eliminating glucose and galactose from the diet. Mutations in the SGLT1 protein, mediating Na^+-dependent sugar uptake at the apical membrane, have been shown to be the basis for at least a portion of glucose–galactose malabsorption cases.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Carbohydrate Digestion and Absorption • Celiac Disease • Diarrhea • Diarrhea, Infectious • Epithelial Barrier Function • Epithelium, Proliferation of • Epithelium, Repair of • Fat Digestion and Absorption • Malabsorption • Small Intestinal Motility • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Barrett, K. E., and Donowitz, M. (eds.) (2002). "Gastrointestinal Transport: Molecular Physiology, Current Topics in Membranes," Vol. 50, Chapters 2, 6–8, 10, 12, and 13. Academic Press, New York.

Johnson, L. R. (ed.). (1994). "Physiology of the Gastrointestinal Tract," 3rd Ed., pp. 1271–1453, Chapters 50–62. Raven Press, New York.

Yamada, T. (ed.). (1999). "Textbook of Gastroenterology," 3rd Ed., Chapter 15. Lippincott, Williams & Wilkins, New York.

Small Intestine, Anatomy

GAËLLE BOUDRY, PING-CHANG YANG, AND MARY H. PERDUE
McMaster University, Hamilton, Canada

enteric nervous system Neural elements distributed within the wall of the digestive tract and organized in two plexuses: myenteric and submucosal. This system can function independently of the central nervous system, to coordinate gastrointestinal function, or can interact with the extrinsic nervous system, acting as the first level of integration.

enterocytes Major cell type in the intestinal epithelium. These tall and columnar cells are highly differentiated to fulfill absorptive and secretory functions and also form a barrier against penetration of bacteria and dietary antigens into the mucosa.

mucosa Layer of the small intestine in contact with the luminal contents. It is formed by upward "villus" projections into the lumen surrounded by downward "crypt" invaginations.

muscularis externa Smooth muscle layer surrounding the mucosa and submucosa. It is composed of an inner circular and outer longitudinal layer, which function in coordination to propel and mix luminal contents.

Peyer's patches Organized lymphoid aggregates disseminated along the small intestine in the mucosa and submucosa; contain functional T and B lymphocytes as well as specialized epithelial microfold cells, which sample luminal antigens and bacteria.

The small intestine is a convoluted tube ranging from 3 to 9 m in length, depending on the tone and the degree of stretch induced during measurement. It extends from the pylorus to the large intestine at the ileocecal junction. It is classically divided into three segments, the duodenum, jejunum, and ileum. Although regional differences exist

Copyright 2004, Elsevier (USA). All rights reserved.

Genetic Defects in Cl^- Absorption

Congenital chloride diarrhea (CLD) is an extremely rare, recessive, autosomal genetic disorder that produces Cl^--rich, acidic diarrhea commencing at birth. The transport defect is limited to the ileum and colon. The gene that causes congenital chloride diarrhea (DRA, SLC26A3) encodes a transmembrane protein that can perform Cl^-/HCO_3^- exchange when the protein is normal, but cannot transport Cl^- when DRA contains mutations found in patients suffering from CLD.

Genetic Defects in Na^+-Dependent Sugar Absorption

Glucose–galactose malabsorption is a rare, autosomal recessive disorder in which Na^+-coupled uptake of glucose and galactose is defective. Sugar ingestion leads to osmotic diarrhea, which can be treated by eliminating glucose and galactose from the diet. Mutations in the SGLT1 protein, mediating Na^+-dependent sugar uptake at the apical membrane, have been shown to be the basis for at least a portion of glucose–galactose malabsorption cases.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Carbohydrate Digestion and Absorption • Celiac Disease • Diarrhea • Diarrhea, Infectious • Epithelial Barrier Function • Epithelium, Proliferation of • Epithelium, Repair of • Fat Digestion and Absorption • Malabsorption • Small Intestinal Motility • Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

Further Reading

Barrett, K. E., and Donowitz, M. (eds.) (2002). "Gastrointestinal Transport: Molecular Physiology, Current Topics in Membranes," Vol. 50, Chapters 2, 6–8, 10, 12, and 13. Academic Press, New York.

Johnson, L. R. (ed.). (1994). "Physiology of the Gastrointestinal Tract," 3rd Ed., pp. 1271–1453, Chapters 50–62. Raven Press, New York.

Yamada, T. (ed.). (1999). "Textbook of Gastroenterology," 3rd Ed., Chapter 15. Lippincott, Williams & Wilkins, New York.

Small Intestine, Anatomy

GAËLLE BOUDRY, PING-CHANG YANG, AND MARY H. PERDUE
McMaster University, Hamilton, Canada

enteric nervous system Neural elements distributed within the wall of the digestive tract and organized in two plexuses: myenteric and submucosal. This system can function independently of the central nervous system, to coordinate gastrointestinal function, or can interact with the extrinsic nervous system, acting as the first level of integration.

enterocytes Major cell type in the intestinal epithelium. These tall and columnar cells are highly differentiated to fulfill absorptive and secretory functions and also form a barrier against penetration of bacteria and dietary antigens into the mucosa.

mucosa Layer of the small intestine in contact with the luminal contents. It is formed by upward "villus" projections into the lumen surrounded by downward "crypt" invaginations.

muscularis externa Smooth muscle layer surrounding the mucosa and submucosa. It is composed of an inner circular and outer longitudinal layer, which function in coordination to propel and mix luminal contents.

Peyer's patches Organized lymphoid aggregates disseminated along the small intestine in the mucosa and submucosa; contain functional T and B lymphocytes as well as specialized epithelial microfold cells, which sample luminal antigens and bacteria.

The small intestine is a convoluted tube ranging from 3 to 9 m in length, depending on the tone and the degree of stretch induced during measurement. It extends from the pylorus to the large intestine at the ileocecal junction. It is classically divided into three segments, the duodenum, jejunum, and ileum. Although regional differences exist

Copyright 2004, Elsevier (USA). All rights reserved.

within the small intestine, the general structure is similar throughout its length.

GROSS ANATOMY

The following discussions focus on the anatomy and histology of two segments of the small intestine, the jejunum and the ileum.

External and Internal Appearance

The jejunum starts right after the duodenojejunal flexure, a configuration that is supported by a fibromuscular band, the ligament of Treitz, which is attached to the diaphragm. The jejunum and ileum lie in coils in the abdominal and pelvic cavities, suspended in a fan-shaped manner by the mesentery, from the posterior abdominal wall. The distance from the posterior abdominal wall to the intestinal wall varies from about 15 to 20 cm, increasing distally. This architecture allows each coil to accommodate to changes in form and position. The jejunum and ileum are not separated by any distinct anatomic landmark, but several features become gradually more apparent from proximal to distal. Classically, the proximal 40% of the mobile small intestine is designated as the jejunum and the distal 60% as the ileum.

The internal surface of the jejunum and proximal ileum shows numerous transverse folds (the plicae circulae, or valves of Kerchkring). These mucosal and submucosal folds are permanent structures, varying from 3 to 10 mm in height and running transversely around the small intestine. Their number gradually decreases through the distal jejunum and they may be entirely absent from the distal ileum. Small lymphoid nodules, the Peyer's patches, of varying size are also apparent at regular intervals. They become more numerous distally and form a large, continuous antimesenteric Peyer's patch in the distal ileum.

Blood and Lymph Supply

From the ligament of Treitz to the ileocecal junction, the small intestine is supplied by a dozen or more branches arising from the left branch of the superior mesenteric artery. These branches run between the two folds of the peritoneum composing the mesentery. As they approach the intestinal wall, they branch and anastomose with one another to form a series of arcades or arches. The straight arteries arising from the final arches send branches to either side of the intestine and encircle it. In turn, these arterial circles emit branches of diminishing size, which penetrate the intestinal wall. Veins accompanying the arterial vessels supplying the jejunum and ileum drain into the superior mesenteric vein, which merges with the splenic vein to form the portal vein. Lymph nodes are arranged in three groups: one near the intestinal wall, one at the level of the arterial arcades, and the third along the upper part of the superior mesenteric artery. These nodes and those in the intestinal wall interconnect by many lymphatic vessels.

Innervation

The parasympathetic nerve supply of the small intestine comes from the dorsal nucleus of the vagus. The fibers from these cell bodies enter the abdominal cavity as the anterior and posterior vagal trunks with the esophagus and pass into the wall of the gut with its blood vessels via the celiac and superior mesenteric ganglia, in which these fibers do not synapse. The fibers end in the intestinal wall, where they synapse with cell bodies in the submucosal and myenteric plexuses. The sympathetic preganglionic fibers have their cell bodies in the spinal cord segment T9 and T10 and enter the sympathetic trunk by white rami communicantes. They leave as the splanchnic nerves and pass the celiac and superior mesenteric ganglia, where they synapse. Postganglionic fibers enter the small intestine with its blood vessels.

SEROSA

The jejunum and ileum are almost completely enveloped by a thin extension of the peritoneum, a monolayer of flattened mesothelial cells and loose connective tissue (Fig. 1). The only exception is the mesenteric border, where the peritoneal folds are separated to allow the entry and exit of blood vessels, lymphatics, and nerves. These latter ramify in a subserosal zone of connective tissue lying between the mesothelial layer and the muscularis externa.

MUSCULARIS EXTERNA

The muscularis externa (or propria) consists of an outer longitudinal and an inner circular layer of smooth muscle separated by the myenteric plexus (the plexus of Auerbach) (Fig. 1). Smooth muscle cells are densely packed in characteristics bundles and packets, separated by small spaces consisting mainly of collagen fibrils and larger spaces occupied by nerves, capillaries, interstitial cells of Cajal, and intramuscular septa of connective tissue. Intestinal smooth muscle cells are uninucleated and spindle-shaped cells measuring

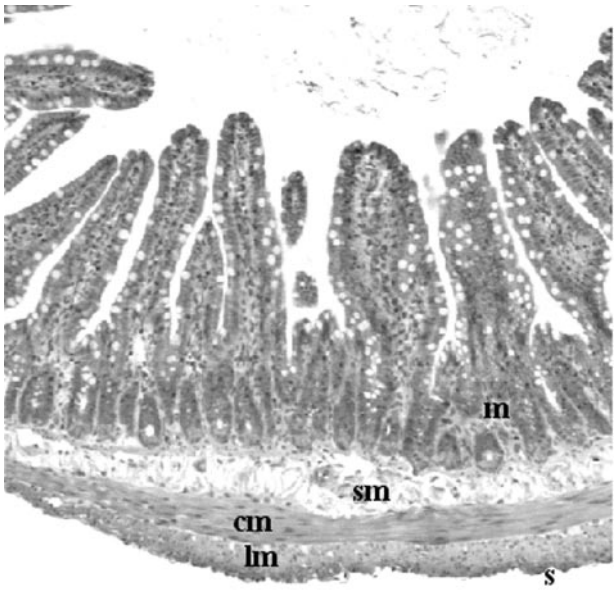


FIGURE 1 Transverse section of rat jejunum (× 200). The wall of the small intestine can be divided in four different layers: the mucosa (m), facing the lumen; the submucosa (sm); the two layers of circular (cm) and longitudinal (lm) muscles forming the muscularis externa; and the serosa (s).

500–700 μm in length at rest, but are able to shorten to less than a quarter of their resting length when contracting. Important features of smooth muscle cells are cell-to-cell junctions (intermediate and gap junctions) involving specializations of the membrane between two adjacent cells. These junctions are essential to produce mechanical and electrical cooperation and communication between cells, which form a muscular syncitium. The intestinal muscular layers and the circular layer, in particular, are richly innervated. There are no classical neuromuscular junctions, rather intramuscular axons tend to run parallel to the smooth muscle cells, sending varicosities from which neurotransmitters are released. Commonly found associated with smooth muscle cells and nerve varicosities are the interstitial cells of Cajal, which communicate with both types of cells. Other cells possibly located within smooth muscle layers are macrophages, mast cells, and myoblast-like cells.

The myenteric plexus, as well as the submucosal plexus located in the submucosa (discussed later), are the two major integrative systems of the enteric nervous system. Plexuses are networks of wide and thin ganglia interconnecting through neuronal processes. Ganglia are composed of two cell types, nerve cells and glial cells, associated with a neuropil formed by closely packed neuronal and glial processes. Each ganglion and the connecting strands are surrounded by a basal lamina and collagen fibrils. Small blood vessels can lie near but do not enter ganglia of the myenteric plexus. Neural connections exist between the myenteric and the submucosal plexuses as well as with extrinsic neurones.

SUBMUCOSA

The submucosa extends from the muscularis externa to the mucosa (Fig. 1). It consists of connective tissue, blood and lymphatic vessels, and the submucosal plexus (plexus of Meissner). The submucosa may also contain scattered migratory cells, of which mast cells are frequently the predominating cell type. Relatively large arterioles, venules, and lymphatic vessels form extensive networks within the submucosa. From these networks, numerous penetrating capillary vessels supply and drain most of the mucosa and muscularis externa. These dense vascular networks enable the submucosa to play a role in vascular routing and related distribution of regional blood and lymphatic flow. The composition of the submucosal plexus is similar to that of the myenteric plexus. However, in humans, and contrary to the composition in rodents, this plexus is composed of two distinct layers, the inner and outer submucosal plexuses. The submucosal plexus has smaller ganglia, less neurons, and a less compact network compared to the myenteric plexus.

MUCOSA

Architecture

The mucosa is composed of three components: the muscularis mucosae, which delimit the mucosa from the submucosa, the lamina propria, and the eptithelial cells, which face the intestinal lumen. A transverse section from any part of the small intestine reveals the unique architecture of the small intestinal mucosa, involving the crypts and villi that cover the entire length of the small intestine. The straight and tubelike crypts (crypts of Lieberkühn) are downward projections of the epithelium into the lamina propria whereas the villi are upward leaflike or fingerlike projections of the surface epithelium and the lamina propria. Villi and crypts are contiguous, in that the epithelium composing these two structures is continuous. However, the number of villi is considerably less than the number of crypts because several crypts may open into a single mouth and more than one mouth can open between two adjacent villi. Fingerlike villi predominate in the jejunum and ileum whereas leaflike villi are more common in the duodenum. The villi are tallest (about 0.8–1 mm) in

the distal part of the duodenum and the proximal jejunum but become progressively shorter. Moreover, jejunal and ileal villi differ also by the series of indentations along the side of the villi, which are marked in the proximal small intestine but disappear in the terminal ileum.

Vascular Pattern of the Intestinal Villi

The vascular pattern of the intestinal villi is essentially the same from the duodenum to the ileum. Each villus is supplied by a central artery, which runs through the center of the villus toward the tip. It then divides into two narrow arterial trunks, which run directly below the epithelial cells downward toward the base of the villus. These arteries give rise to capillary branches that form a dense anastomosing network extending between the arterial trunks and the two veins, which drain each villus. Each intestinal villus contains also a central blind-ended lymph vessel, the lacteal.

Mucosal Components

Muscularis Mucosa

The muscularis mucosa is the outermost layer of the mucosa. It is composed of elastic fibers and 3–10 smooth muscle cells, generally arranged in an outer longitudinal and inner circular layer. Smooth muscle cells may radiate from the muscularis mucosa into the lamina propria and extend in the villi.

Lamina Propria

The lamina propria forms the connective tissue core of the villi and surrounds the crypt epithelium. The crypt and villus epithelial cells and the lamina propria are separated by a distinct basement membrane composed of an ultrastructurally apparent basal lamina and a deeper network of collagenous fibers. The lamina propria is composed of noncellular connective tissue elements, i.e., collagen and elastin, blood and lymphatic vessels, and myofibroblasts supporting villi. However, the main characteristic of the lamina propria is to contain numerous immunologically competent cells as well as nerve endings. The most numerous cell types are mononuclear cells, plasma cells, and lymphocytes. Plasma cells contain immunoglobulin (IgA or IgM) and are concentrated more in the intercrypt region. Lymphocytes, both B and T cells, are found throughout the lamina propria but often form more dense infiltrates just above the muscularis mucosae. Other cell types are more sparsely distributed, including eosinophils, macrophages, and mast cells. However, macrophages are mostly located along the superior part of the lamina propria near the tip of the villi. Numerous nerves endings are present in the lamina propria, many of which are in close contact with mast cells.

Epithelium

The four major cell types of the epithelium (enterocytes, goblet, endocrine, and Paneth cells) arise from stem cells, which divide continuously, located just above the crypt base. Differentiation and maturation occur in 5–6 days as the cells migrate from the crypt to the villus tips, where they are subsequently sloughed into the lumen (Fig. 2).

Crypt epithelium The most abundant cell types in the crypts are undifferentiated enterocytes. However, crypts also contain two distinct cell types, Paneth cells and endocrine cells. Paneth cells do not migrate and are located within the crypt base. Their number increases from the duodenum to the ileum. They derive from undifferentiated crypt cells and seem to undergo replacement at a relatively slow rate. They have a pyramidal shape, the nucleus being located in the basal half of the cell. Their outstanding feature is their content of secretory granules of the zymogen variety, located in the cytoplasm between the nucleus and the apical border of the cell. As secretory cells, Paneth cells show rough endoplasmic reticulum and extensive Golgi apparatus. They contain lysozyme, immunoglobulin, and defensin and seem capable of phagocytic activity, suggesting a role in the regulation of the intestinal microbial flora.

Endocrine cells are relatively abundant in the crypts, although they are also present in the villi. They are of two morphological types, either "open," with a pyramidal

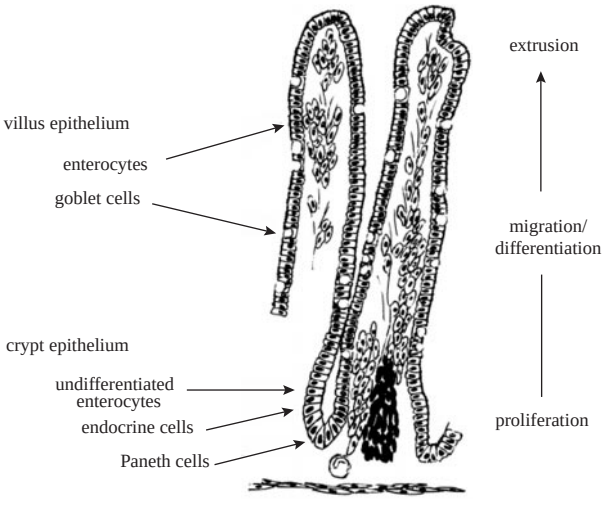


FIGURE 2 Schematic diagram of rat jejunal mucosa, illustrating the different types of cells encountered in the epithelium.

shape and connection with the lumen, or "closed," with a spindle shape and no connection with the lumen. Secretory granules are concentrated in the basal part of the cytoplasm whereas Golgi apparatus elements are supranuclear. The apical surface shows regular tufts of microvilli longer than those of enterocytes (described later). At least 16 distinct types of endocrine cells have been described along the small intestine, each with a characteristic regional distribution and composition.

Villus epithelium Enterocytes are the major villus epithelial cell type. They are highly specialized tall and columnar cells, with an oval nucleus located basally (Fig. 3). The usual organelles are all represented: the endoplasmic reticulum is not prominent but the Golgi apparatus is, the mitochondria are usually long and filamentous, and a number of pinocytic vesicles may be seen near the apical region. The apical surface is composed of microvilli and glycocalyx. Microvilli are close projections of cytoplasm covered by the cell membrane. They are about 1 μm high and 0.1 μm in diameter, providing a 14- to 40-fold increase in the apical surface area.

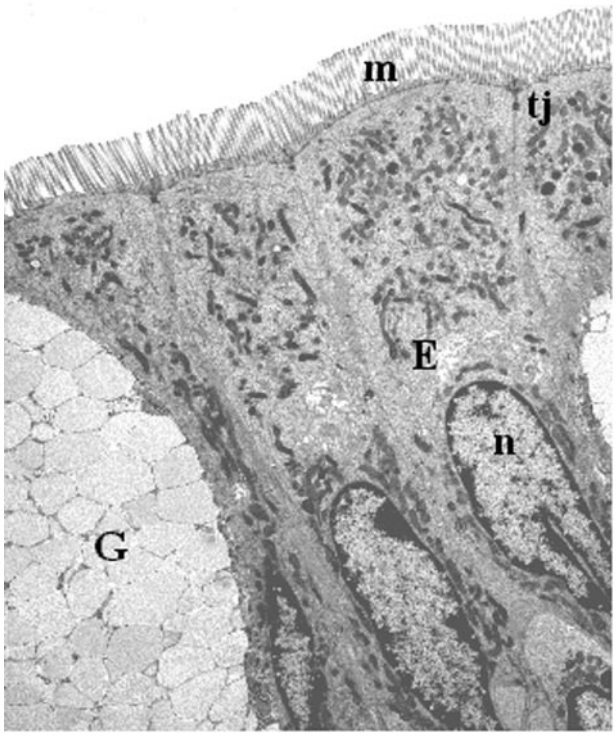


FIGURE 3 High-magnification micrograph of rat villus epithelial cells (× 3000). Enterocytes (E) are tall columnar cells with an oval, basally located nucleus (n). Their characteristic features are the presence of microvilli (m), which enhance apical surface area, and a tight junction (tj), allowing a tight and continuous junction between two adjacent cells. Goblet cells (G) are mucous-secreting cells interspersed among enterocytes.

The glycocalyx is a thin, filamentous layer of mucopolysaccharide material, housing important enzymes that function in terminal digestive processes. Enterocytes (and the other epithelial cells) are surrounded at their apical side by typical 0.5- to 2-μm intercellular junctional complexes. They consist of two distinct structures, the tight junction and the adherens junction, that encircle the cells as a belt, allowing a tight and continuous contact between two adjacent cells. These two structures are associations of transmembrane and cytoplasmic proteins linked with the cytoskeleton.

Goblet cells are interspersed among the enterocytes. They are polarized, mucous-secreting cells, increasing in number distally. They are easily recognizable by their pear-shaped region containing the mass of mature mucigen granules. Often a small "puff" of mucous can be distinguished emerging from the apex. Apical microvilli are sparse and irregular in size and shape. Cytoplasmic organelles, including a well developed network of rough endoplasmic reticulum, free ribosomes, mitochondria, and lysosomes, lie around the mass of mucous granules. Last, intraepithelial lymphocytes, mainly T cells, lie between individual epithelial cells, usually just above the basement membrane. The ratio of epithelial cells to intraepithelial lymphocytes is about 6 : 1.

PEYER'S PATCHES

Peyer's patches are distinct small lymphoid nodules located along the antimesenteric border. Their number increases distally. They are composed of lymphoid follicles that are located in the mucosa but that may also extend into the submucosa. Follicles are composed of a lymphoid nodule with a germinal center, a follicle-associated epithelium, and a dome, situated between these two structures. The germinal center contains mainly IgA-positive B cells. It is surrounded by the dome, populated by B lymphocytes, macrophages, and plasma cells. Facing the lumen is the follicle-associated epithelium, which contains few goblet cells and epithelial cells specialized in the transport of luminal antigens, the microfold cells (M cells). These cells are different from enterocytes in that they have short microvilli and no glycocalyx at the apical side. Moreover, their basolateral membrane is invaginated to form a pocket containing immunocompetent cells.

See Also the Following Articles

Circulation, Overview • Duodenum, Anatomy • Enteric Nervous System • Gastrointestinal Tract Anatomy, Overview •

Interstitial Cells of Cajal • Gastrointestinal Matrix, Organization and Significance

Further Reading

Madara, J. L., and Trier, J. S. (1994). The functional morphology of the mucosa of the small intestine. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 4th Ed., pp. 1577–1622. Raven Press, New York.

Porter, E. M., Bevins, C. L., Ghosh, D., and Ganz, T. (2002). The multifaceted Paneth cell. *Cell. Mol. Life Sci.* **59**, 156–170.

Segal, G. H., and Petras, R. E. (1997). Small intestine. *In* "Histology for Pathologists" (S. S. Sternberg, ed.), 2nd Ed., pp. 495–518. Lippincott-Raven Publ., Philadelphia.

Timmermans, J. P., Hens, J., and Adriaensen, D. (2001). Outer submucous plexus: An intrinsic nerve network involved in both secretory and motility processes in the intestine of large mammals and humans. *Anat. Rec.* **262**, 71–78.

Van Itallie, C. M., and Anderson, J. M. (2001). Molecular structure and regulation of tight junctions. *Curr. Top. Membr.* **50**, 163–186.

Van Kruiningen, H. J., West, A. B., Freda, B. J., and Holmes, K. A. (2002). Distribution of Peyer's patches in the distal ileum. *Inflamm. Bowel Dis.* **8**(3), 180–185.

Small Intestine, Benign and Malignant Neoplasms of the

J. GUSTAV KULDAU AND JOHN M. CARETHERS
University of California, San Diego

ampulla of Vater The orifice in the duodenum through which the common bile duct and pancreatic duct drain.

celiac sprue Intestinal malabsorption disease characterized by diarrhea and malnutrition, caused by immune-mediated gluten sensitivity.

Crohn's disease A disease of chronic inflammation potentially involving any portion of the gastrointestinal tract but with a propensity for the terminal ileum. The etiology of this disease is incompletely understood.

familial adenomatous polyposis (FAP) An inherited syndrome in which thousands of polyps develop in the colon, as well as in the stomach and upper intestine (duodenum). Bony tumors, known as osteomas, and other soft tissue tumors can also occur (Gardner's variant).

Peutz-Jeghers syndrome Familial syndrome consisting of mucocutaneous pigmentation and gastrointestinal polyp (usually harmartomas) formation.

Tumors of the small intestine are rare, accounting for only 1% of all gastrointestinal (GI) neoplasms, even though the small intestine accounts for 90% of the surface area and 75% of the length of intestinal tract. Approximately 5300 new cases of small intestinal malignancies are diagnosed each year. Two-thirds of small bowel neoplasms are malignant and these account for approximately 2% of all GI malignancies. The most common malignancies of the small bowel are adenocarcinomas, carcinoid tumors, lymphomas, and sarcomas. There is a slight male predominance and the average age at disease presentation is 55–60 years. This article will discuss malignant and benign small bowel neoplasms.

PATHOGENESIS

Tumors of the small bowel can form from any of the cells that are found in this organ. Adenomas and adenocarcinomas develop from epithelial and glandular tissue, lymphomas develop from lymphocytes found in lymphoid tissue, sarcomas develop from smooth muscle, and carcinoid tumors arise from the Kulchitsky

 Copyright 2004, Elsevier (USA). All rights reserved.

Interstitial Cells of Cajal • Gastrointestinal Matrix, Organization and Significance

Further Reading

Madara, J. L., and Trier, J. S. (1994). The functional morphology of the mucosa of the small intestine. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 4th Ed., pp. 1577–1622. Raven Press, New York.

Porter, E. M., Bevins, C. L., Ghosh, D., and Ganz, T. (2002). The multifaceted Paneth cell. *Cell. Mol. Life Sci.* **59**, 156–170.

Segal, G. H., and Petras, R. E. (1997). Small intestine. *In* "Histology for Pathologists" (S. S. Sternberg, ed.), 2nd Ed., pp. 495–518. Lippincott-Raven Publ., Philadelphia.

Timmermans, J. P., Hens, J., and Adriaensen, D. (2001). Outer submucous plexus: An intrinsic nerve network involved in both secretory and motility processes in the intestine of large mammals and humans. *Anat. Rec.* **262**, 71–78.

Van Itallie, C. M., and Anderson, J. M. (2001). Molecular structure and regulation of tight junctions. *Curr. Top. Membr.* **50**, 163–186.

Van Kruiningen, H. J., West, A. B., Freda, B. J., and Holmes, K. A. (2002). Distribution of Peyer's patches in the distal ileum. *Inflamm. Bowel Dis.* **8**(3), 180–185.

Small Intestine, Benign and Malignant Neoplasms of the

J. GUSTAV KULDAU AND JOHN M. CARETHERS
University of California, San Diego

ampulla of Vater The orifice in the duodenum through which the common bile duct and pancreatic duct drain.

celiac sprue Intestinal malabsorption disease characterized by diarrhea and malnutrition, caused by immune-mediated gluten sensitivity.

Crohn's disease A disease of chronic inflammation potentially involving any portion of the gastrointestinal tract but with a propensity for the terminal ileum. The etiology of this disease is incompletely understood.

familial adenomatous polyposis (FAP) An inherited syndrome in which thousands of polyps develop in the colon, as well as in the stomach and upper intestine (duodenum). Bony tumors, known as osteomas, and other soft tissue tumors can also occur (Gardner's variant).

Peutz-Jeghers syndrome Familial syndrome consisting of mucocutaneous pigmentation and gastrointestinal polyp (usually harmartomas) formation.

Tumors of the small intestine are rare, accounting for only 1% of all gastrointestinal (GI) neoplasms, even though the small intestine accounts for 90% of the surface area and 75% of the length of intestinal tract. Approximately 5300 new cases of small intestinal malignancies are diagnosed each year. Two-thirds of small bowel neoplasms are malignant and these account for approximately 2% of all GI malignancies. The most common malignancies of the small bowel are adenocarcinomas, carcinoid tumors, lymphomas, and sarcomas. There is a slight male predominance and the average age at disease presentation is 55–60 years. This article will discuss malignant and benign small bowel neoplasms.

PATHOGENESIS

Tumors of the small bowel can form from any of the cells that are found in this organ. Adenomas and adenocarcinomas develop from epithelial and glandular tissue, lymphomas develop from lymphocytes found in lymphoid tissue, sarcomas develop from smooth muscle, and carcinoid tumors arise from the Kulchitsky

 Copyright 2004, Elsevier (USA). All rights reserved.

cell, an enterochromaffin cell located in the crypts of Leiberkuhn.

Many theories have been proposed to explain the low rate of malignancy in the small bowel as compared to the colon and stomach:

1. liquid contents of the small bowel cause less mucosal irritation and dilute carcinogens;
2. rapid transit of small bowel contents decreases exposure to carcinogens;
3. decreased bacterial load in the small bowel (especially anaerobic bacteria) results in less conversion of bile acids into carcinogens;
4. benzpyrene, a known carcinogen found in food, is converted to less toxic metabolites by benzpyrene hydrolase, which is found in higher concentrations in the small intestine than in the stomach and colon; and
5. increased lymphoid tissue and secretory immunoglobulin A found in the small bowel may be protective against cancer formation.

PREDISPOSING CONDITIONS

Many diseases have known associations with small bowel neoplasms:

1. Crohn's disease is a risk factor for adenocarcinoma of the small bowel.
2. Celiac disease is associated with lymphoma and adenocarcinoma.
3. Patients with Peutz-Jeghers syndrome have hamartomas in the jejunum or ileum which may undergo malignant transformation.
4. Patients with FAP have increased rates of small bowel adenomas, especially in the duodenum.

CLINICAL PRESENTATION

Patients with small bowel malignancies often present late in the course of the disease, when the tumor has spread beyond the point of cure. The most common symptoms are abdominal pain, intestinal obstruction, intussusception, gastrointestinal bleeding, palpable abdominal mass, anorexia, and weight loss. Over 50% of benign small bowel tumors remain asymptomatic, whereas 70–90% of malignant tumors lead to symptoms. With malignant tumors, abdominal pain is the most common symptom, present in as many of 80% of patients. Weight loss and anorexia occur in approximately half of patients with malignant tumors and 25% of patients develop intestinal obstruction. Benign tumors lead to gastrointestinal (GI) bleeding more commonly than malignant tumors (29% versus 6%). Abdominal pain, often fluctuating in nature, is the most common symptom of benign lesions.

DIAGNOSIS

Diagnosing small bowel neoplasms is often difficult due to the rarity of the disease and the nonspecific nature of signs, symptoms, and physical exam findings. It is important to consider a small bowel neoplasm in patients with unexplained abdominal pain, weight loss, or occult GI bleeding. Late detection of small bowel tumors is common, with the average time from onset of symptoms to diagnosis being 5 months. Up to 50% of patients will have metastatic disease at presentation. Unfortunately, early detection and treatment are the most significant variables for favorable outcome in these patients.

Patients suspected of having a small bowel neoplasm should undergo a complete history and physical examination, fecal occult blood testing, a complete blood count, measurement of serum electrolytes, and liver function tests.

Radiographic examination is often employed in the evaluation of suspected small bowel tumors. Plain films of the abdomen are of little use unless the lesion has led to intestinal obstruction. Upper GI series with small bowel follow-through (SBFT) are commonly ordered and can show mass lesions in over 50% of patients. A superior test but one that is more technically difficult to perform is enteroclysis, a double-contrast study performed by inserting a tube in the duodenum and infusing barium. The sensitivity of enteroclysis is approximately 90% versus 50% for SBFT. Computed tomography (CT) is not effective for detecting small intralumenal or mucosal tumors, but is excellent at detecting metastasis, especially to the liver. CT enteroclysis is performed by placing a tube in the proximal small bowel and infusing barium while obtaining CT images. The accuracy of this study for detecting small bowel tumors is still being determined.

Endoscopy is often useful in the diagnosis of small bowel tumors. Esophagogastroduodenoscopy (EGD) can visualize only to the second to third portion of the duodenum, limiting its usefulness. EGD is good for evaluating patients with FAP for duodenal adenomas or adenocarcinomas. Lesions past the ligament of Treitz are more difficult to reach with an endoscope. Push enteroscopy, a type of upper endoscopy using a longer than normal endoscope, can allow the user to visualize up to 60 cm of proximal jejunum. A newer technique for visualizing the small intestine is wireless capsule endoscopy. In this method, the patient swallows a capsule

that takes video images as it passes through the length of the small intestine. Capsule endoscopy will likely be helpful in diagnosis of small intestinal tumors, and experience with this technique is growing.

MALIGNANT LESIONS OF THE SMALL INTESTINE

Small Bowel Adenocarcinoma

Adenocarcinoma is the most common type of malignant small bowel carcinoma, constituting 40–50% of small bowel cancers. Approximately 50% of small bowel adenocarcinomas occur in the duodenum, especially at the ampulla of Vater, whereas 30% form in the jejunum and 20% occur in the ileum. The higher concentrations of bile acids and their metabolites in the duodenum have been postulated to explain the increased incidence of carcinoma in this portion of bowel. Adenocarcinomas usually occur between the ages of 50 and 70, with a male:female ratio of approximately 1.4 : 1. Risk factors include Crohn's disease; high animal fat intake; heavy consumption of red meat, salt-cured foods, and smoked foods; prior peptic ulcer disease; FAP; prior colon cancer; celiac sprue; and cystic fibrosis.

Early adenocarcinoma of the small bowel is usually asymptomatic, as with other small bowel lesions. When symptoms occur, they are usually nonspecific. A high degree of suspicion is often required to make an early diagnosis. If the lesion is in the duodenum, endoscopy can be utilized to make a diagnosis.

Surgical intervention provides the only hope for cure in patients with adenocarcinoma of the small bowel. Patients with lesions in the first or second part of the duodenum should undergo pancreaticoduodenectomy (Whipple procedure). Tumors occurring in the distal duodenum, jejunum, or ileum may be treated with wide local excision with lymphadenectomy. Prognosis is determined by resectability, pathologic status of resected margins, histologic grade, and presence or absence of lymph node involvement. Lesions confined to mucosa and submucosa have a good prognosis, whereas lesions that invade through the serosa have a very poor prognosis. The overall 5-year survival rate ranges from 20 to 35%. Small bowel adenocarcinomas are generally considered to be radio-resistant and thus radiation therapy does not appear to be effective. To complicate matters, the small bowel cannot tolerate high doses of radiation without sustaining significant injury. For these reasons, radiotherapy is not widely used for these tumors. Chemotherapy has been used to treat these tumors, but due to the low incidence of small bowel adenocarcinoma, literature on its effectiveness is lacking.

Carcinoid Tumors

Carcinoid tumors are a rare type of neuroendocrine tumor that can be located in the stomach, small intestine, colon, or rectum, although they are most common in the ileum and appendix. Carcinoid tumors make up approximately 40% of primary small bowel malignancies.

These tumors arise from the Kulchitsky cell, an enterochromaffin cell located in the crypts of Lieberkuhn. Grossly, these tumors appear yellow, due to their high lipid content, and are firm. Carcinoid tumors are usually slow-growing and are often diagnosed incidentally. Inflammation with shortening and thickening of the mesentery, the so-called desmoplastic reaction, can lead to intestinal obstruction due to kinking of the intestine. Tumor cells often secrete bioactive products, such as serotonin, which can lead to carcinoid syndrome in approximately 10% of patients. Symptoms of carcinoid syndrome include diarrhea, flushing, and hypotension. Surgery to completely remove the tumor and any involved lymph nodes is considered the best treatment.

Primary Small Bowel Lymphomas

Lymphomas are solid malignancies of lymphoid tissue. Lymphomas often arise in lymph nodes; however, lymphomas may also arise in extranodal sites, with the GI tract being the most common location for extranodal involvement. The diagnosis of primary GI lymphoma requires that no peripheral lymphadenopathy be present, that the patient have a normal white blood cell count, that tumor involvement be predominantly in the GI tract, and that there is no involvement of the liver or spleen. Approximately 10% of GI tract lymphomas arise in the small intestine. The incidence of small bowel lymphoma peaks in the seventh decade of life and there is a slight male predominance. Most tumors arise in the distal small bowel, most likely because there is more lymphoid tissue in these areas, especially in the terminal ileum. Predisposing conditions for small bowel lymphoma include autoimmune disease, immunodeficiency syndromes (especially acquired immunodeficiency syndrome), immunosuppressive therapy after organ transplantation, Crohn's disease, celiac disease, and radiation therapy. Patients most often present with abdominal pain, weight loss, anorexia, and, less commonly, gastrointestinal bleeding or iron deficiency anemia.

There are many types of small intestinal lymphomas. The B-cell lymphomas include marginal B-cell lymphomas, diffuse large B-cell lymphomas, mantle cell lymphoma, follicular lymphoma, and Burkitt's lymphoma. Immunoproliferative small intestinal disease (IPSID), also known as α heavy-chain disease or Mediterranean lymphoma, is a B-cell lymphoma that occurs mainly in the Middle East. This lymphoma is most common in the second or third decade of life. IPSID usually occurs in patients of lower socioeconomic class. There is an association between infectious agents, such as *Giardia lamblia*, and IPSID. Treatment of early-stage disease often consists of antibiotic therapy alone, although chemotherapy is needed for more advanced cases.

The most common T-cell lymphoma is enteropathy-associated intestinal T-cell lymphoma. This lymphoma is a complication of celiac sprue, a genetic disease characterized by gluten sensitivity. These tumors often arise in the jejunum. This tumor should be considered in patients with celiac sprue who are adhering to a strict gluten-free diet but have persistent symptoms of diarrhea and weight loss.

Treatment for small intestinal lymphomas usually include a combination of surgery and chemotherapy. Surgery alone results in a cure for approximately 30% of patients. The overall 5-year survival for small intestinal lymphomas is approximately 20–30%, depending on the type of tumor.

Sarcomas

Sarcomas are malignant tumors that arise from smooth muscle tissue and can occur anywhere in the body where smooth muscle is present. Sarcomas account for 10% of small bowel tumors. The most common type of small bowel sarcoma is the leiomyosarcoma, with less common varieties including fibrosarcoma, liposarcoma, and angiosarcoma. Small bowel sarcomas tend to grow slowly and are usually large at presentation, with many being greater than 5 cm in size. Signs and symptoms at presentation include weight loss, gastrointestinal bleeding, bowel perforation, or palpable mass on abdominal exam. These tumors often grow extralumenally and therefore bowel obstruction is a rare presentation. Sarcomas look very similar to leiomyomas, a benign smooth muscle tumor, and it can be hard to tell the difference between them, even on histological examination. Criteria for malignancy include the following: more than two mitoses per high-powered field; nuclear atypia; and necrosis. Treatment for sarcomas is surgery. Sarcomas rarely metastasize to lymph nodes, so lymph node resection is usually not indicated and will not improve survival. Chemotherapy and radiation therapy have not shown benefit. Survival is approximately 50% at 5 years.

Metastatic Lesions

Even though primary small bowel carcinomas are quite rare, the small intestine is a relatively common site for metastatic disease. Malignant melanoma commonly metastasizes to the small intestine, with 60% of patients with melanoma having gastrointestinal metastases. Primary carcinomas from the breast, lung, and kidney can also metastasize to the small intestine.

BENIGN SMALL INTESTINAL TUMORS

Adenomas

Adenomas of the small intestine can be villous or tubular. Villous adenomas have significant malignant potential. Up to 40% of villous adenomas are found to have malignant cells within them when removed. Periampullary villous adenomas may require a Whipple procedure for resection. Tubular adenomas have lower malignant potential.

Leiomyomas

Leiomyomas are firm, are gray or white in color, and arise from the submucosal layer. On endoscopy, they appear as submucosal lesions with normal overlying mucosa. These tumors consist of well-differentiated smooth muscle cells. Like sarcomas, these lesions often grow extralumenally and therefore can achieve a large size before they are discovered. Central necrosis of the lesion can occur when the tumor outgrows its blood supply, which may lead to ulceration and bleeding into the bowel lumen.

Lipomas

Lipomas are benign tumors that consist mainly of fat. They usually occur in the submucosal layer and have normal overlying mucosa. They are most commonly found in the ileum and jejunum.

See Also the Following Articles

Cancer, Overview • Celiac Disease • Crohn's Disease • Familial Adenomatous Polyposis (FAP) • Hamartomatous Polyposis Syndromes • Lymphomas

Further Reading

Cusack, J. (2002). Diagnosis and management of small bowel neoplasms. UpToDate, Version 10.3. Available at www.uptodate.com.

Cusack, J. (2002). Epidemiology, clinical features, and types of small bowel neoplasms. UpToDate, Version 10.3. Available at www.uptodate.com.

DiSario, J., Randall, W., Vargas, H., and McWhorter, W. (1994). Small bowel cancer: Epidemiological and clinical characteristics from a population-based registry. *Am. J. Gastroenterol.* **89**, 699–701.

Lance, P. (1999). Tumors and other neoplastic diseases of the small intestine. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 2, pp. 1722–1738. Lippincott, Williams & Wilkins, Philadelphia, PA.

Neugut, A., Marvin, M., Rella, V., and Chabot, J. (1997). An overview of adenocarcinoma of the small intestine. *Oncology* **11**, 529–536.

Rustgi, A. (2002). Small intestinal neoplasms. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., Vol. II, pp. 2169–2172. W. B. Saunders, Philadelphia, PA.

Small Intestine, Development

DEBORAH C. RUBIN

Washington University School of Medicine, Saint Louis, Missouri

crypt–villus axis morphogenesis Formation of the crypt–villus unit, the principal functional component of the small intestine.

endodermal specification Process by which the endoderm forms; distinct from the ectodermal and mesodermal layers of the early embryo.

homeobox genes Transcription factors that contain a homeobox or homeodomain, a conserved nucleotide sequence that encodes a DNA binding domain.

The small intestine is a complex epithelial-lined tubular organ that grows to ~6 m in length in the adult human. It is a heterogeneous tissue that exhibits regional differences in morphology and function along the horizontal axis (from duodenum to ileum) and along the vertical axis (from crypt to villus tip). Its epithelium is derived from the primitive endoderm, and connective tissue, muscle, and hematopoietic components are mesodermal in origin. The duodenum, up to the opening for the common bile duct (ampulla of Vater), is derived from the foregut (which also gives rise to the liver, lungs, and pancreas). The midgut gives rise to the duodenum beyond the ampulla, and includes the jejunum, ileum, and large bowel, up to and including the proximal transverse colon.

EMBRYOLOGY OF THE SMALL INTESTINE

The gastrointestinal epithelium is derived from the embryonic endoderm, and the smooth muscle, hematopoietic elements, and connective tissue are derived from mesoderm. The endoderm, ectoderm, and mesoderm are formed during gastrulation, which occurs in the human by week 3 of gestation. During this process, the primitive streak appears in the bilaminar germ disk of the blastocyst, which is suspended between the yolk sac and amnion. Cells migrate ventrally and laterally, forming the three germ layers. The gut tube forms and closes as the endodermal layer folds. In rodents, folding begins as the anterior intestinal portal forms at the anterior tip of the endoderm. Cells located at this portal move posteriorly. The caudal intestinal portal similarly emerges at the posterior endodermal tip, and cells located at this position move anteriorly. The gut tube closes as a result of this folding process. The foregut and hindgut initially form closed sacs, but the midgut communicates with the yolk sac, by the vitelline duct. Aberrant persistence of this duct leads to the congenital anomaly called Meckel's diverticulum. The endoderm contacts ectoderm at the closed ends of the foregut and hindgut, and fuses to form the buccopharyngeal and cloacal membranes.

By weeks 4–5 the gut tube elongates rapidly, and by week 7, this expansion forces its herniation into the umbilicus. By weeks 8–9, intestinal villi have formed. As the gut returns to the abdominal cavity by week 10, it undergoes a complex series of rotational events (Fig. 1). This rotation and subsequent migration of the cecum to the right lower quadrant results in the proper anatomic location of the small and large bowels.

Cusack, J. (2002). Epidemiology, clinical features, and types of small bowel neoplasms. UpToDate, Version 10.3. Available at www.uptodate.com.

DiSario, J., Randall, W., Vargas, H., and McWhorter, W. (1994). Small bowel cancer: Epidemiological and clinical characteristics from a population-based registry. *Am. J. Gastroenterol.* **89**, 699–701.

Lance, P. (1999). Tumors and other neoplastic diseases of the small intestine. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 2, pp. 1722–1738. Lippincott, Williams & Wilkins, Philadelphia, PA.

Neugut, A., Marvin, M., Rella, V., and Chabot, J. (1997). An overview of adenocarcinoma of the small intestine. *Oncology* **11**, 529–536.

Rustgi, A. (2002). Small intestinal neoplasms. *In* "Sleisenger & Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. Friedman, and M. Sleisenger, eds.), 7th Ed., Vol. II, pp. 2169–2172. W. B. Saunders, Philadelphia, PA.

Small Intestine, Development

DEBORAH C. RUBIN

Washington University School of Medicine, Saint Louis, Missouri

crypt–villus axis morphogenesis Formation of the crypt–villus unit, the principal functional component of the small intestine.

endodermal specification Process by which the endoderm forms; distinct from the ectodermal and mesodermal layers of the early embryo.

homeobox genes Transcription factors that contain a homeobox or homeodomain, a conserved nucleotide sequence that encodes a DNA binding domain.

The small intestine is a complex epithelial-lined tubular organ that grows to ~6 m in length in the adult human. It is a heterogeneous tissue that exhibits regional differences in morphology and function along the horizontal axis (from duodenum to ileum) and along the vertical axis (from crypt to villus tip). Its epithelium is derived from the primitive endoderm, and connective tissue, muscle, and hematopoietic components are mesodermal in origin. The duodenum, up to the opening for the common bile duct (ampulla of Vater), is derived from the foregut (which also gives rise to the liver, lungs, and pancreas). The midgut gives rise to the duodenum beyond the ampulla, and includes the jejunum, ileum, and large bowel, up to and including the proximal transverse colon.

EMBRYOLOGY OF THE SMALL INTESTINE

The gastrointestinal epithelium is derived from the embryonic endoderm, and the smooth muscle, hematopoietic elements, and connective tissue are derived from mesoderm. The endoderm, ectoderm, and mesoderm are formed during gastrulation, which occurs in the human by week 3 of gestation. During this process, the primitive streak appears in the bilaminar germ disk of the blastocyst, which is suspended between the yolk sac and amnion. Cells migrate ventrally and laterally, forming the three germ layers. The gut tube forms and closes as the endodermal layer folds. In rodents, folding begins as the anterior intestinal portal forms at the anterior tip of the endoderm. Cells located at this portal move posteriorly. The caudal intestinal portal similarly emerges at the posterior endodermal tip, and cells located at this position move anteriorly. The gut tube closes as a result of this folding process. The foregut and hindgut initially form closed sacs, but the midgut communicates with the yolk sac, by the vitelline duct. Aberrant persistence of this duct leads to the congenital anomaly called Meckel's diverticulum. The endoderm contacts ectoderm at the closed ends of the foregut and hindgut, and fuses to form the buccopharyngeal and cloacal membranes.

By weeks 4–5 the gut tube elongates rapidly, and by week 7, this expansion forces its herniation into the umbilicus. By weeks 8–9, intestinal villi have formed. As the gut returns to the abdominal cavity by week 10, it undergoes a complex series of rotational events (Fig. 1). This rotation and subsequent migration of the cecum to the right lower quadrant results in the proper anatomic location of the small and large bowels.

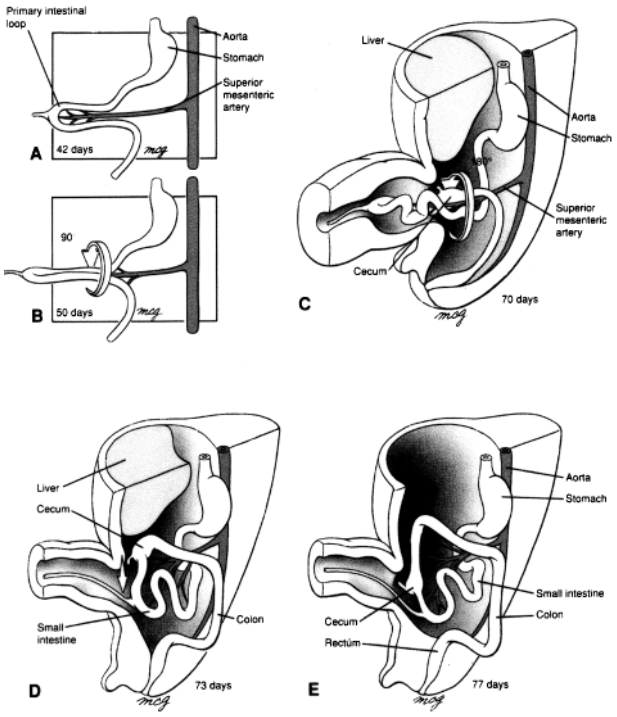


FIGURE 1 Intestinal growth, herniation, and rotation. (A, B) At the end of the sixth week, the primary intestinal loop herniates into the umbilicus, rotating through 90° counterclockwise. (C) The small intestine elongates to form jejunal-ileal loops, the cecum and appendix grow, and at the end of the 10th week, the primary intestinal loop retracts into the abdominal cavity, rotating an additional 180° counterclockwise. (D, E) During the 11th week, the retracting midgut completes this rotation as the cecum is positioned just inferior to the liver. The cecum is then displaced inferiorly, pulling down the proximal hindgut to form the ascending colon. The descending colon is simultaneously fixed on the left side of the posterior abdominal wall. (Reprinted with permission from Larsen, W. J. (1997). "Human Embryology," p. 241. W. B. Saunders Co.

IN THE BEGINNING: EARLY SPECIFICATION OF THE ENDODERM AND FOLDING/FORMATION OF THE GUT TUBE

How is the early endodermal layer formed, distinct from mesoderm and ectoderm? An explosion of information about the regulation of early endodermal specification has come from studies in organisms such as the frog (*Xenopus laevis*), the zebrafish (*Danio rerio*), and the mouse, although many parts of the mammalian puzzle still remain unsolved. A cascade of regulatory transcription factors appears to be responsible for this process. Genes related to *Xenopus VegT*, a "T box" transcription factor, are likely to be at the top of the transcriptional regulatory hierarchy, activating Nodal signaling and other transforming growth factor-β (TGF-β) family proteins, and inducing expression of proteins such as xSox17α, the Mix-like paired homeodomain proteins and GATA proteins. Fibroblast growth factors, members of the Wnt growth factor family, and retinoic acid are soluble factors involved in gastrulation. Experiments in knockout (null) mice, in which specific genes have been deleted to determine their function in early embryonic life, have shown the critical role of some of these transcription factors. For example, *Mixl1* null (*Mixl1*−/−) mice exhibit a thickened primitive streak, arrested embryo development, and absent heart tube and gut, among other abnormalities. Another important set of proteins belongs to the hepatic nuclear factor 3 (HNF3) family, which are involved in endodermal differentiation at later stages of development. Many of these proteins play multiple roles in the early embryo (for example, in anterior−posterior pattern formation and the establishment of left−right asymmetry), thus it is likely that cell-specific differences in receptor populations, concentrations of morphogens, and dosage over time, as well as space, may determine the specific end result of their actions.

The next step in the process of gut morphogenesis is folding and formation of the gut tube. GATA-4 and several of the bone morphogenetic proteins appear to be important regulators in mice. For example, *Gata4*−/− mice exhibit abnormal ventral morphogenesis, lacking a primitive foregut and heart tube. The bone morphogenetic proteins (BMP-2, -4, -5, and -7) are expressed in the anterior intestinal portal and are required for proper folding of the embryo. These events have not been studied in the human.

Regional Differentiation in the Small Bowel

The molecular factors that regulate regional specification along the horizontal axis of the small intestine and colon (from duodenum to rectum) are the subject of intensive investigation. An important class of candidate genes is the homeobox-containing (*hox*) transcription factors, which are involved in the specification of body pattern and which are expressed in specific regions of the gut mesenchyme during embryonic development (Fig. 2). The *hox* genes are also expressed in the adult human small intestine. These genes demonstrate spatial colinearity; i.e., the pattern of expression along the body axis reflects the chromosomal order of the genes. Their importance in small bowel ontogeny has been demonstrated in knockout mice or by experiments in which *hox* genes are "misexpressed" in an aberrant location (i.e., more caudal or ventral than the normal expression pattern). For example, targeted deletion of *Hoxd12* or

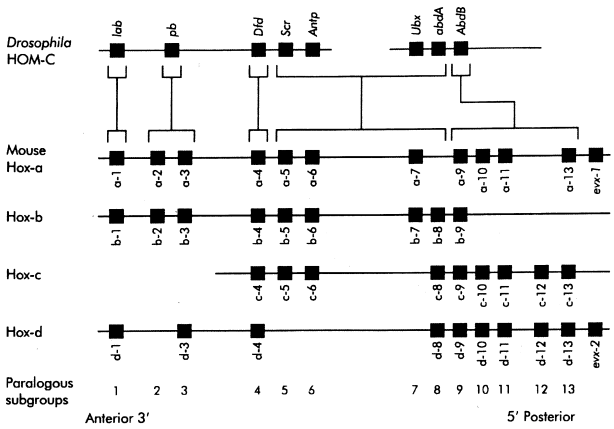


FIGURE 2 Diagrammatic representation of the *HOM-C* complex in *Drosophila* and its phylogenetic homology in the form of four *Hox* paralogues. Reproduced from Beck, F. (2002). Homeobox genes in gut development. *Gut* **51**, p. 451. With permission from the BMJ Publishing Group.

Hoxd13 produces anal sphincter anomalies, and deletion of the *Hoxd4-d13* gene cluster yields abnormalities in the ileocecal valve and the pylorus, as well as the anal sphincter. Virally mediated misexpression of *Hoxd13* in midgut mesoderm, (which is normally expressed in hindgut mesoderm) induces a hindgut endodermal phenotype. These fascinating experiments indicate the importance of the *hox* genes in normal gut morphogenesis and support the critical role of mesenchymal–endodermal interactions in determining cell fate (see later).

Members of the related *Para-hox* gene cluster (which are dispersed homeobox-containing genes), are also important in gut patterning and organ specification. Just as the *Para-hox* gene *Pdx1* has been shown to play a critical role in pancreatic and duodenal development, members of the *Cdx* family, particularly *Cdx2*, have a crucial role in small bowel epithelial differentiation. *Cdx2*−/− (null) mice die at implantation; however, *Cdx2*+/− (heterozygous) mice develop hamartomatous lesions in the small bowel and colon. These lesions contain gastric mucosa from various parts of the stomach, including squamous forestomach epithelium, suggesting that the development of a proximal gastric epithelium is a "default pathway" that is followed when *Cdx2* is absent. *Cdx* genes may also regulate the expression of *hox* genes. *Cdx2* expression persists into adulthood, when it clearly plays a role in the terminal differentiation of the gut epithelium (see later).

The hedgehog signaling pathway has been implicated in gut morphogenesis, as a mediator of endoderm–mesenchyme interactions. Mice in which the *Sonic hedgehog* (*Shh*) gene, or in which its downstream transcriptional regulators, Gli 2 and Gli 3, have been deleted, show abnormalities in foregut formation, with anomalies in lung, trachea, and esophagus formation. *Shh* null mice also show reduced intestinal smooth muscle, gut malrotation, annular pancreas, intestinal transformation of the stomach, duodenal stenosis, and imperforate anus. Ectopic expression of *Shh* induces *Bmp4* and *Hoxd13* expression and results in muscle hypertrophy in the gut, suggesting a role in radial (from villus tip to smooth muscle/serosa) differentiation.

CRYPT–VILLUS AXIS FORMATION, EPITHELIAL–MESENCHYMAL INTERACTIONS, AND EPITHELIAL CELL DIFFERENTIATION

Crypt–Villus Axis Morphogenesis

The principal anatomic and functional (absorptive) unit of the small intestine is the crypt–villus axis (Fig. 3). The crypts of Lieberkuhn contain presumptive gut epithelial stem cells that are the source of the four

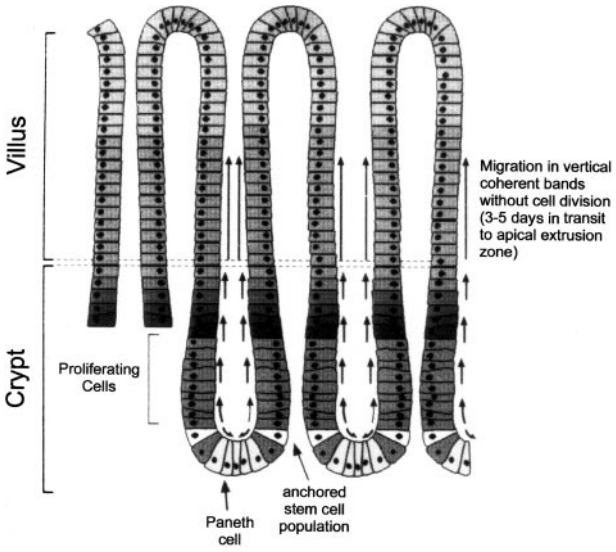


FIGURE 3 Model of organization of the small intestinal crypt–villus axis. The small intestinal crypt contains approximately 250 cells. Anchored stem cells (white) are located in the lower crypt and give rise to proliferating daughter cells that migrate up toward the villus. These differentiate into enterocytes, goblet cells, and enteroendocrine cells. Paneth cells also arise from the stem cell and differentiate during downward translocation to the base. Senescent cells are extruded near the villus tips. Reproduced with permission from Rubin, D. C. (1999). Small intestine: Anatomy and structural anomalies. *In* "Textbook of Gastroenterology" (Yamada, T., Alpers, D. H., Laine, L., Owyang, C., and Powell, D. W., eds.). 3rd Ed., Vol. 2, pp. 1561–1583. Copyright Lippincott Williams and Wilkins. Adapted from Gordon, J. I. (1989). Intestinal epithelial differentiation: New insights from chimeric and transgenic mice. *J. Cell Biol.* **108**, 1187.

major cell types of the gut mucosa. The stem cells, located near the base of the crypts, give rise to proliferating daughter cells in the midcrypt region; daughter cells in turn migrate onto the villus to become terminally differentiated (nonproliferating) absorptive enterocytes, mucus-secreting goblet cells, or enteroendocrine cells, or to the bottom of the crypt to become Paneth cells. Two additional, rarer cell types thought to arise from the same stem cell are the epithelial M cell, which overlies Peyer's patches, and the tuft or caveolated cell, a villus-associated cell.

The small intestinal epithelial crypt–villus axis is formed in humans by week 12 of gestation. The complex process of crypt–villus morphogenesis begins with marked proliferation of the undifferentiated endoderm and surrounding mesenchyme of the gut tube. The endoderm and mesenchyme become stratified, and then undergo a remodeling process that results in the formation of a lumen and villi. This proceeds over time in a rostral-to-caudal direction. Intestinal villi form by weeks 8–9, followed by the crypts by weeks 10–12. Mesenchymal cells invaginate into the endoderm to form crypts and villi. Cells with the appropriate morphologic characteristics of enterocytes, goblet cells, enteroendocrine, and Paneth cells can be identified by week 12. Before birth, the villus-associated enterocytes acquire all the brush border enzymes (e.g., sucrase– isomaltase, lactase, and peptidases) and transporters (for glucose, fructose, etc.) required for normal nutrient absorption. Hormonal factors such as glucocorticoids, and growth factors such as epidermal growth factor and the TGF-β family, are likely to play an important tropic and maturational role.

Epithelial–Mesenchymal Interactions

Many years of experimentation have shown that mesenchymal–epithelial interactions are critical for the formation of the crypt–villus axis. For example, primitive endoderm cannot differentiate in culture without the presence of mesenchyme. Endoderm and mesoderm have inductive capabilities (i.e., each can direct the differentiation of the other tissue in coculture models). As indicated previously, Shh–BMP signaling from endoderm to mesenchyme is important in early gut formation. The requirement for mesenchyme in the maintenance of the crypt–villus axis has been confirmed by studies of Forkhead 6 (Fkh 6; now Foxl1), Nkx2.3, and epimorphin, indicating that the absence of these mesenchymal factors dramatically alters the normal morphology of the crypts and villi.

Crypt Stem Cells and Epithelial Cell Differentiation

Precise isolation and identification of the intestinal crypt stem cell have remained elusive, because, unlike other progenitor cells, such as the hematopoietic stem cell, genetic markers have yet to be characterized. However, some of the factors that maintain the crypt's proliferative compartment have been discovered. The Wnt/β-catenin/adenomatous polyposis coli (APC) signaling pathway, critically important in colon carcinogenesis, also appears to be required for the maintenance of the normal crypt–villus axis. Mice with deletion of the gene encoding Tcf-4, a member of the Tcf/Lef family of hydroxymethylglutaryl coenzyme A (HMG) box transcription factors that are downstream effectors of this pathway, show the absence of proliferative compartments in the prospective crypt region. Mice with this defect die soon after birth. A fascinating study has provided some insight into how cellular compartmentalization and migration along the crypt–villus axis are also regulated by this pathway. β-Catenin/T cell factor (TCF) complexes regulate the expression of ephrin B (EphB) receptor tyrosine kinases. Mice containing null mutations for both the EphB2 and EphB3 receptors show inappropriate mixing on the villus of proliferative cells, which are normally strictly confined to the crypts. The migration of Paneth cells to the base of the crypt also appears dependent on EphB receptor tyrosine kinases, because deletion of the EphB3 receptor tyrosine kinase results in localization of Paneth cells on the villi as well as in the crypts.

A key challenge for the future is to delineate the pathways that determine how intestinal stem cells commit to become enterocytes, enteroendocrine, Paneth, or goblet cells. Evidence indicates that the Notch signaling pathway plays an important role in this process. Targeted deletion in mice of a component of the Notch signaling pathway, called Math 1, produces a crypt–villus axis that contains only enterocytes, suggesting that expression of Math 1 forces the stem cell into a "prosecretory" phenotype (Paneth, enteroendocrine, and goblet cells are secretory cells). Another member of this pathway, Hes1, appears to function as a basal inhibitor of enteroendocrine cell differentiation. It has been hypothesized that high levels of Notch expression induce increased Hes1 expression, which in turn inhibits Math 1, thus leading to the "default" enterocytic pathway. Finally, the *Para-hox* gene *Cdx2* has a crucial role in enterocytic differentiation. It is expressed in the developing and the adult gut. When transfected into a cryptlike epithelial cell line, it induces an enterocyte-like phenotype. *Cdx2* also regulates the expression of a

variety of enterocyte-specific genes, such as sucrase–isomaltase.

In conclusion, although the descriptive embryology of the human small intestine has been well characterized, its molecular basis is still the subject of intensive investigation. Studies using sophisticated techniques to analyze gut development in transgenic and knockout mice, as well as in *Drosophila*, zebrafish, and *Xenopus,* have provided a wealth of new information. These studies plus knowledge of the human genome should soon permit the identification of the molecular basis of a variety of intestinal birth defects.

See Also the Following Articles

Development, Overview • Large Intestine, Development

Further Reading

Babyatsky, M. W., and Podolsky, D. K. (1999). Growth and development of the gastrointestinal tract. *In* "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, L. Laine, C. Owyang, and D. W. Powell, eds.), 3rd Ed., Vol. 1, pp. 547–584. Lippincott Williams & Wilkins, Philadelphia.

Batle, E., Henderson, J. T., Beghtel, H., van den Born, M. M. W., Sancho, E., Huls, G., Meeldijk, H., Robertson, J., van de Wetering, M., Pawson, T., and Clevers, H. (2002). β-Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263.

Beck, F. (2002). Homeobox genes in gut development. *Gut* **51**, 450–454.

De Santa Barbara, P., Van Den Brink, G. R., and Roberts, D. J. (2002). Molecular etiology of gut malformations and diseases. *Am. J. Med. Genet.* (*Semin. Med. Genet.*) **115**, 221–230.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* **19**, 379–383.

Montgomery, R. K., Mulberg, A. E., and Grand, R. J. (1999). Development of the human gastrointestinal tract: Twenty years of progress. *Gastroenterology* **116**, 702–731.

Roberts, D. J. (2000). Molecular mechanisms of development of the gastrointestinal tract. *Dev. Dynam.* **219**, 109–120.

Rubin, D. C. (1999). Small intestine: Anatomy and structural anomalies. *In* "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, L. Laine, C. Owyang, and D. W. Powell, eds.), 3rd Ed., Vol. 2, pp. 1561–1583. Lippincott Williams & Wilkins, Philadelphia.

Shivdasani, R. A. (2002). Molecular regulation of vertebrate early endoderm development. *Dev. Biol.* **249**, 191–203.

Wells, J. M., and Melton, D. A. (1999). Vertebrate endoderm development. *Annu. Rev. Cell Dev. Biol.* **15**, 393–410.

Yang, Q., Bermingham, N. A., Finegold, M. J., and Zoghbi, H. Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* **294**, 2155–2158.

Smoking, Implications of

WILLIAM EVANS AND RICHARD A. WRIGHT
University of Louisville, Kentucky

apoptosis The morphologic changes of programmed cell death.
pouchitis A clinical syndrome of abdominal cramps, frequent watery stools, urgency, incontinence, malaise, and fever in those who have had an ileal–anal anastomosis or a Koch pouch (continent ileostomy).
regional enteritis Inflammation of a region of intestine.

In the United States, 23% of adults smoke cigarettes. Approximately 22% of the female population and 30% of males currently smoke. It is estimated that greater than 400,000 Americans die prematurely each year because of smoking, making it the most preventable cause of death in the United States.

INTRODUCTION

The leading causes of smoking-related death in the United States are lung cancer and heart disease, but smoking contributes to diseases involving all organ systems. Cigarette smoking primarily and secondarily alters the natural history and treatment of a variety of gastrointestinal diseases. The tobacco, tar, and smoke from cigarettes contain carcinogens such as aromatic amines, lactanes, halo-ethers, *n*-nitroso compounds, polycyclic and peroxy compounds. Tobacco smoke has shown to affect gastrointestinal motility, secretion, blood flow, and immunological function. These alterations have a cumulative role in gastrointestinal diseases.

GASTROESOPHAGEAL REFLUX DISEASE

Gastroesophageal reflux disease (GERD) is the most common disease in adult Americans. Pyrosis, or heartburn, is a symptom experienced by 20% of Americans weekly and by up to 50% of Americans monthly. The pathophysiologic process of GERD is multifaceted. Transient lower esophageal sphincter relaxation, hiatal hernia, impaired esophageal acid clearance, and delayed gastric emptying all contribute to GERD. Smoking alters some of these mechanisms and may contribute to more severe GERD. Cigarette smokers have been shown to hyposalivate compared to nonsmokers. The hyposalivation may significantly prolong esophageal acid clearance, decrease acid neutralization by bicarbonate, and decrease epidermal growth factor secretion (a tropic hormone for mucosal regeneration). Smokers have an increase in the number of reflux events due to a decrease in the lower esophageal sphincter pressure and an increase in the number of reflux events from coughing and deep inspiration.

PEPTIC ULCER DISEASE

Peptic ulcer disease (PUD) refers to the formation of excavated defects in the gastrointestinal mucosa of the stomach or duodenum, or both. It is estimated that there are 500,000 new cases and four million recurrences of ulcers in the United States each year. The three major risk factors are nonsteroidal anti-inflammatory drugs, *Helicobacter pylori*, and cigarette smoking. There exists a significant association between active smoking and PUD. Tobacco smoke may contribute to the development of one-fifth of ulcers in the general population. The onset of PUD is twice as high for smokers versus nonsmokers, with no gender difference. There is a dose–response relationship between smoking and the development of gastric and duodenal ulcers for which multiple theories and observations have been proposed. In general, nicotine augments factors that lead to the formation of ulcers, such as acid and pepsin secretion, motility, duodenogastric reflux, and free radical generation. It weakens local mucosal defense mechanisms such as decreasing mucosal blood flow, prostaglandin synthesis, and mucus secretion. With respect to *H. pylori* infection, smoking increases the susceptibility to infection and impairs the eradication rate to standard antibiotic therapy.

Smoking affects the natural history of PUD. Smokers have more recurrences of ulcers than do nonsmokers after healing, and the more they smoke, the greater the risk for ulcer recurrence. After successful antibiotic eradication of *H. pylori*, smoking does not increase the risk of peptic ulcers. Smoking is also associated

 Copyright 2004, Elsevier (USA). All rights reserved.

with more complications of PUD. Smoking is a direct risk factor for perforation of duodenal ulcers but protects against bleeding from PUD. More deaths occur from PUD in smokers in a dose-dependent manner. With traditional treatment of PUD, smoking decreases healing rates of duodenal ulcers but not gastric ulcers.

Smokers are more likely to develop both initial and recurrent ulcers, suffer from ulcer complications, and have delays in healing of their ulcers.

INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease encompasses those conditions with chronic or relapsing immune activation and inflammation within the gastrointestinal tract. The two major forms are regional enteritis and ulcerative colitis (UC). The pathogenesis of inflammatory bowel disease is complex and not well understood, but there is strong evidence that supports smoking as an important environmental factor in its incidence, natural history, and treatment. Smoking has the opposite effects in regional enteritis and ulcerative colitis. Smoking has been thought to be involved in many factors through its effects on thrombogenesis, intestinal defenses, intestinal blood flow, prostaglandin production, and immunological alterations.

A history of current or former smoking increases the risk for developing regional enteritis. The risk is stronger in women than men. Smoking also affects the clinical course of the disease. Relapse of the disease is increased in smokers versus nonsmokers. It has been shown as an independent risk factor for the clinical, surgical, and endoscopic recurrence of regional enteritis. Smoking cessation leads to a more benign natural history of the disease. Smoking impairs response to medical therapy. Female smokers are more likely to require immunosuppressive therapy. Likewise, smokers with regional enteritis are at increased risk for requiring surgery compared to those who have never smoked. Additionally, female smokers are more likely to have repeat operations and this is directly related to the number of cigarettes smoked.

In contrast to regional enteritis, ulcerative colitis is more common in nonsmokers. Smoking may be protective against UC in both sexes. Smoking increases the production of mucus, which acts as a protective barrier in the intestine. Mucus production has been studied in smokers and nonsmokers with UC and compared to control populations. Nonsmoking UC patients had lower levels of mucus production than nonsmoking controls; the level of production in smokers with UC was the same as that in normal individuals.

The risk of developing UC is lower in smokers than in those who have never smoked or in former smokers. The degree of protection increases with the number of cigarettes smoked. The risk for developing UC in former smokers is highest in the first 2 years after cessation. The risk for UC is highest for the heaviest former smokers and lowest for current heavy smokers.

This relationship between smoking and UC has led to investigation of nicotine in the treatment of UC patients. In patients with active colitis, nicotine replacement therapy in addition to conventional therapy improved clinical, histological, and endoscopic scores. Also, UC patients treated with transdermal nicotine patches were more likely to achieve complete remission. However, nicotine therapy was not better than standard therapy for maintaining remission. There is controversy about using an addictive drug such as nicotine to treat a chronic disease. In patients who have had a total colectomy with an ileal–anal anastomosis, smokers have less risk of developing pouchitis than ex-smokers and nonsmokers. Nonetheless, because of other adverse effects of cigarette smoking, it cannot be recommended as routine therapy for ulcerative colitis.

GASTROINTESTINAL CANCERS

Cancer is a complex multistage process involving gene mutation and altered apoptosis, which lead to the formation of a clinical cancer. Genetic make-up and exposure to both dietary risk factors and protective factors regulate this process. Smoking has been most strongly linked to lung cancer, but it is a risk factor for cancer in other organ systems, especially cancers involving the gastrointestinal tract.

Oral cancers, which include cancer of the lip, tongue, gum, salivary glands, mouth, and oropharynx, are mostly of squamous cell origin. The strongest risk factors are cigarette and pipe smoking and chewing tobacco.

The two major types of esophageal cancer are squamous cell cancer and adenocarcinoma. Squamous cell cancer is the most common, but over the past three decades, adenocarcinoma has dramatically increased. The major risk factors for squamous cell cancer are alcohol and tobacco. It is postulated that the chemical carcinogens in cigarette smoke may lead to esophageal epithelial hyperproliferation, progressing to malignant transformation. Independent of alcohol use, smokers have an increased risk for squamous cell cancer in a dose-dependent relationship. This risk may be higher for pipe and cigar smoking. Cessation of smoking will lower the risk to that of a non-smoking population 10 years after quitting.

Chronic gastroesophageal reflux leads to the formation of intestinal metaplasia or Barrett's esophagus. Patients with Barrett's esophagus may develop dysplasia with an increased risk for esophageal adenocarcinoma. Smoking increases the amount of gastroesophageal reflux predisposing to Barrett's esophagus. Smokers with Barrett's are at increased risk of developing adenocarcinoma. The risk rises with the amount and duration of smoking and does not decrease until 30 years postcessation. Tobacco use is definitely responsible for esophageal squamous cell cancer and adenocarcinoma.

Gastric adenocarcinoma is twice as common in heavy smokers, current smokers, and those that started smoking at a younger age.

Cigarette smoking is a risk factor for the development of colorectal adenomas and colorectal carcinoma, most markedly after 20 years of use. Also, it is associated with increased risk of mortality in both men and women with colon cancer.

The most important environmental risk factor strongly associated with pancreatic cancer is smoking. The risk rises with increased amounts of smoking and decreases to that of the normal population 15 years after cessation of smoking. Pancreatic tumors have been found after long-term consumption of tobacco-specific nitrosamines in drinking water.

In hepatocellular carcinoma, smoking as a risk factor is controversial, but most of the current evidence suggests that it has a minor role.

See Also the Following Articles

Barrett's Esophagus • Colitis, Ulcerative • Duodenal Ulcer • Esophageal Cancer • Gastric Ulcer • Gastroesophageal Reflux Disease (GERD) • *Helicobacter pylori* • Pouchitis

Further Reading

Ashley, M. (1997). Smoking and diseases of the gastrointestinal system: An epidemiological review with special reference to sex differences. *Can. J. Gastroenterol.* **11**, 345–352.

Boyko, E., and Koepsell, T. (1987). Risk of ulcerative colitis among former and current smokers. *N. Engl. J. Med.* **316**, 707–710.

Calkins, B. (1989). A meta-analysis of the role of smoking in inflammatory bowel disease. *Digest. Dis. Sci.* **34**, 1841–1854.

Hensrud, D., and Heimburger, D. (1998). Diet, nutrients, and gastrointestinal cancer. *Gastroenterol. Clin. North Am. Clin. Nutr.* **27**, 325–346.

Kahrilas, P., and Gupta, R. (1990). Mechanisms of acid reflux associated with cigarette smoking. *Gut* **31**, 4.

Kato, I., and Nomura, A. (1992). A prospective study of gastric and duodenal ulcer and its relation to smoking, alcohol, and diet. *Am. J. Epidemiol.* **135**, 521–530.

Kurata, J., and Nogawa, A. (1997). Meta-analysis of risk factors for peptic ulcer: Non-steroidal anti-inflammatory drugs, *Helicobacter pylori*, and smoking. *J. Clin. Gastroenterol.* **24**, 2–17.

Pullan, R., and Rhodes, J. (1994). Transdermal nicotine for active ulcerative colitis. *N. Engl. J. Med.* **330**, 811–815.

United States Department of Health and Human Services (1990). "The Health Benefits of Smoking Cessation." A Report of the Surgeon General DHHS Publication No. (CDC) 90–8416. United States Department of Health and Human Services, Office of Smoking and Health, Rockville, MD.

Solitary Rectal Ulcer Syndrome

MARCUS J. BURNSTEIN* AND ROBERT H. RIDDELL†
St. Michael's Hospital and †Mount Sinai Hospital, Toronto

Delorme procedure Perineal approach to the repair of rectal procidentia in which there is circumferential mucosectomy followed by longitudinal plication of the rectal wall, and mucosal reapproximation.

internal intussusception Rectal intussusception in which the lead point remains above the pelvic floor (also called occult prolapse).

intussusception Invagination or prolapse of one segment of bowel into the lumen of the immediately adjacent segment.

rectal procidentia Circumferential full-thickness intussusception of the rectum in which the lead point of the intussusception descends through the anal canal.

rectopexy Fixation by suture or prosthetic mesh of the mobilized rectum to the presacral fascia (to prevent rectal intussusception).

solitary rectal ulcer syndrome Condition that is clinically characterized by a disturbance of defecation, blood and mucus per rectum, and abnormalities of the rectal wall ranging from erythema to polyp formation to ulceration; histologically characterized by obliteration of the lamina propria by fibromuscular proliferation derived from the muscularis mucosae, thickening of the muscularis mucosae, and sometimes misplaced mucosal glands deep to the muscularis mucosae.

Solitary rectal ulcer syndrome (SRUS) is a spectrum of clinicopathological abnormalities of uncertain etiology. The condition is poorly understood and even the name is misleading—the lesion on the rectal wall may be neither solitary nor ulcerative. The essential features of the syndrome are a disturbance of defecation, the passage of blood and mucus per rectum, and abnormalities of the rectal wall, ranging from erythema to polyp formation and ulceration. There is an association of SRUS with failure of relaxation of the pelvic floor during defecation, with internal intussusception of the rectum, and with overt rectal prolapse. Optimal therapy of SRUS is unclear, and the role of surgery is primarily guided by the presence or absence of either overt or occult (internal intussusception) rectal prolapse. In fact, the understanding of SRUS and the management of this condition may be enhanced by recognizing two discrete entities—SRUS with and without intussusception of the rectal wall.

CLINICAL FEATURES

Solitary rectal ulcer syndrome (SRUS) affects men and woman approximately equally, although some studies have found a female predominance. It is most often seen in young adults, but the condition has presented in children as well as in patients over the age of 60 years. The typical symptoms are straining at evacuation of stool, often for prolonged periods of time, a feeling of incomplete rectal emptying, and the passage of blood and mucus per rectum. Fecal incontinence is rare, but mucus leakage is common. Many patients resort to self-digitation either to aid evacuation or, less commonly, to relieve a sense of obstruction. The percentage of patients who suffer anorectal, perineal, or lower abdominal pain is highly variable. Most series report delays in diagnosis of several years.

The symptoms and signs are closely linked to the presence or absence of overt rectal prolapse, which is present in about 25% of patients. The general physical examination is unremarkable, but anorectal evaluation is always abnormal. On straining, roughly half of the patients will show marked perineal descent. In patients with overt rectal prolapse, the symptoms and physical findings of prolapse may predominate, including fecal incontinence and a low tone, patulous anus. In the absence of overt rectal prolapse, anal tone and squeeze are likely to be normal. Digital rectal examination may reveal induration and thickening of the rectal wall, or a polypoid mass may be felt. The lesion is usually mobile on the underlying muscle.

Sigmoidoscopy is abnormal and may reveal a wide range of findings. The classical ulcer is shallow with a whitish base, a thin erythematous rim, and with normal surrounding mucosa. The ulcers may vary in size, shape, number, and location. The lesions are multiple in about one-third of cases. A polypoid lesion may be present in 25% of the cases, and in some patients, there may only be patchy, granular erythema of the mucosa. The changes of SRUS are most frequently seen anteriorly, but may occur at any point on the rectal wall, even circumferentially. The usual level is between 6 and 10 cm from the anal verge but may occur from 4 to 15 cm. Confluent

 Copyright 2004, Elsevier (USA). All rights reserved.

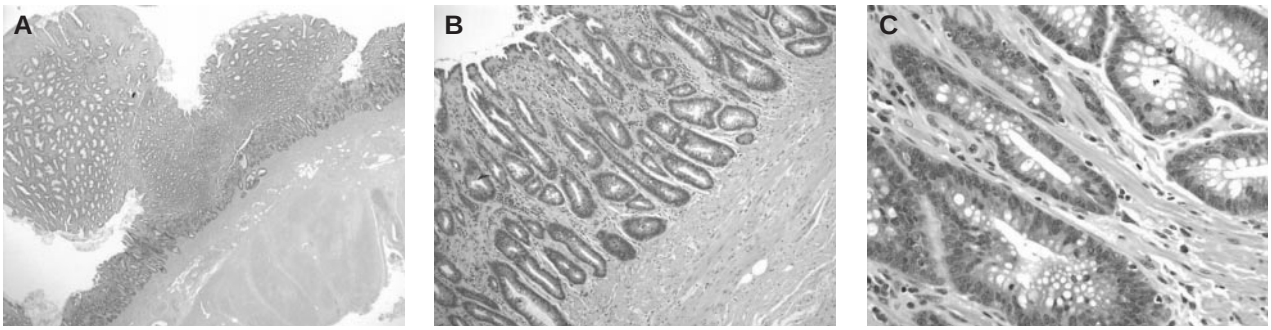


FIGURE 1 (A) Overview of SRUS with a polypoid mucosa. (B) Detail of mucosa, showing a hyaline-like lamina propria with fibromuscular replacement of the lamina propria and a markedly thickened muscularis mucosae. (C) Detail of the loss of inflammatory cells in the lamina propria, with bundles of smooth muscle fibers running parallel to the crypts.

circumferential changes may rarely produce rectal stenosis.

HISTOLOGY

Biopsies must be taken to confirm the diagnosis. These should be taken from the edge of an ulcer or from the hyperemic or polypoid mucosa. The main histologic features are obliteration of the lamina propria by fibromuscular proliferation of the muscularis mucosae, muscle fibers streaming up between the crypts, thickening of the muscularis mucosae (Fig. 1), and misplaced mucosal glands deep to the muscularis mucosae. Similar findings have been described in biopsies of prolapsing mucosa occurring elsewhere in the gastrointestinal tract when prolapse occurs. In the large bowel, these include "ostomy" sites, hemorrhoids and inflammatory cloacogenic polyps (which are virtually identical to SRUS but occur at the dentate line), and around the orifices of diverticula.

Complications that occur are most frequently superficial erosions or ulcers that can bleed (Fig. 2A), ischemia with a typical pseudomembrane and fibrin thrombi in the superficial capillaries, and reactive changes in the underlying epithelium (Fig. 3). If there is hemorrhage into the stalk, there may be misplaced glands in the submucosa similar to that found in the overlying mucosa although without the features of prolapse (colitis cystica profunda), and often accompanied by surrounding lamina propria, hemorrhage, and hemosiderin-laden macrophages. Immediately beneath mucosae with ulceration or erosions, the crypts may have a stellate or serrated appearance similar to that found in hyperplastic polyps (Fig. 2B). Sometimes the reactive changes may be very marked and mimic adenomatous change; proliferation to the surface may occur. The biopsies shown in Fig. 4 are deemed sufficiently adenoma-like that the patient is being followed regularly to ensure that there is no progression, because the natural history of unusual changes such as these is unknown. There is, of course, no reason why adenomas should not occur in SRUS, but typical adenomas appear to be exceedingly rare.

INVESTIGATIONS

The diagnosis of SRUS is made by history, physical examination, endoscopy, and biopsy. Radiologic and physiologic studies are of minimal diagnostic value. However, investigations may help to define the pathogenesis of this condition in an individual patient and

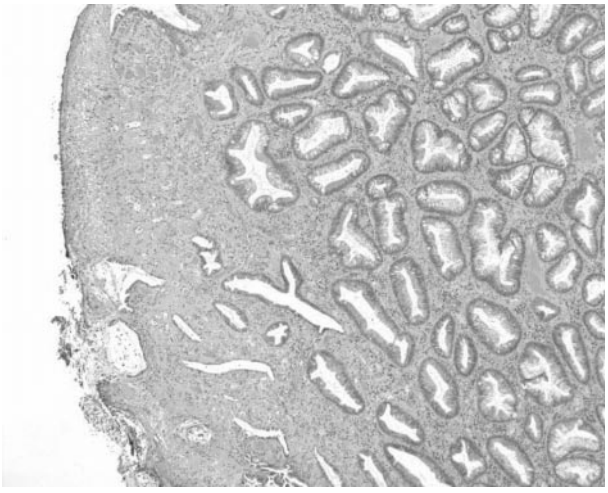


FIGURE 2 Area with superficial ulceration (left). Note that the crypts immediately beneath the ulcerated surface can take on an appearance that mimics that seen in a hyperplastic polyp with serrated crypts, a nonspecific pattern that can occur beneath any ulcerated surface in the large bowel.

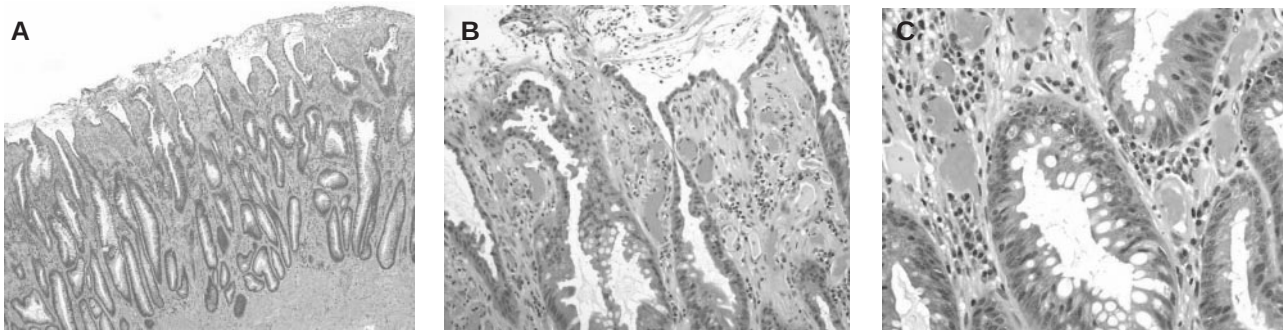


FIGURE 3 (A) Ischemic changes with a superficial pseudomembrane are common. (B) Detail of pseudomembrane and fibrin thrombi in capillaries. (C) Reactive changes in underlying epithelium and fibrin thrombi in capillaries.

may aid in therapeutic planning. The colon should be evaluated to exclude coexisting pathology.

Defecating Proctography

Evacuation proctography (defecography) can demonstrate abnormal rectal mechanics during defecation, including failure of the anorectal angle to open, incomplete emptying of the rectum, perineal descent, and intussusception of the rectal wall, with the intussusception remaining above the pelvic floor (occult rectal prolapse or internal intussusception) or exiting via the anus (overt rectal prolapse).

Physiologic Studies

The results of physiologic studies have been variable. Physiologic parameters correlate with the presence or absence of rectal prolapse and the degree of prolapse. Anal manometry has been reported to be normal, or to demonstrate reduced resting and squeeze pressures, primarily related to whether overt rectal prolapse is present. Pudendal neuropathy, perineal descent, and inability to expel a rectal balloon are all common findings. Electrophysiological studies have demonstrated failure of puborectalis muscle relaxation and inappropriate contraction of the puborectalis muscle during attempts at rectal evacuation.

Endoanal Ultrasound

Characteristic findings are thickening of the muscularis propria and an indistinct transition between the mucosa and muscularis propria. Thickening of the internal sphincter is commonly seen. Less frequently, the submucosa and external sphincter may also appear to be thickened. Ultrasound may also be used to confirm nonrelaxation of the puborectalis.

PATHOGENESIS

The etiology of SRUS is unknown, and it is likely that there are a variety of causes. Two factors appear to predominate, nonrelaxation (or paradoxical contraction) of the puborectalis muscle and rectal prolapse. The evidence suggests that patients with nonrelaxing puborectalis and patients with prolapse are distinct clinical groups and that there is not a progression from "spastic pelvic floor" to rectal intussusception, or from internal intussusception to overt full-thickness prolapse.

Straining against a nonrelaxing pelvic floor may predispose to a cascade of direct trauma to the mucosa against the closed anal canal, mucosal prolapse, mucosal ischemia, and mucosal ulceration. Nonrelaxing puborectalis has been described in one-half to three-fourths of patients with SRUS. The incidence of overt or occult rectal prolapse is highly variable in the literature; a degree of prolapse may be seen in 80% of patients with SRUS.

Direct trauma from self-digitation has been suggested as a causative factor, but this seems very unlikely; lesions are often located beyond the reach of a finger, many patients who self-digitate do not develop the syndrome, and cessation of self-digitation does not lead to mucosal healing.

TREATMENT

Conservative Treatment

Topical therapies, of which many have been tried, are generally unsuccessful. The only effective nonoperative treatments address the underlying defecation disorder. Patients must learn to avoid straining, to restrict the number of visits to the toilet, and to decrease the duration of each visit. A high-fiber diet may be

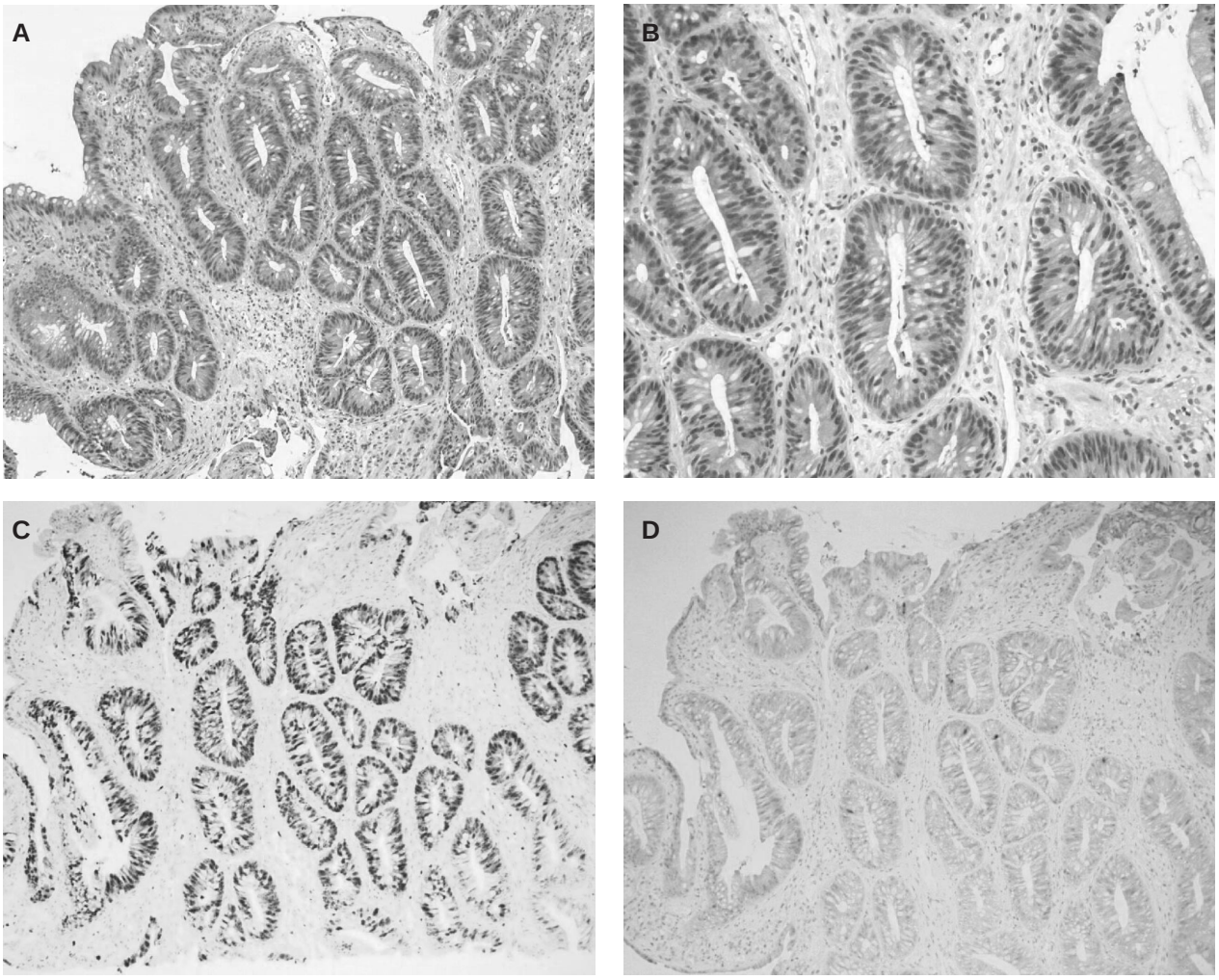


FIGURE 4 (A, B) Marked reactive changes mimicking dysplasia, with stratified nuclei suggesting low-grade dysplasia. (C) Mib-1 (Ki67), a proliferation marker, shows extension of the proliferative zone to the surface. (D) p53 immunostaining is unremarkable, with only occasional nuclei staining in the usual distribution (normal pattern).

helpful. Biofeedback retraining of the defecation mechanism, in combination with patient education, has had variable rates of success, but if the pattern of abnormal defecation can be corrected, most patients will have at least symptomatic improvement if not complete healing of the rectal lesion. The majority of patients can be managed nonoperatively. Biofeedback and counseling may need to be repeated because the benefits appear to deteriorate over time. Biofeedback can be used alone or as an adjunct to operative treatment.

Operative Treatment

SRUS in association with overt rectal prolapse should be treated by operative repair of the prolapse. Patients with overt rectal prolapse are not helped by nonoperative approaches. The optimal therapeutic approach to occult prolapse is less clear, but if conservative treatment fails, then operative correction of the intussusception is reasonable.

The best way to repair rectal prolapse remains an extremely difficult and contentious issue. Many operations have been described and there is no consensus regarding which is best. It is increasingly accepted that no single operation is best in all circumstances. In selecting the procedure for an individual patient, the surgeon must consider the general health and age of the patient, the morbidity, mortality, and recurrence rate for the operation, and the effect of the operation on concomitant symptoms such as fecal incontinence, constipation, and ineffective emptying. Whatever operative approach is taken, it is essential to recognize that

correction of prolapse is not equivalent to symptom control—fecal incontinence, mucus discharge, tenesmus, and straining may all persist despite control of the prolapse.

The main operative approaches are abdominal and perineal. In general, the abdominal operations have lower recurrence rates compared to the perineal operations, but are associated with greater morbidity, including bleeding, infection, bowel obstruction, and sexual dysfunction secondary to injury of pelvic autonomic nerves. Morbidity may be reduced with a laparoscopic approach.

Abdominal Operations

The medically fit patient has conventionally been treated by an abdominal approach. The popular abdominal operations incorporate fixation of the mobilized rectum to the presacral fascia, with or without resection.

Rectopexy The rectum is fully mobilized and as the rectum is retracted in a cephalad direction, the lateral ligaments of the rectum (or lateral rectal tissues) are fixed with sutures to the presacral fascia. The degree to which the rectum should be mobilized anteriorly and laterally is controversial, but there is complete agreement that full posterior mobilization to the tip of the coccyx is important in reducing recurrence. This is a simple procedure with a recurrence rate in the range of 5% in most series. The mobilized rectum may be anchored to the presacral fascia with the use of a prosthetic mesh that partially encircles the rectum. A partial wrap, leaving the anterior one-fourth to one-half of the rectal circumference free, prevents obstruction. The peritoneum is closed over the mesh to avoid adhesion of small bowel loops. The recurrence rate with this approach is also very low, less than 5% in many large series. Unfortunately, rectopexy is often complicated by ongoing problems with evacuation, and this may be especially true in the SRUS population.

Resection The resectional procedures are combination rectopexy–resection, anterior resection, and low anterior resection. When resection of the redundant sigmoid colon is added to rectopexy, prosthetic mesh is avoided because of the risk of infection. The colonic resection is generally extended proximally to eliminate the redundant sigmoid colon. Recurrence is in the range of 5–10%. Resection–rectopexy may improve constipation, but the quality of this evidence is poor. Extending the resection below the peritoneal reflection appears to increase morbidity without decreasing the recurrence rate.

Perineal approaches The principal attraction of the perineal operations is that they are well tolerated, even by frail patients. The perineal operations can be performed under regional or local anesthesia, and these operations are especially attractive when there has been previous pelvic surgery. As with the abdominal operations, full mechanical and antibiotic bowel preparation is used. The operations may be done in the lithotomy or prone jackknife position.

Delorme procedure In this operation, the mucosa is circumferentially stripped off the underlying muscle from 1.5 cm proximal to the dentate line to the tip of the prolapsed rectum; the mucosal tube is dissected until resistance is encountered. The denuded rectal wall is longitudinally plicated with a series of absorbable sutures, the mucosal tube is excised, and the mucosa is reapproximated. The Delorme operation may not be ideal in patients with solitary rectal ulcer syndrome if there is a lot of induration of the rectal wall. Recurrence rates range from 5 to 27%.

Perineal proctosigmoidectomy With the rectum prolapsed, a circumferential incision 2 cm proximal to the dentate line is deepened through the full thickness of the bowel wall. The peritoneum is opened and the rectosigmoid is mobilized until the redundant bowel cannot be pulled down any further. Anterior and posterior plication of the levator muscles may be added at this point. About 2 cm distal to the anus, the inner tube of the rectosigmoid is transected along with its mesentery. An anastomosis is performed 1–2 cm proximal to the dentate line. Morbidity and mortality are low, but are probably slightly higher than with the Delorme procedure. Recurrence rates range from 0 to 60%, with most recent reviews reporting recurrence rates of 5–10%.

With all of these operations, healing of the SRUS is expected in the majority of patients; however, long-term symptomatic improvement is only 50–60%. Patients with SRUS who do not have overt prolapse or internal intussusception are not helped by operations. For some patients, a permanent stoma is ultimately established, but even a stoma does not guarantee symptom control.

SUMMARY

SRUS is an uncommon disorder with distinct clinical, histological, radiographic, and physiologic findings. The pathogenesis is unclear, but probably varies among patients. Correction of an underlying defecation disorder will help many patients. If occult rectal prolapse is present, and if symptoms have persisted after behavior modification, then rectopexy is reasonable. If overt rectal prolapse is present, then this should be repaired.

See Also the Following Articles

Colitis Cystica Profunda and Solitary Rectal Ulcer Syndrome • Intussusception • Rectal Ulcers

Further Reading

Felt-Bersma, R. J. F., and Cuesta, M. A. (2001). Rectal prolapse, rectal intussusception, rectocele, and solitary rectal ulcer syndrome. *Gastroenterol. Clin. North Am.* **30**(1), 199–222.

Lobert, P. F., and Appelman, H. D. (1981). Inflammatory cloacogenic polyp. A unique inflammatory lesion of the anal transitional zone. *Am. J. Surg. Pathol.* **5**(8), 761–766.

Malouf, A. J., Vaizey, C. J., and Kamm, M. A. (2001). Results of behavioral treatment (biofeedback) for solitary rectal ulcer syndrome. *Dis. Colon Rectum* **44**, 72–76.

Sitzler, P. J., Kamm, M. A., Nicholls, R. J., and McKee, R. F. (1998). Long-term clinical outcome of surgery for solitary rectal ulcer syndrome. *Br. J. Surg.* **85**, 1246–1250.

Tendler, D. A., Aboudola, S., Zacks, J. F., O'Brien, M. J., and Kelly, C. P. (2002). Prolapsing mucosal polyps: An underrecognized form of colonic polyp—A clinicopathological study of 15 cases. *Am. J. Gastroenterol.* **97**(2), 370–376.

Tjandra, J. J., Fazio, V. W., Petras, R. E., Lavery, I. C., Oakley, J. R., Milsom, J. W., and Church, J. M. (1993). Clinical and pathologic factors associated with delayed diagnosis in solitary rectal ulcer syndrome. *Dis. Colon Rectum* **36**, 146–153.

Vaizey, C. J., Roy, A. J., and Kamm, M. A. (1997). Prospective evaluation of the treatment of solitary rectal ulcer syndrome with biofeedback. *Gut* **41**, 817–820.

Vaizey, C. J., van den Bogaerde, J. B., Emmanuel, A. V., Talbot, I. C., Nicholls, R. J., and Kamm, M. A. (1998). Solitary rectal ulcer syndrome. *Br. J. Surg.* **85**, 1617–1623.

van Tets, W. F., and Kuijpers, J. H. C. (1995). Internal rectal intussusception—Fact or fancy? *Dis. Colon Rectum* **38**, 1080–1083.

Somatostatin

VICENTE MARTINEZ*,† AND YVETTE TACHÉ*

**Veterans Administration Greater Los Angeles Healthcare System and University of California, Los Angeles, and †Astra Zeneca R&D, Mölndal, Switzerland*

enterochromaffin-like cell Synthesizes, stores, and secretes histamine as a paracrine regulator in the gastric mucosa.

gastrin Gastrointestinal hormone produced by G cells in the gastric antral mucosa; stimulates gastric acid secretion.

somatostatin Tetradecapeptide that is an inhibitor of growth hormone release.

Somatostatin (SST), a tetradecapeptide isolated in the early 1970s from sheep hypothalami, was identified as an inhibitor of growth hormone release. SST has been found in virtually every organ and is recognized as a peptide that exerts an inhibitory action on a variety of physiological functions, acting either as a classical endocrine hormone, a local (paracrine) regulatory factor, or a neurotransmitter. In mammals, the gastrointestinal tract, the brain, and the pancreas contain the highest amounts of SST. In the rat, the gastrointestinal tract accounts for ~65% of the total body SST, the brain for ~25%, the pancreas for ~5%, and the remaining organs for ~5%. In the brain, SST functions as a neurotransmitter, affecting cognitive, locomotor, sensory, and autonomic functions. In the gastrointestinal tract, SST displays a wide array of physiological functions and pharmacological effects, mostly inhibitory, of which the suppression of gastric acid secretion is a landmark response. In view of these properties, different analogues have been developed for the treatment of various human diseases. During the past few years, SST receptor subtypes have been cloned and pharmacologically characterized. Selective SST agonists and antagonists and genetic animal models with selective deletion of receptor subtype have been developed. These new tools are of critical importance to the understanding of SST signaling pathways in the regulation of gastrointestinal function.

SOMATOSTATIN GENE AND GENE PRODUCTS

Somatostatin-14 and -28

In mammals, SST derives from the expression of a single gene that, in humans, maps to the long arm of chromosome 3. The amino acid sequence of SST has

 Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Felt-Bersma, R. J. F., and Cuesta, M. A. (2001). Rectal prolapse, rectal intussusception, rectocele, and solitary rectal ulcer syndrome. *Gastroenterol. Clin. North Am.* **30**(1), 199–222.

Lobert, P. F., and Appelman, H. D. (1981). Inflammatory cloacogenic polyp. A unique inflammatory lesion of the anal transitional zone. *Am. J. Surg. Pathol.* **5**(8), 761–766.

Malouf, A. J., Vaizey, C. J., and Kamm, M. A. (2001). Results of behavioral treatment (biofeedback) for solitary rectal ulcer syndrome. *Dis. Colon Rectum* **44**, 72–76.

Sitzler, P. J., Kamm, M. A., Nicholls, R. J., and McKee, R. F. (1998). Long-term clinical outcome of surgery for solitary rectal ulcer syndrome. *Br. J. Surg.* **85**, 1246–1250.

Tendler, D. A., Aboudola, S., Zacks, J. F., O'Brien, M. J., and Kelly, C. P. (2002). Prolapsing mucosal polyps: An underrecognized form of colonic polyp—A clinicopathological study of 15 cases. *Am. J. Gastroenterol.* **97**(2), 370–376.

Tjandra, J. J., Fazio, V. W., Petras, R. E., Lavery, I. C., Oakley, J. R., Milsom, J. W., and Church, J. M. (1993). Clinical and pathologic factors associated with delayed diagnosis in solitary rectal ulcer syndrome. *Dis. Colon Rectum* **36**, 146–153.

Vaizey, C. J., Roy, A. J., and Kamm, M. A. (1997). Prospective evaluation of the treatment of solitary rectal ulcer syndrome with biofeedback. *Gut* **41**, 817–820.

Vaizey, C. J., van den Bogaerde, J. B., Emmanuel, A. V., Talbot, I. C., Nicholls, R. J., and Kamm, M. A. (1998). Solitary rectal ulcer syndrome. *Br. J. Surg.* **85**, 1617–1623.

van Tets, W. F., and Kuijpers, J. H. C. (1995). Internal rectal intussusception—Fact or fancy? *Dis. Colon Rectum* **38**, 1080–1083.

Somatostatin

VICENTE MARTINEZ*,† AND YVETTE TACHÉ*

**Veterans Administration Greater Los Angeles Healthcare System and University of California, Los Angeles, and †Astra Zeneca R&D, Mölndal, Switzerland*

enterochromaffin-like cell Synthesizes, stores, and secretes histamine as a paracrine regulator in the gastric mucosa.

gastrin Gastrointestinal hormone produced by G cells in the gastric antral mucosa; stimulates gastric acid secretion.

somatostatin Tetradecapeptide that is an inhibitor of growth hormone release.

Somatostatin (SST), a tetradecapeptide isolated in the early 1970s from sheep hypothalami, was identified as an inhibitor of growth hormone release. SST has been found in virtually every organ and is recognized as a peptide that exerts an inhibitory action on a variety of physiological functions, acting either as a classical endocrine hormone, a local (paracrine) regulatory factor, or a neurotransmitter. In mammals, the gastrointestinal tract, the brain, and the pancreas contain the highest amounts of SST. In the rat, the gastrointestinal tract accounts for ~65% of the total body SST, the brain for ~25%, the pancreas for ~5%, and the remaining organs for ~5%. In the brain, SST functions as a neurotransmitter, affecting cognitive, locomotor, sensory, and autonomic functions. In the gastrointestinal tract, SST displays a wide array of physiological functions and pharmacological effects, mostly inhibitory, of which the suppression of gastric acid secretion is a landmark response. In view of these properties, different analogues have been developed for the treatment of various human diseases. During the past few years, SST receptor subtypes have been cloned and pharmacologically characterized. Selective SST agonists and antagonists and genetic animal models with selective deletion of receptor subtype have been developed. These new tools are of critical importance to the understanding of SST signaling pathways in the regulation of gastrointestinal function.

SOMATOSTATIN GENE AND GENE PRODUCTS

Somatostatin-14 and -28

In mammals, SST derives from the expression of a single gene that, in humans, maps to the long arm of chromosome 3. The amino acid sequence of SST has

 Copyright 2004, Elsevier (USA). All rights reserved.

been determined from multiple tissues and has been shown to have remarkable phylogenetic conservation across species. SST is synthesized as prepro-SST and is posttranslationally processed into two different molecular forms, SST-14 and SST-28, with 14 and 28 amino acid residues, respectively. In fish, two separate genes encoding for SST peptides have been identified. One corresponds to the mammalian gene and gives rise only to SST-14. The second generates extended forms of SST, namely, anglerfish-28 (homologue of mammalian SST-28) and catfish-22. The SST-14 sequence is totally conserved, whereas there is only a 40–66% homology between SST-28 and its fish counterparts.

In mammals, SST-28 contains the SST-14 moiety, responsible for the biological activity, at its carboxyl-terminal end. The processing machinery required for the synthesis is present in various tissues with different relative activities. This leads to different tissue-specific types of synthesis and release of SST-14 and -28. SST-14 is the predominant molecular form in the stomach, duodenum, colon, and pancreas. In contrast, in the small intestine, over 50% of the content corresponds to SST-28.

Cortistatins Are Somatostatin-Like Peptides

A second SST-like gene, called cortistatin, has been cloned in humans and rats. This gene gives rise to two cleavage products—human cortistatin-17 and the rat homologue cortistatin-14, and human and rat cortistatin-29. Unlike the broad tissue expression of the SST gene, expression of the corticostatin gene is restricted to the cerebral cortex.

TISSUE DISTRIBUTION OF SOMATOSTATIN

SST is distributed throughout the gastrointestinal (GI) tract, being present in specialized endocrine cells and neurons. In mammals, SST is found mainly in the mucosal layer, namely, in endocrine cells termed D cells, which contain 90% of SST in the stomach and duodenum. The remaining SST is present in neuronal structures contributing to both the intrinsic and extrinsic innervation of the gut.

Endocrine D Cells Are the Main Source of Gastric Somatostatin

In the stomach, D cells are localized both in the antral and the fundic mucosa. These cells have a characteristic morphology, with cytoplasmatic processes that terminate in close proximity to the target cells. Such morphology supports the concept of a paracrine or local regulatory action of SST in the gastric mucosa. In the fundic mucosa, D cells are of a "closed type," without luminal contact, whereas in the antrum they are of an "open type," with their apical membrane exposed to the gastric lumen. This organization correlates with functional responses to different stimuli.

Endocrine D cells can also be found in the intestine and pancreas. Intestinal D cells are flask-shaped cells with their apical membranes open to the lumen, and, therefore, respond to luminal stimuli. In the pancreas, endocrine D cells are a functional component of the islets, participating in the control of the endocrine functions. Table I summarizes the main endocrine sources of gastrointestinal SST in the context of their regulation and biological role.

Intestinal Somatostatin Is Mainly of Neural Origin

Whereas gastric SST is mainly of endocrine origin, intestinal SST is mainly present in neurons. SST immunoreactivity is detected in the submucosa and muscle layers of the entire gut, with the lowest levels observed in the colon. The majority of the neuronal SST-like immunoreactivity is present in the intrinsic innervation of the gut. For instance, about 20% of the cell bodies in the submucous plexus of the guinea pig small intestine contain SST-like immunoreactivity, corresponding to cholinergic secretomotor neurons. Pancreatic ganglia and nerves also contain significant amounts of SST.

REGULATION OF SOMATOSTATIN SYNTHESIS AND RELEASE

Anatomical and biochemical studies support the notion that SST is secreted mostly in a paracrine/neurocrine fashion, and therefore circulating levels are relatively low and may not reflect release at local sites. In addition, to avoid spreading of actions, the peptide half-life is relatively short, about 3 minutes in circulation. All these factors make it difficult to study SST release and regulation *in vivo*. Consequently, isolated preparations have been widely used to assess SST release.

Intraluminal Stimuli Influence Gastric Somatostatin Release

Intraluminal factors associated with the diet, food components, and low pH are the main stimulants of SST release from open-type endocrine D cells in the stomach and intestine. For instance, luminal acid is the main stimulant of gastric SST release from antral D cells.

TABLE I Sources and Biological Actions of SST of Endocrine Origin in the Gastrointestinal Tract[a]

Localization	Cellular type	Releasing stimuli	Targets	Biological actions	Mechanism of action
Fundic mucosa	Closed	Neural input (ACh, GRP), endocrine (gastrin, CCK)	ECL cells, parietal cells	Inhibition of histamine release, direct inhibition of secretion	Paracrine
Antral mucosa	Open	Luminal (nutrients, pH), sensory neurons	G cells	Inhibition of gastrin synthesis and release	Paracrine
Intestinal epithelium	Flasklike, open	Luminal (nutrients)	Various cell types	Inhibition of neuroendocrine secretion, modulation of motility, inhibition of intestinal transport, inhibition of splanchnic blood flow, inhibition of tissue growth/proliferation, modulation of food intake	Paracrine Endocrine Neurocrine
Pancreatic islets	Closed	Endocrine, circulating nutrients (glucose)	Other islet cell types	Inhibition of insulin and glucagon release	Neurocrine Endocrine

[a] Abbreviations: ACh, acetylcholine; GRP, gastrin-releasing peptide; CCK, cholecystokinin; ECL, enterochromaffin-like.

Similarly, the presence of nutrients and acid in the duodenum stimulates the release of gastric SST, an effect probably mediated through the release of intestinal enterogastrones. SST release through these mechanisms is of physiological importance for the feedback regulation of gastric acid secretion. Physiological postprandial changes of SST levels in the peripheral circulation or portal blood originate from gastric and intestinal SST release. Other sources do not contribute to the peripheral plasma SST levels.

The Autonomic Nervous System Differentially Affects Somatostatin Release

The autonomic nervous system also exerts an important role in the regulation of SST release from the gut. As in other systems, the sympathetic and parasympathetic components have opposite effects on SST release. Adrenergic agonists, acting through β-adrenergic receptors, stimulate gastric SST release. In contrast, muscarinic cholinergic agents inhibit SST release. However, the existence of vagal nonadrenergic stimulatory pathways for SST release has also been described in some species. These actions explain, at least in part, the inhibitory and stimulatory effects that the sympathetic and parasympathetic nervous systems have, respectively, on acid secretion.

Gut Peptides Modulate Somatostatin Release

In addition to the luminal and autonomic control, several regulatory peptides, acting as classical hormones or as neuropeptides, modulate gastrointestinal SST release. These can be categorized as stimulatory and inhibitory agents. The most important stimulants of SST release are gastrin and cholecystokinin (CCK). Other peptides, such as peptides of the secretin family, bombesin/gastrin-releasing peptide (GRP), glucose-dependent insulinotropic peptide, oxyntomodulin, or glucagon-like peptide-1 (GLP-1), which consistently inhibit gastric acid secretion, may function by stimulating SST release. On the other hand, substance P, opioids, insulin, and glucagon are potent inhibitors of SST release. In addition, *in vitro* studies suggest that fundic D cells are under negative feedback control by their own product through an autocrine mechanism of action. However, so far, no SST receptors have been localized on D cells.

SOMATOSTATIN RECEPTORS

SST acts on its target cells through high-affinity plasma membrane receptors (Table II). To date, five different SST receptor (SSTR) subtypes, SSTR1−5, belonging to the superfamily of G protein-coupled receptors with seven transmembrane domains, have been cloned and pharmacologically characterized in different systems. In addition, two splice variants of SSTR2, SSTR2a and SSTR2b, which differ in length and composition of their intracellular carboxy termini, have been isolated and cloned in the mouse and the rat. Overall, there is a 39−57% sequence identity among the different receptor subtypes. The five receptors also have a remarkable degree of structural

TABLE II Characteristics of the Cloned Human SST Receptor Subtypes

Characteristic[a]	SSTR1	SSTR2a	SSTR3	SSTR4	SSTR5
Chromosomal localization	14q13	17q24	22q13.1	20p11.2	16p13.3
Amino acids	391	369	418	388	363
G protein coupling	+	+	+	+	+
Effector coupling					
Adenylyl cyclase	↓	↓	↓	↓	↓
Tyrosine phosphatase	↑	↑	↑	↑	↑
MAP kinase	↑	↓	↓↑	↑	↓
K^+ channels		↑	↑	↑	↑
Ca^{2+} channels	↓	↓			
Na^+/H^+ exchanger	↑				
AMPA/kainate glutamate channels	↑	↓			
Phospholipase C/IP_3		↑			↓↑
Phospholipase A2				↑	
Tissue distribution	Brain, pituitary, GI tract, kidney	Brain, pituitary, GI tract, kidney	Brain, pituitary, GI tract, kidney	Brain, pituitary, GI tract, lungs, placenta	Brain, pituitary, GI tract

[a] Abbreviations: MAP, mitogen-activated protein; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid.

conservation across species, with between 80 and 99% homology among the human, rat, and mouse isoforms. The nearest relatives of SST receptors are the opioid receptors that share about 37% sequence similarity to the mouse SSTR1.

Early studies, based on receptor autoradiography using nonselective ligands, showed a high density of SST binding sites all along the GI mucosa as well as in neural structures. In recent years, the anatomical and cellular distribution of SST receptor gene expression has been established in rodent and human tissues as well as in different tumors and cell lines. Lately, the development of SST receptor subtype-selective antibodies has allowed the direct localization of the SST receptor at the protein level by immunohistochemistry. These studies reveal an intricate pattern of receptor expression throughout the central nervous system and the periphery, with an overlapping but characteristic pattern that is receptor subtype selective and tissue and species specific.

G Protein Coupling and Signal Transduction of Somatostatin Receptors

The activation of SST receptors by SST ligands elicits cellular responses through G protein-linked modulation of multiple second-messenger systems, including adenylyl cyclase, Ca^{2+} and K^+ ion channels, the Na^+/H^+ antiporter, guanylate cyclase, phospholipase C, phospholipase A2, mitogen-activated protein (MAP) kinase, and serine, threonine, and phosphotyrosyl protein phosphatase. These pathways show only partial receptor selectivity (Table II). In general, all five receptor subtypes, and especially the human isoforms, are potent inhibitors of adenylyl cyclase and cyclic adenosine monophosphate (cAMP) formation via pertussis toxin-sensitive G proteins. Subtypes 1, 2, 3, and 5 display acute desensitization of adenylyl cyclase coupling.

In the presence of the agonist, receptor subtypes 2, 3, 4, and 5 undergo rapid internalization. Receptor subtype 1 fails to be internalized but is instead up-regulated at the membrane in response to continued agonist exposure.

The SSTR2 Is the Receptor Subtype Predominant in the Stomach

The presence of the five SST receptor subtypes in the stomach, with a predominance of SSTR2 receptor mRNA, has been established by ribonuclease protection assays and reverse transcription polymerase chain reaction (RT-PCR). At a cellular level, *in situ* hybridization studies demonstrate the presence of SSTR2 mRNA-containing cells in the gastric mucosa and submucosa as well as in the myenteric plexus and the external muscle layer. In addition, RT-PCR techniques show that the mRNAs for both SSTR2a and SSTR2b isoforms are present in the stomach. Immunohistochemical studies using antipeptide antibodies specific for the SSTR2a and SSTR2b receptors show that a small population of enterochromaffin-like (ECL) cells in the stomach expresses SSTR2a receptors. On the other hand, the SSTR2b receptor is mainly localized in parietal cells and in high concentration in ECL cells. So far, no SST receptors have been found in other cellular types of the gastric mucosa.

SSTR2a has also been localized in nerve fibers that arise from the nervous plexuses and innervate the mucosa with fibers running parallel to the epithelia. Many of these fibers are in close proximity to D cells of the gastric mucosa. This distribution strongly suggests that SSTR2a receptors might mediate SST effects in the GI tract via neuronal and paracrine pathways.

All SSTR Receptor Subtypes Are Present in the Intestine

In situ hybridization studies have identified all five receptor subtypes in the rat duodenum, jejunum, ileum, and colon. All receptor subtypes are present in the epithelium, submucosal plexus, and muscle layers, whereas only SSTR1−3 are found in the myenteric plexus. Immunohistochemical studies have demonstrated the existence of SSTR2a receptors in the myenteric and submucosal plexuses, with a predominance in the submucosal plexus, and in fibers innervating the muscle, mucosa, and vasculature. The SSTR2a receptors are present in the neuronal soma and processes as well as in axon terminals, suggesting both pre- and postsynaptic effects of SST in the gut. Moreover, this distribution is consistent with SSTR2a receptors being expressed by functionally distinct enteric neurons, indicating that SST plays a significant role in the regulation of intestinal motor and secretory functions. In addition, SSTR2a receptors are present in interstitial cells of Cajal, indicative of other target cells for SST to regulate smooth muscle activity.

SSTR2 and SSTR5 Are the Main Receptor Subtypes in the Pancreas

Early autoradiographic studies established the presence of SST binding sites in alpha and beta cells from pancreatic islets. Molecular studies have shown that rat islets contain mRNA for all five receptor subtypes, although this does not imply the presence of functional receptors. In humans, immunohistochemical studies have revealed a predominant expression of SSTR2 and SSTR5, with a specific cellular distribution: beta cells are rich in SSTR5, alpha cells contain mostly SSTR2, and delta cells contain SSTR5.

Neuroendocrine Enteropancreatic Tumors Express Somatostatin Receptors

In addition to normal tissues, the majority of human tumors, either benign or malignant, express high concentrations of SST receptors, frequently featuring more than one isotype. SST receptors are especially present in almost all neuroendocrine enteropancreatic tumors [gastrinoma, vasoactive intestinal peptide-producing tumor (VIPoma), carcinoid, insulinoma, glucagonoma, SSToma, pancreatic polypeptide-producing tumor (PPoma), and growth hormone releasing factor-producing tumor (GRFoma)]. The expression appears to be tumor specific, but in general the pattern is SSTR2 > SSTR1 > SSTR3 > SSTR4 > SSTR5. However, studies carried out so far are mainly based on mRNA expression, which may not necessarily reflect functional receptor levels. SSTR2 predominance is found in about 90% of carcinoid tumors and 80% of endocrine pancreatic tumors.

BIOLOGICAL ACTIONS OF SOMATOSTATIN IN THE GASTROINTESTINAL TRACT: RECEPTOR SUBTYPE SELECTIVITY

Somatostatin Is the Main Inhibitor of Gastric Acid Secretion

SST is a potent inhibitor of gastric acid secretion and probably the main inhibitory regulator during the cephalic, gastric, and intestinal phases of secretion. SST administered peripherally inhibits gastric acid secretion at doses producing plasma increments of the peptide similar to those observed during postprandial states in dogs and humans. The SST-dependent regulation of gastric acid secretion involves gastric SST produced by fundic and antral endocrine D cells, acting locally through paracrine, rather than endocrine, mechanisms. Fundic D cells are stimulated by gastrin and by acetylcholine, as well as by other neuropeptides; they are closely associated to parietal cells and ECL cells through cytoplasmic extensions. Antral D cells respond to changes in gastric luminal acidity and are closely associated, by way of cytoplasmic extensions, with their target cells, i.e., G cells, resulting in the local release of SST, which, in turn, inhibits gastrin synthesis and release from G cells.

During both the cephalic and gastric phases of a meal, gastric acid secretion is stimulated mainly through the release of gastrin. Attenuation of excessive acid secretion can be attributed to the local release of SST, which, in turn, regulates the activity of G cells, ECL cells, and parietal cells. Physiological effects of SST result from a direct inhibition of parietal cell secretion and indirectly by inhibition of histamine release from ECL cells and gastrin release from G cells. The role of SST in the control of basal (interdigestive) gastric acid secretion is less clear. *In vitro* studies suggest that SST

tonically inhibits acid secretion; however, this has not been fully demonstrated in *in vivo* conditions.

SSTR2 Receptors Mediate the Effects of Somatostatin on Gastric Acid Secretion

Several *in vivo* and *in vitro* studies in rats, mice, dogs, and humans using different relatively selective receptor peptide analogues of SST have established that the SSTR2 receptor subtype mediates the inhibitory effects of SST on gastric acid secretion (Table III). Further functional evidence for the involvement of SSTR2 receptors on SST-dependent control of gastric acid secretion comes from the use of the selective SSTR2 antagonist PRL-2903. Infused intravenously, PRL-2903 antagonizes SST-induced inhibition of acid secretion. Conclusive evidence for the involvement of SSTR2 receptors in the regulation of gastric acid secretion also comes from functional studies in mice with targeted disruption of the SSTR2a receptor gene. Because the SSTR2b variant is a spliced product of the SSTR2a variant, these animals lack both forms of the receptor, therefore none of the splice variants is expressed and no information regarding the SSTR2 isoform mediating SST actions on acid secretion can be derived.

SSTR2 Receptors Mediate Somatostatin-Dependent Inhibition of Gastrin and Histamine Release

SST-induced inhibition of gastrin and histamine secretion is also mediated by the activation of SSTR2 receptors located, respectively, in G cells and ECL cells. Effects on gastrin release are further supported by studies with the SSTR2 antagonist, PRL-2903, showing that the antagonist increases plasma levels of gastrin.

D Cells as an Integrator of Inhibitory Responses in the Stomach

Several gut peptides [i.e., amylin, adrenomedullin, calcitonin gene-related peptide (CGRP), bombesin, glucose-dependent insulinotropic polypeptide, GLP-1, or pituitary adenylate cyclase-activating polypeptide (PACAP)] inhibit gastric acid secretion through the release of SST or have SST-dependent mechanisms of action. In addition, receptors for several of these neuropeptides have been localized in gastric D cells. These functional and morphological observations suggest that gastric D cells may function as a common target for a variety of gut peptides. Activation of D cell by these peptides translates into the release of SST, and in turn the activation of SSTR2 receptors on ECL and parietal cells leading to an inhibition of acid output.

Somatostatin and Small Intestinal Function

SST inhibits small bowel motility and intestinal absorption of nutrients. In the small intestine, neuronal SST is released locally by intestinal distension and participates, as a neuromodulator, in the coordination of the descending relaxation of the peristaltic reflex. Pharmacological studies and the use of SSTR2 knockout mice have indicated that this receptor subtype is involved in peristalsis, however the implication of other receptor subtypes cannot be ruled out.

Somatostatin Regulates Pancreatic Function through SSTR2 and SSTR5 Receptors

SST affects both exocrine and endocrine pancreatic secretion, acting through paracrine and endocrine mechanisms. Pharmacological studies in normal and SSTR2 knockout mice have shown that pancreatic actions of SST are mediated mainly through SSTR2 and SSTR5 receptors.

Somatostatin Inhibits Pancreatic Amylase Release through SSTR5 Receptors

Modulation of exocrine pancreatic function is associated with changes in arterial SST of gastric or intestinal origin (endocrine action). Physiological changes in plasma SST levels inhibit pancreatic enzyme and bicarbonate secretion. Pharmacological studies using receptor-selective analogues of SST have identified the SSTR5 as the main receptor subtype inhibiting amylase release.

TABLE III Receptor Selectivity of SST Actions in the Gastrointestinal Tract

Biological activity	Receptor subtype	Localization
Inhibition of gastric acid secretion	SSTR2	Parietal cells
Inhibition of histamine release	SSTR2	Enterochromaffin-like cells
Modulation of the peristaltic reflex	SSTR2/other undetermined	Myenteric plexus
Inhibition of pancreatic amylase release	SSTR5	Acinar cells
Inhibition of insulin release	SSTR5	Pancreatic beta cells
Inhibition of glucagon release	SSTR2	Pancreatic alpha cells
Inhibition of cell proliferation	SSTR1, 2, 3, 4, and 5	Tumor cells

Supporting this observation, binding studies have also identified the presence of SSTR5 receptors on pancreatic acinar cells.

Somatostatin Inhibits Pancreatic Endocrine Secretion through SSTR2 and SSTR5 Receptors

SST from D cells in the pancreatic islets acts locally (paracrine) on alpha, beta, and PP cells, inhibiting the release of glucagons, insulin, and PP, respectively. This interaction provides a basic intraislet control of insulin release. Pharmacological studies using receptor-selective analogues of SST in normal and SSTR2 and SSTR5 knockout mice have shown that islet regulation is receptor and hormone specific. Stimulation of SSTR5 receptors located on beta cells is the main pathway to inhibit insulin secretion, whereas the inhibition of glucagon release from alpha cells depends on the stimulation of SSTR2 receptors.

PATHOPHYSIOLOGICAL ASPECTS OF SST

A clear pathophysiological role of SST has been demonstrated only in cases with SST-producing tumors (SSTomas). In these patients, the clinical signs are associated with an overproduction and high levels of plasma SST. The resulting syndrome is characterized by diabetes, gallstones, malabsorption, and pancreatic insufficiency.

Alterations in the responsiveness and number of gastric D cells might be related to the hypersecretory states observed in patients with nonatrophic gastritis associated with *Helicobacter pylori* infection and in patients with duodenal or gastric ulcers. Under these conditions, D cell responsiveness to gastrin and other stimulants will be reduced. This will lead to lower levels of SST, which, in turn, results in an increased gastrin release and gastric acid output. However, although these mechanisms have been established in several animal models, their importance for human pathophysiology remains to be elucidated.

THERAPEUTIC USE OF SST IN GASTROENTEROLOGICAL DISORDERS

SST and its stable analogue octreotide have been consistently used in the treatment of esophageal variceal bleeding, pathological states of the exocrine pancreas, and neuroendocrine enteropancreatic tumors. Gut neuroendocrine tumors contain abundant SST receptors, and this has been used as a diagnostic tool for tumor localization. In addition, receptors have been demonstrated to be functional in most cases and either SST or its analogue, octreotide, has been used as symptomatic therapy. The mechanisms for the symptomatic improvement are due to SSTR-mediated inhibition of mediators release that leads to the major associated symptoms. In addition, SST and octreotide seem to have an inhibitory effect on cell growth and proliferation, reducing tumor size. So far, all receptor subtypes have been shown to induce cell cycle arrest in different tumor cell lines, although through different mechanisms (Table III).

SST and the analogue octreotide reduce intestinal secretion and have been used to treat diarrhea. SST was originally shown to be effective in treating diarrhea caused by neuroendocrine tumors that produced VIP (VIPoma). In this condition, SST reduced both the production of VIP and its effects on the intestine. It has since been recognized that SST analogues are also effective in treating other diarrheal conditions.

See Also the Following Articles

Anti-Diarrheal Drugs • Cholecystokinin (CCK) • Gastric Acid Secretion • Gastrin • Growth Hormone • Histamine • Pancreatic Enzyme Secretion (Physiology) • Portal Hypertension and Esophageal Varices • Variceal Bleeding • Vasoactive Intestinal Peptide (VIP)

Further Reading

Abdu, F., Hicks, G. A., Hennig, G., Allen, J. P., and Grundy, D. (2002). Somatostatin sst2 receptors inhibit peristalsis in the rat and mouse jejunum. *Am. J. Physiol.* **282**, G624–G633.

Arnold, R., Wied, M., and Behr, T. H. (2002). Somatostatin analogues in the treatment of endocrine tumors of the gastrointestinal tract. *Expert Opin. Pharmacother.* **3**, 643–656.

Chiba, T., and Yamada, T. (1994). Gut somatostatin. *In* "Gut Peptides" (J. H. Walsh and G. J. Dockray, eds.), pp. 123–145. Raven Press, New York.

Hofland, L. J., and Lamberts, S. W. (2001). Somatostatin receptor subtype expression in human tumors. *Ann. Oncol.* **12**(Suppl. 2), S31–S36.

Lloyd, K. C. K., and Walsh, J. (1994). Gastric secretion. *In* "Gut Peptides" (J. H. Walsh and G. J. Dockray, eds.), pp. 633–645. Raven Press, New York.

Martinez, V. (2002). Somatostatin receptors and the control of gastric acid secretion. *In* "Gut–Brain Peptides in the New Millennium" (Y. Taché, Y. Goto, G. Ohning, and T. Yamada, eds.), pp. 167–178. CURE Foundation, Los Angeles.

Martinez, V., Curi, A. P., Torkian, B., Schaeffer, J. M., Wilkinson, H. A., Walsh, J. H., and Taché, Y. (1998). High basal gastric acid secretion in somatostatin receptor subtype 2 knockout mice. *Gastroenterology* **114**, 1125–1132.

Patel, Y. C. (1999). Somatostatin and its receptor family. *Front. Neuroendocrinol.* **20**, 157–198.

Piqueras, L., Taché, Y., and Martinez, V. (2003). Somatostatin receptor type 2 mediates bombesin-induced inhibition of gastric acid secretion in mice. *J. Physiol. (Lond.)* (in press).

Schusdziarra, V. (1996). Somatostatin: Biological actions and pathophysiology. *In* "Octreotide: From Basic Science to Clinical Medicine" (C. Scarpignato, ed.), pp. 35–53. Karger, Basel.

Sternini, C., Wong, H., Wu, S. V., de Giorgio, R., Yang, M., Reeve, J., Jr., Brecha, N. C., and Walsh, J. H. (1997). Somatostatin 2A receptor is expressed by enteric neurons, and by interstitial cells of Cajal and enterochromaffin-like cells of the gastrointestinal tract. *J. Comp. Neurol.* **386**, 396–408.

Strowski, M. Z., Kohler, M., Chen, H. Y., Trumbauer, M. E., Li, Z., Szalkowski, D., Gopal-Truter, S., Fisher, J. K., Schaeffer, J. M., Blake A. D., Zhang, B. B., and Wilkinson, H. A. (2003). Somatostatin receptor subtype 5 regulates insulin secretion and glucose homeostasis. *Mol. Endocrinol.* **17**, 93–106.

Strowski, M. Z., Parmar, R. M., Blake, A. D., and Schaeffer, J. M. (2000). Somatostatin inhibits insulin and glucagon secretion via two receptor subtypes: An *in vitro* study of pancreatic islets from somatostatin receptor 2 knockout mice. *Endocrinology* **141**, 111–117.

Vuyyuru, L., Harrington, L., Arimura, A. S., and Schubert, M. L. (1997). Reciprocal inhibitory paracrine pathways link histamine and somatostatin secretion in the fundus of the stomach. *Am. J. Physiol.* **273**, G106–G111.

Sphincter of Oddi Dysfunction

JAMES WATKINS AND STUART SHERMAN
Indiana University Medical Center, Indianapolis

ampullary stenosis Narrowing of the ampulla, often causing obstruction.

botulinum toxin Protein produced by *Clostridium botulinum* bacteria, used therapeutically to paralyze or weaken the muscle into which it is injected.

endoscopic retrograde cholangiopancreaticography Fiber-optic endoscope insertion into the duodenum and dye injection via the ampulla of Vater, to visualize the biliary and pancreatic ducts.

endoscopic sphincterotomy Cutting of the sphincter via an endoscopic approach.

postcholecystectomy pain Painful attacks that occur after removal of the gallbladder.

quantitative hepatobiliary scintigraphy Nuclear medicine study that assesses bile flow through the biliary tract.

recurrent pancreatitis Recurrent inflammation of the pancreas.

Rome II criteria Consensus criteria developed to diagnose sphincter of Oddi dysfunction.

sphincter ablation Surgical removal of the sphincter.

sphincter of Oddi dysfunction Abnormality in the contractility of the sphincter of Oddi.

sphincter of Oddi manometry Manometric studies of the sphincter of Oddi.

Since its original description by Rugero Oddi in 1887, the sphincter of Oddi has been the subject of much study and controversy. Its very existence as a distinct anatomic or physiologic entity has been disputed. Hence, it is not surprising that the clinical syndrome of sphincter of Oddi dysfunction and its therapy are controversial areas. Nevertheless, this dysfunction is commonly diagnosed and treated by physicians, requiring both knowledge of the anatomy and physiology of the sphincter of Oddi and clinical presentations and methods to diagnose and treat sphincter of Oddi dysfunction.

DEFINITIONS

Sphincter of Oddi dysfunction (SOD) refers to an abnormality of sphincter of Oddi contractility. Dysfunction manifests as a benign, noncalculous obstruction to flow of bile or pancreatic juice through the pancreaticobiliary junction, i.e., the sphincter of Oddi (SO). SOD may be manifested clinically by pancreaticobiliary pain, pancreatitis, or deranged liver function tests. The dysfunction is actually made up of two entities. SO dyskinesia refers to a primary motor abnormality of the SO that may result in a hypotonic sphincter but more commonly is seen as a hypertonic sphincter. In contrast, SO stenosis refers to a structural alteration of the sphincter, probably from an inflammatory process with subsequent fibrosis. Because it is often impossible to distinguish patients with SO dyskinesia from those with SO stenosis, the diagnosis "sphincter of Oddi

 Copyright 2004, Elsevier (USA). All rights reserved.

Schusdziarra, V. (1996). Somatostatin: Biological actions and pathophysiology. *In* "Octreotide: From Basic Science to Clinical Medicine" (C. Scarpignato, ed.), pp. 35–53. Karger, Basel.

Sternini, C., Wong, H., Wu, S. V., de Giorgio, R., Yang, M., Reeve, J., Jr., Brecha, N. C., and Walsh, J. H. (1997). Somatostatin 2A receptor is expressed by enteric neurons, and by interstitial cells of Cajal and enterochromaffin-like cells of the gastrointestinal tract. *J. Comp. Neurol.* **386**, 396–408.

Strowski, M. Z., Kohler, M., Chen, H. Y., Trumbauer, M. E., Li, Z., Szalkowski, D., Gopal-Truter, S., Fisher, J. K., Schaeffer, J. M., Blake A. D., Zhang, B. B., and Wilkinson, H. A. (2003). Somatostatin receptor subtype 5 regulates insulin secretion and glucose homeostasis. *Mol. Endocrinol.* **17**, 93–106.

Strowski, M. Z., Parmar, R. M., Blake, A. D., and Schaeffer, J. M. (2000). Somatostatin inhibits insulin and glucagon secretion via two receptor subtypes: An *in vitro* study of pancreatic islets from somatostatin receptor 2 knockout mice. *Endocrinology* **141**, 111–117.

Vuyyuru, L., Harrington, L., Arimura, A. S., and Schubert, M. L. (1997). Reciprocal inhibitory paracrine pathways link histamine and somatostatin secretion in the fundus of the stomach. *Am. J. Physiol.* **273**, G106–G111.

Sphincter of Oddi Dysfunction

JAMES WATKINS AND STUART SHERMAN
Indiana University Medical Center, Indianapolis

ampullary stenosis Narrowing of the ampulla, often causing obstruction.

botulinum toxin Protein produced by *Clostridium botulinum* bacteria, used therapeutically to paralyze or weaken the muscle into which it is injected.

endoscopic retrograde cholangiopancreaticography Fiber-optic endoscope insertion into the duodenum and dye injection via the ampulla of Vater, to visualize the biliary and pancreatic ducts.

endoscopic sphincterotomy Cutting of the sphincter via an endoscopic approach.

postcholecystectomy pain Painful attacks that occur after removal of the gallbladder.

quantitative hepatobiliary scintigraphy Nuclear medicine study that assesses bile flow through the biliary tract.

recurrent pancreatitis Recurrent inflammation of the pancreas.

Rome II criteria Consensus criteria developed to diagnose sphincter of Oddi dysfunction.

sphincter ablation Surgical removal of the sphincter.

sphincter of Oddi dysfunction Abnormality in the contractility of the sphincter of Oddi.

sphincter of Oddi manometry Manometric studies of the sphincter of Oddi.

Since its original description by Rugero Oddi in 1887, the sphincter of Oddi has been the subject of much study and controversy. Its very existence as a distinct anatomic or physiologic entity has been disputed. Hence, it is not surprising that the clinical syndrome of sphincter of Oddi dysfunction and its therapy are controversial areas. Nevertheless, this dysfunction is commonly diagnosed and treated by physicians, requiring both knowledge of the anatomy and physiology of the sphincter of Oddi and clinical presentations and methods to diagnose and treat sphincter of Oddi dysfunction.

DEFINITIONS

Sphincter of Oddi dysfunction (SOD) refers to an abnormality of sphincter of Oddi contractility. Dysfunction manifests as a benign, noncalculous obstruction to flow of bile or pancreatic juice through the pancreaticobiliary junction, i.e., the sphincter of Oddi (SO). SOD may be manifested clinically by pancreaticobiliary pain, pancreatitis, or deranged liver function tests. The dysfunction is actually made up of two entities. SO dyskinesia refers to a primary motor abnormality of the SO that may result in a hypotonic sphincter but more commonly is seen as a hypertonic sphincter. In contrast, SO stenosis refers to a structural alteration of the sphincter, probably from an inflammatory process with subsequent fibrosis. Because it is often impossible to distinguish patients with SO dyskinesia from those with SO stenosis, the diagnosis "sphincter of Oddi

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Hogan–Geenen Sphincter of Oddi Classification System Related to the Frequency of Abnormal Sphincter of Oddi Manometry and Pain Relief by Biliary Sphincterotomy

Patient group classification		Approximate frequency of abnormal sphincter manometry	Probability of pain relief by sphincterotomy if manometry is		Manometry before sphincter ablation
Type	Description		Abnormal	Normal	
Biliary I	Patients with biliary-type pain, abnormal aspartate transaminase or alkaline phosphatase > twice normal, documented on two or more occasions, delayed drainage of ERCP contrast from the biliary tree >45 minutes, and dilated common bile duct >12 mm in diameter	75–95%	90–95%	90–95%	Unnecessary
Biliary II	Patients with biliary-type pain but only one or two of the above criteria	55–65%	85%	35%	Highly recommended
Biliary III	Patients with only biliary-type pain and no other abnormalities	25–60%	55–65%	<10%	Mandatory

dysfunction" has been generally applied to both groups of patients. In an attempt to deal with this overlap in etiology, and also to determine the appropriate utilization of SO manometry (SOM), the Hogan–Geenen clinical SOD classification system has been developed for patients with suspected SOD (Table I); this classification is based on clinical history, laboratory results, and endoscopic retrograde cholangiopancreaticography (ERCP) findings.

A variety of less accurate clinical and diagnostic terms are sometimes used in the medical literature to describe SOD, including papillary stenosis, ampullary stenosis, biliary dyskinesia, and postcholecystectomy syndrome (even though SOD may occur with the gallbladder intact).

ANATOMY, PHYSIOLOGY, AND PATHOPHYSIOLOGY

The sphincter of Oddi is a small complex of smooth muscles surrounding the terminal common bile duct, main (ventral) pancreatic duct (of Wirsung), and the common channel (ampulla of Vater), when present. It has both circular and "figure-eight" components. The high-pressure zone generated by the sphincter is variably 4–10 mm. in length. Its role is to regulate bile and pancreatic exocrine juice flow and to prevent duodenum-to-duct reflux (i.e., to maintain a sterile intraductal environment). The SO possesses both a variable basal pressure and phasic contractile activity. The former appears to be the predominant mechanism, regulating outflow of pancreaticobiliary secretion into the intestine. Although phasic SO contractions may aid in regulating bile and pancreatic juice flow, their primary role appears to be maintaining a sterile intraductal milieu. Sphincter regulation is under both neural and hormonal control. Phasic wave activity of the sphincter is closely tied to the migrating motor complex (MMC) of the duodenum. Innervation of the bile duct does not appear to be essential, based on reports that sphincter function is preserved following liver transplantation. Although regulatory processes vary among species, cholecystokinin and secretin appear to be most important in causing sphincter relaxation, whereas nonadrenergic, noncholinergic neurons, which at least in part transmit vasoactive intestinal peptide (VIP) and nitric oxide, also relax the sphincter. The role of cholecystectomy in altering these neural pathways needs further definition. Luman and colleagues reported that cholecystectomy, at least in the short-term, suppresses the normal inhibitory effect of pharmacological doses of cholecystokinin (CCK) on the sphincter of Oddi. However, the mechanism of this effect is unknown.

Wedge specimens of the SO obtained at surgical sphincteroplasty from patients with SOD show evidence of inflammation, muscular hypertrophy, fibrosis, or adenomyosis within the papillary zone in approximately 60% of patients. In the remaining 40% of patients with normal histology, a motor disorder is suggested.

Less commonly, infections with cytomegalovirus or *Cryptosporidium*, as may occur in AIDS patients, or *Strongyloides* have caused SOD.

How does SOD cause pain? From a theoretical point of view, abnormalities of the SO can give rise to pain by impeding the flow of bile and pancreatic juice, resulting in ductal hypertension, ischemia arising from spastic contractions, and "hypersensitivity" of the papilla. Although unproved, these mechanisms may act alone or in concert to explain the genesis of pain.

EPIDEMIOLOGY

SOD may occur in pediatric or adult patients of any age; however, patients with SOD are typically middle-aged females. A survey on functional gastrointestinal disorders confirmed that SOD affects females more frequently than males and indicated a high association with work absenteeism, disability, and health care use. Although SOD most commonly occurs after cholecystectomy, it may be present with the gallbladder *in situ*.

Postcholecystectomy pain resembling the patient's preoperative biliary colic occurs in at least 10–20% of patients. The frequency of diagnosing SOD in reported series varies considerably with the patient selection criteria, the definition of SOD utilized, and the diagnostic tools employed. In a British report, sphincter of Oddi dysfunction was diagnosed in 9% of 451 consecutive patients being evaluated for postcholecystectomy pain. Roberts-Thomson evaluated 431 similar patients and found SOD in 11%. In a subpopulation of such patients with a normal ERCP (except dilated ducts in 28%) and recurrent pain of more than 3 months' duration, SOD was diagnosed in 68%. Sherman and colleagues used SOM to evaluate 115 patients with pancreaticobiliary pain with and without liver function test abnormalities. Patients with bile duct stones and tumors were excluded from analysis. Fifty-nine of 115 patients (51%) had an abnormal basal sphincter of Oddi pressure greater than 40 mmHg. These patients were further categorized by the Hogan–Geenen SOD classification system (Table I). The frequency of abnormal manometry was 86, 55, and 28%, for types I, II, and III patients, respectively. These abnormal manometric frequencies are very similar to those reported by others for type I and type II patients. In type III patients, the finding of an abnormal basal sphincter pressure has varied from 12 to 55%. As noted, patient selection factors may be one explanation for this great variability.

SOD can involve abnormalities in either the biliary sphincter, pancreatic sphincter, or both. The true frequency of SOD would then depend on whether one or both sphincters were studied. To fully assess the sphincter by SOM, both the bile duct and pancreatic ducts must be evaluated. In a series of 360 patients with pancreaticobiliary pain, 19% had abnormal pancreatic sphincter basal pressure alone, 11% had abnormal biliary basal sphincter pressure alone, and in 31%, the basal pressure was abnormal for both sphincters (overall frequency of SOD was 61%). Dysfunction may occur in the pancreatic duct portion of the SO and cause recurrent pancreatitis and pancreatic-type pain. Although a pancreatic SOD classification system has been developed (similar to the biliary SOD classification system), it has not been widely utilized. Manometrically documented SOD has been reported in 15–72% of patients with recurrent pancreatitis, previously labeled as idiopathic.

CLINICAL PRESENTATION

Abdominal pain is the most common presenting symptom of patients with SOD. The pain is usually epigastric or right upper quadrant, may be disabling, and lasts for 30 minutes to several hours. In some patients, the pain is continuous with episodic exacerbations. It may radiate to the back or shoulder and be accompanied by nausea and vomiting. Food or narcotics may precipitate the pain. The pain may begin several years after a cholecystectomy was performed for a gallbladder dysmotility or stone disease and is similar in character to the pain leading to the cholecystectomy. Alternatively, patients may have continued pain that was not relieved by a cholecystectomy. Jaundice, fever, or chills are rarely observed. The Rome II diagnostic criteria for SOD are episodes of severe steady pain located in the epigastrium and right upper quadrant, and all of the following criteria: (1) symptom episodes last 30 minutes or more with pain-free intervals, (2) symptoms have occurred on one or more occasions in the previous 12 months, (3) the pain is steady and interrupts daily activities or requires consultation with a physician, and (4) there is no evidence of structural abnormalities to explain the symptoms. The pain is not relieved by trial medications for acid–peptic disease or irritable bowel syndrome. Laboratory abnormalities consisting of transient elevation of liver function tests, typically during episodes of pain, are present in less than 50% of patients. After initial evaluation, patients are commonly categorized according to the Hogan–Geenen SOD classification system (Table I). Patients with SOD may present with typical pancreatic pain (epigastric and/or left upper quadrant radiating to the back) and recurrent pancreatitis.

Clinical Evaluation

The diagnostic approach to suspected SOD may be influenced by the presence of key clinical features.

However, the clinical manifestations of functional abnormalities of the SO may not always be easily distinguishable from those caused by organic ones (e.g., common bile duct stones) or other functional nonpancreaticobiliary disorders (e.g., irritable bowel syndrome).

General Initial Evaluation

Evaluation of patients with suspected SOD (i.e., patients with upper abdominal pain with characteristics suggestive of a pancreatobiliary origin) should be initiated with standard serum liver chemistries, serum amylase and/or lipase, abdominal ultrasonography, and/or computed tomography (CT) scans. The serum enzyme studies should be drawn during bouts of pain, if possible. Mild elevations (less than twice the upper limits of normal) are frequent in SOD, whereas greater abnormalities are more suggestive of stones, tumors, and liver parenchymal disease. Although the diagnostic sensitivity and specificity of abnormal serum liver chemistries are relatively low, recent evidence indicates that the finding of abnormal liver tests in biliary type II patients may predict a favorable response to endoscopic sphincterotomy. CT scans and abdominal ultrasounds are usually normal but occasionally a dilated bile duct or pancreatic duct may be found (particularly in patients with type I SOD). Standard evaluation and treatment of other more common upper gastrointestinal conditions, such as peptic ulcer disease and gastroesophageal reflux, should be done simultaneously. In the absence of mass lesions, stones, or response to acid-suppression therapeutic trials, the suspicion for sphincter disease is heightened.

DIAGNOSTIC METHODS (NONINVASIVE)

Because SOM (considered by most authorities to be the gold standard for diagnosing SOD) is difficult to perform, invasive, not widely available, and associated with a relatively high complication rate, several noninvasive and provocative tests have been designed in an attempt to identify patients with SOD.

Morphine–Prostigmin Provocative Test (Nardi Test)

Morphine has been shown to cause sphincter of Oddi contraction. Prostigmin (neostigmine), 1 mg subcutaneously, is added as a vigorous cholinergic secretory stimulant to morphine (10 mg subcutaneously) to make this challenge test. The morphine–Prostigmin test, historically, had been used extensively to diagnose SOD. Reproduction of the patient's typical pain associated with a fourfold increase in aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase, amylase, or lipase constitutes a positive response. The usefulness of this test is limited by its low sensitivity and specificity in predicting the presence of SOD and its poor correlation with outcome after sphincter ablation. This test has largely been replaced by tests believed to be more sensitive.

Ultrasonographic Assessment of Extrahepatic Bile Duct and Main Pancreatic Duct Diameter after Secretory Stimulation

After a lipid-rich meal or cholecystokinin administration, the gallbladder contracts, bile flow from the hepatocytes increases, and the sphincter of Oddi relaxes, resulting in bile entry into the duodenum. Similarly, after a lipid-rich meal or secretin administration, pancreatic exocrine juice flow is stimulated and the sphincter of Oddi relaxes. If the sphincter of Oddi is dysfunctional and causes obstruction to flow, the common bile duct or main pancreatic duct may dilate under secretory pressure. This can be monitored by transcutaneous ultrasonography. Sphincter and terminal duct obstruction from other causes (stones, tumors, strictures, etc.) may similarly cause ductal dilation and need to be excluded. Pain provocation should also be noted if present. To date, limited studies comparing these noninvasive tests with sphincter of Oddi manometry or outcome after sphincter ablation show only modest correlation. Because of intestinal gas, the pancreatic duct many not be visualized on standard transcutaneous ultrasound. Despite the superiority of endoscopic ultrasound in visualizing the pancreas, Catalano *et al.* report the sensitivity of secretin-stimulated endoscopic ultrasound in detecting SOD to be only 57%.

Quantitative Hepatobiliary Scintigraphy

Hepatobiliary scintigraphy assesses bile flow through the biliary tract. Impairment to bile flow from sphincter disease, tumors, or stones (as well as parenchymal liver disease) results in impaired radionuclide flow. The precise criteria to define a positive (abnormal) study remain controversial, but duodenal arrival time greater than 20 minutes and hilum to duodenum time greater than 10 minutes are most widely used. Most studies are flawed by lack of correlation with SOM or outcome after sphincter ablation. However, four studies have reported the performance characteristics of hepatobiliary scintigraphy in 105 patients using the results of SOM as the reference standard. The overall

sensitivity was 78% (range, 44–100%), specificity was 90% (range 80–100%), positive predictive value was 92% (range, 82–100%), and negative predictive value was 81% (range, 62–100%). However, these promising results have not been reproduced by other investigators. Overall, it appears that patients with dilated bile ducts and high-grade obstruction are likely to have a positive scintigraphic study. Esber and colleagues found that patients with lower grade obstruction (Hogan–Geenen classification types II and III) generally have normal scintigraphy, even if done after cholecystokinin provocation. Moreover, Pineau *et al.* reported that 8 of 20 asymptomatic control subjects had an abnormal cholecystokinin-stimulated study. The value of hepatobiliary scintigraphy is also limited by the fact that it does not evaluate the pancreatic sphincter.

In the absence of more definitive data, the current conclusion is that noninvasive testing for sphincter of Oddi dysfunction has a relatively low or undefined sensitivity and specificity and is, therefore, not recommended for general clinical use, except in situations when more definitive testing (manometry) is unsuccessful or unavailable.

DIAGNOSTIC METHODS (INVASIVE)

Because of their associated risks, invasive testing with ERCP and manometry should be reserved for patients with clinically significant or disabling symptoms. In general, invasive assessment of patients for SOD is not recommended unless definitive therapy (sphincter ablation) is planned if abnormal sphincter function is found.

Cholangiography

Cholangiography is essential to rule out stones, tumors, or other obstructing processes of the biliary tree that may cause symptoms identical to those of SOD. Once such lesions are ruled out by a good quality cholangiographic study, ducts that are dilated and/or drain slowly suggest obstruction at the level of the sphincter. Although some controversy exists, extrahepatic ducts that are greater than 12 mm in diameter (postcholecystectomy), when corrected for magnification, are considered dilated. Although definitive normal supine drainage times have not been well defined, a postcholecystectomy biliary tree that fails to empty all contrast media by 45 minutes is generally considered abnormal.

Endoscopic evaluation of the papilla and peripapillary area can yield important information that can influence the diagnosis and treatment of patients with suspected SOD. Occasionally, ampullary cancer may simulate SOD. The endoscopist should do tissue sampling of the papilla (preferably after sphincterotomy) in suspicious cases. Radiographic features of the pancreatic duct are also important to assess in the patient with suspected SOD. Dilation of the pancreatic duct (>6 mm in the pancreatic head and >5 mm in the body) and delayed contrast drainage time ($\geq$9 minutes in the prone position) may give indirect evidence for the presence of SOD.

Sphincter of Oddi Manometry

The most definitive development in the understanding of the pressure dynamics of the SO came with the advent of sphincter of Oddi manometry. SOM is the only available method to measure SO motor activity directly. Although SOM can be performed intraoperatively and percutaneously, it is most commonly done in the ERCP setting. SOM is considered by most authorities to be the gold standard for evaluating patients for sphincter dysfunction. The use of manometry to detect motility disorders of the sphincter of Oddi is similar to its use in other parts of the gastrointestinal tract. Unlike other areas of the gut, SOM is more technically demanding, invasive, and hazardous. Questions remain as to whether these short-term observations (2- to 10-minute recordings per pull-through) reflect the "24-hour pathophysiology" of the sphincter.

SOM is recommended in patients with idiopathic pancreatitis or unexplained disabling pancreaticobiliary pain with or without hepatic and pancreatic enzyme abnormalities, and to assess for sphincterotomy stenosis or residual sphincter hypertension in symptomatic patients after sphincterotomy is performed for SOD. An ERCP is usually performed (if an adequate study is not available) immediately before the SOM to exclude other potential structural causes for the patient's symptoms. Indications for the use of SOM have also been developed according to the Hogan–Geenen SOD classification System (Table I). In patients with type I, there is a general consensus that a structural disorder of the sphincter (i.e., sphincter stenosis) exists. Although SOM may be useful in documenting SOD, it is not an essential diagnostic study prior to endoscopic or surgical sphincter ablation. Such patients uniformly benefit from sphincter ablation regardless of the SOM results (see later). Patients with type II demonstrate SO motor dysfunction in 55–65% of cases. In this group of patients, SOM is highly recommended because the results of the study predict outcome from sphincter ablation. Patients with type III have pancreaticobiliary pain without other objective evidence of sphincter outflow obstruction. SOM is mandatory to confirm the presence of SOD. Although

data are limited, it appears that the results of SOM may predict outcome from sphincter ablation in these patients.

SOM is performed by advancing a specialized triple-lumen #5 French water-perfused catheter selectively into the bile duct and/or pancreatic duct. The catheter is pulled across the sphincter zone while pressures are recorded with the aid of a pressure transducer. A variety of modifications of the standard #5 French manometry catheter are commercially available. (The technique of SOM and the interpretation of the tracings are beyond the scope of this discussion.)

It is important to emphasize that complete sphincter assessment requires manometric evaluation of both the biliary and pancreatic sphincters. Current data indicate that an abnormal basal sphincter pressure may be confined to one side of the sphincter in 35–65% of patients. Thus, one sphincter may be dysfunctional whereas the other is normal. Failure to appreciate the anatomically separate pancreatic and biliary sphincters and the potential for discordant manometric results may result in misdiagnosis and improper therapy.

Pancreatitis is the most common major complication after SOM. Using standard perfused catheters, pancreatitis rates as high as 31% have been reported. Such high complication rates have initially limited more widespread use of SOM. These data also emphasize that manometric evaluation of the pancreatic duct is associated with a high complication rate. A variety of methods to decrease the incidence of postmanometry pancreatitis have been developed. In a prospective randomized study, Sherman and colleagues found that the aspirating catheter (this catheter allows for aspiration of the perfused fluid from end and side holes while accurately recording pressure from the two remaining side ports) reduces the frequency of pancreatic duct manometry-induced pancreatitis from 31 to 4%. The reduction in pancreatitis with use of this catheter in the pancreatic duct and the very low incidence of pancreatitis after bile duct manometry lend support to the notion that increased pancreatic duct hydrostatic pressure is a major cause of this complication. Thus, routine aspiration of pancreatic juice should accompany study of the pancreatic duct sphincter by SOM.

Stent Trial as Diagnostic Test

Placement of a pancreatic or biliary stent on a trial basis in hope of achieving pain relief and predicting the response to more definitive therapy, i.e., sphincter ablation, has received only limited application. Pancreatic stent trials, especially in patients with normal pancreatic ducts, are strongly discouraged because serious ductal and parenchymal injury may occur if stents are left in place for more than a few days. Goff reported a biliary stent trial in 21 type II and type III patients with SOD with normal biliary manometry. Seven French stents were left in place for at least 2 months if symptoms resolved and were removed sooner if they were judged ineffective. Relief of pain with the stent was predictive of long-term pain relief after biliary sphincterotomy. Unfortunately, 38% of the patients developed pancreatitis (14% were graded severe) following stent placement. Because of this high rate of complications, biliary stent trials are strongly discouraged. Rolny and colleagues also reported a series of bile duct stent placements as predictor of outcome following endoscopic sphincterotomy in 23 postcholecystectomy patients (7 type II and 16 type III). Similar to the study by Goff, resolution of pain during at least 12 weeks of stenting predicted a favorable outcome from sphincterotomy irrespective of sphincter of Oddi pressure. In this series there were no complications related to stent placement.

THERAPY FOR SOD

The therapeutic approach in patients with SOD is aimed at reducing the resistance caused by the sphincter of Oddi to the flow of bile and/or pancreatic juice. Historically, most emphasis has been placed on definitive intervention, i.e., surgical sphincteroplasty or endoscopic sphincterotomy. This appears appropriate for patients with high-grade obstruction (type I as per Hogan–Geenen criteria). In patients with lesser degrees of obstruction, the clinician must carefully weigh the risks and benefits before recommending invasive therapy.

Medical Therapy

Medical therapy for documented or suspected SOD has received only limited study. Because the SO is a smooth muscle structure, it is reasonable to assume that drugs that relax smooth muscle might be an effective treatment for SOD. Sublingual nifedipine and nitrates have been shown to reduce the basal sphincter pressures in asymptomatic volunteers and symptomatic patients with SOD. Although medical therapy may be an attractive initial approach in patients with SOD, several drawbacks exist. First, medication side effects may be seen in up to one-third of patients. Second, smooth muscle relaxants are unlikely to be of any benefit in patients with the structural form of SOD (i.e., SO ste-

nosis) and the response is incomplete in patients with a primary motor abnormality of the SO (i.e., SO dyskinesia). Finally, long-term outcome from medical therapy has not been reported. Nevertheless, because of the "relative safety" of medical therapy and the benign (though painful) character of SOD, this approach should be considered in all type III and less severely symptomatic type II SOD patients before considering more aggressive sphincter ablation therapy.

Surgical Therapy

The surgical approach, most commonly, is a transduodenal biliary sphincteroplasty with a transampullary septoplasty (pancreatic septoplasty). Some 60–70% of patients were reported to have benefited from this therapy during a 1- to 10-year followup. Patients with an elevated basal sphincter pressure determined by intraoperative SOM were more likely to improve from surgical sphincter ablation than were those with a normal basal pressure. Some reports have suggested that patients with biliary-type pain have a better outcome than do patients with idiopathic pancreatitis, whereas others have suggested no difference. However, most studies found that symptom improvement following surgical sphincter ablation alone was relatively uncommon in patients with established chronic pancreatitis. The surgical approach for SOD has largely been replaced by endoscopic therapy. Patient tolerance, cost of care, morbidity, mortality, and cosmetic results are some of the factors that favor an initial endoscopic approach. At present, surgical therapy is reserved for patients with restenosis following endoscopic sphincterotomy and when endoscopic evaluation and/or therapy are not available or technically feasible.

Endoscopic Therapy

Endoscopic sphincterotomy is the current standard therapy for patients with SOD. Most data on endoscopic sphincterotomy relate to biliary sphincter ablation alone. Clinical improvement following therapy has been reported to occur in 55–95% of patients (Table I). These variable outcomes are reflective of the different criteria used to document SOD, the degree of obstruction (type I biliary patients appear to have a better outcome than do types II and III), the methods of data collection (retrospective vs. prospective), and the techniques used to determine benefit. Rolny and colleagues studied 17 type I postcholecystectomy biliary patients by SOM. In this series, 65% had an abnormal SOM (although not specifically stated, it appears that the biliary sphincter was studied alone). Nevertheless, during a mean followup interval of 2.3 years, all patients benefited from biliary sphincterotomy. The results of this study suggest that because type I biliary patients invariably benefit from biliary sphincterotomy, SOM in this patient group is not only unnecessary, but it may also be misleading. The results of this study, however, have never been validated at another center.

Although most of the studies reporting efficacy of endoscopic therapy in SOD have been retrospective, three notable randomized trials have now been reported. In a landmark study by Geenen and associates, 47 postcholecystectomy type II biliary patients were randomized to biliary sphincterotomy or sham sphincterotomy. SOM was performed in all patients but was not used as a criterion for randomization. During a 4-year followup, 95% of patients with an elevated basal sphincter benefited from sphincterotomy. In contrast, only 30–40% of patients with an elevated sphincter pressure treated by sham sphincterotomy or with a normal sphincter pressure treated by endoscopic sphincterotomy or sham sphincterotomy benefited from this therapy. The two important findings of this study were that SOM predicted the outcome from endoscopic sphincterotomy and that endoscopic sphincterotomy offered long-term benefit in type II biliary patients with SOD.

Sherman and associates reported their preliminary results of a randomized study comparing endoscopic sphincterotomy, surgical biliary sphincteroplasty with pancreatic septoplasty (with or without cholecystectomy), to sham sphincterotomy for type II and type III biliary patients with manometrically documented SOD. During a 3.0-year followup period, 69% of patients undergoing endoscopic or surgical sphincter ablation improved, compared to 24% in the sham sphincterotomy group ($p = 0.009$). There was a trend for type II patients to benefit more frequently from sphincter ablation, compared to type III patients [13/16 (81%) vs. 11/19 (58%); $p = 0.14$]. Evidence is now accumulating that the addition of a pancreatic sphincterotomy to an endoscopic biliary sphincterotomy in such patients may improve the outcome (see later).

In a third study, postcholecystectomy patients with biliary-type pain (mostly type II) were prospectively randomized to endoscopic sphincterotomy or sham following stratification according to SOM. Of patients with elevated basal pressure, 85% (11 of 13) improved at 2 years after endoscopic sphincterotomy, whereas 38% (5 of 13) of patients improved after a sham procedure ($p = 0.041$).

There are numerous nonrandomized trials reporting benefit rates of 56–78% in type II patients with manometrically documented SOD treated with biliary sphincterotomy. The benefit rate in type III patients was more variable but the number of patients treated was relatively small. Untreated pancreatic sphincter hypertension appears to be one of the reasons that patients with SOD fail to benefit from biliary sphincterotomy alone. Because there are separate biliary and pancreatic sphincters, ablation of the biliary sphincter alone usually leaves the pancreatic sphincter pressure unaltered. Surgeons have appreciated this for years, and this accounts for the common surgical approach of a biliary and pancreatic sphincteroplasty. Eversman and colleagues found that 90% of SOD patients with persistent pain or pancreatitis after biliary sphincterotomy had residual abnormally elevated pancreatic sphincter pressure. This same group reported that 80% of SOD patients with isolated elevated biliary sphincter pressure (normal pancreatic sphincter pressure) were clinically improved at 5-year followup after biliary sphincterotomy alone, whereas only 48% of patients with elevated basal pancreatic sphincter pressure (with normal or abnormal basal biliary sphincter pressure) benefited from this therapy. Evidence is now accumulating that the addition of an endoscopic pancreatic sphincterotomy to a biliary sphincterotomy in patients with SOD may improve outcome. This will be discussed further in the following sections.

These results clearly indicate that the response rate and enthusiasm for sphincter ablation must be correlated with patient presentation and balanced against the high complication rates reported for endoscopic therapy of SOD. Most studies indicate that patients undergoing endoscopic sphincterotomy for SOD have complication rates two to five times higher compared to patients undergoing endoscopic sphincterotomy for ductal stones. Pancreatitis is the most common complication, occurring in up to 20% of patients. Endoscopic techniques are being developed (e.g., pancreatic duct stenting prior to combined pancreaticobiliary sphincterotomy and pancreatic stenting after biliary sphincterotomy) to limit such complications.

Balloon Dilation and Stenting

In an attempt to be less invasive and possibly preserve sphincter function, adaptation of this technique to treat SOD has been described. Unfortunately, because of the unacceptably high complication rates, primarily pancreatitis, this technology has little role in the management of SOD. Similarly, although biliary stenting might offer short-term symptom benefit in patients with SOD and predict outcome from sphincter ablation, it, too, has unacceptably high complication rates and cannot be advocated in this setting based on the available data.

Botulinum Toxin Injection

Botulinum toxin (Botox), a potent inhibitor of acetylcholine release from nerve endings, has been successfully applied to smooth muscle disorders of the gastrointestinal tract such as achalasia. In a preliminary clinical trial, Botox injection into the SO resulted in a 50% reduction in the basal sphincter pressure and improved bile flow. This reduction in pressure may be accompanied by symptom improvement in some patients. It follows that Botox may serve as a therapeutic trial for SOD, with responders undergoing permanent sphincter ablation. One such study was recently reported. Twenty-two postcholecystectomy type III patients with manometric evidence of SOD underwent Botox injection into the intraduodenal sphincter segment. Overall, 11 of the 12 patients who responded to Botox, versus 2 of 10 patients who did not gain pain relief, later benefited from endoscopic sphincterotomy ($p < 0.01$). Such an approach, however, does require at least two procedures, each with their associated complications.

SOD IN RECURRENT PANCREATITIS

SOD has been manometrically documented in 15–72% of patients with recurrent pancreatitis, previously labeled as idiopathic. Biliary sphincterotomy alone has been reported to prevent further pancreatitis episodes in more than 50% of such patients. The value of ERCP, SOM, and sphincter ablation therapy was studied in 51 patients with idiopathic pancreatitis; 24 (47.1%) had an elevated basal sphincter pressure and 30 were treated by either biliary sphincterotomy ($n = 20$) or surgical sphincteroplasty with pancreatic septoplasty ($n = 10$). Of these patients, 15 of 18 (83%) with an elevated basal sphincter pressure had long-term benefit (mean followup, 38 months) from sphincter ablation therapy (including 10 of 11 treated by biliary sphincterotomy), in contrast to only 4 of 12 (33.3%, $p < 0.05$) with a normal basal sphincter pressure (including 4 of 9 treated by biliary sphincterotomy). Guelrud *et al.*, however, found that severance of the pancreatic sphincter was necessary to resolve the pancreatitis (Table II). In this series, 69 patients with idiopathic pancreatitis due to SOD underwent treatment by standard biliary

TABLE II Pancreatic Sphincter Dysfunction and Recurrent Pancreatitis: Response to Sphincter Therapy

Treatment	Number of patients improved/total number of patients
Biliary sphincterotomy alone	5/18 (28%)
Biliary sphincterotomy followed by pancreatic sphincter balloon dilation	13/24 (54%)
Biliary sphincterotomy plus pancreatic sphincterotomy at later session	10/13 (77%)[a]
Biliary sphincterotomy and pancreatic sphincterotomy at same session	12/14 (86%)[a]

[a] $p < 0.005$ vs. biliary sphincterotomy alone.

sphincterotomy ($n = 18$), biliary sphincterotomy with pancreatic sphincter balloon dilation ($n = 24$), biliary sphincterotomy followed by pancreatic sphincterotomy in separate sessions ($n = 13$), or combined pancreatic and biliary sphincterotomy in the same session ($n = 14$). Eighty-one percent of patients undergoing pancreatic and biliary sphincterotomy had resolution of their pancreatitis compared to 28% of patients undergoing biliary sphincterotomy alone ($p < 0.005$). These data are consistent with the theory that many such patients who benefit from biliary sphincterotomy alone have subtle gallstone pancreatitis. The results of Guelrud *et al*. also support the anatomic findings of separate biliary and pancreatic sphincters, and the manometry findings of residual pancreatic sphincter hypertension in more than 50% of persistently symptomatic patients who undergo biliary sphincterotomy alone. Toouli *et al*. also demonstrated the importance of pancreatic and biliary sphincter ablation in patients with idiopathic pancreatitis. In this series, 23 of 26 patients (88%) undergoing surgical ablation of both the biliary and pancreatic sphincter were either asymptomatic or had minimal symptoms at a median followup of 24 months (9−105 months). A different group retrospectively evaluated the long-term results of endoscopic pancreatic sphincterotomy in 55 patients with presumed (recurrent pancreatitis with pancreatic duct dilation and contrast medium drainage time from the pancreatic duct greater than 10 minutes) or manometrically documented pancreatic sphincter dysfunction. During a median followup of 16 months (range, 3−52 months), 34 patients (62%) reported significant pain improvement. Patients with normal pancreatograms were more likely to respond to therapy than were those with pancreatographic evidence of chronic pancreatitis (73 vs. 58%).

Currently, the best method to treat residual pancreatic sphincter stenosis after biliary sphincterotomy awaits further study. Patients with idiopathic pancreatitis who fail to respond to biliary sphincterotomy alone should have their pancreatic sphincter reevaluated and be considered for sphincter ablation if residual high pressure is found.

SOD IN PATIENTS WITH INTACT GALLBLADDER

SOD may exist in the presence of an intact biliary tract with the gallbladder *in situ*. Because the symptoms of SO or gallbladder dysfunction cannot be readily separated, the diagnosis of SOD is commonly made after cholecystectomy or less frequently after proper investigations have excluded gallbladder abnormalities. The frequency of manometrically documented SOD in patients prior to cholecystectomy has received limited study. Guelrud and colleagues studied 121 patients with symptomatic gallstones and a normal common bile duct diameter (by transcutaneous ultrasound) by SOM prior to cholecystectomy. An elevated basal sphincter pressure was found in 14 patients (11.6%). SOD was diagnosed in 4.1% of patients with a normal serum alkaline phosphatase (4 of 96) and in 40% with an elevated serum alkaline phosphatase (10 of 25). Ruffolo *et al*. evaluated 81 patients with symptoms suggestive of biliary disease but normal ERCP and no gallbladder stones on transcutaneous ultrasound by scintigraphic gallbladder ejection fraction and endoscopic SOM. Fifty-three percent of patients had SOD and 49% had an abnormal gallbladder ejection fraction. SOD occurred with a similar frequency in patients with an abnormal gallbladder ejection fraction (50%) and a normal ejection fraction (57%).

The approach to patients with suspected SOD and intact gallbladder is frequently a challenging problem. In patients that have a clearly abnormal gallbladder ejection fraction (GBEF) of less than 35%, it seems prudent to proceed with laparoscopic cholecystectomy. In patients with a low GBEF, this approach may provide relief of symptoms for 70−80% of patients. Although there is some evidence that cholecystectomy may exacerbate SOD by loss of the gallbladder reservoir function and decreased compliance, the risks of laparoscopic cholecystectomy are less than those of ERCP with SOM. Patients with a borderline or normal GBEF present a diagnostic and therapeutic dilemma as well. A recent study has suggested that symptom reproduction

after CCK infusion may be as important as GBEF in the evaluation of ultrasound-negative biliary colic. These results would tend to favor a trial of cholecystectomy before proceeding with SOM in patients with intact gallbladder and suspected SOM. The GBEF may be misleading in patients who have had previous biliary sphincterotomy, with a lack of gallbladder filling seen in 75% postsphincterotomy.

FAILURE TO ACHIEVE SYMPTOMATIC IMPROVEMENT AFTER BILIARY SPHINCTEROTOMY

There are several potential explanations as to why patients may fail to achieve symptom relief after biliary sphincterotomy is performed for well-documented sphincter of Oddi dysfunction. First, the biliary sphincterotomy may have been inadequate or restenosis may have occurred. Although the biliary sphincter is commonly not totally ablated, one study indicates that clinically significant biliary restenosis occurs relatively infrequently. If no "cutting space" remains in such a patient, balloon dilation to 8–10 mm. may suffice, but long-term outcome from such therapy is unknown.

Second, the importance of pancreatic sphincter ablation is being increasingly recognized. One report has shown that 25 of 26 patients (mostly type II), who failed to respond to biliary sphincterotomy, had elevated pancreatic sphincter pressure. Endoscopic pancreatic sphincterotomy was performed with overall symptomatic improvement in two-thirds of patients. Another group performed pancreatic sphincterotomy on 43 type I and type II SOD patients who failed to benefit from biliary sphincterotomy alone. During the followup period, 72% were symptom free and 19% were partially or transiently improved. A study by a different group also presented data demonstrating that response to sphincterotomy depends on treating the diseased sphincter segment. Specifically, the outcome from biliary sphincterotomy alone depended on whether the biliary, pancreatic, or both sphincters were abnormal. Patients with pancreatic sphincter hypertension who fail to improve from biliary sphincterotomy alone can be "rescued" by undergoing pancreatic sphincterotomy. In this study, 80% of patients with isolated biliary sphincter hypertension benefited from biliary sphincterotomy during a 17-month followup period. In contrast, 15% of patients with isolated pancreatic sphincter hypertension and 50% with combined sphincter disease improved. However, when patients with persistent symptoms and elevated basal pancreatic sphincter pressure were then treated with a pancreatic sphincterotomy, the improvement rate was 77% for the group with isolated pancreatic sphincter hypertension and 80% for the group with combined sphincter disease. Overall, 79% of patients improved after a biliary sphincterotomy and selective pancreatic sphincterotomy, compared with a 45% benefit rate if only biliary sphincterotomy was done.

A third explanation as to why patients may fail to respond to sphincterotomy is because they have chronic pancreatitis. These patients may or may not have abnormal pancreatograms. Endoscopic ultrasound may show parenchymal and ductular changes of the pancreas in some of these patients, suggesting chronic pancreatitis. Finally, some patients may be having pain from altered gut motility of the stomach, small bowel, or colon (irritable bowel or pseudo-obstruction variants). There is increasing evidence that upper gastrointestinal motility disorders may masquerade as pancreatobiliary-type pain (i.e., discrete right upper quadrant pain). Multiple preliminary studies show disordered duodenal motility in such patients. This area needs much more study to determine the frequency, significance, and/or coexistence of these motor disorders along with SOD. A recent study suggested that type III patients have duodenal specific visceral hyperalgesia with pain reproduction by duodenal distension. These patients were also shown to have high levels of somatization, depression, obsessive–compulsive behavior, and anxiety compared to control subjects.

SUMMARY

In summary, our knowledge of sphincter of Oddi dysfunction and manometric techniques to assist in this diagnosis are evolving. Successful endoscopic SOM requires good general ERCP skills and careful attention to the main details involved. If SOD is suspected in a type III or mild to moderate pain level type II patient, medical therapy should generally be tried. If medical therapy fails or is bypassed, ERCP and manometric evaluation are recommended. The role of less invasive studies remains uncertain due to undefined sensitivity and specificity. Sphincter ablation is generally warranted in symptomatic type I patients and type II and type III patients with abnormal manometry. The symptom relief rate varies from 55 to 95%, depending on the patient presentation and selection. Initial nonresponders require thorough pancreatic sphincter and pancreatic parenchymal evaluation. SOD patients have relatively high complication rates after invasive studies or therapy. Thorough review of the risk:benefit ratio with individual patients is mandatory.

Acknowledgment

The authors are grateful to Tina Jackson for the technical preparation of this document.

See Also the Following Articles

Biliary Tract, Anatomy • Manometry • Pancreatitis, Chronic • Sphincters • Sphincterotomy

Further Reading

Chen, J. W. C., Saccone, G. T. P., and Toouli, J. (1998). Sphincter of Oddi manometry in the diagnosis and therapy of pancreatic disease. *Gastrointest. Endosc. Clin. North Am.* **8**, 79–85.

Eversman, D., Fogel, E., Rusche, M., Sherman, S., and Lehman, G. A. (1999). Frequency of abnormal pancreatic and biliary sphincter manometry compared with clinical suspicion of sphincter of Oddi dysfunction. *Gastrointest. Endosc.* **50**, 637–641.

Fogel, E. L., and Sherman, S. (2000). Performance of sphincter of Oddi manometry. *Clin. Perspect. Gastroenterol.* **4**, 165–173.

Freeman, M. L., Nelson, D. B., Sherman, S., *et al.* Complications of endoscopic biliary sphincterotomy. (1996). *N. Engl. J. Med.* **335**, 909–918.

Geenen, J. E., Hogan, W. J., Dodds, W. J., Toouli, J., and Venu, R. P. (1989). The efficacy of endoscopic sphincterotomy after cholecystectomy in patients with suspected sphincter of Oddi dysfunction. *N. Engl. J. Med.* **320**, 82–87.

Hogan, W., Sherman, S., Pasricha, P., and Carr-Locke, D. L. (1997). Position paper on sphincter of Oddi manometry. *Gastrointest. Endosc.* **45**, 342–348.

Kalloo, A. N., and Pasricha, P. J. (1996). Therapy of sphincter of Oddi dysfunction. *Gastrointest. Endosc. Clin. North Am.* **6**, 117–125.

Kuo, W. H., Pasricha, P. J., and Kalloo, A. N. (1998). The role of sphincter of Oddi manometry in the diagnosis and therapy of pancreatic disease. *Gastrointest. Endosc. Clin. North Am.* **8**, 79–85.

Sherman, S. (2002). What is the role of ERCP in the setting of abdominal pain of pancreatic or biliary origin (suspected sphincter of Oddi dysfunction)? *Gastrointest. Endosc.* **56**, S258–S266.

Sherman, S., Troiano, F. P., Hawes, R. H., *et al.* (1990). Sphincter of Oddi manometry—Decreased risk of clinical pancreatitis with the use of modified aspirating catheter. *Gastrointest. Endosc.* **36**, 462–466.

Sostre, S., Kalloo, A. N., Spiegler, E. J., Camargo, E. E., and Wagner, H. N. (1992). A noninvasive test of sphincter of Oddi dysfunction in postcholecystectomy patients: The scintigraphic score. *J. Nucl. Med.* **33**, 1216–1222.

Wehrmann, T., Seifert, H., Seipp, M., Lembeke, B., and Caspary, W. F. (1998). Endoscopic injection of botulinum toxin for biliary sphincter of Oddi dysfunction. *Endoscopy* **30**, 702–707.

Sphincterotomy

RUSSELL J. NAUTA AND ALLISON C. NAUTA
Harvard University and Mount Auburn Hospital

retroperitoneum The space behind the peritoneal cavity and anterior to the muscles and bones of the back, wherein reside the blood vessels, nerves, and lymph nodes associated with the abdominal viscera.
splanchnic nerve Nerves transmitting pain of visceral origin.

Sphincterotomy is the division of the specialized muscle fibers defining an aperture. In the gastrointestinal tract, the two sphincters typically divided are the sphincter of Oddi and the internal anal sphincter. The aperture (ampulla) defined by the sphincter of Oddi is that at which the common bile duct and pancreatic duct enter the duodenum. The internal anal sphincter preserves fecal continence and is located between the anal canal and the external environment.

DIVISION OF THE SPHINCTER OF ODDI

Patients in whom division of the sphincter of Oddi is performed typically have symptoms referable to obstruction of the biliary tree (jaundice with or without pain, dark urine, light stools) and laboratory studies suggesting obstruction (elevation of the direct bilirubin, alkaline phosphatase, and/or amylase levels). Inflammation of the pancreas or stones of biliary or pancreatic origin causing obstruction typically have associated pain and fever. Obstructions secondary to benign or malignant tumor occluding the sphincter of Oddi classically present with painless jaundice and weight loss from malabsorptive malnutrition. Pain associated with malignant obstruction typically indicates surgically unresectable, advanced disease and splanchnic nerve involvement from tumor ingrowth into the splanchnic nerves of the retroperitoneum.

Division of the sphincter of Oddi (sphincterotomy) is most typically performed with cold knife or electrocautery endoscopically introduced into the sphincteral lumen and withdrawn through one side of the lumen. If the sphincter is tight and precludes cannulation by this method, a "pre-cut" technique may be used, where endoscopic incision in the bile duct is made through the common wall between the duodenum and the bile duct. This incision is then carried to the ampulla. Sphincterotomy both facilitates bile flow and allows access to the common duct so that impacted stones might be removed or radiographic visualization of the biliary and pancreatic ducts might be accomplished. The radiologic visualization of these ducts through endoscopically guided retrograde injection of contract is known as endoscopic retrograde cholangiopancreatography. Visualization of the bile duct may be diagnostic and determine whether surgery is necessary for stone or stricture; if obstruction of the bile duct has occurred from a benign process, the sphincterotomy itself may also be therapeutic and obviate the need for open surgical intervention. In such instances, the sphincterotomy or associated stone extraction relieves the patient's obstructive symptoms.

Resting sphincter tone is decreased following sphincterotomy. Although this often leads to some low level of continuous flow of bile into the duodenum, additional relaxation of the sphincter still accompanies the ingestion of a meal.

DIVISION OF THE LATERAL INTERNAL ANAL SPHINCTER

The continence mechanism for fecal material involves an anatomic and physiologic fusion of the anorectum's longitudinal and circular muscle fibers with the levator ani muscles and those of the puborectalis sling. Anal fissure is a painful tear in the anal sphincter. Typically seen in patients with constipation, the established fissure is recognizable by the typical linear tear in the posterior midline and associated skin tag (sentinel pile) just distal to the tear.

Division of the internal anal sphincter, as described by Parks, breaks the cycle of pain, spasm, and recurrent tearing, reducing resting anal pressures by 25–50% for at least 4–6 years postoperatively.

The procedure may be performed under a general, regional, or local anesthetic. Rather than dividing the internal sphincter directly over the fissure, a lateral site for the incision is chosen. The sphincter may be divided by direct visualization through an incision made over it or as a blind procedure wherein a knife is insinuated

 Copyright 2004, Elsevier (USA). All rights reserved.

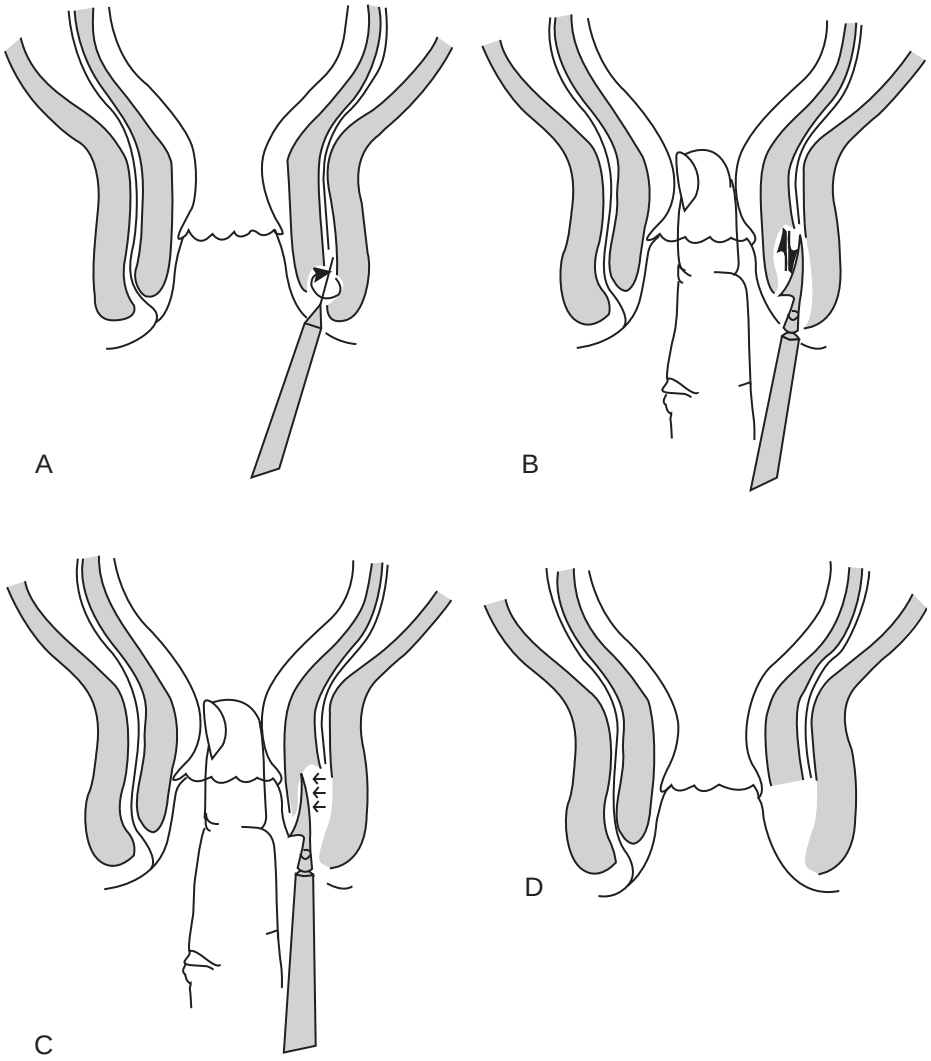


FIGURE 1 The scalpel is placed through a lateral stab wound made between the external and internal anal sphincters (A). It is rotated (B) and advanced in a sawing motion to approach the mucosa and the examining finger (C). The completed distal lateral sphincterotomy is shown in D.

between the external and internal sphincters (the intersphincteric groove) and the internal sphincter is divided with a finger in the rectum to avoid entry into the anal canal and creation of a fistula-in-ano (Fig. 1). The resultant decrease in tone facilitates spontaneous healing of both the surgical wound and the original fissure. The procedure is rarely associated with incontinence of solid stool or gas; fewer than 1% of patients experience incontinence of mucous or of liquid stool.

See Also the Following Articles

Anal Sphincter • Bile Duct Injuries and Fistulas • Sphincter of Oddi Dysfunction • Stents

Further Reading

American Society for Gastrointestinal Endoscopy (1999). The role of ERCP in diseases of the biliary tract and pancreas: Guidelines for clinical application. *Gastrointest. Endosc.* **50**, 915–920.

Bennett, R. C., and Goligher, J. C. (1962). Results of internal sphincterotomy for anal fissure. *Br. Med. J.*, 1500–1503.

Hoffman, D. C., and Goligher, J. C. (1970). Lateral subcutaneous internal sphincterotomy in the treatment of anal fissure. *Br. Med. J.* **3**, 673–675.

Parkes, A. G. (1967). The management of fissure-in-ano. *Hosp. Med.* **1**, 737.

Sphincters

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

reflux Backward flow or movement, as in reflux of contents from the stomach into the esophagus.
sphincter Ring of circular muscle surrounding and closing an orifice.

Sphincters are present in a variety of locations along the digestive tract. Some of the sphincters are composed of skeletal muscle, others are composed of smooth muscle. Skeletal muscle sphincters are controlled by the central nervous system. Smooth muscle sphincters are controlled by the enteric nervous system. There are two skeletal muscle sphincters, the upper esophageal sphincter and the external anal sphincter. Smooth muscle sphincters are found at the gastroesophageal junction, gastroduodenal junction, ileocolonic junction, and termination of the large intestine in the anus.

SMOOTH MUSCLE SPHINCTERS

Smooth muscle sphincters consist of rings of muscle that remain in a continuous state of contraction. Specialized "latch" mechanisms in the contractile filaments enable sphincters to maintain contractile tone for extended periods with minimal expenditure of energy. The effect of the tonic contractile state is to occlude the lumen in a region that separates two specialized compartments. With the exception of the internal anal sphincter, sphincters function to prevent the backward movement of intraluminal contents. The internal anal sphincter prevents uncontrolled movement of intraluminal contents through the anus.

The lower esophageal sphincter prevents reflux of gastric acid into the esophagus. Incompetence results in chronic exposure of the esophageal mucosa to acid, which can lead to heartburn and dysplastic changes that may become cancerous. The gastroduodenal sphincter, which is sometimes called the pyloric sphincter, prevents excessive reflux of duodenal contents into the stomach. Incompetence of this sphincter can result in the reflux of bile acids and proteolytic enzymes from the duodenum. Bile acids and proteolytic enzymes are damaging to the protective barrier in the gastric mucosa; prolonged exposure can lead to gastritis and ulceration.

The ileocolonic sphincter prevents reflux of colonic contents into the ileum. Incompetence can allow entry of bacteria into the ileum from the colon and may result in bacterial overgrowth. Bacterial counts are normally low in the small intestine.

ENTERIC NERVOUS CONTROL

The ongoing contractile tone in the smooth muscle sphincters is generated by myogenic mechanisms. The contractile state is an inherent property of the muscle and is independent of the nervous system. Transient relaxation of the sphincter to permit the forward passage of material is accomplished by activation of enteric inhibitory motor neurons that innervate the sphincteric musculature.

The inhibitory innervation of the smooth muscle sphincters is transiently activated for timed opening and passage of luminal contents. Smooth muscle sphincters remain tonically contracted, occluding the lumen and thereby preventing the passage of contents between adjacent compartments (e.g., between stomach and esophagus). The inhibitory neurons that innervate the sphincter are normally inactive and are switched on with timing appropriate for coordination of the opening of the sphincter with physiological events in adjacent regions. When this occurs, the inhibitory neurotransmitter relaxes the ongoing muscle contraction in the sphincteric muscle and prevents excitation-contraction in the adjacent muscle from spreading into and closing the sphincter. Damage to or loss of the inhibitory innervation of smooth muscle sphincters results in failure of the sphincter to relax. This results in the sphincter becoming a barrier to the onward passage of the luminal contents from compartment to compartment.

See Also the Following Articles

Achalasia • Barrett's Esophagus • Belching • Defecation • Disinhibitory Motor Disorder • Dysphagia • Enteric Nervous System • Fecal Incontinence • Flatulence • Gastroesophageal Reflux Disease (GERD) • Pylorus • Rumination Syndrome • Swallowing

 Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Papasova, M. (1989). Sphincteric function. *In* "Handbook of Physiology, The Gastrointestinal System I" (J. Wood, ed.), Sect. 6, Vol. 1, Part 2, pp. 987–1023. American Physiological Society, Bethesda.

Wood, J. (2003). Neurogastroenterology and gastrointestinal motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480, Lippincott Williams & Wilkins, Baltimore.

Splenectomy

RUSSELL J. NAUTA

Harvard University and Mount Auburn Hospital

autoantibody An antibody directed against oneself.

cross-sectional imaging Radiologic studies such as computed tomography or magnetic resonance imaging, which display the relevant anatomy as a series of "slices" from which three-dimensional relationships can be inferred.

encapsulated organisms Bacteria with a well-developed cell wall (e.g., pneumococcus).

hematopoiesis Creation of the formed elements of the blood (e.g., red corpuscles).

sequestration Retention of formed elements of the blood as they pass through the spleen, often leading to their destruction or to enlargement of the spleen.

The normal spleen is a fist-sized organ located beneath the left hemidiaphragm. It has immunologic, hematopoietic, and filtering functions, which vary with the patient's age; alteration in these functions in disease may cause recognizable syndromes characterized by changes in splenic size, synthesis and sequestration of the formed elements of the blood, and rate of removal of formed elements of the blood from the circulation. Splenectomy may be indicated for irreparable splenic injury, multiple organ injury, hyperfunction, increase in size, splenic neoplasm, splenic abscess, or aneurysm of the splenic artery. Removal of the spleen is accompanied by increased systemic susceptibility to infection, particularly with encapsulated organisms and particularly in young patients.

INJURY TO THE SPLEEN

The spleen may be injured with a blunt blow to the abdomen with or without overlying rib fracture. Rarely, it may rupture with infectious diseases such as infectious mononucleosis. In cases of rib fracture, the injury caused by the penetrating fragments may be similar to that occasioned by a traversing missile or by impalement. In penetrating injury, suspicion of splenic injury is further aroused by signs and symptoms of blood loss in the presence of broken ribs or an appropriate trajectory of the penetrating foreign body. In blunt trauma, rib fracture, mechanism of injury, or Kehr's sign (left shoulder pain from ipsilateral diaphragmatic irritation) may also raise suspicion. Significant red blood cells in the peritoneum on lavage or fluid on diagnostic ultrasound with a plausible mechanism of injury may suggest the need for laparotomy, as may more specific evidence of organ injury on cross-sectional imaging. Isolated splenic injuries do not require splenectomy in all cases; some that do not involve the hilum may be managed without laparotomy and some injuries may be successfully repaired at laparotomy. When the spleen is one of multiple intra-abdominal organs injured, particularly when one of the others is the colon, splenectomy is preferable to splenic preservation. Although splenectomy for some elective indications may be performed laparoscopically, almost all splenectomies for trauma are performed by open laparotomy because of the hemodynamic urgency of the situation (hemorrhagic shock) and the concomitant difficulties in inspecting both the spleen and the other organs by a laparoscopic approach in the presence of significant blood loss.

Because of its intimate relationship with the colon as this viscus passes through the left upper quadrant and because of the proximity of the spleen to the left kidney, left adrenal, pancreatic tail, and greater curvature of the

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Papasova, M. (1989). Sphincteric function. *In* "Handbook of Physiology, The Gastrointestinal System I" (J. Wood, ed.), Sect. 6, Vol. 1, Part 2, pp. 987–1023. American Physiological Society, Bethesda.

Wood, J. (2003). Neurogastroenterology and gastrointestinal motility. *In* "Medical Physiology" (R. Rhoades and G. Tanner, eds.), 2nd Ed., Chap. 26, pp. 449–480, Lippincott Williams & Wilkins, Baltimore.

Splenectomy

RUSSELL J. NAUTA

Harvard University and Mount Auburn Hospital

autoantibody An antibody directed against oneself.

cross-sectional imaging Radiologic studies such as computed tomography or magnetic resonance imaging, which display the relevant anatomy as a series of "slices" from which three-dimensional relationships can be inferred.

encapsulated organisms Bacteria with a well-developed cell wall (e.g., pneumococcus).

hematopoiesis Creation of the formed elements of the blood (e.g., red corpuscles).

sequestration Retention of formed elements of the blood as they pass through the spleen, often leading to their destruction or to enlargement of the spleen.

The normal spleen is a fist-sized organ located beneath the left hemidiaphragm. It has immunologic, hematopoietic, and filtering functions, which vary with the patient's age; alteration in these functions in disease may cause recognizable syndromes characterized by changes in splenic size, synthesis and sequestration of the formed elements of the blood, and rate of removal of formed elements of the blood from the circulation. Splenectomy may be indicated for irreparable splenic injury, multiple organ injury, hyperfunction, increase in size, splenic neoplasm, splenic abscess, or aneurysm of the splenic artery. Removal of the spleen is accompanied by increased systemic susceptibility to infection, particularly with encapsulated organisms and particularly in young patients.

INJURY TO THE SPLEEN

The spleen may be injured with a blunt blow to the abdomen with or without overlying rib fracture. Rarely, it may rupture with infectious diseases such as infectious mononucleosis. In cases of rib fracture, the injury caused by the penetrating fragments may be similar to that occasioned by a traversing missile or by impalement. In penetrating injury, suspicion of splenic injury is further aroused by signs and symptoms of blood loss in the presence of broken ribs or an appropriate trajectory of the penetrating foreign body. In blunt trauma, rib fracture, mechanism of injury, or Kehr's sign (left shoulder pain from ipsilateral diaphragmatic irritation) may also raise suspicion. Significant red blood cells in the peritoneum on lavage or fluid on diagnostic ultrasound with a plausible mechanism of injury may suggest the need for laparotomy, as may more specific evidence of organ injury on cross-sectional imaging. Isolated splenic injuries do not require splenectomy in all cases; some that do not involve the hilum may be managed without laparotomy and some injuries may be successfully repaired at laparotomy. When the spleen is one of multiple intra-abdominal organs injured, particularly when one of the others is the colon, splenectomy is preferable to splenic preservation. Although splenectomy for some elective indications may be performed laparoscopically, almost all splenectomies for trauma are performed by open laparotomy because of the hemodynamic urgency of the situation (hemorrhagic shock) and the concomitant difficulties in inspecting both the spleen and the other organs by a laparoscopic approach in the presence of significant blood loss.

Because of its intimate relationship with the colon as this viscus passes through the left upper quadrant and because of the proximity of the spleen to the left kidney, left adrenal, pancreatic tail, and greater curvature of the

Copyright 2004, Elsevier (USA). All rights reserved.

stomach, surgery on these organs may produce splenic injury. Most such injuries can be repaired, but persistent bleeding or deep instrumental injury may require splenectomy for hemorrhage control.

SPLENECTOMY FOR ANEMIA

Hereditary spherocytosis is an autosomal dominant disease related to bonding deficiencies of the proteins of the red blood cell membrane. Red cells lose their deformability due to this abnormality and become round (spherocytes) or elliptical (elliptocytes) when viewed on smears of the peripheral blood. Shearing of the red blood cell membrane on passage of blood through the spleen results in an anemia with associated jaundice, reticulocytosis, splenomegaly, and occasionally pigmented gallstones as the aberrant cells are removed from the circulation. Diagnosis is made by examination of the peripheral smear in the proper clinical setting. Splenectomy relieves the anemia, but the cellular deformity persists.

Another autosomally inherited disease, thalassemia, also produces cellular membrane rigidity, but in this case by precipitation of excessive α-hemoglobin chains in the presence of decreased production of β-chains. Sequestration of blood in the splenic cords results in intrasplenic hemolysis of red cells, enlargement of the spleen, and tenderness. Removing the spleen surgically addresses the component of the anemia occasioned by the mechanical lysis, but not that occasioned by the production of the abnormal hemoglobin.

In congenital anemias, an attempt is made to defer removal of the organ well into childhood because of concerns related to overwhelming postsplenectomy sepsis. Removal of the organ before the age of 4 years is justified only by severe disease manifestations.

IMMUNOLOGIC REASONS FOR SPLENECTOMY

The immunologic "tagging" of the formed elements of the blood for premature removal by the spleen is the descriptive explanation offered for the low platelet counts of idiopathic thrombocytopenic purpura (ITP) and the acquired hemolytic anemias. In ITP, determination of the circulating platelet count in patients with petechial hemorrhage in the skin or mucous membranes, bleeding without trauma, or bleeding out of proportion to injury establishes the diagnosis of thrombocytopenia; bone marrow aspirates showing megakaryocytes and the absence of a pharmacologic reason for thrombocytopenia support the diagnosis. Splenectomy is reserved for patients failing courses of steroids or other medical treatment or those requiring chronic steroid administration to maintain an adequate platelet count. Splenectomy is successful in raising the platelet count to asymptomatic levels in 70% of patients; the other 30% presumably retain immunoglobulin G (IgG)-mediated platelet-destructive mechanisms in other organs.

If the IgG autoantibody tags red blood cell membrane proteins, red cells adhere to splenic macrophages, which destroy them. Steroids again are the front-line treatment, with salvage splenectomy in the event of failure of medical management producing favorable responses in the ranges reported for ITP.

A splenectomy is effective for hemolytic anemia only if the antibodies are warm-reacting; hemolysis for cold-reacting antibodies is IgM-mediated and occurs intravascularly rather than within the spleen. As with ITP, the persistence of IgG (warm antibody)-mediated hemolysis after splenectomy indicates a robust macrophage depot elsewhere in the reticuloendothelial system, liver, or bone marrow.

MECHANICAL HYPERSPLENISM

Hypersplenism is a condition characterized by pancytopenia and splenomegaly in the presence of an active bone marrow. Primary treatment is administration of corticosteroids, with splenectomy reserved for pharmacologic failure.

When the abnormal cells of a myeloproliferative disorder, lipid storage disease, or lymphoproliferative disorder infiltrate the splenic red pulp, blood flow through the organ is diminished, leading to red blood cell glucose deprivation and cell death by hemolysis. To the extent that the marrow is infiltrated as well, failure of red cell production may augment the anemia. Production of other formed elements may also be suppressed by the infiltrate. Splenectomy produces variable results, depending on whether extramedullary hematopoiesis augments formed element counts, and is usually preceded by a course of steroids or chemotherapy. When salutary, it provides symptomatic relief (as the palpable and symptomatic enlargement of such spleens leads to early satiety) and may facilitate the administration of drugs to treat the primary disorder.

The spleen is often enlarged as part of the rise in tributary pressures of the valveless portal system in patients with portal hypertension. When asymptomatic, splenectomy for enlargement alone is not indicated. Splenic vein decompression with portasystemic shunt or transjugular intrahepatic portasystemic shunt, rather than splenectomy, is the treatment is of choice for thrombocytopenic portal hypertensive patients if the splenic vein is patent. In cases where bleeding gastric varices occur because of splenic vein thrombosis,

splenectomy ameliorates the bleeding in controlling back pressure from the thrombosed splenic vein, which is transmitted to the gastric varix. It is by this decompression that the hemorrhage abates.

Splenectomy for mechanical hypersplenism carries an operative mortality rate of 15–30%, as opposed to less than 5% for elective splenectomy for nonmechanical causes.

MISCELLANEOUS REASONS FOR SPLENECTOMY

The spleen may be an embolic site for bacteria from an infected heart valve or traumatic site and splenic abscess may ensue. Diagnosis is made by noting the abscess on cross-sectioned imaging or observing a defect on liver–spleen scan performed for persistent infection. Splenic abscess is treated by splenectomy.

Aneurysms of the arterial blood supply to the spleen may rupture and require splenectomy. Rupture, however, is too rare a complication to produce a uniform recommendation for prophylactic splenectomy. Women having a splenic artery aneurysm and contemplating pregnancy should be advised of the increased risk of rupture during pregnancy.

TECHNIQUES OF SPLENECTOMY

The spleen may be removed by open surgery or by a laparoscopic technique. In both techniques, the separation of the greater curvature of the stomach from the splenic flexure of the colon occurs by dividing the fatty tissue in the gastrocolic omentum. As the dissection is carried cephalad along the greater curvature of the stomach, the greater curvature may be brought forward and folded upward to expose the splenic artery originating from the celiac axis; the short gastric vessels are seen and divided high on the greater curvature. In operations performed for low platelet counts (e.g., ITP), the splenic artery may be secured on visualization, allowing platelet transfusion before division of the remaining vessels and mobilization of the spleen.

STAGING OF HODGKIN'S DISEASE

The spleen is sometimes involved with malignant lymphoma. Operative removal of the spleen, liver biopsy, and lymph node sampling were commonplace before computed tomography and lymphangiogram were shown to accurately stage Hodgkin's disease in most cases. Such "staging laparotomy" is now reserved for individuals with equivocal imaging where differences in treatment result from knowing whether or not the spleen is involved.

COMPLICATIONS OF SPLENECTOMY

Removal of the spleen may be accompanied by bleeding. This may be from a named vessel or from surgical division of splenic capsular adhesions to the abdominal side wall or diaphragm. Thrombocytopenic conditions increase the likelihood of bleeding from such sites.

Damage to the pancreatic tail during splenectomy may result in pseudocyst formation as pancreatic ductal integrity is lost and the fibroblastic containment is incomplete. When the pancreatic tail is thought to have been potentially injured during splenectomy, a drain is left in the splenic fossa; high amylase levels from the drainage effluent suggest that the drain be retained. The resultant pancreaticocutaneous fistula will often close without surgery. Left subphrenic abscess from superinfection of blood or pseudocyst contents may require radiologic or open drainage. Howell-Jolly bodies (nucleated red cells) and thrombocytosis (sometimes to greater than 1,000,000 platelets per cubic centimeter) may occur following splenectomy. Antibody production by the spleen stops with splenectomy, with a concomitant decrease in opsonization of encapsulated bacteria.

Overwhelming postsplenectomy sepsis represents the end result of a predisposition to infection with pneumococcus, meningococcus, and *Haemophilus influenza* bacteria as a consequence of splenectomy. Most fatalities occur in children and concerns for the development of postsplenectomy sepsis have prompted some to recommend long-term penicillin therapy following splenic removal in the pediatric population. In the adult population, increased awareness of the potential for infection and lower threshold for antibiotic treatment of upper respiratory infections are suggested in postsplenectomy patients.

In most cases of elective splenectomy, the operation is preceded by vaccination for the encapsulated organisms; the resultant antibodies may be at least transiently protective.

See Also the Following Article

Laparoscopy

Further Reading

Asknes, J., Adelnour, M., and Mathisen, O. (1995). Risk factor associated with mortality and morbidity after elective splenectomy. *Eur. J. Surg.* **161**, 253–258.

George, J., El-Harake, M., and Raskob, G. (1994). Chronic idiopathic thrombocytopenic purpura. *N. Engl. J. Med.* **331**, 1207–1211.

Lambert, C., and Williamson, J. (1990). Splenic artery aneurysm: A rare case of upper gastrointestinal bleeding. *Am. Surg.* **56**, 543.

Stents

RUSSELL J. NAUTA AND ALLISON C. NAUTA
Harvard University and Mount Auburn Hospital

ampulla of Vater The aperture surrounded by the sphincter of Oddi.
sphincter of Oddi Circular muscle located at the aperture where common bile duct and pancreatic ducts join the duodenum.
T-tube Synthetic tube in the shape of a "T" whose transverse limb stents the site of an incision in the bile duct and whose vertical limb is brought through that incision and through the abdominal wall. Biliary anatomy may be defined by the injection of contrast material through the vertical arm and stones may be manipulated through it.

A stent is a tubular prosthesis introduced into a hollow viscus to ensure that the viscus retains a minimum functional diameter despite obstruction occasioned by occlusion of the visceral lumen by stone, tumor, or extrinsic compression. Stents may be used temporarily in situations in which resolution of the obstruction is likely to occur or to ensure minimal visceral diameter in diverse processes (e.g., cancer) wherein the compromise of visceral diameter is anticipated to be permanent or progressive. Typical sites for visceral stent placement in decreasing order of frequency are the biliary tree, the esophagus, the pancreatic duct, and the sigmoid colon. Neoplastic disease of gastrointestinal origin may occasionally lead to permanent placement of stents in the ureter(s) to permit continued passage of urine from the upper urinary tracts to the bladder. Some surgeons use temporary ureteral stents as a means of palpably identifying the ureters during operations on the sigmoid colon.

Stents (named for a 19th century dentist, Dr. William Stent) are typically introduced under endoscopic or fluoroscopic guidance and placed in a position where the prosthesis will allow normal passage of food or secretions through the narrowed segment because both of its ends are in undiseased or unobstructed regions of the viscus (Fig. 1). Visceral stents may be solid or webbed and are typically fashioned of plastic or metal.

THE BILE DUCT

The unobstructed biliary tree is "self-flushing" by the four cooperative processes of ongoing bile production, relaxation of the sphincter of Oddi with meal ingestion, propulsion of bile by gallbladder contraction, and intrinsic peristalsis of the biliary tree. Stasis of bile behind malignant strictures, benign strictures, or stones predisposes to cholangitis, as indigenous bacteria, failing to be expelled into the gut, proliferate. This may lead to local or systemic infection. Instrumentation of the biliary tree endoscopically or transhepatically may introduce exogenous bacteria into an obstructed system as an unwanted by-product of therapy.

Biliary Obstruction Secondary to Stone Disease

The endoscopic release of biliary obstruction due to stones involves retrograde manipulation of the biliary tree and extraction of stones through the ampulla of

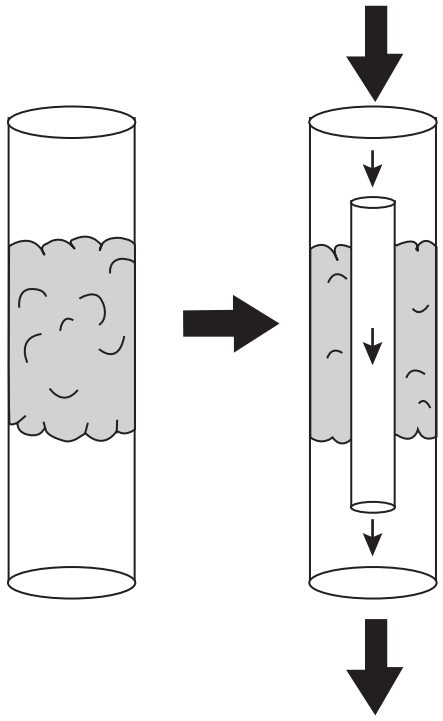


FIGURE 1 A stent traverses an area of obstruction, permitting flow through it. The ends of the stent are in normal areas of the obstructed organ.

 Copyright 2004, Elsevier (USA). All rights reserved.

Vater; an operating endoscope traverses the mouth, pharynx, esophagus, and stomach to reach the duodenum. Endoscopic access to the ampulla for stone extraction and biliary drainage is facilitated by enlargement of the ampulla of Vater by cutting the surrounding sphincter of Oddi (cf. sphincterotomy). Confidence that swelling or a retained stone would not result in cholangitis from endogenous or introduced bacteria is enhanced by the introduction of temporary stents through the endoscope to span the ampulla at the end of the extraction procedure. One end of such a stent rests in the bile duct and the other rests in the duodenum. As with T-tubes left in the bile duct after surgical manipulation, stents are typically removed when stones are documented to be absent on a subsequent radiologic study and/or when swelling from the manipulation of sphincterotomy has subsided and free drainage of bile from the duct can be documented.

Biliary Obstruction Secondary to Pancreatitis

The most common reason for extrinsic compression of the bile duct is inflammation or injury of the pancreatic head, with resultant swelling of the ampulla. Such swelling can cause derangement of liver function tests similar to that seen in obstructing stone or tumor. Diagnostic endoscopic retrograde cholangiopancreatogram (ERCP) with therapeutic stenting of a bile duct obstructed from pancreatitis may exclude stones as the cause of the inflammation and normalize elevated levels of alkaline phosphatase, transaminases, total bilirubin, and conjugated bilirubin. Treatment of the pancreatitis itself is supportive (cf. pancreatitis) and some such structures resolve with the resolution of the inflammation.

Malignant Obstruction of the Biliary Tree and Esophagus

Malignant obstruction of the common bile duct secondary to neoplasm of the distal bile duct, ampulla of Vater, or pancreatic head produces elevation of liver function tests. ERCP shows neither a silhouetted stone in transit nor the smooth stricture of pancreatitis, but an abrupt end of both pancreatic and bile ducts in the region of the ampulla (the so-called double duct sign). Endoscopic manipulation of the ampulla for diagnostic cytology or retrograde injection of the contrast to define the anatomy during ERCP may introduce bacteria into an obstructed system. Stent placement may allow normalization of liver function tests in anticipation of an operation performed to resect or bypass the neoplasm with curative or palliative intent. Alternatively, in patients at high risk or with metastatic disease, stent placement is intended to be permanent and to provide a palliative alternative to open surgery. Because of the stent's small caliber and resultant tendency to obstruct with biliary debris, endoscopic replacement is often required or performed prophylactically at 3-month intervals.

With esophageal malignancies, lifetime stenting is considered as an alternative to surgery because a majority of patients subjected to "curative" surgery for malignant stricture in this region do not survive 5 years.

STENTING OF THE PANCREAS

Repeated bouts of alcoholic pancreatitis produce segmental stricturing of the pancreatic duct with stone formation and pain. ERCP defines whether the fibrotic ductal changes as the gland heals are obliterative or obstructive. Surgical decompression of the duct with longitudinal pancreaticojejunostomy (Puestow procedure) or ablative procedures of the tail [distal pancreatectomy or retrograde drainage of the pancreatic duct (the Duval procedure)] have wide variation in success of relieving pain and allowing withdrawal of narcotics depending on whether the pain is due to obstruction or inflammation. Interventional failures may be multifactorial in this population. Alcoholic recidivism may cause painful recurrent attacks of acute pancreatitis.

Permanent stenting of the pancreatic duct as an alternative to surgery to relieve the pain of chronic obstruction is controversial, as the small caliber of the stents requires frequent changes for occlusion. Some have suggested stents as a predictive tool to define who would benefit from a surgical decompressive approach and traverse the obstruction in the pancreatic duct. Such stents are introduced through the ampulla during ERCP. The absence of relief from stenting predicts a minimal impact of surgical duct decompression on narcotic requirements.

STENTING OF THE COLON

Placement of a stent in the sigmoid colon to facilitate bowel preparation for an obstructing cancer or as an alternative to surgical resection or diversion is controversial. Placed retrograde through a colonoscope, the prosthesis clogs easily with solid fecal material indigenous to the left colon; it may also dislodge, erode, cause bleeding, or pass with the stool. The use of stents to ensure a minimal diameter for the sigmoid colon competes unfavorably with surgery; even with metastatic disease, most patients with sigmoid colon cancer can be surgically resected or diverted. Few patients able to

have sigmoid resection are so obstructed that no mechanical preparation of the bowel can be carried out. Intraoperative alternatives to stenting, such as on-the-table bowel irrigation, further limit the role of stenting.

STENTING AS AN ADJUNCT TO SURGERY

When surgical resection or bypass involves anastomosis of a small duct, such as the pancreatic duct or bile duct, to the small intestine, some surgeons advocate temporary stents to ensure a guaranteed anastomotic diameter, thereby ensuring patency and facilitating drainage as anastomoses heal. The absence of controlled studies makes it impossible to determine whether such an approach is beneficial; for larger anastomoses (e.g., of ducts enlarged by chronic obstruction), most surgeons do not use such a prosthesis.

URETERAL STENTS AND THE GASTROINTESTINAL TRACT

Ureteral stents are used in association with treatment of gastrointestinal disease in two circumstances: to identify and preserve ureters during operations on the gastrointestinal tract in which such identification might be difficult and to ensure renal function in obstructing pelvic cancers.

Stents placed to avoid ureteral injury are introduced through retrograde cystoscopic cannulation of the ureteral orifices with rigid tubes. In operations on the lower gastrointestinal tract (e.g., sigmoid colon for diverticulitis or malignant disease), such identification is said to help some surgeons avoid injury.

For advanced pelvic malignancy, the need for stents is often identified by alteration in renal function or demonstration of distal ureteral obstruction on a radiographic study such as ultrasound or computed tomography. Stenting is preferable to anuria in circumstances in which a therapeutic impact on the obstructing malignancy is likely. When obstructing malignancy is unlikely to be able to be successfully treated, ureteral obstruction may be left unstented, as the toxicity of renal failure occasions a death preferable to that of a painfully growing untreatable pelvic malignancy. When chronic ureteral stenting is used, stent changes are often required because of debris collection in the small-lumen tubes.

See Also the Following Articles

Bile Duct Injuries and Fistulas • Sphincterotomy

Further Reading

Assimus, D. G., Patterson, L. C., and Taylor, C. L. (1994). Changing incidence and etiology of iatrogenic ureteral injuries. *J. Urol.* **152**, 2240–2246.

Carr-Locke, D. (1999). Endoscopic therapy of chronic pancreatitis. *Gastrointest. Endosc.* **49**, 577–580.

Ho, H. S., and Frey, C. (1997). Current approaches to the surgical management of chronic pancreatitis. *The Gastroenterologist* **5**, 128–136.

Savage, P. T. (1967). Immediate resection with end-to-end anastomosis for carcinoma of the large bowel presenting with acute obstruction. *Proc. R. Soc. Med.* **60**, 207–209.

Stear, M., Waxman, I., and Freedman, S. (1995). Chronic pancreatitis. *N. Engl. J. Med.* **332**, 1482–1490.

Stomach, Adenomas and Carcinomas of the

GORDON D. LUK
Digestive Health Associates of Texas, Bedford

acanthosis nigricans Darkly pigmented velvety patches of skin folds found in association with gastrointestinal malignancies.
chromoendoscopy Spraying of a dye that can be seen endoscopically and that may highlight surface irregularities or specific histologic features.
dermatomyositis Inflammatory myopathy manifested by symmetric proximal muscle weakness and skin rash, often in association with malignancies.
endoscopic ultrasound (endosonography) Use of an ultrasound transducer at the endoscope tip to obtain sonographic images of anatomic structures adjacent to the endoscope.
incidence Number of new cases of a specific disease found in a defined population over a specific time period (such as the number of new cases over a 1-year period).
Leser–Trelat sign Sudden appearance of multiple seborrheic keratoses, often concomitant with acanthosis nigricans, found in association with gastrointestinal malignancies.
premalignant polyp Cell growth that has the potential of progressing to a malignancy over time.
prevalence Number of existing cases of a specific disease found in a defined population at one specific point in time.
Trousseau's syndrome Migratory superficial thrombophlebitis occurring as a result of a hypercoagulable state, often in association with or preceding the diagnosis of cancer.

Gastric polyps are nonmalignant, protruding, intraluminal masses inside the stomach, found in 1–2% of patients. Adenomas account for about 10% of gastric polyps. Adenomas are premalignant polyps, and are believed to progress to malignant gastric neoplasms over time. Nonadenomatous polyps generally have no documented malignant potential (with the possible exception of gastric fundic polyps in patients with familial adenomatous polyposis). Gastric carcinomas (also called gastric cancers), or adenocarcinomas of the stomach, include carcinomas in the cardia, which also involve the esophagogastric junction. They do not include squamous cell carcinomas (almost universally with an esophageal primary) involving the esophagogastric junction. Carcinomas of the stomach account for about 90% of all malignant gastric neoplasms, which also include lymphomas, stromal cell tumors, and carcinoids.

ADENOMAS OF THE STOMACH

Adenomas of the stomach are found throughout the stomach, but are more common in the antrum and along anastomoses. They are variable in size, ranging from a few millimeters to several centimeters in diameter. They are true premalignant gastric polyps. Their malignant potential increases with increasing size, increasing villous component, higher degree of dysplasia, and multiplicity in number. In adenomas larger than 2 cm, there is about a 10% frequency of carcinoma within the adenoma.

Although there have been no large-scale prospective studies, some reports suggest an association of gastric adenomas with colon adenomatous polyps. Patients with gastric adenomas may have a fourfold higher risk of having colon adenomatous polyps, and those with colon adenomatous polyps may have up to a 20-fold increase in gastric adenomas. A special clinical setting involves the patient with familial adenomatous polyposis. These patients often have multiple fundic gland polyps. Sporadic fundic gland polyps in patients without familial adenomatous polyposis appear to have no malignant potential. However, in patients with familial adenomatous polyposis, their fundic gland polyps may be pathogenetically distinct from sporadic fundic gland polyps, carry genetic alterations, and have malignant potential. In patients with familial adenomatous polyposis, polypectomy or biopsies of multiple fundic gland polyps are needed to detect adenomatous or malignant changes. In addition, because of the increased frequency of gastroduodenal cancers in such patients, periodic surveillance endoscopy is indicated.

Small gastric adenomas less than 1 cm in size, like most other small gastric polyps, are usually asymptomatic. But larger adenomas, because of their propensity to be located in the antrum, may cause gastric outlet obstruction, epigastric distress or pain, as well as occult or overt bleeding. With the wide use of endoscopy examinations, most gastric adenomas are diagnosed during endoscopic examinations, usually for abdominal pain or occult blood loss. When found, all gastric polyps should be excised totally by polypectomy if feasible. If total

 Copyright 2004, Elsevier (USA). All rights reserved.

excision is not possible, a focus of invasive carcinoma could very well be missed. For multiple gastric polyps, the largest five or six should be removed totally by polypectomy, followed by biopsy of a representative sample of the remainder. For adenomas larger than 2 cm, endoscopic ultrasound may be helpful in determining which layers of the stomach wall are involved. If invasive carcinoma is found or if there are multiple large gastric adenomas, a subtotal gastrectomy might be indicated. Once gastric adenomas are found, the patient probably should have a surveillance endoscopy at 1 year. If there is recurrence of adenomatous polyps or if adenomatous tissue is present on biopsy, surveillance endoscopy 1 year later is again indicated. If the repeat examination at 1 year is negative, the next endoscopy could be delayed for 3 to 5 years, analogous to the current recommendation for the surveillance of colon adenomatous polyps. No surveillance is indicated if the gastric polyp is nonadenomatous.

Several small-scale studies suggest that *Helicobacter* eradication may lead to regression of gastric adenomas and delay or inhibit their progression to cancer. Eradication of *Helicobacter* in patients with gastric adenomas and concomitant *Helicobacter* infection would appear warranted.

CARCINOMAS OF THE STOMACH

Introduction

In the United States, carcinoma of the stomach is the eleventh most common type of cancer and is the fourteenth leading cause of cancer death. For the year 2003, 22,400 new cases of, and 12,100 deaths from, carcinoma of the stomach were projected. Worldwide, carcinoma of the stomach remains the second most common cancer and the second most common cause of cancer death, despite a continuing worldwide decline in prevalence and death rate over the past seven decades. Even with the decline in the overall incidence of carcinomas of the stomach, the decline is primarily that of carcinomas in the distal stomach; whereas the incidence of carcinomas in the gastric cardia has been increasing rapidly in the past three decades. This appears to be accounted for primarily by an increase in carcinomas associated with Barrett's esophagus.

Worldwide, there is also a 10-fold or greater difference in the incidence and prevalence of carcinomas of the stomach, with the highest incidence in Japan, Korea, China, and eastern Europe, and the lowest incidence in North America, western Europe, Australia, and New Zealand. Even within the same country, incidence can be markedly variable. For example, there is a high incidence in the mountainous regions of Columbia, but not in the coastal regions. This regional difference has been attributed to environmental factors. Japanese immigrants in the United States have about a 25% reduction in incidence of gastric carcinomas compared to the general Japanese population. The second generation has more than a 50% reduction, and subsequent generations have an incidence comparable with that of the general United States population.

Etiology and Risk Factors

The pathogenesis of carcinomas of the stomach is most likely multifactorial. There has long been postulated a sequence of histologic premalignant changes, progressing from atrophic gastritis to intestinal metaplasia and ultimately to carcinoma. These premalignant histologic changes may be necessary but clearly are not sufficient. The existence of a genetic predisposition is suggested by the finding of carcinoma of the stomach in patients with Lynch syndrome II, one of the hereditary nonpolyposis colorectal cancer syndromes. Many patients with familial adenomatous polyposis also develop adenomas and carcinomas of the stomach. In addition, studies have shown an approximately twofold increase in the relative risk of carcinomas of the stomach in twins and in first-degree relatives of patients with gastric carcinomas.

Currently, it appears that the most clinically significant risk factor is *Helicobacter pylori* infection. Patients with *Helicobacter* infection have a three- to eight-fold increased risk of developing carcinomas of the stomach. The exact pathogenetic mechanisms have not been defined. With the increasingly frequent detection of *Helicobacter* infection, specifically in association with peptic ulcer disease, there is ongoing use of antibiotics to eradicate chronic *Helicobacter* infections in developed countries. This may eventually help fuel further declines in the prevalence and incidence of carcinomas of the stomach.

Many other risk factors have also been proposed for carcinomas of the stomach. However, most of these other risk factors have been associated with only small and inconsistent increased risks. The strongest risk factors, other than heredity and *Helicobacter* infection, include pernicious anemia, previous gastrectomy or gastric surgery, chronic atrophic gastritis, and intestinal metaplasia.

Clinical Manifestations

The clinical presentation of carcinomas of the stomach is dramatically different in Japan compared to the rest of the world. Japan has very high incidence and has had a

national screening program in place since the 1960s. Perhaps as a result of the aggressive screening program, carcinomas of the stomach tend to be at an early stage when diagnosed in Japan. Early gastric cancer, defined as carcinoma limited to only the mucosa and submucosa, accounts for up to 50% of cases of gastric carcinomas in Japan. However, early gastric cancer accounts for fewer than 20% of cases in the United States.

Early gastric cancer tends to be asymptomatic in up to 80% of patients. In advanced gastric carcinoma, i.e., in patients with later stage disease, the most common findings are weight loss and abdominal pain. In addition, nausea and vomiting, anorexia, dysphagia, occult or overt gastrointestinal blood loss, early satiety, and symptoms of peptic ulcer disease are found in more than one-quarter of patients. All of these symptoms are, of course, nonspecific. In addition, advanced gastric carcinomas may also present with nodal metastases and intraabdominal metastases, and their associated signs and symptoms, including adenopathy, ascites, and gastrointestinal obstruction. Paraneoplastic conditions, such as Trousseau's syndrome, acanthosis nigricans, Leser–Trelat sign, and dermatomyositis, have been described.

Diagnosis

The diagnosis of carcinoma of the stomach is best made by endoscopy with biopsy. The endoscopic appearance of carcinomas can range from polypoid masses, often with ulcerations, to ulcerating masses, and to superficial and infiltrating lesions. Both early and advanced carcinomas have been classified morphologically, but the morphologic classifications have limited clinical value for staging, treatment, or prognosis.

Definitive tissue diagnosis requires endoscopic biopsy. Diagnostic accuracy is highest when at least six biopsies are taken. If an ulcer is present, biopsies should be taken from the edge and base of the ulcer and not from the necrotic debris. Chromoendoscopy with dye staining may help highlight suspicious areas to guide targeted biopsy. Contrast radiological studies may be used for diagnosis, but a followup endoscopic examination with biopsy is still indicated to confirm tissue diagnosis and guide subsequent treatment and management.

Imaging studies may also be helpful in staging. Computed tomography (CT) scans are especially helpful in detecting pulmonary or hepatic metastases, as well as intraabdominal and large peritoneal metastases. CT scans often underestimate the size and extent of involvement of the primary tumor and can miss small intraabdominal and peritoneal metastases of 5 mm or less, and are therefore prone to understaging. Endoscopic ultrasound is most helpful for defining the extent of the primary tumor. It can have up to 90% accuracy in defining the depth of invasion, and this is especially helpful in differentiating early gastric cancer from more advanced disease. Endoscopic ultrasound is also helpful in detecting perigastric lymph node involvement. In addition, endoscopic ultrasound-guided fine needle aspiration can enhance staging accuracy. However, distant lymph nodes, the liver, and the lungs cannot be imaged satisfactorily. In patients who present higher surgical risks, some surgeons also prefer preoperative laparoscopy to detect small peritoneal metastases and to determine surgical resectability.

Screening

Japan is the only country that has published results of large nationwide population screening programs for carcinomas of the stomach. Screening was initially by double-contrast radiographs, then by intragastric cameras, but is now done totally by endoscopy. About 7 million persons are screened and about 7000 cases of carcinomas of the stomach are identified annually. This accounts for more than 100 cases per 100,000 persons screened. More than 50% of gastric carcinomas detected by screening are early gastric cancers. About 90% have no lymph node involvement, undergo curative resection, and have 5- and 10-year survival rates of greater than 90%. Further analysis shows that persons undergoing screening have 50% lower risks of dying from carcinomas of the stomach compared to those who do not participate in screening.

In the rest of the world, including the United States, the incidence of carcinomas of the stomach is markedly lower and renders screening less clinically effective and thus much less cost-effective. In the United States, even in persons with identified predisposing risk factors, such as *Helicobacter* infection, pernicious anemia, previous gastrectomy, chronic atrophic gastritis, and intestinal metaplasia, screening does not appear to have value. Although there are no published data or clinical trials, relative risk data suggest that screening can be recommended only for patients with identified adenomas of the stomach, as well as those with familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer.

Staging

The most important factor in determining the curative resectability and prognosis of carcinomas of the stomach is clinical stage. There are minor differences in the staging systems currently in use, but most systems are comparable. Most generally accepted staging

systems are based on the primary tumor (T), nodal involvement (N), and metastasis (M) (TNM) system, whereby these factors are used to determine a clinical stage. Patients with carcinoma (tumor) *in situ* (Tis) or intraepithelial neoplasia have stage 0 disease, and have a chance of 5-year survival of close to 100%. Those with T1 tumors, limited to the mucosa and submucosa (also termed early gastric cancer), are generally stage I, and have 5-year survival rates of higher than 90%. T2 tumors, with involvement of the muscularis propria, generally fall into stages II and III, and patients have 5-year survival rates of about 50%. Patients with T3 tumors, which involve the serosa and are generally stage III, have 5-year survival rates of about 20%. Patients with T4 tumors, which involve adjacent organs and structures, almost always have nodal involvement or distant metastases. They usually have stage IV disease and have virtually a 0% 5-year survival rate (Figs. 1 and 2).

Treatment

Surgical resection of the primary carcinoma and excision of adjacent involved lymph nodes remain the standard of treatment. Even with advanced disease that is not amenable for cure, surgical resection with palliative intent remains the most effective way of providing symptomatic relief, with relief of abdominal pain and obstructive symptoms in more than 50% of patients. Therefore, laparotomy or laparoscopy with curative intent and to optimize palliation should be carried out in virtually all patients except those who are not surgical candidates or those with advanced distant metastases and no local obstructive symptoms. For patients with polypoid early gastric cancers, Japanese endoscopists will perform endoscopic mucosal resection, although this technique has not seen wide use in the United States. For those patients not amenable to surgical resection or palliation, or for those in whom surgical palliation is not possible, endoscopic balloon dilation, stent placement, thermal therapy, laser therapy, or photodynamic therapy can provide palliation in selected patients.

The high recurrence and relapse rates after curative and palliative resections have prompted numerous studies on preoperative and postoperative adjuvant therapy. So far, there is no evidence that preoperative adjuvant radiation therapy or chemotherapy has any survival benefit. Similarly, postoperative radiation therapy has also not shown any survival benefit. Postoperative adjuvant therapy, as well as chemotherapy for advanced nonresectable disease, has shown a small survival benefit. More than 30 chemotherapeutic agents, used as single agents or in combination, have been tested in clinical trials. Successful single-agent or combination chemotherapy regimens have shown about a 30% response rate, almost always partial responses, but no or only limited survival benefit. The current standard

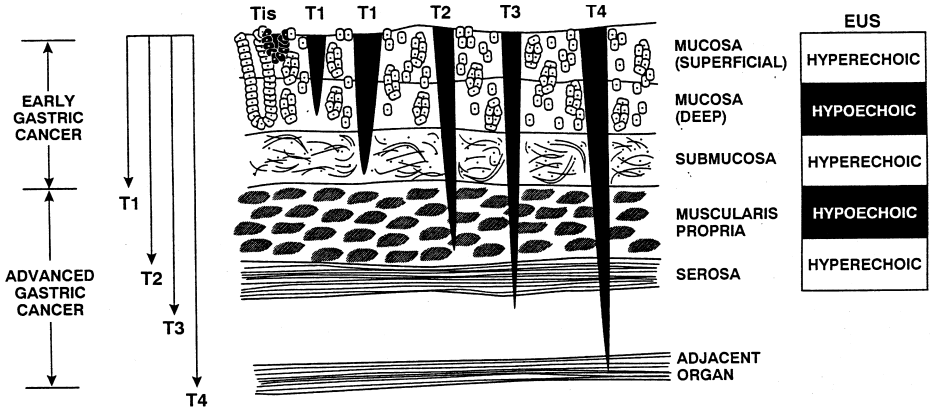
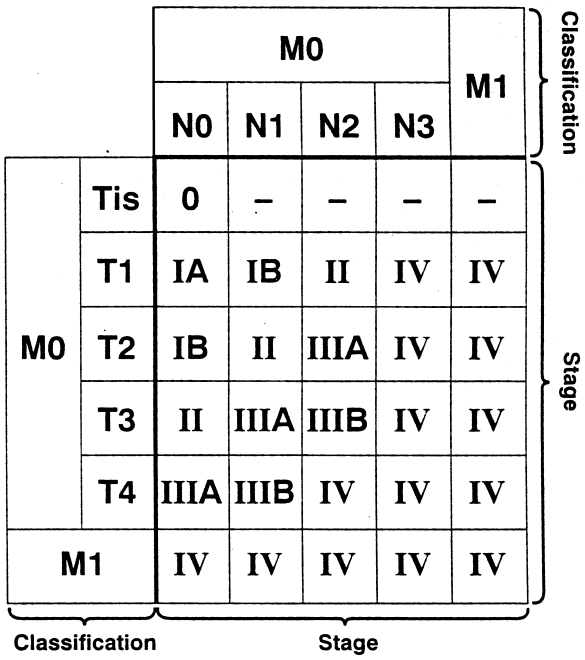


FIGURE 1 Classification of gastric carcinomas by depth of primary tumor invasion (T classification). In the tumor/node/metastasis (TNM) classification, T denotes depth of invasion. Tis designates carcinoma (tumor) *in situ* (intraepithelial neoplasia); T1 tumors are confined to the mucosa and submucosa, T2 tumors penetrate the muscularis propria but not the serosa, T3 tumors penetrate the serosa without involving contiguous structures, and T4 tumors penetrate the serosa and involve adjacent organs and tissues. In early gastric cancer, the disease is confined to the mucosa and submucosa, without regard to size or nodal involvement, and is equivalent to T1 tumors. The layers of the gastric wall may be visualized by endoscopic ultrasound (EUS) as five layers, alternately hyperechoic (bright) and hypoechoic (dark). Reproduced with permission from Figure Luk, G. D. (1998). Tumors of the stomach. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.) 6th edition, page 743. Copyright W. B. Saunders.



Stage	5-Year Survival
0	100%
IA	95%
IB	82%
II	55%
IIIA	30%
IIIB	15%
IV	2%
EGC (T1)	90%

FIGURE 2 Tumor/nodule/metastasis (TNM) staging of gastric carcinoma and relationship to survival. This common staging system is based on the TNM classification. T classification is by depth of primary tumor invasion (see Fig. 1), and N classification is by nodal involvement; N0 represents no nodal involvement, N1 represents involvement of perigastric nodes within 3 cm of the primary tumor, N2 represents involvement of more distant perigastric nodes and regional nodes that are amenable to removal at gastrectomy, and N3 represents involvement of more distant intraabdominal nodes that are not removable at surgery. Stage 0 represents Tis, carcinoma (tumor) *in situ*, which is not a true clinical malignancy. Stage IA represents the earliest stage of malignancy, progressively advancing through stages IB, II, IIIA, IIIB, and finally stage IV. Stage IV represents disseminated metastatic disease, with the presence of M1 at any T or N stage. Although it is theoretically possible to have stage IV disease that is T1N0M1, most stage IV disease is T3 or T4 with at least N2 involvement. It is also important to note that early gastric cancer, which is T1 disease, may be a separate clinical entity with a highly favorable prognosis and for which gastrectomy is curative, at least in Japan. Reproduced with permission from Figure Luk, G. D. (1998). Tumors of the stomach. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.) 6th Ed., p. 744. Copyright W. B. Saunders.

combination regimen is that of epirubicin, cisplatin, and fluorouracil (ECF). This has been shown to have about a 50% response rate, with a few complete responses, and to confer about a 3-month survival advantage. For patients who cannot tolerate or have failed standard therapy, there are generally ongoing clinical trials sponsored by The National Cancer Institute, and patients who meet eligibility criteria may benefit by being enrolled in these clinical trials. It appears that most new advances in the early detection and management of adenomas and carcinomas of the stomach will emanate from Japan, where the disease is much more prevalent and where there are well-established national screening and research programs.

See Also the Following Articles

Atrophic Gastritis • Cancer, Overview • Familial Adenomatous Polyposis (FAP) • Gastrectomy • Gastric Cancer Surveillance • Gastric Polyps • Gastric Surgery • *Helicobacter pylori* • Pernicious Anemia

Further Reading

Abraham, S. C., Nobukawa, B., Giardiello, F. M., Hamilton, S. R., and Wu, T. T. (2000). Fundic gland polyps in familial adenomatous polyposis: Neoplasms with frequent somatic adenomatous polyposis coli gene alterations. *Am. J. Pathol.* **157**, 747–754.

American Cancer Society. (2003). Stomach cancer web site: http://www.cancer.org/docroot/home/index.asp.

Archimandritis, A., Spiliadis, C., Tzivras, M., Vamvakousis, B., Davaris, P., Manika, Z., and Scandalis, N. (1996). Gastric epithelial polyps: A retrospective endoscopic study of 12974 symptomatic patients. *Ital. J. Gastroenterol.* **28**, 387–390.

Boland, C. R., and Savides, T. J. (1999). Tumors of the stomach. *In* "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, C. Owyang, L. Laine, and D. W. Powell, eds.), 3rd Ed., pp. 1500–1529. Lippincott Williams & Wilkins, Philadelphia.

Conlon, K. C., and Brennan, M. F. (2000). Laparoscopy for staging abdominal malignancies. *Adv. Surg.* **34**, 331–350.

De Vivo, R., Pignata, S., Palaia, R., Parisi, V., and Daniele, B. (2000). The role of chemotherapy in the management of gastric cancer. *J. Clin. Gastroenterol.* **30**, 364–371.

Greenlee, R. T., Hill-Harmon, M. B., Murray, T., and Thun, M. (2001). Cancer statistics 2001. *CA Cancer J. Clin.* **51**, 15–36.

Luk, G. D. (1998). Tumors of the stomach. *In* "Gastrointestinal and Liver Disease" (M. Feldman, B. F. Scharschmidt, and M. H. Sleisenger, eds.), 6th Ed., pp 733–757. W. B. Saunders, Philadelphia.

National Cancer Institute. (2003). NCI CancerNet web sites: Stomach (Gastric) Cancer, http://cancernet.nci.nih.gov/Cancer_Types/Stomach_(Gastric)_Cancer.shtml; What You Need to Know about Stomach Cancer, http://cancernet.nci.nih.gov/wyntk_pubs/stomach.htm; NCI Clinical Trials, News and Search, http://cancertrials.nci.nih.gov/index.html.

National Institutes of Health (2003). NIH Medline Plus stomach cancer web site: http://www.nlm.nih.gov/medlineplus/stomachcancer.html.

Saito, K., Arai, K., Mori, M., Kobayashi, R., and Ohki, I. (2000). Effect of *Helicobacter pylori* eradication on malignant transformation of gastric adenoma. *Gastrointest. Endosc.* **52**, 27–32.

Yoshikane, H., Sakakibara, A., Hidano, H., Niwa, Y., Goto, H., and Yokoi, T. (2001). Piecemeal endoscopic aspiration mucosectomy for large superficial intramucosal tumors of the stomach. *Endoscopy* **33**, 795–799.

Stomach, Anatomy

DANIEL R. CLAYBURGH AND JERROLD R. TURNER
The University of Chicago

hiatal hernia Protrusion of part of the stomach, usually the cardia, into the thoracic cavity through the esophageal opening of the diaphragm.

mesothelium A layer of cells lining an internal cavity of the body, such as the peritoneal or pericardial cavities.

pepsin A protein secreted in the stomach that begins cleaving ingested proteins into smaller polypeptides. It functions optimally in the acid environment of the stomach, at pH 1–3, and is inactivated when the acid is neutralized, at pH 5 or higher, in the duodenum.

pernicious anemia An anemic condition characterized by larger than normal (megaloblastic) red blood cells. Insufficient gastric production of intrinsic factor leads to deficient ileal absorption of vitamin B12. Deficiency of B12 leads to ineffective red blood cell production in the bone marrow.

plexus A network or joining together of multiple nerves, blood vessels, or lymphatic vessels.

vagus nerve Cranial nerve X. The paired vagus nerves provide parasympathetic innervation to the heart, lungs, and gastrointestinal tract to the level of the left colic flexure.

The stomach is a distensible sac connected to the esophagus proximally and the duodenum distally. It serves two major functions in the digestive system. First, it is a temporary holding area for food, slowly portioning the mass of consumed food into the duodenum for further digestion and absorption. Second, food is mixed with gastric secretions and churned to be broken down into a semiliquid form, termed chyme. Both the gross anatomy and the microscopic anatomy of the stomach reflect these functions.

GROSS ANATOMY

Location and Anatomic Divisions

The stomach is located in the left upper region of the abdomen, just beneath the diaphragm. Connected proximally to the esophagus and distally to the duodenum, the stomach is distensible, with a potential volume between 1200 and 3000 ml depending on the volume of food present. It is roughly shaped like the letter J; the medial concave side is known as the lesser curvature and the lateral convex side is known as the greater curvature (Fig. 1). The stomach can be divided into five parts. The

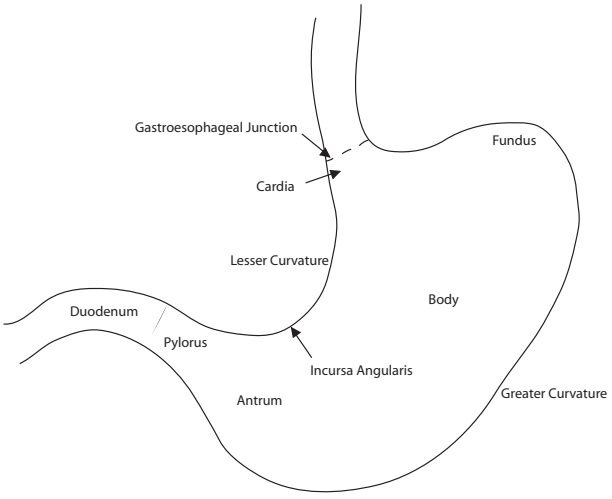


FIGURE 1 Anatomy of the stomach.

Copyright 2004, Elsevier (USA). All rights reserved.

Saito, K., Arai, K., Mori, M., Kobayashi, R., and Ohki, I. (2000). Effect of *Helicobacter pylori* eradication on malignant transformation of gastric adenoma. *Gastrointest. Endosc.* **52**, 27–32.

Yoshikane, H., Sakakibara, A., Hidano, H., Niwa, Y., Goto, H., and Yokoi, T. (2001). Piecemeal endoscopic aspiration mucosectomy for large superficial intramucosal tumors of the stomach. *Endoscopy* **33**, 795–799.

Stomach, Anatomy

DANIEL R. CLAYBURGH AND JERROLD R. TURNER
The University of Chicago

hiatal hernia Protrusion of part of the stomach, usually the cardia, into the thoracic cavity through the esophageal opening of the diaphragm.
mesothelium A layer of cells lining an internal cavity of the body, such as the peritoneal or pericardial cavities.
pepsin A protein secreted in the stomach that begins cleaving ingested proteins into smaller polypeptides. It functions optimally in the acid environment of the stomach, at pH 1–3, and is inactivated when the acid is neutralized, at pH 5 or higher, in the duodenum.
pernicious anemia An anemic condition characterized by larger than normal (megaloblastic) red blood cells. Insufficient gastric production of intrinsic factor leads to deficient ileal absorption of vitamin B12. Deficiency of B12 leads to ineffective red blood cell production in the bone marrow.
plexus A network or joining together of multiple nerves, blood vessels, or lymphatic vessels.
vagus nerve Cranial nerve X. The paired vagus nerves provide parasympathetic innervation to the heart, lungs, and gastrointestinal tract to the level of the left colic flexure.

The stomach is a distensible sac connected to the esophagus proximally and the duodenum distally. It serves two major functions in the digestive system. First, it is a temporary holding area for food, slowly portioning the mass of consumed food into the duodenum for further digestion and absorption. Second, food is mixed with gastric secretions and churned to be broken down into a semiliquid form, termed chyme. Both the gross anatomy and the microscopic anatomy of the stomach reflect these functions.

GROSS ANATOMY

Location and Anatomic Divisions

The stomach is located in the left upper region of the abdomen, just beneath the diaphragm. Connected proximally to the esophagus and distally to the duodenum, the stomach is distensible, with a potential volume between 1200 and 3000 ml depending on the volume of food present. It is roughly shaped like the letter J; the medial concave side is known as the lesser curvature and the lateral convex side is known as the greater curvature (Fig. 1). The stomach can be divided into five parts. The

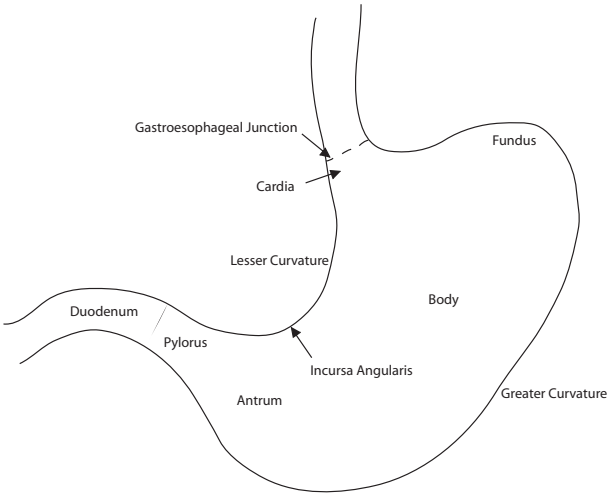


FIGURE 1 Anatomy of the stomach.

Copyright 2004, Elsevier (USA). All rights reserved.

cardia is an ill-defined region beginning at the gastroesophageal junction and extending into the first to 2–3 cm of the stomach. The fundus is the section of stomach that lies superior to the gastroesophageal junction. Approximately two-thirds of the way along the lesser curvature lies a sharp angle known as the incisura angularis; above this angle to the level of gastroesophageal junction lies the corpus or body of the stomach. The distal third of the stomach, the antrum, extends below the incisura angularis. Finally, the pylorus is a narrow 1 to 2 cm long channel connecting the stomach to the duodenum. A relatively common alteration in gastric anatomy is a hiatal hernia. This condition increases the occurrence of esophageal reflux and may eventually lead to esophageal adenocarcinoma.

Circulation

The stomach receives the majority of its blood supply from the celiac trunk, a branch off of the descending aorta that arises at the level of the T12 vertebra. The celiac trunk splits into three major divisions: the left gastric artery, the splenic artery, and the hepatic artery. The left gastric artery runs along the lesser curvature and anastamoses with the right gastric artery, a branch of the hepatic artery. The hepatic artery also gives rise to the gastroduodenal artery, which later becomes the right gastroepiploic artery. This artery runs along the greater curvature and supplies the lateral stomach, with two branches of the splenic artery, the left gastroepiploic and short gastric arteries. There are numerous interconnections, or anastamoses, among the arteries supplying the stomach, making ischemic infarction a very uncommon event.

Innervation

The stomach receives both sympathetic and parasympathetic innervation. Sympathetic nerves arise from the thoracic spinal cord and synapse in the celiac ganglia. From there, the sympathetic nerves follow the gastric and gastroepiploic arteries to enter the stomach. Pain sensation from the stomach is relayed to the central nervous system by way of the afferent sympathetic fibers. Parasympathetic innervation comes from the vagus nerves. Both the paired anterior and posterior vagus nerves bifurcate as they enter the abdomen. The anterior vagus divides into the hepatic and anterior gastric branches and the posterior vagus forms the celiac and posterior gastric branches. The anterior and posterior gastric branches innervate the majority of the stomach, although the pylorus is innervated by the hepatic branch of the anterior vagus. In keeping with the "rest and digest" function of the parasympathetic nervous system, these nerves stimulate acid secretion in the body and fundus of the stomach and increase motility in the antrum.

MICROSCOPIC ANATOMY

As in the rest of the gastrointestinal tract, the wall of the stomach is divided into four layers: mucosa, submucosa, muscularis propria, and serosa. These layers are specialized within the stomach to carry out the stomach's role in digestion.

Mucosa

General Appearance

The mucosa is the innermost layer of the wall lining the stomach cavity. The mucosa and submucosa below it are piled into folds, known as rugae, in the contracted stomach. The rugae are haphazardly arranged in the fundus and body, but are arranged longitudinally in the antrum. As the stomach distends, the rugae flatten out to accommodate the increased volume.

On closer inspection, the mucosa consists of pits (foveolae) invaginating from the surface. Beneath these is an extensive network of glands that empty into the pits, with one to seven glands emptying into

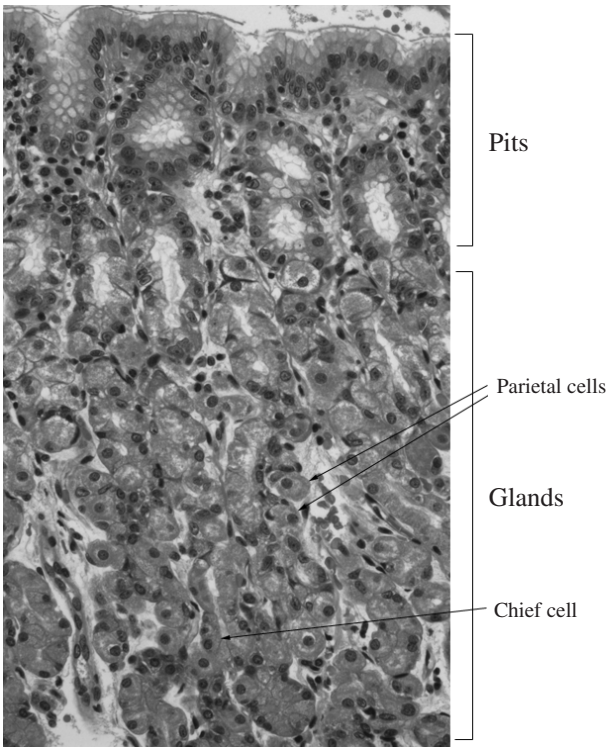


FIGURE 2 Histology of the gastric body and fundus.

each pit (Fig. 2). Beginning at their entry into the pits, each gland is divided into an isthmus, a neck, and a base. The pits and glands are lined by a single layer of columnar epithelial cells. As described below, multiple specialized cells types that carry specific digestive functions are present in the gastric glands. The epithelial cells rest on the basement membrane, which separates the epithelial compartment of the mucosa from the lamina propria. The lamina propria is a loose collection of tissue that supports the epithelium; within it are blood vessels, nerves, lymphatic vessels, immune cells, and connective tissue. Together with the epithelium and the lamina propria, the final component of the mucosa is the muscularis mucosa. The muscularis mucosa is a continuous sheet of smooth muscle that defines the border between the mucosa and the submucosa. The muscularis mucosa is composed of two layers: an inner circular and an outer longitudinal layer.

The gastric mucosa is vital to the function of the stomach as a digestive organ. The epithelium is responsible for the secretion of substances important for digestion, including hydrochloric acid and pepsin. The lumenal pH of the stomach can be as low as 0.9–1.5. This harsh environment is important for the breakdown of food, but also potentially damaging for the lining of the stomach itself. Thus, the mucosa has several protective mechanisms that collaborate to form a barrier that protects the stomach from its own secretions. First, the cells that line that surface and pits secrete a thin layer of mucus that coats the epithelium and forms a barrier to acid diffusion. Additionally, active bicarbonate secretion by the surface epithelium creates a thin neutral pH environment at the epithelial surface. The cell–cell junctions that link gastric epithelial cells also form a barrier to acid movement, with the tight junctions preventing the leak of lumenal acid between epithelial cells. Loss of this barrier function is one of the first effects of nonsteroidal anti-inflammatory drugs, the most common cause of gastric injury in the United States. The rich gastric blood supply is important for the rapid clearance of leaked acid and as the principal source of oxygen and nutrients to the epithelium. Finally, the rapid turnover rate of the foveolar epithelium, which is replaced every 3–6 days, also ensures that any damage to the mucosa is repaired quickly. Given the harsh environment of the gastric lumen, all of these protective mechanisms are necessary for the prevention of serious mucosal damage. For example, decreased gastric blood flow, as in shock, or interruption of epithelial cell division, as occurs with some chemotherapeutic agents, can result in mucosal ulceration.

Body and Fundus

These portions of the stomach mucosa are responsible for the majority of the secretions necessary for the stomach to function properly. As such, the glands form a thick layer that represents approximately three-quarters of the mucosal thickness. The remaining mucosa is devoted to the gastric pits. As is true throughout the stomach, the pits are lined by columnar cells that secrete mucus and bicarbonate.

The glands of the body and fundus contain several cell types. Parietal cells are the primary cells of these glands. These cells are distributed along the length of the gland, are roughly pyramidal in shape, and secrete hydrochloric acid and intrinsic factor. Intrinsic factor forms a complex with vitamin B12 in the gastric lumen and facilitates its adsorption in the distal ileum. Parietal cells (and chief cells, discussed below) are replaced only every 1 to 2 years, in contrast to every few days for the mucous cells. Thus, damage to parietal cells can lead to long-term problems; this is the case with pernicious anemia. Autoimmune destruction of parietal cells leads to a deficiency of intrinsic factor. Over a period of months to years, a shortfall in the level of vitamin B12 develops, leading to ineffective red blood cell production.

The parietal cells have a complex network of canals, or canaliculi, that run through them and open into the gland lumen. These canaliculi are lined with the H^+,K^+-ATPase, the proton pump responsible for acid secretion. These pumps are the target of a common class of therapeutic agents, the "proton pump inhibitors," that are used for the treatment of gastritis, gastric ulcer, and esophageal injury due to reflux of gastric acid. The proton pump is capable of creating a lumenal H^+ concentration 3 million times that found in the cell. Creating such a large gradient is energy-intensive. Thus, parietal cells contain abundant mitochondria, which are essential for the production of ATP. Secretion of acid and intrinsic factor from the parietal cell is stimulated by three different mediators. Vagus nerve stimulation leads to acid secretion, as does histamine release from nearby enterochromaffin-like (ECL) cells. In addition, gastrin released from G cells in the antrum also increases parietal cell secretions and stimulates increases in parietal cell mass.

Chief cells are predominantly confined to the bases of the gastric glands. These cells secrete pepsinogen I and II, the inactive precursors of pepsin. These precursors are cleaved in the low pH of the gastric lumen, at which point they become enzymatically active and are designated pepsin. The pepsinogens are packaged in granules that fill the cytoplasm of the chief cells; fusion

of the vesicle with the cell membrane leads to pepsinogen release. Stimulation via the vagus nerve is the primary activator of pepsinogen secretion.

An assortment of neuroendocrine cell types is also present in the gastric glands. In the body and fundic mucosa, most cells are ECL cells. These cells respond to G-cell-derived gastrin with the secretion of histamine, which, in turn, stimulates acid secretion. Serotonin-producing enterochromaffin cells are also present. Unlike the other cells of the mucosal glands, the endocrine cells secrete their products into the bloodstream within the lamina propria, rather than into the lumen of the gland. These hormones can then act on nearby (paracrine stimulation) or distant (endocrine stimulation) cells.

Mucous neck cells are also scattered within the glands and are most heavily clustered at the gland neck. These cells secrete some mucus, but their primary function is that of a stem cell. They divide to form undifferentiated daughter cells that are capable of several additional rounds of division before they differentiate into the various cell types that constitute the gastric epithelium. During differentiation, these cells migrate up to replace the surface mucous epithelium or migrate down to replace parietal, chief, and neuroendocrine cells.

Antrum

The antrum is easily differentiated from the body and fundus on a microscopic level. This portion of the stomach does not have as great a secretory function as the body and fundus; thus, the antral glands occupy less than half of the mucosal thickness. The pits occupy the remaining majority of the mucosa. The pits are lined with mucus-producing cells similar to those cells found in the pits of the body and fundus. The glands of the antrum differ from those in the body and fundus, as they do not contain parietal or chief cells. Aside from mucus-producing cells and stem cells, endocrine cells dominate the antral glands. Approximately half of these endocrine cells are G cells, which produce gastrin. As mentioned earlier, gastrin is a hormone that stimulates parietal cells in the body and fundus to produce acid; it also enhances gastric motility. Significant numbers of serotonin-producing enterochromaffin cells and somatostatin-producing D cells also populate the antral glands.

Submucosa

The submucosa is a layer of loose connective tissue located between the muscularis mucosa and the muscularis propria. Numerous plexuses of nerves, arteries, veins, and lymphatics are found in this layer.

Muscularis Propria

The stomach differs from other parts of the digestive system in that its muscularis propria contains three distinct muscle layers instead of the usual two. The outer layer of longitudinally arranged fibers and the inner layer of circularly arranged fibers correspond to the two layers found throughout the rest of the digestive tract, but a third layer, consisting of obliquely arranged fibers, is located just interior to the circular layer. The muscularis propria of the stomach is also thicker than the muscularis propria throughout the remainder of the gastrointestinal tract. In addition to the important contractile function carried out by the muscularis propria, this thick muscle layer serves as an impediment to gastric ulcer progression. This is of critical importance, because the underlying serosa offers minimal protection and because ulcer extension through the muscularis propria leads to gastric perforation. The resulting spillage of gastric contents into the abdomen requires emergent surgery and is associated with a high rate of mortality.

At both the pylorus and the gastroesophageal junction, the circular layer is more pronounced, where it forms sphincters that allow the stomach cavity to be sealed off. Failure of the gastroesophageal junction can result in the reflux of gastric acid into the esophagus. Failure of the pylorus can result in unregulated release of incompletely processed food into the small intestine. Congenital hypertrophic pyloric stenosis, a disorder found more commonly in male infants, represents inappropriate expansion of the muscularis propria in the pyloric region with obstruction of the pyloric channel. Ingested food cannot enter the small intestine via the pylorus, resulting in projectile vomiting as the gastric contents are forced through the gastroesophageal junction. This condition is surgically correctable and is the most common surgery performed during the first 6 months of life.

Serosa

The external, or peritoneal, surface of the stomach is covered by the serosa. This thin layer consists of loose connective tissue covered by a single layer of cuboidal mesothelial cells. This covering serves to reduce friction on the stomach during its churning movements.

See Also the Following Articles

Circulation, Overview • Duodenum, Anatomy • Esophagus, Anatomy • Gastrointestinal Matrix, Organization and Significance • Gastrointestinal Tract Anatomy, Overview • Pylorus

Further Reading

Gartner, L. P., and Hiatt, J. L. (1997). "Color Textbook of Histology." W. B. Saunders, Philadelphia, PA.

Gray, H. (1995). "Gray's Anatomy," 15th Ed. Barnes and Noble, New York.

Owen, D. A. (1997). Stomach. *In* "Histology for Pathologists" (S. Sternberg, ed.), 2nd Ed., pp. 481–493. Lippincott-Raven, Philadelphia, PA.

Raufman, J. P. (2003). Stomach: Anatomy and structural anomalies. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 4th Ed., pp. 1279–1291. J. B. Lippincott, Philadelphia, PA.

Turner, J. R., and Odze, R. D. Gastric polyps. *In* "Surgical Pathology of the Gastrointestinal Tract, Liver, Biliary Tract and Pancreas" (R. D. Odze, J. R. Goldblum, and J. M. Crawford, eds.). Harcourt Health Sciences, Baltimore, MD, in press.

Stress

WILLIAM B. SALT II

Mt. Carmel Health, Columbus, Ohio

allostasis Dynamic processes involved in the defense of homeostasis that are generated in response to real or perceived stressors/triggers.

allostatic load Cost of allostasis, which is the wear-and-tear damage resulting from chronic overactivity, underactivity, or mismanagement of allostatic systems, all of which can lead to disease and illness.

homeostasis Internal neurobiological stability and balance that are maintained through allostasis; critical to survival and health.

stress Adaptive physiological response to real or perceived threat to homeostasis; a natural system to protect and restore homeostasis, but can become harmful over time.

stressor External or internal physical, biological, environmental, or situational factors that represent an actual or perceived threat to homeostasis.

The hormones and other physiologic agents that mediate the central stress response of the brain, body, and digestive system (gut) have protective and adaptive immediate effects that are essential for survival and maintenance of homeostasis and health (i.e., allostasis, or "good stress"). However, over longer time intervals, stress response agents exact a cost when they are over produced, under produced, or mismanaged (i.e., allostatic load, or "bad stress"). Alterations in the central stress response and allostatic load related to genetic, environmental, and behavioral influences contribute to the pathophysiology of a broad range of diseases and illnesses; however, scientifically validated strategies for reduction of harmful effects of the stress response and allostatic load are available.

INTRODUCTION

Many people with disease and illness note exacerbation of symptoms correlated with life events that are perceived to be stressful. Others fail to report any such association. What is stress and how common is it? How does the body respond to stress? Is stress good or bad for us? What can be done about stress? New scientific understanding of individual responses to acute and chronic stress and of the mind/brain–body/gut connection has resulted in a reassessment of the role of chronic stress in disease and illness.

DISEASE AND ILLNESS

Disease can be defined as the externally verifiable evidence of a pathological state, whereas illness is defined as a person's perception of ill health, which is evident from symptom reports, beliefs, and behavior. There can be a significant discordance between disease and illness. For example, an individual with the disease of a deep peptic ulcer of the stomach or duodenum may not describe any illness symptoms, such as abdominal pain. By contrast, a person with a small and superficial ulcer may report severe pain. The stress response contributes to the pathophysiology of disease and illness and helps explain their common discordance.

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Gartner, L. P., and Hiatt, J. L. (1997). "Color Textbook of Histology." W. B. Saunders, Philadelphia, PA.

Gray, H. (1995). "Gray's Anatomy," 15th Ed. Barnes and Noble, New York.

Owen, D. A. (1997). Stomach. *In* "Histology for Pathologists" (S. Sternberg, ed.), 2nd Ed., pp. 481–493. Lippincott-Raven, Philadelphia, PA.

Raufman, J. P. (2003). Stomach: Anatomy and structural anomalies. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 4th Ed., pp. 1279–1291. J. B. Lippincott, Philadelphia, PA.

Turner, J. R., and Odze, R. D. Gastric polyps. *In* "Surgical Pathology of the Gastrointestinal Tract, Liver, Biliary Tract and Pancreas" (R. D. Odze, J. R. Goldblum, and J. M. Crawford, eds.). Harcourt Health Sciences, Baltimore, MD, in press.

Stress

WILLIAM B. SALT II

Mt. Carmel Health, Columbus, Ohio

allostasis Dynamic processes involved in the defense of homeostasis that are generated in response to real or perceived stressors/triggers.

allostatic load Cost of allostasis, which is the wear-and-tear damage resulting from chronic overactivity, underactivity, or mismanagement of allostatic systems, all of which can lead to disease and illness.

homeostasis Internal neurobiological stability and balance that are maintained through allostasis; critical to survival and health.

stress Adaptive physiological response to real or perceived threat to homeostasis; a natural system to protect and restore homeostasis, but can become harmful over time.

stressor External or internal physical, biological, environmental, or situational factors that represent an actual or perceived threat to homeostasis.

The hormones and other physiologic agents that mediate the central stress response of the brain, body, and digestive system (gut) have protective and adaptive immediate effects that are essential for survival and maintenance of homeostasis and health (i.e., allostasis, or "good stress"). However, over longer time intervals, stress response agents exact a cost when they are over produced, under produced, or mismanaged (i.e., allostatic load, or "bad stress"). Alterations in the central stress response and allostatic load related to genetic, environmental, and behavioral influences contribute to the pathophysiology of a broad range of diseases and illnesses; however, scientifically validated strategies for reduction of harmful effects of the stress response and allostatic load are available.

INTRODUCTION

Many people with disease and illness note exacerbation of symptoms correlated with life events that are perceived to be stressful. Others fail to report any such association. What is stress and how common is it? How does the body respond to stress? Is stress good or bad for us? What can be done about stress? New scientific understanding of individual responses to acute and chronic stress and of the mind/brain–body/gut connection has resulted in a reassessment of the role of chronic stress in disease and illness.

DISEASE AND ILLNESS

Disease can be defined as the externally verifiable evidence of a pathological state, whereas illness is defined as a person's perception of ill health, which is evident from symptom reports, beliefs, and behavior. There can be a significant discordance between disease and illness. For example, an individual with the disease of a deep peptic ulcer of the stomach or duodenum may not describe any illness symptoms, such as abdominal pain. By contrast, a person with a small and superficial ulcer may report severe pain. The stress response contributes to the pathophysiology of disease and illness and helps explain their common discordance.

Copyright 2004, Elsevier (USA). All rights reserved.

DEFINITION OF STRESS

Stress is an adaptive physiological response to the detection of either external (exteroceptive) or internal (interoceptive) stressors. Bruce S. McEwen developed a new way of conceptualizing and understanding the stress response (Fig. 1). When stressors represent a real or perceived threat to the neurobiological balance (homeostasis) of a person, physiological and behavioral responses are triggered in order to achieve adaptation and survival in the short run (i.e., allostasis). The protective and restorative allostatic processes of the body are mediated through the autonomic nervous system, the hypothalamic–pituitary–adrenal (HPA) axis, and the cardiovascular, metabolic and immune systems. However, over longer periods of time, the same response systems that are designed to protect and restore can be damaging and can cause or exacerbate symptoms, disease, and illness (i.e., allostatic load).

THE EMOTIONAL MOTOR SYSTEM

The central stress response is mediated through the emotional motor system (EMS) of the central nervous system (CNS). The EMS refers to a network that includes the anterior cingulate cortex, hypothalamic nuclei, amygdala, periaqueductal gray, and brain stem nuclei (locus ceruleus, Barrington's nucleus, dorsal motor nucleus of the vagus, and rostral ventral medulla), which form the basis for vagal, parasympathetic, and sympathetic visceral efferent pathways. In Fig. 2, a simplified depiction of the four principal output functions of the EMS-mediated stress response, "ascending aminergic" refers to arousal, attentional, and emotional feeling output.

Although the terms "emotion" and "feeling" are commonly used interchangeably, they can be conceptualized as distinct dimensions of the stress response generated by the EMS, playing a critical role in maintenance of health and in cognitive processes, including perception, learning, and decision-making. Emotion is a collection of encoded stress responses triggered from parts of the brain to the body and to other parts of the brain via both neural and humoral routes, resulting in changes within the body and certain areas of the brain. Examples of emotion include fear, anger, disgust, and joy, each with unique physiological and behavioral expression. Feeling is the conscious emotional experience or mental state that commonly, but not necessarily, accompanies a given emotional stress response. For example, the physiologically expressed emotion of fear with associated gut symptoms of abdominal pain and diarrhea may or may not be accompanied by the emotional feeling of anxiety.

THE MIND/BRAIN–GUT CONNECTION

The bidirectional neurobiological communication of the mind/brain–gut connection is critical to understanding the expression of the emotional stress response in the gut. The central stress response and allostasis can be triggered by both exteroceptive and interoceptive stressors. With an approximate external surface area of 2 square yards, the skin participates with the other senses in detection of exteroceptive stressors. The gut has an even larger internal surface area, 5250 square

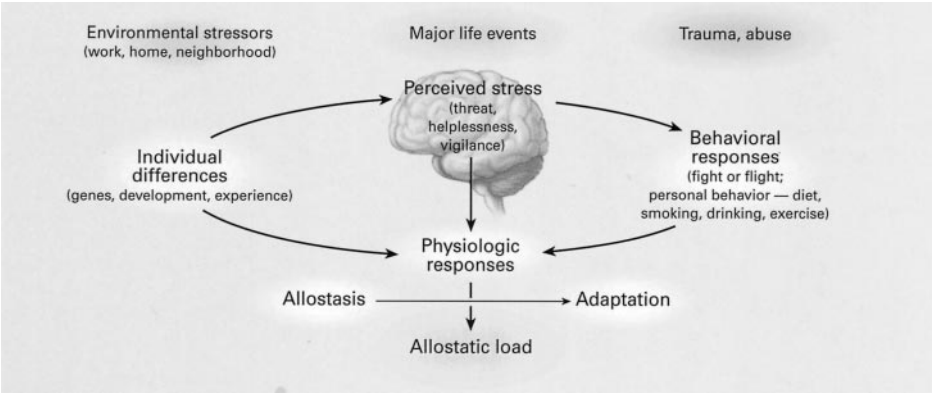


FIGURE 1 The stress response and development of allostatic load. The perception of stress is influenced by experiences, genetics, and behavior. When the brain perceives an experience as stressful, physiologic and behavioral responses are initiated, leading to allostasis and adaptation. Over time, allostatic load can accumulate, and the overexposure to mediators of neural, endocrine, and immune stress can have adverse effects on various organ systems, leading to disease. Reprinted with permission from McEwen, B. S. (1998). Protective and Damaging Effects of Stress Mediators. *N. Engl. J. Med.* **338**, 171–179. Copyright 1998, Massachusetts Medical Society, all rights reserved.

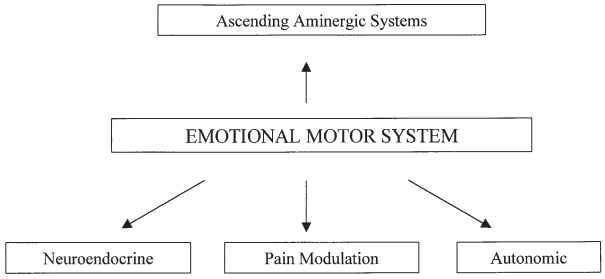


FIGURE 2 Major outputs of the emotional motor system. The emotional motor system refers to a set of parallel output pathways that are activated in response to perceived threat or fear. Alterations in these four pathways have been demonstrated in patients with irritable bowel syndrome. Reprinted from Mayer, E. A. (1999). Emerging Disease Model for Functional Gastrointestinal Disorders. *Am. J. Med.* **107(5A)**, 12S–19S. Copyright 1999, with permission from Excerpta Medica, Inc.

yards (the size of a football field), and detects visceral interoceptive stressors such as inflammation, infection, blood loss, and food intake.

The brain and gut both derive from the neural crest of the human embryo and share many of the same neurons and chemical transmitters, such as serotonin. The enteric nervous system (ENS) embedded in the wall of the gut and extending from the esophagus to the anus is one of three subdivisions of the autonomic nervous system (ANS), together with the sympathetic and parasympathetic divisions. Although bidirectionally linked and integrated with the CNS, the ENS is an enteric minibrain located close to the gut effector systems that it controls, i.e., the musculature, secretory glands, vasculature, and mucosal epithelium. Just as stereotypical emotional stress responses to threatening stressors are encoded within the EMS, the ENS stores a library of programs of defensive stress responses for different patterns of gut motor and secretory behavior that can either be activated by stressors interpreted by the brain or sensed locally within the gut. One example is the "power propulsion" (anal direction) response designed as a protective process to move digestive contents out of the digestive tract (symptoms include diarrhea and abdominal pain).

GUT EMOTION AND FEELING

Joseph E. LeDoux and Emeran A. Mayer emphasized that stressor-triggered and encoded physiological emotional responses, such as fear and anger, are unique and usually generated unconsciously. For example, fear is associated with inhibition of upper gastrointestinal motility and secretion, which can result in symptoms of bloating, loss of appetite, nausea, and even vomiting. By contrast, anger is associated with stimulation of gastric contractions and acid secretion, which may be accompanied by symptoms of upper abdominal pain.

Visceral afferent information related to emotional gut responses modulates both emotional affective and cognitive mental function. Language expressions such as "my stomach is tied in knots" and "I hate your guts" imply common understanding of the linkage between emotional feelings, such as anxiety and anger, and specific and commonly unpleasant visceral gut experiences. Furthermore, the expression "it's my gut feeling" implies that visceral sensations relate to a prerational insight or "emotional intelligence," a term popularized by psychologist Daniel Goleman in his book of the same name.

ROLE OF ALLOSTATIC LOAD IN DISEASE AND ILLNESS

Dysregulation of the central stress response system related to allostatic load contributes to the development of a variety of diseases and illnesses, including hypertension, atherosclerosis, central obesity, diabetes, the insulin resistance syndrome (metabolic syndrome x), and certain disorders of immune function. Chronically elevated corticosteroid levels induced by persisting stress may adversely affect hippocampal structure and function, which can result in deficits of both memory and cognition. Gastrointestinal diseases and illnesses adversely impacted by allostatic load include nonalcoholic fatty liver disease (also associated with the insulin resistance syndrome), gastroesophageal reflux disease (GERD), peptic ulcer disease, inflammatory bowel disease, and functional gastrointestinal disorders (FGID).

FUNCTIONAL MEDICAL SYMPTOMS AND SYNDROMES

Functional illness is related to altered physiological function, absence of an identifiable structural or biochemical cause, and involvement of virtually any organ system with a variety of symptoms (most common symptoms are pain, discomfort, disturbed sleep, and loss of vitality). Such functional symptoms are also called "medically unexplained symptoms." Fibromyalgia (chronic and recurrent widespread aching pain and fatigue) is one example of a related functional syndrome that is a medically unexplained collection of symptoms. The most common FGID is irritable bowel syndrome (IBS), characterized by chronic and recurrent abdominal pain and/or discomfort associated with altered bowel function (constipation, diarrhea, or both).

Douglas A. Drossman pioneered the application to the pathophysiology of IBS a biopsychosocial model that

describes the interacting role of genetic, biological, psychosocial, and cognitive factors. Extending the perspective of this system, Emeran A. Mayer developed a disease model of FGID that is also based on Bruce S. McEwen's concepts of allostasis and allostatic load. An enhanced responsiveness of the central stress/emotion circuits to real or perceived exteroceptive or interoceptive stressors by IBS patients is reflected in altered modulation of gastrointestinal motility, secretion, permeability, and immune function, and altered perceptual and emotional response to visceral events.

AFFECTIVE SYMPTOMS AND DISORDERS

The concepts of allostasis and allostatic load focus on the mind/brain as both the interpreter and responder and the target of real or perceived stressors/triggers. Emerging neurobiological models of the adverse consequences of allostatic load and the dysregulated central stress response help to explain the common association of emotional feelings and affective disorders, such as anxiety and depression, with the emotional gut symptoms of FGID, such as IBS. Furthermore, research confirms that depressive illness and hostility are both associated with cardiovascular and other systemic diseases. Similar therapeutic strategies may be applicable in the management of both affective disorders and functional disorders and certain diseases.

MIND–BODY MEDICINE AND REDUCTION OF ALLOSTATIC LOAD

There is an emerging convergence of ancient, traditional, and modern scientific approaches to disease, illness, and healing. The neurobiological basis for mind–body, complementary and alternative medicine (CAM), and integrative medicine interventions is becoming increasingly understood. Mind–body medicine, which refers to the application of multidisciplinary methods based on the inseparable connection between the mind and the body and the complicated interactions that take place between thoughts, body, and the outside world, has the goal of restoring homeostasis and health by enhancing the natural healing capacities conferred by innate body systems.

Research has confirmed that stress reduction and certain mind–body cognitive behavioral interventions can significantly improve health outcomes and reduce the need for more expensive treatments in a variety of diseases and illnesses. Behavioral responses can be harmful or helpful (see Fig. 1). The objective is reduction of allostatic load (Fig. 3). Helpful strategies include

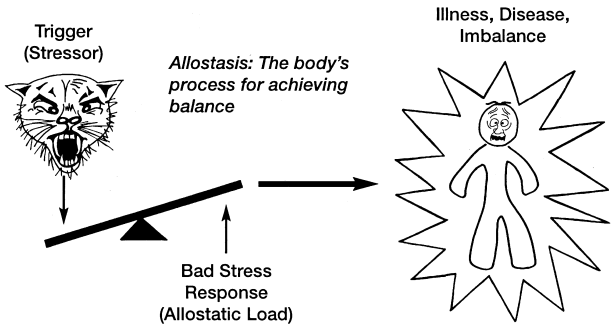


FIGURE 3 Disturbed homeostasis (imbalance). Reprinted with permission from Salt II, W. B. and Neimark, N. F. (2002). Irritable Bowel Syndrome and the MindBodySpirit Connection. Copyright 2002, Parkview Publishing, all rights reserved.

consistent commitment to healthy lifestyle choices (e.g., diet and exercise) and the development of effective coping styles that incorporate positive mental outlook and avoidance of both high-demand/low-control stressors and social isolation. Individuals can learn to activate mechanisms that oppose the stress response and induce what Harvard's Herbert Benson terms "the relaxation response" via various techniques (e.g., progressive relaxation, hypnosis, meditation, yoga, and breathing exercises).

See Also the Following Articles

Brain–Gut Axis • Enteric Nervous System • Irritable Bowel Syndrome • Neurogastroenterology • Psychosociology of Irritable Bowel Syndrome

Further Reading

Benson, H. (2002). Mind/Body Medical Institute web site (affiliated with Harvard Medical School and Beth Israel-Deaconess Medical Center). Retrieved at http://www.mbmi.org/.

Drossman, D. A. (1998). Presidential address: Gastrointestinal illness and the biopsychosocial model. *Psychosom. Med.* **60**(3), 258–267.

LeDoux, J. E. (1996). "The Emotional Brain." Touchstone, New York.

Mayer, E. A. (1999). Emerging disease model for functional gastrointestinal disorders. *Am. J. Med.* **107**(5A), 12S–19S.

Mayer, E. A., and Saper, C. B., eds. (2000). "The Biological Basis for Mind Body Interactions." Elsevier Science, San Diego.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *N. Engl. J. Med.* **338**, 171–179.

McEwen, B. S. (2002). "The End of Stress As We Know It." Joseph Henry Press, New York.

Salt, W. B. 2nd, and Neimark, N. F. (2002). "Irritable Bowel Syndrome and the MindBodySpirit Connection." Parkview Publishing, Columbus.

Van Itallie, T. B. (2002). Stress: A risk factor for serious illness. *Metabolism* **51**(6, Suppl. 1), 40–45.

Wood, J. D. (2002). Neuropathophysiology of the irritable bowel syndrome. *J. Clin. Gastroenterol.* **35**(1, Suppl.), S11–S22.

Stress Ulceration

RAJEEV JAIN
Presbyterian Hospital of Dallas

histamine-2 receptor antagonists Pharmacologic agents that inhibit acid secretion by acting as specific antagonists of the histamine-2 receptor on the parietal cell.

mucosal restitution Early phase of gastrointestinal mucosal repair whereby damaged cells slough off and are replaced by viable cells.

proton pump inhibitors Substituted benzimidazoles that inhibit acid secretion by blocking the parietal cell H^+, K^+-ATPase.

sucralfate Basic aluminum salt of sucrose octasulfate; protects mucosal integrity locally without substantially altering gastric pH.

Stress ulcer is a gastric mucosal injury typically seen in critically ill patients who require mechanical ventilation or have a coagulopathy. These ulcerations develop from a multifactorial process, with the primary insult being gastric ischemia. Patients in the intensive care unit can experience significant bleeding that is associated with increased mortality. Preventive strategies have focused on acid suppression and local mucosal protection.

RISK FACTORS

Patients admitted to the intensive care unit (ICU) are a heterogeneous group with variable risks for bleeding. The two strongest risk factors for stress ulcer bleeding are respiratory failure requiring mechanical ventilation for greater than 48 hours and coagulopathy (defined as platelet count of <50,000 cells/mm^3 or International Normalized Ratio of >1.5). Other risk factors include previous gastrointestinal bleeding, hypotension, trauma, burn injury (>35% of body surface), central nervous system injury, renal failure, hepatic failure, organ transplantation, and sepsis. However, many of these risk factors may simply represent surrogate markers for mechanical ventilation and/or coagulopathy.

PATHOPHYSIOLOGY

Although incompletely understood, the pathophysiology of stress ulcer formation is the result of a multifactorial process. The proximal gastrointestinal tract maintains mucosal integrity with a microcirculation that provides nutrients and a route of elimination for toxins. Additionally, a mucus layer protects the gastric mucosa by forming a physical barrier to acid and other intraluminal irritants.

The major physiologic derangement in stress ulcer development is a loss of mucosal integrity. In the setting of severe physiologic stress, hypoperfusion leads to a cascade of events mediated by the overproduction of nitric oxide synthase. A reperfusion injury ensues in which increased production of oxygen free radicals, local intramural acidosis caused by back diffusion of luminal hydrogen ions, and a loss of washout effect combine to disrupt mucosal integrity.

In intracranial processes, hypergastrinemia may occur and lead to overproduction of acid. Other factors that may contribute to stress ulcer formation include reduced gastric epithelial cell restitution, abnormal gastric and small bowel motility, increased bile reflux, and nutritional disturbances.

CLINICAL PRESENTATION

Natural History

The majority of stress ulcerations are asymptomatic and clinically insignificant. In endoscopic studies, mucosal injury is found in greater than 75% of critically ill patients within 24 hours of admission. The typical mucosal injury seen is multiple, diffuse, superficial erosions in the gastric fundus and body; however, focal, deep ulceration can also be seen in both the stomach and the duodenum.

In clinical studies, the presentation of stress ulcer bleeding has been categorized as overt and clinically important bleeding. Overt bleeding is defined as hematemesis, gross blood or dark material resembling coffee grounds in nasogastric aspirate, hematochezia, or melena. Clinically important bleeding is defined as overt bleeding complicated by any one of the following events within 24 hours: (1) a spontaneous decrease in systolic blood pressure of >20 mmHg, (2) an increase in heart rate of >20 beats/min, (3) a decrease of >10 mmHg of

 Copyright 2004, Elsevier (USA). All rights reserved.

systolic blood pressure shortly after sitting up, (4) a decrease in hemoglobin level of >2 g/dl, and (5) blood transfusion with no appropriate increase in hemoglobin level. In ICU patients, the estimated incidence of overt bleeding is 5% and the incidence of clinically important bleeding is 1–4%.

In a multicenter trial of critically ill patients, 31% of patients with one or two risk factors (mechanical ventilation and/or coagulopathy) had clinically important bleeding whereas only 0.1% of patients without these risk factors bled. Clinically important bleeding typically occurs within the first 2 weeks of ICU admission with a mortality of nearly 50%, a fivefold increase from those patients who did not experience bleeding.

Diagnosis

The diagnosis of stress ulcer bleeding is confirmed by endoscopy when the typical findings of either erosions or ulcer are identified.

THERAPY

If ulceration with high-risk stigmata (nonbleeding visible vessel or active bleeding) is found, endoscopic interventions such as injection of epinephrine and/or thermal therapy can lead to hemostasis with a reduced rebleeding rate. Concurrent medical therapy with intravenous proton pump inhibitor (PPI) is also indicated in patients with high-risk stigmata. Occasionally, angiography with embolization into the vessel that is the source of the bleeding or surgery is required to treat bleeding refractory to medical and endoscopic therapy.

PREVENTION

Prophylaxis against stress ulcer in high-risk patients has become the standard of care in the ICU. Gastric pH-altering regimens such as antacids, histamine-2 receptor antagonists (H2RAs), and PPIs may prevent stress ulcer bleeding by maintaining intragastric pH >4, which inhibits pepsin and helps prevent clot dissolution.

Antacids

Antacids control intragastric pH by buffering hydrogen ions. In a meta-analysis, antacids in comparison to placebo displayed a trend toward reducing both overt and bleeding and clinically important bleeding. No change in mortality was noted. In comparison to H2RAs, antacids displayed a trend toward greater bleeding. In comparison to sucralfate, antacids were somewhat more effective in preventing bleeding.

H2RAs

In a meta-analysis, intravenous H2RAs (cimetidine or ranitidine), in comparison to placebo, significantly decreased both the rate of overt bleeding and clinically important bleeding. No change in mortality was noted. H2RAs displayed a trend toward decreased clinically important bleeding in comparison to both antacids and sucralfate. In a large, multicenter trial comparing ranitidine (50 mg intravenously every 6 hours) versus sucralfate (1 g intragastrically every 6 hours), clinically important bleeding was lower in patients receiving ranitidine in comparison to sucralfate (1.7 versus 3.8%, $p=0.02$). During prolonged intravenous H2RA therapy, the development of tolerance may lead to a reduction in acid suppression.

PPIs

Three small clinical trials have evaluated the use of omeprazole suspension delivered via nasogastric tube in mechanically ventilated patients. Two trials used omeprazole granules dissolved in sodium bicarbonate and one used intact omeprazole granules in water. In the study comparing omeprazole to ranitidine, the prevalence of clinically important bleeding was significantly less in the omeprazole group in comparison to ranitidine (6 versus 31%). All three studies had methodological limitations such as small sample size, open-label design, and variance in stress ulcer risk factors, prompting caution in the interpretation and application of the study results.

From a pharmacologic standpoint, intravenous PPIs suppress gastric acid to a greater extent than do intravenous H2RAs, and without the development of tolerance. Data are not yet available comparing intravenous PPIs to H2RAs in stress ulcer prophylaxis.

Sucralfate

Sucralfate, the basic aluminum salt of sucrose octasulfate, protects mucosal integrity locally without substantially altering gastric pH. In a meta-analysis, sucralfate in comparison to placebo decreased overt bleeding significantly but did not reduce clinically important bleeding. Sucralfate was not superior to antacids or H2RAs in preventing overt or clinically important bleeding. Sucralfate was associated with a lower incidence of nosocomial pneumonia in comparison to antacids and H2RAs.

Summary of Preventive Therapies

Based on the available data, stress ulcer prophylaxis is warranted only in high-risk critically ill patients such as those requiring mechanical ventilation or with significant coagulopathy, although it may be considered in patients with numerous other risk factors. Intravenous H2RAs are the preferred preventive strategy in these high-risk patients, with only 30 patients requiring therapy to prevent one episode of clinically significant bleeding, although PPIs may also prove effective in future trials.

See Also the Following Articles

Antacids • H2-Receptor Antagonists • Proton Pump Inhibitors • Stress

Further Reading

Beejay, U., and Wolfe, M. M. (2000). Acute gastrointestinal bleeding in the intensive care unit: The gastroenterologist's perspective. *Gastrointest. Clin. North Am.* **29**(2), 309–336.

Cash, B. D. (2002). Evidence-based medicine as it applies to acid suppression in the hospitalized patient. *Crit. Care Med.* **30**(Suppl.), S373–S378.

Cook, D. J., Fuller, H. D., Guyatt, G. H., *et al.* (1994). Risk factors for gastrointestinal bleeding in critically ill patients. *N. Engl. J. Med.* **330**, 377–381.

Cook, D. J., Reeve, B. K., Guyatt, G. H., Heyland, D. K., Griffith, L. E., Buckingham, L., and Tryba, M. (1996). Stress ulcer prophylaxis in critically ill patients: Resolving discordant meta-analyses. *JAMA* **275**, 308–314.

Fennerty, M. B. (2002). Pathophysiology of the upper gastrointestinal tract in the critically ill patient: Rationale for the therapeutic benefits of acid suppression. *Crit. Care Med.* **30**(Suppl.), S351–S355.

Silen, W. (1987). Gastric mucosal defense and repair. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), pp. 1055–1069. Raven Press, New York.

Spechler, J. S. (2002). Peptic ulcer disease and its complications. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 748–772. Saunders, Philadelphia.

Steinberg, K. P. (2002). Stress-related mucosal disease in the critically ill patient: Risk factors and strategies to prevent stress-related bleeding in the intensive care unit. *Crit. Care Med.* **30**(Suppl.), S362–S364.

Submucosal Tumors of the Gastrointestinal Tract

THOMAS SAVIDES AND PAUL KEFALIDES
University of California, San Diego

carcinoids Indolent neuroendocrine tumors that originate in the mucosa but grow deep into the submucosa.

duplication cysts Congenital intestinal duplications that can be found in the walls of the esophagus, stomach, or duodenum.

gastrointestinal stromal cell tumors Rare mesenchymal neoplasms that can be found throughout the digestive tract.

granular cell tumors Generally benign growths that are derived from smooth muscle or Schwann cells.

lipomas Rare benign tumors of adipocytes that can occur throughout the gastrointestinal tract from the esophagus to the rectum.

pancreatic rest Ectopic pancreatic tissue located in the wall of the stomach or small intestine.

submucosal tumor A growth of tissue in the intestinal wall that originates from the submucosa, from the muscularis propria, or from extrinsic compression by adjacent structures.

varices Abnormally swollen or dilated blood vessels.

Submucosal tumors (SMTs) are rare with a combined incidence estimated at 0.3% annually. However, SMTs are an increasingly common finding for general internists and gastroenterologists as more patients undergo diagnostic endoscopy or radiographic studies with oral contrast. Most SMTs are asymptomatic, and for some types, the natural history is not well understood. However, a few unique, sometimes subtle, characteristics can distinguish

Copyright 2004, Elsevier (USA). All rights reserved.

Summary of Preventive Therapies

Based on the available data, stress ulcer prophylaxis is warranted only in high-risk critically ill patients such as those requiring mechanical ventilation or with significant coagulopathy, although it may be considered in patients with numerous other risk factors. Intravenous H2RAs are the preferred preventive strategy in these high-risk patients, with only 30 patients requiring therapy to prevent one episode of clinically significant bleeding, although PPIs may also prove effective in future trials.

See Also the Following Articles

Antacids • H2-Receptor Antagonists • Proton Pump Inhibitors • Stress

Further Reading

Beejay, U., and Wolfe, M. M. (2000). Acute gastrointestinal bleeding in the intensive care unit: The gastroenterologist's perspective. *Gastrointest. Clin. North Am.* **29**(2), 309–336.

Cash, B. D. (2002). Evidence-based medicine as it applies to acid suppression in the hospitalized patient. *Crit. Care Med.* **30**(Suppl.), S373–S378.

Cook, D. J., Fuller, H. D., Guyatt, G. H., *et al.* (1994). Risk factors for gastrointestinal bleeding in critically ill patients. *N. Engl. J. Med.* **330**, 377–381.

Cook, D. J., Reeve, B. K., Guyatt, G. H., Heyland, D. K., Griffith, L. E., Buckingham, L., and Tryba, M. (1996). Stress ulcer prophylaxis in critically ill patients: Resolving discordant meta-analyses. *JAMA* **275**, 308–314.

Fennerty, M. B. (2002). Pathophysiology of the upper gastrointestinal tract in the critically ill patient: Rationale for the therapeutic benefits of acid suppression. *Crit. Care Med.* **30**(Suppl.), S351–S355.

Silen, W. (1987). Gastric mucosal defense and repair. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), pp. 1055–1069. Raven Press, New York.

Spechler, J. S. (2002). Peptic ulcer disease and its complications. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 748–772. Saunders, Philadelphia.

Steinberg, K. P. (2002). Stress-related mucosal disease in the critically ill patient: Risk factors and strategies to prevent stress-related bleeding in the intensive care unit. *Crit. Care Med.* **30**(Suppl.), S362–S364.

Submucosal Tumors of the Gastrointestinal Tract

THOMAS SAVIDES AND PAUL KEFALIDES
University of California, San Diego

carcinoids Indolent neuroendocrine tumors that originate in the mucosa but grow deep into the submucosa.

duplication cysts Congenital intestinal duplications that can be found in the walls of the esophagus, stomach, or duodenum.

gastrointestinal stromal cell tumors Rare mesenchymal neoplasms that can be found throughout the digestive tract.

granular cell tumors Generally benign growths that are derived from smooth muscle or Schwann cells.

lipomas Rare benign tumors of adipocytes that can occur throughout the gastrointestinal tract from the esophagus to the rectum.

pancreatic rest Ectopic pancreatic tissue located in the wall of the stomach or small intestine.

submucosal tumor A growth of tissue in the intestinal wall that originates from the submucosa, from the muscularis propria, or from extrinsic compression by adjacent structures.

varices Abnormally swollen or dilated blood vessels.

Submucosal tumors (SMTs) are rare with a combined incidence estimated at 0.3% annually. However, SMTs are an increasingly common finding for general internists and gastroenterologists as more patients undergo diagnostic endoscopy or radiographic studies with oral contrast. Most SMTs are asymptomatic, and for some types, the natural history is not well understood. However, a few unique, sometimes subtle, characteristics can distinguish

Copyright 2004, Elsevier (USA). All rights reserved.

them on radiographs and video endoscopy. Moreover, the emergence of endoscopic ultrasound (EUS) technology is allowing a better understanding of SMTs and is making evaluation of SMTs more accurate. By demonstrating the exact size and point of origin of a SMT, EUS helps to identify patients who may have a malignant lesion and may require surgery. For any given patient with a SMT, the ideal approach depends on its physical characteristics, its histology, and the clinical setting in which it was diagnosed. This article will describe how SMTs initially present and will summarize the distinguishing characteristics of the most common subtypes of SMTs. An approach to tissue diagnosis, treatment, and follow-up for each type of SMT will be emphasized.

IDENTIFICATION AND PRESENTATION

A submucosal tumor (SMT) is defined as a growth of tissue in the intestinal wall that originates from the submucosa, from the muscularis propria, or from extrinsic compression by adjacent structures (see Table I). Most SMTs are found incidentally on upper endoscopy where they appear as a mass or projection into the gastrointestinal lumen. Because they are covered by normal mucosa, some authors suggest that a better name for SMTs would be "subepithelial tumors."

In some cases, additional features can be seen on endoscopy. Stromal cell tumors, for example, frequently show two prominent lobes on video endoscopy. Lipomas are known to have a waxy hue and are very soft and "pillow-like" when explored with an endoscopic biopsy forceps.

SMTs may also be found by contrast radiography or computed tomography (CT) scan. When patients undergo barium radiography, a SMT will cause a well-demarcated filling defect with smooth borders; the lesion will displace the mucosal folds. CT scans can show the intramural mass of a SMT if the cuts are finer than the diameter of the mass. If a CT shows a homogeneous fat-tissue density structure in the wall of the colon, for example, the diagnosis of lipoma can be made with confidence.

TABLE I Common Submucosal Lesions by Layer of Origin and Other EUS Features

Lesion	Borders	Echogenicity	Consistency
Submucosa	Smooth		May be separated
Cyst		Anechoic	May be separated
Lipoma		Hyperechoic	Homogeneous
Carcinoid		Hypoechoic	Heterogeneous
Varices		Anechoic	
Muscularis propria	Smooth, irregular	Hypoechoic	Homogeneous small lesions, homogeneous large lesions, may be heterogeneous
GIST			
Extrinsic	Smooth	Isoechoic	Homogeneous
Spleen			
Liver			

EVALUATION

Although SMTs may be initially identified by endoscopy, radiographs, or CT scan, endoscopic ultrasound (EUS) provides a more complete description by delineating the originating wall layer in the gastrointestinal tract, identifying vascular invasion, and characterizing the internal milieu of the tumor.

A radial ultrasound transducer at the tip of an endoscope can demonstrate five layers of the gastrointestinal tract at 7.5 MHz: superficial mucosa, deep mucosa, submucosa, muscularis propria, and serosa. EUS can also image adjacent organs such as the pancreas, liver, gallbladder, spleen, kidneys, and adrenal glands. By definition, SMTs arise from the submucosa (third layer) or the muscularis propria (fourth layer) or are external to the gastrointestinal (GI) tract. The contents of the SMT will also create a characteristic echo pattern that is either bright (hyperechoic) or dark (hypoechoic) and either homogeneous or heterogeneous in consistency. The borders may be smooth or irregular on EUS. EUS can visualize SMTs sized from 5 to 60 mm. In a series of 15 resected specimens that were preoperatively evaluated by EUS, 87% were correctly sized by the technique. Some SMTs are more difficult to diagnose accurately by EUS. A study of interobserver agreement among nine expert endosonographers showed that most endoscopists agreed on the ultrasound diagnosis of cysts, lipomas, and extrinsic compressions, but for carcinoids, granular cell tumors, pancreatic rest tissue and metastases, the endoscopists tended to disagree.

COMMON SUBMUCOSAL TUMORS

There are multiple types of submucosal tumors. They can be classified by the layer of origin in the gastrointestinal wall and further differentiated based on histology and natural history.

Lesions in the Submucosa

Lipoma

Lipomas are rare benign tumors of adipocytes that can occur throughout the GI tract from the esophagus to the rectum. They are most often found in the colon.

Most do not cause symptoms; however, when they become larger than 1 cm, lipomas may cause abdominal pain, obstruction, bleeding, or dyspepsia.

On video endoscopy, lipomas appear spherical or round and may have a waxy hue rather than the pink of normal mucosa. When biopsy forceps are placed through the endoscope, lipomas show the "tenting" and "cushion" signs. Tenting refers to the mucosa being easily pulled off the lesion. With the cushion sign, pushing the forceps into the lesion causes a pillow-like depression in the lesion.

Superficial biopsies of lipomas will show only the normal overlying mucosa. Multiple biopsies in the same location or a biopsy that follows electrocautery through the mucosa can document the characteristic adipose tissue. Because lipomas contain a generous amount of fat, they may also be diagnosed by CT scan.

EUS images of lipomas show homogeneous, hyperechoic, and well-demarcated lesions that are fully contained in the submucosa (see Fig. 1). When the EUS images are pathognomonic, the diagnosis is secure and fine-needle aspiration (FNA) is not required. Follow-up endoscopy, however, may be considered at 2- to 4-year intervals to confirm that the lesion is not growing rapidly.

Carcinoid Tumor

Carcinoids are indolent neuroendocrine tumors that originate in the mucosa but grow deep into the submucosa. The estimated incidence is 1 to 2 per 100,000 in the United States. They may produce a number of vasoactive substances and neurotransmitters, such as histamine, somatostatin, and catecholamines. An earlier embryologic classification scheme that organized carcinoids as foregut, midgut, or hindgut has given way to a system that emphasizes the organ involved and the presence of nuclear atypia, mitoses, and necrosis.

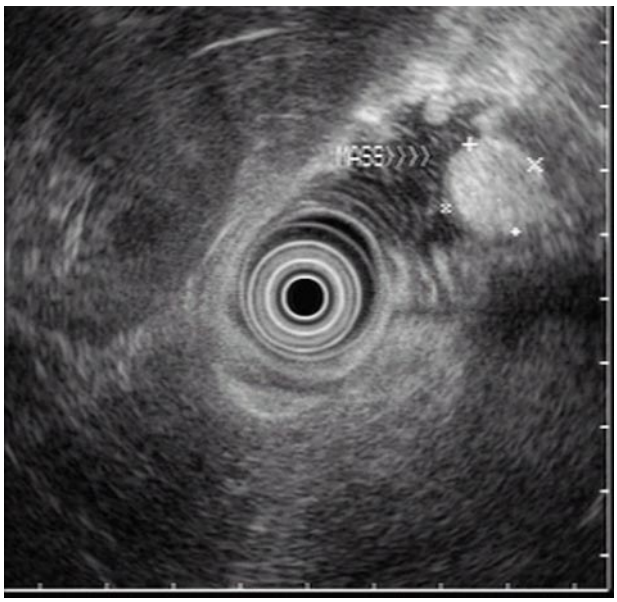


FIGURE 1 EUS image of duodenal lipoma showing hyperechoic lesion in submucosa with intact overlying superficial and deep mucosal layers.

In the United States, the appendix is the most common site for carcinoids. In this location, tumors are usually found incidentally at the time of appendectomy as they rarely produce symptoms. Carcinoids in the small bowel, appendix, and right colon may metastasize to the liver. When these metastases severely injure the liver, patients may experience the constellation of flushing, asthma, and diarrhea that is characteristic of carcinoid syndrome. In a large database review, tumors in the distal two-thirds of the colon rarely produced systemic symptoms but were most likely to present with bleeding. Gastric carcinoids are often multicentric and arise in settings of hypergastrinemia such as Zollinger-Ellison syndrome or atrophic gastritis. Gastric carcinoids also occur sporadically. In the stomach, carcinoids may cause flushing without the other elements of the carcinoid syndrome.

When viewed endoscopically, carcinoids usually appear as a smooth, yellowish mass sometimes with an umbilicated ulceration. Histologic diagnosis can be made with mucosal biopsy. EUS can be used to evaluate the critical features of size, depth of invasion, and lymph node metastases. Typically, carcinoids appear hypoechoic and homogeneous and are located in the submucosa.

Tumors with dimensions greater than 2 cm or penetration of the muscularis propria are associated with increased metastatic risk. Treatment requires surgical excision. For tumors in the colon and small bowel, the resection must include wide margins with regional lymph node dissection. Appendiceal carcinoids that are less than 2 cm can be managed more conservatively. Because these small appendiceal tumors are unlikely to recur locally after resection, local excision is sufficient. For lesions greater than 2 cm, one-third of patients will have nodal or regional metastases at the time of diagnosis and right hemicolectomy should be considered.

Small gastric carcinoids that arise in the setting of pernicious anemia and chronic atrophic gastritis are generally indolent lesions. As fewer than 10% metastasize, these carcinoids may also be managed with local excision. Endoscopic removal of gastric carcinoids is becoming more common. If the excision is complete based on follow-up video endoscopy and EUS, then

periodic endoscopic surveillance is recommended at 1- to 2-year intervals.

A 30-year review of individuals diagnosed with carcinoid tumors at one center showed that patient outcome is influenced by age, stage at diagnosis, and urinary level of 5-hydroxyindoleacetic acid. Because many pancreatic and small bowel carcinoids are metastatic at the time of diagnosis, they tend to have the worst overall outcomes with a 5-year survival of approximately 50% compared to 80–85% for appendiceal and rectal carcinoids.

Pancreatic Rest

Pancreatic rest is ectopic pancreatic tissue located in the wall of the stomach or small intestine. Pancreatic rests are thought to develop early in embryogenesis before the fusion of the dorsal and ventral pancreatic anlage tissue. The estimated prevalence at autopsy is 1–2% but they are seen in fewer than 1 in 1000 endoscopies. Most gastric lesions are seen within 6 cm of the pylorus. The typical endoscopic image of an umbilicated nodule is nonspecific. EUS shows a heterogeneous hyperechoic lesion in the submucosa. An echo-free ductal region in the basal aspect of the lesion is characteristic of ectopic pancreas. Generally, pancreatic rests are asymptomatic but there are rare reports of obstruction, pancreatitis, and development of pancreatic neoplasm.

Varices

Collateral vessels in the stomach, esophagus, small bowel, and colon form as a consequence of portal hypertension. They are seen in approximately 60% of decompensated cirrhotics and 30% of compensated cirrhotics at the time of presentation. Endoscopically, varices present as protuberant columns in the esophagus and are easy to diagnose visually. In the stomach, however, gastric varices may be mistaken for prominent folds in the gastric fundus. EUS can be helpful by providing characteristic images of echo-free intramural and extramural vessels with visible blood flow. Varices may be seen in the mucosa, submucosa, or muscle layer or may be para-esophageal or para-gastric.

Duplication Cyst

Duplication cysts are congenital intestinal duplications that can be found in the walls of the esophagus, stomach, or duodenum. Because they do not communicate with the lumen and can contain actively secreting glands, they can enlarge and produce symptoms or become complex cystic masses as secretions build up. In general, they are asymptomatic and benign. EUS demonstrates a smooth-bordered, submucosal lesion that is anechoic. EUS-guided FNA of duplication cysts is usually not recommended out of concern for introducing infection.

Granular Cell Tumor

Granular cell tumors are generally benign growths that are derived from smooth muscle or Schwann cells. Only 1% involve the gastrointestinal tract, in which case they are found most commonly in the esophagus. They may also be found in the anal canal, where they can be confused with hemorrhoids. They appear as polypoid growths on endoscopy and require deep mucosal biopsy for diagnosis. EUS images usually show a hypoechoic mass that occupies the deep mucosal layer or submucosa. Histologically, they consist of masses of cells that appear like histiocytes with a granular, eosinophilic cytoplasm. Although the probability of malignant transformation is low, endoscopic or surgical resection is advisable especially if serial examinations demonstrate interval changes in size or invasiveness. Metastatases appear rarely. Although researchers have reported a case of multiple esophageal and gastric granular cell tumors with nuclear features of malignancy, the patient was cured by local resection.

Muscularis Propria Lesions

Stromal Cell Tumor

Gastrointestinal stromal cell tumors (GISTs) are rare mesenchymal neoplasms that can be found throughout the digestive tract (Fig. 2). The vast majority (90%) originate in the stomach and small intestine and 10–30% may be malignant with intra-abdominal metastases. Formerly classified as leiomyomas and leiomyosarcomas, GIST is the new unifying nomenclature for all gastrointestinal mesenchymal tumors with unique genetic and histochemical features. GIST describes most of what were formerly considered leiomyomas and leiomyosarcomas.

GISTs are thought to arise from a stem cell with multipotent differentiation. Many GISTs show differentiation toward the pacemaker cells of the gastrointestinal tract, the interstitial cells of Cajal. Others exhibit features seen in smooth muscle and neural tissues. Histologically, GISTs show spindle cells growing in fasicles or sheets of epithelioid-type cells. GISTs are highly vascular and frequently present with gastrointestinal hemorrhage.

On EUS, stromal cell tumors are hypoechoic and arise from the muscularis propria or muscularis

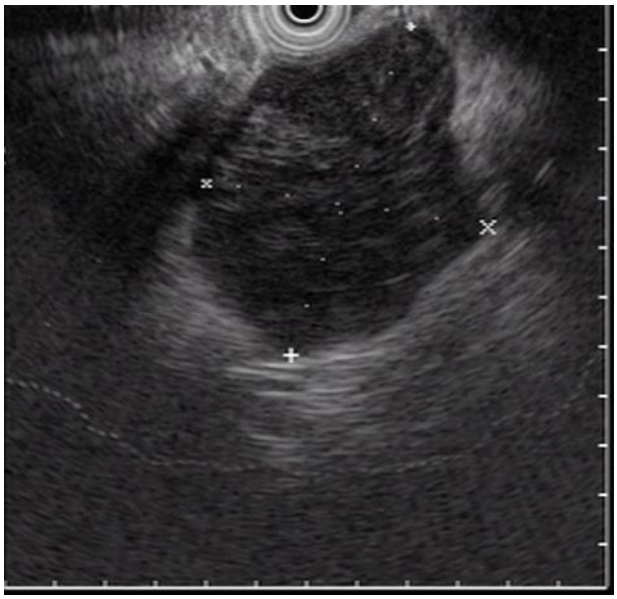


FIGURE 2 EUS image of GIST in gastric wall showing heterogeneous, hypoechoic lesion involving muscularis propria.

mucosa. The most reliable method for determining malignancy of a GIST is immunohistochemical and genetic analysis of the specimen by surgical pathology. GISTs are classified as having low, high, or indeterminate malignant potential based on the presence of several criteria that can be determined by EUS and surgical pathology. Size greater than 3 cm, a high mitotic count per high-powered field (HPF), the presence of tumor necrosis, nuclear pleomorphism, dense cellularity, an alveolar or clustered cell pattern, and microscopic invasion of the lamina propria or blood vessels are predictors of high malignant potential (Table II). Finding any two of these features suggests that a stromal tumor has high malignant potential. A stromal tumor with only one of these traits is considered a tumor with uncertain malignant potential. If these criteria are absent, the tumor is benign. Histopathology is the gold standard for determining malignant potential. As a part of this process, the results of 50 HPFs are tabulated. FNA cytology provides insufficient material to definitively determine malignant potential.

TABLE II Submucosal Tumors: Predictors of Malignancy

EUS	Histologic
Irregular margins	>50 mitoses per HPF
Size >3 cm	Tumor necrosis
	Nuclear pleomorphism
	Dense cellularity
	Alveolar or clustered cell pattern
	Microscopic invasion of the lamina propria or blood vessels
Cystic spaces	Missense mutation of c-kit gene
Lymphadenopathy	

The missense mutation of the c-kit gene, which codes for a growth factor receptor with tyrosine kinase activity, has recently been identified as a molecular biologic marker for increased recurrence of GISTs and higher mortality. Assays for c-kit and CD34 performed on fine-needle aspirates may improve the diagnostic yield of FNA in settings where malignant GIST is suspected. Research is also ongoing to delineate the sensitivity of EUS criteria alone in the evaluation of stromal tumors. In one series of 56 GISTs studied by EUS, the presence of two out of three characteristics—irregular extralumenal margins, cystic spaces, and malignant-appearing lymphadenopathy—had a positive predictive value of 100% for malignant or borderline malignant histology. Other researchers studied the Doppler characteristic of GISTs and reported that turbulent, pulsatile flow was highly sensitive for malignancy (100%).

True leiomyomas are now thought to be rare in the GI tract, occurring most often in the esophagus or gastric cardia. They are composed of spindle cells but are less cellular than GISTs and immunohistochemically they differ by not expressing CD34 or CD117 (c-kit).

Suspected stromal cell tumors that are symptomatic and associated with bleeding, obstruction, or pain should be removed. Additionally, GISTs that are greater than 3 cm or increasing in size when seen in follow-up should be removed surgically to prevent metastatases. Laparoscopic resection of GISTs has been described. Although metastatic GISTs are generally highly resistant to chemotherapy, the new tyrosine kinase inhibitor STI571 has been reported to be highly active against GIST in a patient with metastatic disease.

Extrinsic Compression

A number of adjacent structures can produce extrinsic compression on the gastrointestinal tract and appear as submucosal lesions on video endoscopy. Adjacent organs such as the spleen as well as lymph nodes, and pancreatic pseudocysts should be considered in the differential. EUS can be helpful in these situations. If a lesion is extralumenal, shows no relationship with layers of the gastrointestinal tract, and moves independently of the mucosa with the patient's respirations, then the mass is most likely external to the gastrointestinal tract.

TABLE III Submucosal Tumors: Indications for Resection

Size >3 cm
Symptoms: pain, bleeding, or obstruction
Interval enlargement
Cystic spaces
Lymphadenopathy

BIOPSY AND REMOVAL

Following characterization of an SMT by EUS, consideration should be given to periodic surveillance, biopsy, or excision. There are several methods for biopsy of SMTs. Utilizing the deep-well biopsy technique is one option whereby repeated pinch biopsies of the same point denude the mucosa and yield submucosal tissue. Alternatively, the top layer of mucosa can be removed by snare electrocautery and the submucosa can then be sampled. EUS and FNA of submucosal tumors can help determine tissue type but may not be able to demonstrate malignancy.

Most authors recommend surgical or endoscopic excision if lesions are greater than 3 cm in diameter or if they are producing symptoms (see Table III). An SMT that is shown by EUS to arise from the deep mucosa can be removed safely by conventional endoscopic polypectomy techniques. Saline injection may facilitate snare polypectomy.

In one case series, researchers used EUS to identify the wall layer of origin and to exclude characteristics suspicious for malignancy in 54 SMTs. Endoscopic resection was then performed for all masses arising in submucosa or muscularis mucosa. Muscularis propria lesions underwent mucosectomy followed by deep biopsy. Two of the 54 patients had biopsies showing malignancy. Although this procedure was safe and led to no perforations, the results suggest that EUS criteria alone may not identify all malignant lesions that require surgical removal.

CONCLUSION

In conclusion, gastroenterologists and internists may consider the endoscopic and EUS characteristics of SMTs in order to make a presumptive diagnosis of the lesion's histology and predict its behavior. With this information, a strategy for surgical or endoscopic removal or surveillance can be planned. All lesions with features that suggest malignancy should be removed. Additionally, any SMT that is large, increasing in size, or associated with symptoms should be considered for removal. Fortunately, most SMTs are benign. As imaging and histopathologic techniques improve, there may be enhanced, noninvasive approaches for the diagnosis and treatment of SMTs.

See Also the Following Articles

Barium Radiography • Computed Tomography (CT) • Endoscopic Ultrasonography

Further Reading

Axelrad, A. M., and Fleischer, D. E. (1998). Esophageal tumors. *In* "Gastrointestinal and Liver Disease" (M. Feldman, M. H. Sleisenger, and B. F. Scharschmidt, eds.), Vol. 1, pp. 540–554. W. B. Saunders, Philadelphia, PA.

Chung, S., and Roberts-Thomson, I. C. (2001). Gastrointestinal: Pancreatic rests. *J. Gastroenterol. Hepatol.* **14**, 829.

Dagher, L., and Burroughs, A. (2001). Variceal bleeding and portal hypertensive gastropathy. *Eur. J. Gastroenterol. Hepatol.* **13**, 81–88.

Ernst, S. I., Hubbs, A. E., Przygodzki, R. M., Emory, T. S., Sobin, L. H., and O'Leary, T. J. (1998). KIT mutation portends poor prognosis in gastrointestinal stromal/smooth muscle tumors. *Lab. Invest.* **78**, 1633–1636.

Goldschmiedt, M., and Gutta, K. (1998). Other tumors and diseases of the pancreas. *In* "Gastrointestinal and Liver Disease" (M. Feldman, M. H. Sleisenger, and B. F. Scharschmidt, eds.), Vol. 1, pp. 895–901. W. B. Saunders, Philadelphia, PA.

Gress, F., Schmitt, C., Savides, T., *et al.* (2001). Interobserver agreement for EUS in the evaluation and diagnosis of submucosal masses. *Gastrointest. Endosc.* **53**, 71–76.

Iwase, H., Kusugami, K., Suga, S., and Kyokane, K. (1997). Color Doppler-enhanced endoscopic ultrasonographic diagnosis of upper gastrointestinal submucosal lesions. *Dig. Endosc.* **9**, 116–121.

Joensuu, H., Roberts, P. J., Sarlomo-Rikala, M., *et al.* (2001). Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N. Engl. J. Med.* **344**, 1052–1056.

Kawamoto, K., Yamada, Y., Utsunomiya, T., *et al.* (1997). Gastrointestinal submucosal tumors: Evaluation with endoscopic US. *Radiology* **205**, 733–740.

Kindblom, L. G., Remotti, H. E., Aldenborg, F., and Meis-Kindblom, J. M. (1998). Gastrointestinal pacemaker cell tumor (GIPACT): Gastrointestinal stromal tumors show phenotypic characteristics of the interstitial cells of Cajal. *Am. J. Pathol.* **152**, 1259–1269.

Kojima, T., Takahashi, H., Parra-Blanco, A., Kohsen, K., and Fujita, R. (1999). Diagnosis of submucosal tumor of the upper GI tract by endoscopic resection. *Gastrointest. Endosc.* **50**, 516–522.

Kulke, M. H., and Mayer, R. J. (1999). Carcinoid tumors. *N. Engl. J. Med.* **340**, 858–868.

Lasota, J., Jasinski, M., Sarlomo-Rikala, M., and Miettinen, M. (1999). Mutations in exon 11 of c-Kit occur preferentially in malignant versus benign gastrointestinal stromal tumors and do not occur in leiomyomas or leiomyosarcomas. *Am. J. Pathol.* **154**, 53–60.

Lewin, K. J. (1992). Mesenchymal tumors. *In* "Gastrointestinal Pathology and Its Clinical Implications" (K. J. Lewin, R. H. Riddell, and W. M. Weinstein, eds.), Vol. 1, pp. 284–341. Igaku-Shoin, New York.

Li, S. Q., O'Leary, T. J., Buchner, S. B., *et al.* (2001). Fine needle aspiration of gastrointestinal stromal tumors. *Acta Cytol.* **45**, 9–17.

Miettinen, M., Sarlomo-Rikala, M., and Lasota, J. (1999). Gastrointestinal stromal tumors: Recent advances in understanding of their biology. *Hum. Pathol.* **30**, 1213–1220.

Modlin, I. M., and Sandor, A. (1997). An analysis of 8305 cases of carcinoid tumors. *Cancer* **79**, 813–829.

Onaitis, M. W., Kirshbom, P. M., Hayward, T. Z., *et al.* (2000). Gastrointestinal carcinoids: Characterization by site of origin and hormone production. *Ann. Surg.* **232**, 549–556.

Palazzo, L., Landi, B., Cellier, C., Cuillerier, E., Roseau, G., and Barbier, J. P. (2000). Endosonographic features predictive of benign and malignant gastrointestinal stromal cell tumours. *Gut* **46**, 88–92.

Roesch, T., Kapfer, B., Will, U., *et al.* (1999). Endoscopy alone is not sufficient for a reliable diagnosis of upper GI submucosal tumors (SMT). *Gastrointest. Endosc.* **49**, AB211.

Rosch, T., Lorenz, R., Dancygier, H., von Wickert, A., and Classen, M. (1992). Endosonographic diagnosis of submucosal upper gastrointestinal tract tumors. *Scand. J. Gastroenterol.* **27**, 1–8.

Seidal, T., and Edvardsson, H. (2000). Diagnosis of gastrointestinal stromal tumor by fine-needle aspiration biopsy: A cytological and immunocytochemical study. *Diagn. Cytopathol.* **23**, 397–401.

Wolfsohn, D. M., Savides, T. J., Easter, D. W., and Lyche, K. D. (1997). Laparoscopy-assisted endoscopic removal of a stromal-cell tumor of the stomach. *Endoscopy* **29**, 679–682.

Yasuda, K. (2000). Gastrointestinal lesions. *In* "The Handbook of Endoscopic Ultrasonography in Digestive Tract," Vol. 1, pp. 45–82. Blackwell Science, Malden, MA.

Substance P

STEVEN R. VIGNA

Duke University Medical Center and Durham Veteran's Affairs Medical Center

capsaicin Natural ingredient of hot peppers of the genus *Capsicum*, the pungent ingredient of many "hot" and spicy foods. A subtype of primary afferent neurons is sensitive to capsaicin; at low doses, capsaicin stimulates these neurons and is toxic to the neurons at high doses or after prolonged treatment.

chemical coding Particular combination of neurotransmitters, neuropeptides, and other neuronal chemical markers characteristic of a particular class of neuron.

dorsal root ganglia Series of paired neural ganglia adjacent to the spinal cord; contain the neuronal cell bodies of capsaicin-sensitive primary afferent nerves as well as other afferent nerves.

enteric nervous system Neurons and supporting cells found within the walls of the gastrointestinal tract, including the pancreas and gallbladder.

myenteric plexus (Auerbach's plexus) Network of nerves and small ganglia that lie in the plane between the external longitudinal and inner circular layers of smooth muscle in the muscularis externa of the intestine.

submucous plexus (Meissner's plexus) Network of small ganglia and nerves located in the submucosa of the intestine.

tachykinins Family of neuropeptides sharing the same carboxyl-terminal amino acid sequence comprising the molecular center of biological activity of each peptide (F-x-G-L-M-amide). The mammalian tachykinins include substance P, neurokinin A, neuropeptide K, neuropeptide γ, and neuropeptide B.

Substance P, a peptide composed of 11 amino acids, is expressed by neurons in the central, peripheral, and enteric nervous systems. Substance P is a major neurotransmitter in the gastrointestinal tract; it affects motility, secretion, and inflammation and can cause vomiting by an action in the central nervous system. Most actions of substance P are mediated by the neurokinin-1 receptor, thus excellent receptor antagonists have been developed to aid in differentiation of physiological versus pharmacological actions of substance P. These receptor antagonists may also prove to be of therapeutic value in treating substance P-associated inflammatory bowel disease, acute intestinal inflammation, and chemotherapy-induced nausea and vomiting.

INTRODUCTION

The 11 amino acids of the peptide substance P (SP) are arranged in the sequence RPKPQQFFGLM-amide.

Copyright 2004, Elsevier (USA). All rights reserved.

Li, S. Q., O'Leary, T. J., Buchner, S. B., *et al.* (2001). Fine needle aspiration of gastrointestinal stromal tumors. *Acta Cytol.* **45**, 9–17.

Miettinen, M., Sarlomo-Rikala, M., and Lasota, J. (1999). Gastrointestinal stromal tumors: Recent advances in understanding of their biology. *Hum. Pathol.* **30**, 1213–1220.

Modlin, I. M., and Sandor, A. (1997). An analysis of 8305 cases of carcinoid tumors. *Cancer* **79**, 813–829.

Onaitis, M. W., Kirshbom, P. M., Hayward, T. Z., *et al.* (2000). Gastrointestinal carcinoids: Characterization by site of origin and hormone production. *Ann. Surg.* **232**, 549–556.

Palazzo, L., Landi, B., Cellier, C., Cuillerier, E., Roseau, G., and Barbier, J. P. (2000). Endosonographic features predictive of benign and malignant gastrointestinal stromal cell tumours. *Gut* **46**, 88–92.

Roesch, T., Kapfer, B., Will, U., *et al.* (1999). Endoscopy alone is not sufficient for a reliable diagnosis of upper GI submucosal tumors (SMT). *Gastrointest. Endosc.* **49**, AB211.

Rosch, T., Lorenz, R., Dancygier, H., von Wickert, A., and Classen, M. (1992). Endosonographic diagnosis of submucosal upper gastrointestinal tract tumors. *Scand. J. Gastroenterol.* **27**, 1–8.

Seidal, T., and Edvardsson, H. (2000). Diagnosis of gastrointestinal stromal tumor by fine-needle aspiration biopsy: A cytological and immunocytochemical study. *Diagn. Cytopathol.* **23**, 397–401.

Wolfsohn, D. M., Savides, T. J., Easter, D. W., and Lyche, K. D. (1997). Laparoscopy-assisted endoscopic removal of a stromal-cell tumor of the stomach. *Endoscopy* **29**, 679–682.

Yasuda, K. (2000). Gastrointestinal lesions. *In* "The Handbook of Endoscopic Ultrasonography in Digestive Tract," Vol. 1, pp. 45–82. Blackwell Science, Malden, MA.

Substance P

STEVEN R. VIGNA

Duke University Medical Center and Durham Veteran's Affairs Medical Center

capsaicin Natural ingredient of hot peppers of the genus *Capsicum*, the pungent ingredient of many "hot" and spicy foods. A subtype of primary afferent neurons is sensitive to capsaicin; at low doses, capsaicin stimulates these neurons and is toxic to the neurons at high doses or after prolonged treatment.

chemical coding Particular combination of neurotransmitters, neuropeptides, and other neuronal chemical markers characteristic of a particular class of neuron.

dorsal root ganglia Series of paired neural ganglia adjacent to the spinal cord; contain the neuronal cell bodies of capsaicin-sensitive primary afferent nerves as well as other afferent nerves.

enteric nervous system Neurons and supporting cells found within the walls of the gastrointestinal tract, including the pancreas and gallbladder.

myenteric plexus (Auerbach's plexus) Network of nerves and small ganglia that lie in the plane between the external longitudinal and inner circular layers of smooth muscle in the muscularis externa of the intestine.

submucous plexus (Meissner's plexus) Network of small ganglia and nerves located in the submucosa of the intestine.

tachykinins Family of neuropeptides sharing the same carboxyl-terminal amino acid sequence comprising the molecular center of biological activity of each peptide (F-x-G-L-M-amide). The mammalian tachykinins include substance P, neurokinin A, neuropeptide K, neuropeptide γ, and neuropeptide B.

Substance P, a peptide composed of 11 amino acids, is expressed by neurons in the central, peripheral, and enteric nervous systems. Substance P is a major neurotransmitter in the gastrointestinal tract; it affects motility, secretion, and inflammation and can cause vomiting by an action in the central nervous system. Most actions of substance P are mediated by the neurokinin-1 receptor, thus excellent receptor antagonists have been developed to aid in differentiation of physiological versus pharmacological actions of substance P. These receptor antagonists may also prove to be of therapeutic value in treating substance P-associated inflammatory bowel disease, acute intestinal inflammation, and chemotherapy-induced nausea and vomiting.

INTRODUCTION

The 11 amino acids of the peptide substance P (SP) are arranged in the sequence RPKPQQFFGLM-amide.

Copyright 2004, Elsevier (USA). All rights reserved.

Substance P is a member of the tachykinin family of mammalian neuropeptides. Other mammalian tachykinins include neurokinin A (NKA), the NKA-related peptides neuropeptide K (NPK) and neuropeptide γ (NPγ), and neurokinin B (NKB). Substance P got its unusual name from its discoverers, who identified an intestine-contracting activity extracted from equine intestine (it was then called preparation P). The tachykinins are named for their rapid ability to act on gut tissues, compared to the slower acting bradykinins. Substance P is encoded by the preprotachykinin I (PPT I) gene. Four mRNAs, termed α-PPT I, β-PPT I, γ-PPT I, and δ-PPT I, can be generated from the gene for PPT I by alternative mRNA splicing. The α-PPT I and δ-PPT I mRNAs encode only SP whereas the β-PPT I and γ-PPT I mRNAs encode both SP and NKA. Precursor proteins are synthesized from all four mRNAs and subsequent posttranslational processing generates the tachykinin peptides. β-PPT I and γ-PPT I proteins are subject to multiple proteolytic cleavages, resulting in the generation of NPK and NPγ, respectively, as well as SP and NKA. NKB is encoded by the PPT II gene, which does not encode any other tachykinins. Thus, although the genome contains only one gene for SP, the mRNA transcribed from the gene and the precursor proteins translated from the mRNAs can be processed in multiple ways to result in the expression of various combinations of SP, NKA, NPK, and NPγ in nerves.

In the gastrointestinal tract of the rat, the only species in which PPT I has been examined in detail, γ-PPT I mRNA comprises 80−90% of the total tachykinin RNA and β-PPT I mRNA and α-PPT I mRNA account for 10−20% and less than 1%, respectively. Thus, all members of the tachykinin family of neuropeptides except NKB are probably expressed in the gut. Three receptors correspond to the three major mammalian tachykinins. All three tachykinin receptors are G-protein-coupled receptors. Each receptor exhibits its highest affinity binding to a different tachykinin: SP has highest affinity for the neurokinin-1 (NK-1) receptor, NKA has highest affinity for the neurokinin-2 (NK-2) receptor, and NKB has highest affinity for the neurokinin-3 (NK-3) receptor. However, all three receptors can interact with all three mammalian tachykinins pharmacologically and it is possible that they do so physiologically as well. Substance P has been shown to stimulate several second-messenger systems coupled to the NK-1 receptor, i.e., inositol 1,4,5-trisphosphate, calcium, cyclic adenosine monophosphate (cAMP), mitogen-activated protein (MAP) kinase, and arachidonic acid.

Substance P was originally discovered in extracts of gastrointestinal tissue and was shown to contract gastrointestinal muscle in 1931. Studies of the role of SP in regulating the gastrointestinal tract have continued unabated since the 1930s and are still contributing insights into the nature of SP. Because of its potential importance in understanding the mechanisms of normal physiology and serving as the basis for potential new therapies for various diseases, SP is one of the best studied neuropeptides. In addition to detailed knowledge of the molecular biology of the peptide and its receptor, great strides have been made in understanding the pharmacology of SP, including the development of multiple selective agonists and antagonists for the NK-1 receptor. In the gut, SP is an important neurotransmitter both in enteric neurons (those intrinsic to the gut) and in extrinsic primary afferent nerves (the cell bodies of which are located adjacent to the spinal cord in the dorsal root ganglia). Various specific stimuli cause the release of SP and it has a predominantly excitatory action on gut nerve, muscle, and glands.

GASTROINTESTINAL MOTILITY

The muscle layers of the gut (muscularis externa and muscularis mucosae) are innervated by a dense array of SP-containing nerve fibers, most of which originate from the enteric nervous system. The neuronal cell bodies in the enteric nervous system are in the myenteric plexus (MP), located between the inner circular and the outer longitudinal layers of smooth muscle of the muscularis externa, and the submucosal plexus (SMP), located in the submucosa. The SP-expressing enteric neurons have been mapped in detail in the guinea pig intestine such that their projections within and between the nerve plexuses and to the muscle layers and mucosa are well established. Although the neuroanatomy of gastrointestinal SP is not as well known in species other than the guinea pig, there do not appear to be major functional differences among mammalian species. The physiological functions of neurons depend on their chemical coding (the combination of neurotransmitters, neuropeptides, and other neuronal chemical markers characteristic of a particular neuron) and morphological characteristics, and studies of these features of enteric SP neurons have revealed roles as sensory neurons, interneurons, and motor neurons. Almost all MP and SMP SP neurons in the guinea pig intestine coexpress acetylcholine (ACh) whereas the extrinsic SP-expressing neurons coexpress calcitonin gene-related peptide (CGRP) and are sensitive to capsaicin, the pungent ingredient of hot peppers. Enteric SP neurons are not sensitive to capsaicin. Thus, it is possible to determine the intrinsic versus extrinsic source of SP in the gut by administering capsaicin, which acts as

an excitotoxin by first selectively stimulating and then, at higher doses or longer times, defunctionalizing extrinsic primary afferent, but not intrinsic, enteric neurons.

Various stimuli, including electrical depolarization, distension of the gut wall, intraluminal acid in the stomach, inflammation, various neurotransmitters and hormones, and capsaicin, can elicit SP release in the gut. Tachykinins enhance motor activity in virtually all regions and layers of the mammalian gut. Sometimes this action is a direct one on muscle, but in other cases, it reflects activation of enteric neurons that stimulate the muscle by release of ACh. A further complication arises when it is remembered that nerves that express the PPT I gene products most often corelease NKA and/or the NKA-like peptides NPK or NPγ along with SP. Furthermore, NK-1 receptors are expressed by both nerves and muscle cells whereas NK-2 receptors are mostly limited to smooth muscle cells. As a final complication, even though most tachykinin actions on gut motility are stimulatory, they can also inhibit gut muscle by stimulating inhibitory neural pathways or by interrupting stimulatory pathways. Most of the stimulatory actions of SP and NKA on gastrointestinal motility, especially in the case of peristalsis, are in synergism with ACh, with which they are coreleased from enteric neurons.

In studies of the human esophagus *in vitro*, tachykinins stimulate contraction of the lower esophageal sphincter (LES) and esophageal body via NK-2 receptors. In animal studies, tachykinins facilitate swallowing of fluids and increase LES pressure. In some species, SP causes only a transient LES contraction that is followed by a prolonged relaxation mediated by NK-1 receptors. It is not known if SP plays a major physiological role in normal human esophageal function.

Most of the information available about the effects of SP on gastric motility comes from animal studies. In the rat, injection of SP can either stimulate or inhibit gastric emptying and transit of a liquid meal, depending on whether parasympathetic or sympathetic reflexes are simultaneously activated. In the cat and dog, systemic administration of SP induces phasic contraction of the stomach via ACh release. Overall, SP is primarily stimulatory to gastric motility.

In the human small intestine and colon, NK-1 receptors are expressed by smooth muscle cells in the muscularis externa and muscularis mucosae and on neurons in the MP, and NK-2 receptors are expressed by smooth muscle but not by enteric neurons. This distribution correlates well with the observations that NKA directly stimulates human intestinal muscle contraction and that the effects of SP seem to occur indirectly via nerves. Intestinal peristalsis is a form of integrated propulsive motility resulting from the sequential activation of ascending excitatory and descending inhibitory reflexes initiated by local stretch of the intestinal wall, and resulting in propulsion of the contents of the intestine aborally. Because local stretch of the intestinal wall is the major stimulus for peristalsis, the role of tachykinins in the ascending contraction and descending relaxation of intestinal circular muscle in response to stretch or radial distension of the gut wall has been examined in detail. The peristaltic reflex is complex, consisting of distension-sensitive enteric sensory nerves, interneurons, orally projecting excitatory motor neurons to cause ascending contraction, and aborally projecting inhibitory motor neurons to cause descending relaxation. A variety of experimental approaches, including the administration of tachykinin receptor antagonists, SP desensitization, and SP immunoneutralization, have been used to demonstrate that SP and ACh are involved in the ascending reflex contraction caused by gut wall distension and that the descending relaxation is independent of tachykinins.

Additional analyses with antagonists selective for the NK-1 or NK-2 receptors have revealed that NK-2 receptors primarily mediate ascending reflex contraction in the guinea pig intestine, although some involvement of NK-1 receptors is also evident. It has also been demonstrated that stretch-induced ascending contraction, but not descending relaxation, of the human jejunum and rat colon is associated with release of endogenous SP and NKA. The concept that tachykinins are involved in the regulation of gastrointestinal propulsive motility *in vivo* is supported by the inhibition of gastric emptying and gastrointestinal transit in rats after administration of a tachykinin receptor antagonist.

Substance P contracts the human gallbladder by a direct effect on smooth muscle cells, but the physiological significance of this effect is not known.

GASTROINTESTINAL SECRETION

There is a large body of evidence that SP participates, along with several other sialogogic messengers, in the parasympathetic regulation of salivary secretion. In rats, release of SP (and NKA) from the parotid glands is increased after nerve stimulation. In addition, the tissue content of SP is greatly reduced after prolonged stimulation of the parasympathetic but not sympathetic nerves supplying the salivary glands. Salivation, which also decreases after prolonged nerve stimulation, is restored by infusion of subthreshold doses of SP.

Tachykinin receptor antagonists reduce salivary output to parasympathetic nerve stimulation in the rat. The salivation evoked by sympathetic nerve stimulation in the rat is not affected by tachykinin receptor antagonists. The stimulatory effect of SP on salivation in rats and other species is potentiated by the coreleased neuropeptide vasoactive intestinal polypeptide (VIP). It is likely that SP is a physiologically important regulator of salivation in animals and people.

In contrast, the weak and variable effects of SP on gastric secretion suggest that the peptide does not play a major role in regulating gastric secretion. The currently available evidence supports only a potential modulatory influence of SP on the secretory functions of the stomach. In addition, there appear to be major species differences in the effects of SP on gastric secretion. For example, SP does not affect basal gastric acid secretion in rats or dogs but SP does stimulate gastric acid secretion in cats. Similar differences have been noted in the effects of SP on stimulated gastric acid secretion. The actions of SP on gastric acid secretion are so divergent that the only conclusion possible is that its effects depend on the species studied, the experimental conditions used, and the particular stimulus used to stimulate secretion. A good way to test the hypothesis that SP plays a physiological role in regulating gastric acid secretion would be to test the effects of NK-1 receptor antagonists, but this does not appear to have been done yet. Based on the evidence that SP is a potent pepsigogue in dogs and guinea pigs, it is possible that SP is a physiologically significant regulator of pepsinogen secretion. Pepsinogen is the inactive precursor of the proteolytic enzyme, pepsin, in the stomach. Pepsinogen is secreted by gastric chief cells, which have been shown to express NK-1 receptors, and is converted to pepsin in the gastric lumen by gastric acid.

There is strong evidence that SP is an important stimulant of secretion in the small intestine and colon. Systemic administration of SP induces a net secretion of water, sodium, potassium, and chloride into the small intestine of dogs, cats, and ferrets. SP stimulates an enteric neural reflex that results in the activation of secretomotor VIP neurons. A component of the SP effect can also be accounted for by a direct effect of SP on the mucosal epithelium. SP stimulates secretion by increasing the short circuit current (I_{sc}), an index of electrogenic anion secretion. The increased I_{sc} is accompanied by a net secretion of sodium, chloride, and bicarbonate. The relative degree to which the SP effect is mediated by nerves versus direct effects on the mucosal epithelium varies with the type of tachykinin agonist used and the region of intestine studied. In the porcine and guinea pig small intestines, both the neural and epithelial responses are mediated by NK-1 receptors. The situation may be different in other species. However, the involvement of NK-1 receptors is consistent with the observation that secretomotor neurons in both the MP and SMP express NK-1 receptors. The stimulant action of SP on I_{sc} in the small intestine requires calcium influx into the tissue because it is inhibited by removal of extracellular calcium or by pretreatment with verapamil, a calcium channel-blocking drug. The sodium/potassium/chloride cotransport mechanism has also been demonstrated to be important in SP-evoked stimulation of I_{sc}.

In the rat colon, SP converts fluid absorption into net fluid secretion. The secretory responses to NK-1 and NK-2 receptor agonists are suppressed by tetrodotoxin (TTX), an inhibitor of neural transmission, and blockade of nitric oxide (NO) synthesis. These findings have led to the proposal that SP evokes secretion in the rat colon by a sequential activation of NK-1 and NK-2 receptors and the formation of NO in a complex enteric neural reflex. The I_{sc} response in the mucosa of the guinea pig colon is mediated by NK-1 receptors. The prominence of NK-1 receptor actions differentiates the guinea pig colon from the rat colon, in which all three tachykinin receptors act to increase I_{sc}.

These observations and others have given rise to the concept that stimulation of MP and SMP neurons that innervate the mucosa of the guinea pig small intestine releases SP and NKA within the enteric ganglia, activating VIP-mediated secretomotor pathways, and within the mucosa, stimulating epithelial cell secretion directly. This conclusion is supported by experiments showing that SP desensitization, SP immunoneutralization, and NK-1 receptor antagonists inhibit the increase in I_{sc} caused by electrical field stimulation of the guinea pig small intestinal and colonic mucosae. Thus, it is clear that SP influences the secretory activity of the small intestine and colon by switching function from net absorption to net secretion of fluid and electrolytes. As discussed below, there is also increasing evidence that SP-stimulated intestinal secretion may play a role in various intestinal inflammatory diseases.

Substance P has also been demonstrated to reduce basal and hormone-stimulated bile output and to increase digestive enzyme secretion from the exocrine pancreas, but the physiological significance of these actions is unknown.

INTESTINAL INFLAMMATION

One of the sources of SP in the gastrointestinal tract is primary afferent neurons, which have cell bodies in the

dorsal root ganglia (DRG) adjacent to the spinal cord. A subset of these sensory neurons express SP as well as other neurotransmitters (such as CGRP) and are sensitive to the excitotoxin, capsaicin. Many observations in recent years have supported the concept that these extrinsic sensory neurons expressing SP are involved in both conveying nociceptive information to the spinal cord and in regulating the inflammatory and immune responses in the peripheral tissues they innervate, including the gut. Thus, SP released by extrinsic sensory neurons in response to peripheral tissue damage may signal pain in the spinal cord (eventually resulting in pain perception centrally) and may also participate in regulating inflammation, immune responses, and, ultimately, wound healing in the affected peripheral tissue. In this set of responses, the same system (primary afferent neurons) that innervates the entire body (except the brain) is also involved in sensing tissue damage (using the pain response to avoid further damage) and in initiating repair responses such as inflammation and wound healing. In this scenario, acute inflammation is adaptive because it is an initial step in wound healing. However, in susceptible individuals, the inflammation may become chronic, for unknown reasons, resulting in debilitating chronic inflammatory bowel diseases (IBDs) such as Crohn's disease (CD) and ulcerative colitis (UC) in the gastrointestinal tract. In addition, some acute inflammatory conditions in the gut, such as pseudomembranous colitis, appendicitis, and cholecystitis, often require medical intervention because of the intensity of the inflammation.

Animal studies support the conclusion that SP released from DRG fibers innervating the intestinal mucosa plays a significant role in intestinal inflammation. In rats, NK-1 receptor antagonists have been shown to inhibit intestinal inflammation in both acute and chronic inflammation models. Pretreatment with neurotoxic doses of capsaicin to eliminate SP-expressing primary sensory afferent neurons also inhibits acute intestinal inflammation in the rat intestine; acute treatment with excitatory intraluminal doses of capsaicin elicits inflammation. In addition, targeted deletion of the gene encoding the NK-1 receptor in mice also strongly inhibits acute intestinal inflammation. These observations have given rise to the concept that inflammatory agents in the gut somehow activate capsaicin-sensitive primary sensory nerves in the mucosa, resulting in local SP release that subsequently is involved in stimulating an inflammatory cascade. Substance P is well known to cause activation of immune system cells, to increase permeability of the vascular endothelium (resulting in plasma extravasation), and to stimulate leukocyte chemotaxis.

Although it is not known if SP is involved in acute or chronic gastrointestinal inflammation in humans, indirect evidence exists supporting this concept. Tissue levels of SP have been shown to be both higher and lower than normal in UC and CD. These discrepancies may reflect tissue sampling in different phases of the disease process or other factors. In addition, changes in tachykinin receptor binding in UC and CD have been examined in detail. Examination of several neurotransmitter and hormone receptors has revealed that only NK-1 receptor binding is significantly affected in surgically resected specimens of IBD tissue. In tissue from both UC and CD patients, NK-1 receptor binding is dramatically increased. The increased NK-1 receptor binding occurs in arteriolar endothelium and in lymph nodules, sites of activation of inflammatory and immune responses. Neurokinin-2 receptor binding is not different in normal control versus IBD tissue, and NK-3 receptors are not detected in the human intestine.

To assess whether the increased NK-1 receptor binding is restricted to chronic gut inflammatory conditions, appendix tissue from patients with appendicitis, negative appendicitis (in which the patient presents with the symptoms of appendicitis but pathological analysis indicates the surgically resected tissue is normal), and normal controls were analyzed as representative of acute gut inflammation. In normal appendix, similar to normal colon, low levels of NK-1 receptor binding are observed in the circular layer of smooth muscle of the muscularis externa. In the inflamed appendicitis tissue, additional NK-1 receptor binding sites are evident in small arterioles and lymph nodules in the appendix. Interestingly, the negative appendix tissue also shows additional NK-1 binding sites associated with blood vessels and lymph nodules, suggesting that increased NK-1 receptor binding occurs prior to the infiltration of mononuclear inflammatory cells characteristic of the inflamed appendix. Whether SP and NK-1 receptors initiate the inflammatory changes in IBD and appendicitis as they do in animal models of chronic and acute intestinal inflammation is unknown, but the data from the negative appendicitis cases suggest that NK-1 receptors may indeed mediate the plasma extravasation and infiltration of mononuclear inflammatory cells observed in these diseases.

CHEMOTHERAPY-INDUCED NAUSEA AND VOMITING

Cancer patients treated with chemotherapeutic agents often suffer from acute and/or delayed nausea and vomiting (emesis) caused by the chemotherapy drug. The

cornerstone of current antiemetic therapy consists of treatment with a serotonin (5-hydroxytryptamine; 5-HT) type 3 receptor (5-HT3) antagonist. The combination of a 5-HT3 antagonist and a corticosteroid such as dexamethasone is the standard of care for cancer patients receiving moderate to intense chemotherapy. These agents appear to exert their antiemetic actions at peripheral sites in the body. When this treatment is optimized, a majority of cancer patients receiving chemotherapy experience no emesis during their treatment. However, incompletely controlled nausea and vomiting is a significant problem for a minority of patients, such as those undergoing high-dose chemotherapy regimens and those developing emesis more than 24 hours after chemotherapy (delayed emesis). In these patients, 80–95% suffer from nausea and vomiting in the 7 days following the start of chemotherapy despite the use of standard antiemetic treatment.

It is clear that more effective antiemetic treatments are needed. In animal studies conducted in the mid-1990s, it was found that injection of SP induces emesis and that treatment with NK-1 receptor antagonists inhibits emesis. In fact, compared to 5-HT3 receptor antagonists, NK-1 receptor antagonists inhibited a much broader spectrum of emetic stimuli, including morphine, apomorphine, nicotine, copper sulfate, ipecac, radiation, cyclophosphamide, cisplatin, motion, and anesthesia. The NK-1 receptor antagonists also differed from the serotonin receptor antagonists in their site of action by acting in the central nervous system, presumably to antagonize the actions of endogenously released SP. Another advantage of NK-1 receptor antagonists over serotonin receptor antagonists relates to the activity of these agents in acute versus delayed emesis. The primary activity of the serotonin receptor antagonists has been observed on acute emesis in animals and humans, with little effect on cisplatin-induced delayed emesis. In contrast, NK-1 receptor antagonists exhibit activity in both acute and delayed emesis.

In early clinical trials, NK-1 receptor antagonists proved to be safe and effective in preventing acute cisplatin-induced nausea and vomiting, although when used alone they may be no more active than the 5-HT3 receptor antagonists. When an NK-1 receptor antagonist is combined with a 5-HT3 receptor antagonist and dexamethasone, control of acute cisplatin-induced emesis is improved by 20–30% over that obtained with the combination of a 5-HT3 receptor antagonist and dexamethasone without the NK-1 receptor antagonist. Most significantly, complete control of cisplatin-induced delayed emesis is better by 30–40% when using the triple drug combination compared to placebo.

It is clear that additional clinical trials are necessary to determine the potential value of NK-1 receptor antagonists in treating chemotherapy-induced nausea and vomiting. The early findings are very promising and it is likely that NK-1 receptor antagonist therapy in some form will find a place in the future treatment of these conditions.

SUMMARY

The known physiological actions of SP in the gut include a role in peristalsis in the small intestine and colon and the stimulation of salivation. Although less well studied, SP may also prove to be important in esophageal motility, pepsinogen secretion, gallbladder contraction, intestinal secretion, bile output, and pancreatic secretion, but the physiological significance of these actions has yet to be determined.

See Also the Following Articles

Colitis, Ulcerative • Crohn's Disease • Emesis • Enteric Nervous System • Gastric Motility • Nausea • Parasympathetic Innervation • Pepsin • Salivary Glands, Physiology • Vasoactive Intestinal Peptide (VIP)

Acknowledgments

This work was supported by NIH Grant DK-50265 and by a Merit Review grant from the Department of Veterans Affairs.

Further Reading

Fong, T. M. (1996). Molecular biology of tachykinins. *In* "Neurogenic Inflammation" (P. Geppetti and P. Holzer, eds.), pp. 3–14. CRC Press, Boca Raton.

Helke, C. J., Krause, J. E., Mantyh, P. W., Couture, R., and Bannon, M. J. (1990). Diversity in mammalian tachykinin peptidergic neurons: Multiple peptides, receptors, and regulatory mechanisms. *FASEB J.* **4**, 1606–1615.

Hesketh, P. J. (2001). Potential role of the NK_1 receptor antagonists in chemotherapy-induced nausea and vomiting. *Support Care Cancer* **9**, 350–354.

Holzer, P., and Holzer-Petsche, U. (1997). Tachykinins in the gut. Part I. Expression, release and motor function. *Pharmacol. Ther.* **73**, 173–217.

Holzer, P., and Holzer-Petsche, U. (1997). Tachykinins in the gut. Part II. Roles in neural excitation, secretion and inflammation. *Pharmacol. Ther.* **73**, 219–263.

Maggi, C. A., Patacchini, R., Rovero, P., and Giachetti, A. (1993). Tachykinin receptors and tachykinin receptor antagonists. *J. Autonom. Pharmacol.* **13**, 23–93.

Mantyh, P. W., Vigna, S. R., and Maggio, J. E. (1994). Receptor involvement in pathology and disease. *In* "The Tachykinin Receptors" (S. H. Buck, ed.), pp. 581–610. Humana Press, Totowa, New Jersey.

Pothoulakis, C., Castagliuolo, I., and LaMont, J. T. (1998). Nerves and intestinal mast cells modulate responses to enterotoxins. *News Physiol. Sci.* **13**, 58–63.

Swallowing

DEVANG N. PRAJAPATI AND REZA SHAKER
Medical College of Wisconsin, Milwaukee

cricopharyngeous Striated muscle attached to the posterior aspect of the lamina of the cricoid cartilage; it forms the major component of the upper esophageal sphincter.

pharyngeal swallow Reflexive swallows independent of volitional control largely responsible for clearance of residual gastric refluxate and swallowed contents.

Swallowing, or deglutition, is a highly coordinated activity that results in the transport of food and secretions from the oral cavity into the stomach via the esophagus. Since the respiratory and digestive tracts cross in the pharynx, the deglutitive sequence includes mechanisms that prevent aspiration of swallowed contents into the airway. As a result, with swallowing, respiration is suspended and a conformational change occurs in the structure of the oropharynx from a primarily respiratory pathway to a digestive pathway. An intricate network of neuromuscular signals controls the complex deglutitive process.

INTRODUCTION

Initiation of a swallow may be volitional or reflexive. Once initiated, however, every swallow leads to a predictable, involuntary sequence of events during which the bolus clears the pharynx and is transported to the stomach. On average, an individual swallows once per minute, usually to clear secretions, and approximately 1000 times per day, not including swallows with meals. Since the entire oropharyngeal component of the swallowing process takes less than a couple of seconds, evaluating this sequence has in the past been difficult. With the recent advent of newer imaging technologies, a greater understanding of this complex process is being obtained. Normal deglutition is composed of four phases: the preparatory phase, the oral phase, the pharyngeal phase, and the esophageal phase.

OROPHARYNGEAL STRUCTURES

Many oropharyngeal structures take part in the deglutitive process. Figure 1 depicts the various muscles and surrounding tissues that are involved in the swallowing process. A total of 30 paired striated muscles and 6 pairs of cranial and cervical nerves participate in oropharyngeal swallowing. Deglutitive muscle groups are attached to cartilage (i.e., epiglottic, arytenoid, cricoid, thyroid) and the hyoid bone to provide a conduit for bolus transit and to close the airway during swallowing. The tongue is the most important muscle in the preparatory and oral phases of swallowing. In the pharynx, various groups of muscles combine to perform actions that propel the bolus through the esophagus and prevent aspiration or nasal regurgitation. The intrinsic muscles of the pharynx, the superior, middle, and inferior constrictors, are largely responsible for a peristaltic wave that aids bolus transport to the esophagus and clears the pharynx of residue. The upper esophageal sphincter is composed of a group of muscles that prevent gastric refluxate from entering the oropharynx. The cricopharyngeus is the major muscle of the sphincter along with the inferior pharyngeal constrictors, the most proximal part of the esophagus, and the thyropharyngeus, all of which maintain tonic contraction of the sphincter at rest. Protection of the airway from aspiration is achieved by closure of the larynx, as the upper esophageal sphincter is opened.

NEUROLOGIC PATHWAYS

Much has been discovered about the central and peripheral neural pathways of deglutition in the past several years. The major components of the neural control of swallowing include sensory afferent and motor efferent fibers, located in the cranial nerves, and central organizing centers, located in the brainstem.

Sensory afferents are carried by branches of the glossopharyngeal (IX) and vagus (X) nerves to the nucleus tractus solitarius in the medulla. These fibers also carry sensory information from pulmonary stretch receptors and chemoreceptors located in the carotid and aortic bodies. This common neural pathway allows for control of respiration during deglutition. Afferent stimulation of receptors in the oropharynx is capable of initiating an automatic or reflexive/pharyngeal swallow, independent of volitional control. The pharyngeal swallow appears to be an important mechanism for pharyngeal

 Copyright 2004, Elsevier (USA). All rights reserved.

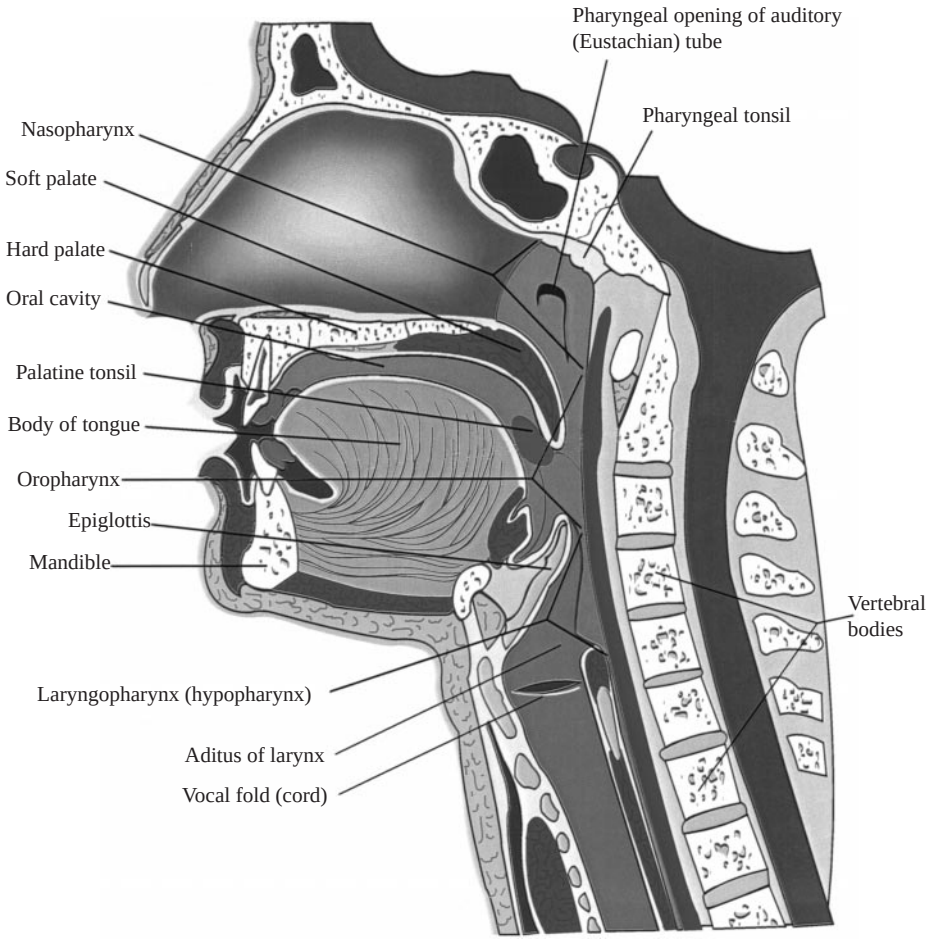


FIGURE 1 Muscles and surrounding tissues involved in the swallowing process.

clearance for airway protection. Stimulation of the superior laryngeal nerve is most effective in producing a swallow, followed by the glossopharyngeal nerve. Other mechanoreceptors on the tongue, soft palate, and tonsillar pillars are also able to initiate a swallow if mechanically stimulated by liquid.

The brainstem plays an important role in deglutition. The motor nuclei of the cranial nerves innervating the pharyngeal muscles are located in the pons and medulla. These include the motor nuclei of the trigeminal (V), facial (VII), glossopharyngeal (IX), and vagus (X) nerves. Although stimulation of the individual nuclei will result in contraction of the various deglutitive muscles, they alone are unable to initiate the swallow reflex. Concurrent stimulation of multiple sensory afferents and summation of sensory inputs to interneurons above these nuclei, which coordinate this activity, are required. Central paired swallowing centers are located in the medulla oblongata. These poorly defined areas include the nucleus tractus solitarius, ventromedian reticular formation, and the nucleus ambiguous and are responsible for processing afferent sensory signals and programming the motor swallowing sequence. This area also receives cortical input for volitional control of deglutition.

Recent studies have shown that cortical inputs have a significant influence on the brainstem swallowing centers and a complex network of interneurons coordinates neuromuscular activity between these two levels of neuronal function. Functional magnetic resonance imaging has demonstrated cortical activation in the anterior cingulate gyrus, the motor/premotor cortex, the insula and occipital/parietal lobes in Brodmann's areas 7, 19, and 31 with deglutition. These areas, however, are not unique to deglutition and also are activated in nondeglutitive oral actions.

PHASES OF DEGLUTITION

Preparatory and Oral Phases

When an individual takes in food, the food is broken down mechanically by the muscles of mastication

and chemically by salivary secretions, into a bolus that is transportable by the tongue to the oropharynx during the preparatory phase. The subsequent oral phase is involved in the transfer of the bolus from the oral cavity to the pharynx. The tongue is the major muscle involved in this phase of deglutition and is able to adapt its shape and propulsive force depending on the size of the bolus. Sequential squeezing of the tongue against the hard and soft palate generates a peristaltic pressure wave that propels the bolus from the oral cavity to the pharynx.

Pharyngeal Phase

The pharyngeal phase lasts less than 1 s; however, it is extremely complex and largely reflexive. Since the respiratory and digestive tracts cross in the pharynx, respiration is stopped during the pharyngeal swallow. Although the exact triggers of the pharyngeal swallow are not known, it usually starts as the bolus is being transported by the tongue to the pharynx, presumably by stimulation of oropharyngeal receptors of the superior laryngeal nerve. Studies have shown that although this may be true with direct stimulation of the various oropharyngeal structures, in the process of normal eating, food often enters the pharynx prior to initiation of a swallow, suggesting that either volitional control may override this stimulus or a summation of sensory input will inhibit this reflex. Alternatively, the pharyngeal phase during normal eating can be triggered as part of a seamless centrally initiated stereotyped sequence without the need for peripheral input. Figure 2 illustrates the pharyngeal phase of swallowing during evaluation with an oropharyngeal esophagram.

Swallows have been categorized as primary, when they are volitional, or secondary, when they are reflexive (i.e., without voluntary control). These secondary swallows have been referred to as pharyngeal swallows. In general, they have the same biomechanics and characteristics as primary swallows, except for the fact there is no sequential contact of the tongue with the hard palate. It is believed that these secondary swallows are important for clearing the hypopharynx of any residue or refluxate from the stomach. Pharyngeal swallows can be triggered by stimulation of many areas of the oropharynx, including the anterior faucial pillars, posterior tongue, epiglottis, posterior pharynx, and larynx. The threshold for mechanical stimulation increases with age and this may in part explain why older patients are at increased risk for oropharyngeal dysphagia and aspiration.

Once the tongue has moved the bolus to its base by lingual peristalsis, multiple well-coordinated events take place virtually simultaneously. The goal of these motor activities is to change the pharynx from a respiratory vessel to a digestive vessel, while preventing aspiration into the airway. Laryngeal closure appears to be the first event that occurs. The larynx is elevated and closed to prevent aspiration into the airway. Laryngeal closure is a stepwise process that results in first closure of the true vocal cords followed by the false cords. The epiglottis is then brought down to cover the glottic area. Studies have demonstrated that vocal cord adduction starts prior to any other deglutitive event including movement of the food from the mouth. There are four steps observed in laryngeal closure, starting with adduction of the vocal cords with horizontal approximation of the arytenoids, vertical approximation of the arytenoids to the base of the epiglottis, laryngeal ascent, and epiglottal descent. In order to prevent nasopharyngeal regurgitation, the nasopharynx is closed off by the velopharyngeal contraction and midline contraction of superior pharyngeal constrictors.

The upper esophageal sphincter is an area of high pressure between the pharynx and the esophagus. The main component of the sphincter is the cricopharyngeus muscle, with contribution of the pharyngeal constrictors and proximal esophagus. The cricopharyngeus receives innervation through the pharyngoesophageal, superior laryngeal, and recurrent laryngeal nerves and sensory input from the glossopharyngeus nerve. The upper esophageal sphincter relaxes for approximately 0.5 s with deglutition to allow bolus passage. In addition to intrinsic relaxation, the upper esophageal sphincter is opened by anterior hyoid/laryngeal traction by the suprahyoid muscles. These are the same muscles involved in laryngeal displacement and closure as described above. After the bolus is transferred to the posterior pharynx, pharyngeal peristalsis and posterior tongue thrust carry it across the upper esophageal sphincter and into the cervical esophagus. The propagated pharyngeal contraction moves from the superior, middle, and inferior pharyngeal constrictors, emptying the pharynx of any residue.

Esophageal Phase

Once the bolus has been transferred to the esophagus across the upper esophageal sphincter, esophageal peristalsis carries it down to the stomach within a span of 6–10 s. Passage of the bolus into the proximal esophagus is followed by contraction of the upper esophageal sphincter, which initiates a progressive wave of circular muscle contraction, or peristalsis, which propels the food bolus along the esophagus. Transport of liquid is aided by gravity and liquids tend to reach the stomach

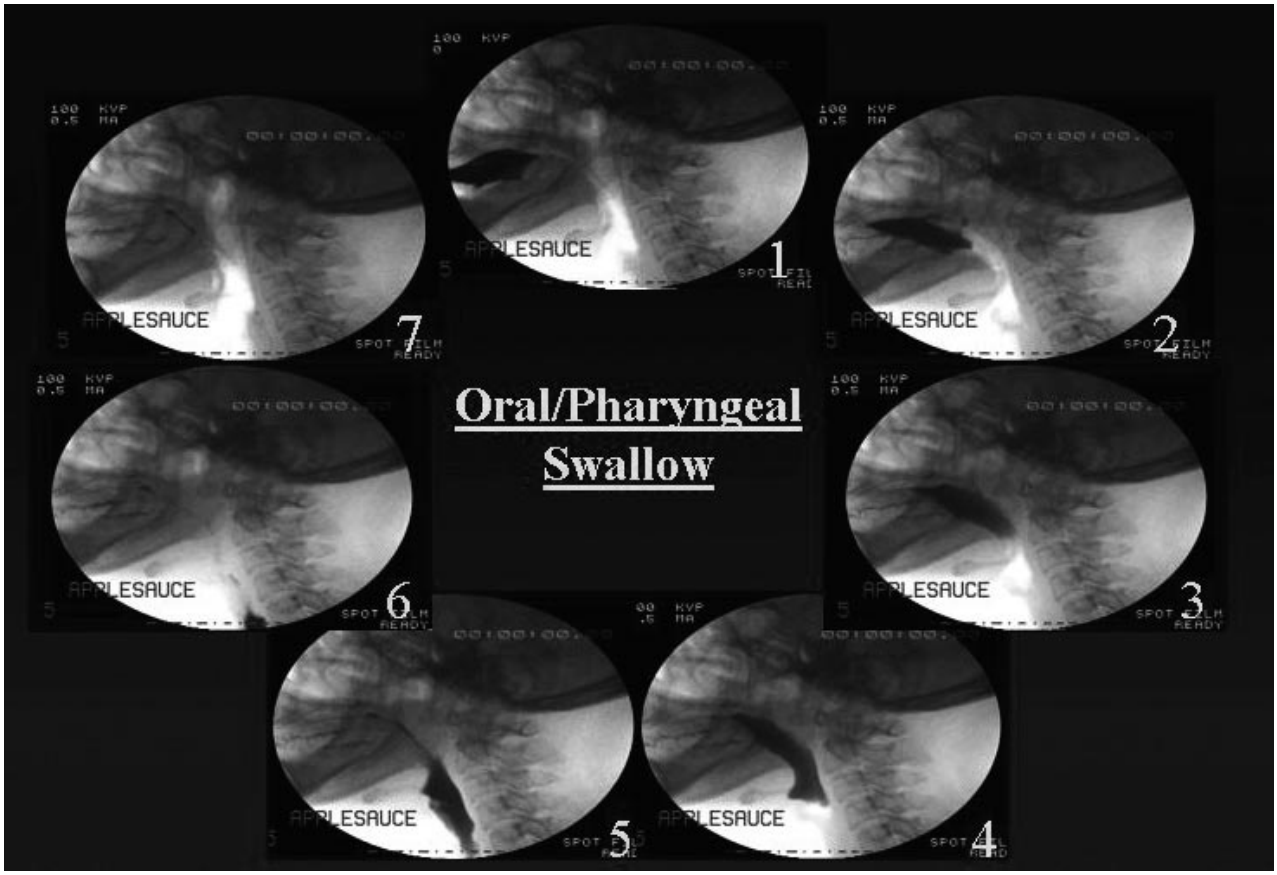


FIGURE 2 Pharyngeal phase of swallowing during an evaluation with an oropharyngeal esophagram.

prior to solids. Prior to the arrival of the bolus at the distal esophagus, the lower esophageal sphincter relaxes to allow passage of the bolus into the stomach.

SUMMARY

The act of deglutition is a repetitive, uniform action that belies the level of complexity of its control. Since the respiratory tract and the digestive tract converge in the pharynx, airway protective mechanisms are programmed into swallowing mechanisms. Due to the many facets of neuromuscular activity that are responsible for deglutition, a wide array of medical problems can result in deglutitive dysfunction. Newer imaging modalities are now able to provide much greater insight into the swallowing process, which in turn, it is hoped, will lead to more effective therapies for people who have swallowing disorders.

See Also the Following Articles

Dysphagia • Esophageal Strictures • Esophagus, Anatomy • Hiccups (Singultus) • Rumination Syndrome • Sphincters

Further Reading

Cook, I. J., Dodds, W. J., and Dantas, R. O. (1989). Opening mechanisms of the human upper esophageal sphincter. *Am. J. Physiol.* **257**, G748–G751.

Dodds, W. J. (1989). The physiology of swallowing. *Dysphagia* **3**, 171–178.

Dua, K. S., Ren, J., Bardan, E., Xie, P., and Shaker, R. (1997). Coordination of deglutitive function and pharyngeal bolus transit during normal eating. *Gastroenterology* **112**, 73–83.

Kahrilas, P. J. (1993). Pharyngeal structure and function. *Dysphagia* **8**, 303–307.

Kahrilas, P. J., Dodds, W. J., Dent, J., Logemann, J. A., and Shaker, R. (1988). Upper esophageal sphincter function during deglutition. *Gastroenterology* **95**, 52–62.

Kern, M., Birn, R., Jaradeh, S., Jesmanowicz, A., Cox, R., Hyde, J., and Shaker, R. (2001). Swallow related cerebral cortical activity maps are not specific to deglutition. *Am. J. Physiol. (Gastrointest. Liver Physiol.)* **280**, G531–G538.

Lang, I. M., and Shaker, R. (1997). Anatomy and physiology of the upper esophageal sphincter. *Am. J. Med.* **103**, 50S–55S.

Logemann, J. A., Kahrilas, P. J., Cheng, J., Pauloski, B. R., Gibbons, P. J., Rademaker, A. W., and Lin, S. (1992). Closure mechanism of the laryngeal vestibule during swallow. *Am. J. Physiol. (Gastrointest. Liver Physiol.* **25**) **262**, G338–G344.

Martin, R. E., and Sessle, B. J. (1993). The role of the cerebral cortex in swallowing. *Dysphagia* **8**, 195–202.

Shaker, R. (1992). Oropharyngeal dysphagia: Practical approach to diagnosis and management. *Semin. Gastrointest. Dis.* **3**, 115–128.
Shaker, R., Dodds, W. J., Hogan, W. J., and Arndorfer, R. C. (1990). Coordination of deglutitive glottic closure with oropharyngeal swallowing. *Gastroenterology* **98**, 1478–1484.
Shaker, R., Ren, J., Zamir, Z., Sarna, A., and Sui, Z. (1994). Effect of aging, position and temperature on the threshold volume triggering pharyngeal swallows. *Gastroenterology* **107**, 396–402.

Sympathetic Innervation

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

paravertebral ganglia Ganglia located alongside and parallel to either side of the vertebral column. Also known as sympathetic chain ganglia.
postganglionic sympathetic neurons Neurons of the sympathetic nervous system that have their cell bodies in peripheral ganglia and send axonal projections to the digestive tract.
preganglionic sympathetic neurons Neurons of the sympathetic nervous system that have their cell bodies in the spinal cord and send axonal projections to sympathetic ganglia in the periphery.
prevertebral ganglia The celiac, superior mesenteric, and inferior mesenteric ganglia of the sympathetic nervous system located in the abdomen.
secretomotor neurons The neurons in the enteric nervous system, which innervate the intestinal crypts of Lieberkühn to evoke the secretion of water, electrolytes, and mucus.

The sympathetic nervous system is one of the three divisions of the autonomic nervous system that innervate the digestive tract. The other two divisions are the parasympathetic and enteric nervous systems.

INTRODUCTION

Neuronal cell bodies of the central nervous system component of the sympathetic innervation are positioned in the intermediolateral horn in the thoracic and lumbar regions of the spinal cord (Fig. 1). Efferent sympathetic fibers leave the spinal cord in the ventral roots to make their first synaptic connections with neurons in prevertebral sympathetic ganglia located in the abdomen. These are termed preganglionic sympathetic neurons. The prevertebral ganglia, which are the targets of the preganglionic neurons, are the celiac, the superior mesenteric, and the inferior mesenteric ganglia. Cell

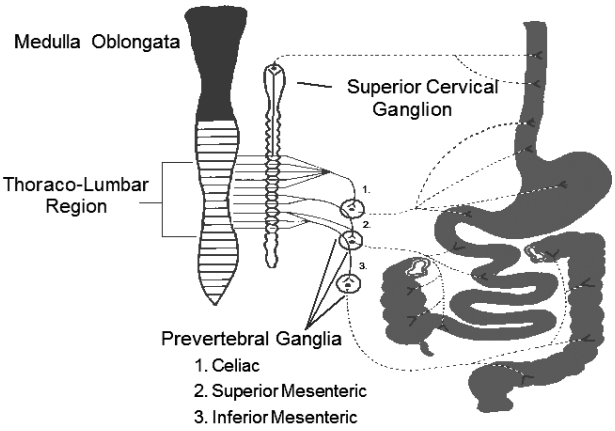


FIGURE 1 The sympathetic division of the autonomic nervous system innervates all levels of the digestive tract. Sympathetic pathways to the gut start with neurons in the thoracic and lumbar regions of the spinal cord. Efferent sympathetic fibers leave the spinal cord in the ventral roots to make their first synaptic connections with neurons in prevertebral sympathetic ganglia located in the abdomen. The prevertebral ganglia are the celiac, the superior mesenteric, and the inferior mesenteric ganglia. Cell bodies in the prevertebral ganglia project to the digestive tract, where they form synapses with neurons of the enteric nervous system in addition to innervating the blood vessels, mucosa, and specialized regions of the musculature. The upper esophagus receives innervation from the superior cervical ganglion, which is the uppermost of the paravertebral sympathetic ganglia.

Copyright 2004, Elsevier (USA). All rights reserved.

Shaker, R. (1992). Oropharyngeal dysphagia: Practical approach to diagnosis and management. *Semin. Gastrointest. Dis.* **3**, 115–128.
Shaker, R., Dodds, W. J., Hogan, W. J., and Arndorfer, R. C. (1990). Coordination of deglutitive glottic closure with oropharyngeal swallowing. *Gastroenterology* **98**, 1478–1484.
Shaker, R., Ren, J., Zamir, Z., Sarna, A., and Sui, Z. (1994). Effect of aging, position and temperature on the threshold volume triggering pharyngeal swallows. *Gastroenterology* **107**, 396–402.

Sympathetic Innervation

JACKIE D. WOOD
The Ohio State University College of Medicine and Public Health

paravertebral ganglia Ganglia located alongside and parallel to either side of the vertebral column. Also known as sympathetic chain ganglia.
postganglionic sympathetic neurons Neurons of the sympathetic nervous system that have their cell bodies in peripheral ganglia and send axonal projections to the digestive tract.
preganglionic sympathetic neurons Neurons of the sympathetic nervous system that have their cell bodies in the spinal cord and send axonal projections to sympathetic ganglia in the periphery.
prevertebral ganglia The celiac, superior mesenteric, and inferior mesenteric ganglia of the sympathetic nervous system located in the abdomen.
secretomotor neurons The neurons in the enteric nervous system, which innervate the intestinal crypts of Lieberkühn to evoke the secretion of water, electrolytes, and mucus.

The sympathetic nervous system is one of the three divisions of the autonomic nervous system that innervate the digestive tract. The other two divisions are the parasympathetic and enteric nervous systems.

INTRODUCTION

Neuronal cell bodies of the central nervous system component of the sympathetic innervation are positioned in the intermediolateral horn in the thoracic and lumbar regions of the spinal cord (Fig. 1). Efferent sympathetic fibers leave the spinal cord in the ventral roots to make their first synaptic connections with neurons in prevertebral sympathetic ganglia located in the abdomen. These are termed preganglionic sympathetic neurons. The prevertebral ganglia, which are the targets of the preganglionic neurons, are the celiac, the superior mesenteric, and the inferior mesenteric ganglia. Cell

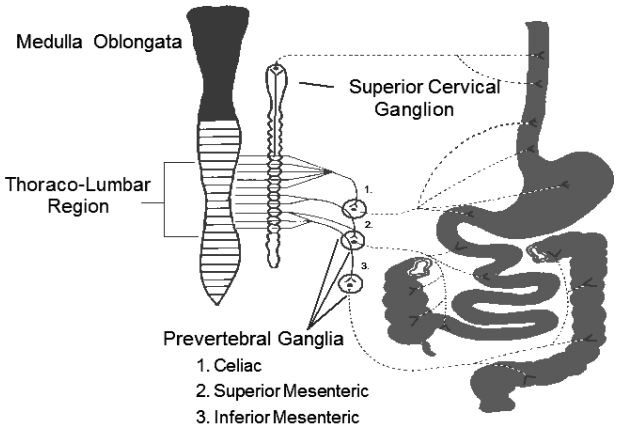


FIGURE 1 The sympathetic division of the autonomic nervous system innervates all levels of the digestive tract. Sympathetic pathways to the gut start with neurons in the thoracic and lumbar regions of the spinal cord. Efferent sympathetic fibers leave the spinal cord in the ventral roots to make their first synaptic connections with neurons in prevertebral sympathetic ganglia located in the abdomen. The prevertebral ganglia are the celiac, the superior mesenteric, and the inferior mesenteric ganglia. Cell bodies in the prevertebral ganglia project to the digestive tract, where they form synapses with neurons of the enteric nervous system in addition to innervating the blood vessels, mucosa, and specialized regions of the musculature. The upper esophagus receives innervation from the superior cervical ganglion, which is the uppermost of the paravertebral sympathetic ganglia.

Copyright 2004, Elsevier (USA). All rights reserved.

bodies in the prevertebral ganglia project their axons to the digestive tract, where they form synapses with neurons of the enteric nervous system in addition to innervating the blood vessels, mucosa, and sphincteric regions of the musculature. Neurons in the prevertebral ganglia are termed postganglionic sympathetic neurons. Postganglionic sympathetic neurons express nicotinic-type receptors for acetylcholine. Preganglionic neurons release acetylcholine as a neurotransmitter in the prevertebral ganglia.

Sympathetic input to the digestive tract generally functions to shunt blood from the splanchnic to the systemic circulation during exercise and stressful environmental encounters. This occurs coincident with suppression of digestive functions, including propulsive motility and secretion.

SYMPATHETIC NEUROTRANSMISSION

Norepinephrine released from sympathetic postganglionic neurons is the principal mediator of sympathetic actions in the gut. Norepinephrine acts directly on sphincteric muscles to increase tension and keep the sphincter closed and it also acts on the vasculature to decrease blood flow. The inhibitory action of norepinephrine at synapses in the neural control circuitry is primarily responsible for sympathetic inactivation of motility. Inhibitory action on secretomotor neurons suppresses secretion of electrolytes and water from the intestinal crypts of Lieberkühn.

The synaptic interface between the postganglionic fibers of the sympathetic nervous system and the enteric nervous system is at presynaptic α_2 adrenoceptors. Norepinephrine released from sympathetic fibers suppresses the release of excitatory neurotransmitters at both enteric synapses and neuro-effector junctions. Suppression of synaptic transmission by the sympathetic innervation occurs at most excitatory synapses in the enteric neural networks. This inactivates the neural circuits that generate intestinal motor behavior. Activation of the sympathetic inputs allows only continuous discharge of inhibitory motor neurons to the nonsphincteric muscles. The overall effect is to suspend intestinal motility in conjunction with reduced intestinal blood flow. This state is called physiologic ileus when it occurs transiently and pathologic ileus when it persists and produces symptoms of intestinal obstruction.

See Also the Following Articles

Autonomic Innervation • Enteric Nervous System • Parasympathetic Innervation

Further Reading

Appenzeller, O. (1982). "The Autonomic Nervous System." Elsevier, Amsterdam.

Gabella, G. (1976). "Structure of the Autonomic Nervous System." Wiley, New York.

Wood, J. (1994). Physiology of the enteric nervous system. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, D. Alpers, J. Christensen, E. Jacobson, and J. Walsh, eds.), 3rd Ed., Vol. 1, pp. 423–482. Raven Press, New York.

Wood, J. (1999). "Neurogastroenterology and Motility." Milner-Fenwick, Timonium, MD.

Taste and Smell

ANDREW I. SPIELMAN* AND FRITZ W. LISCHKA†

New York University College of Dentistry and †Monell Chemical Senses Center and University of Pennsylvania

chemosensory systems Biological systems that detect soluble and volatile chemicals.
gustatory Related to taste.
olfactory Related to smell.
pheromone Chemical that, when emitted by members of a species, will affect the behavior or physiology of other members of that species.
taste bud Cluster of 80–150 specialized epithelial cells; responsible for the initial events of taste reception.
umami Savory taste; basic taste quality elicited by monosodium glutamate.
vomeronasal organ Sensory organ specialized to detect pheromones in certain animals.

Chemosensory systems, of which taste and smell are specialized forms, detect both soluble and volatile chemicals. The senses of taste and olfaction can affect social behaviors, including feeding, territoriality, and mating. Taste and smell are also used in selection and evaluation of flavor and in avoidance of potentially harmful compounds. Through the cephalic phase of digestion, taste also affects certain exocrine and endocrine secretions, thus affecting nutrition and metabolism and the overall quality of life.

OVERVIEW OF TASTE AND SMELL

All animals respond to various chemicals in nature; not all chemicals, however, are detected exclusively by chemosensory taste and smell systems. Painful, irritating, and pungent chemicals, for example, are also detected by the trigeminal system, and chemicals associated with sexual and social signals (pheromones) are detected by the vomeronasal organ in certain animals. Although the vomeronasal organ is physically present in humans, its functionality is controversial (see further).

Receptors for taste and olfaction are located at the entry port of each governing system, i.e., the gastrointestinal tract for taste and the respiratory tract for olfaction. Unlike other sensory systems, the taste and olfaction sensory systems have specialized peripheral chemosensory receptors that interact with the soluble and volatile chemicals that are subsequently rejected, ingested, or inhaled. Intake of chemicals can be either beneficial or harmful, and taste and smell are important discriminatory screening mechanisms for avoiding potentially harmful chemicals.

PERIPHERAL ORGANIZATION OF THE TASTE SYSTEM

The peripheral gustatory system is exposed to a variety of physical, chemical, and biological insults. Extremely hot, cold, irritating, acidic, and nonsterile stimuli may have damaging effects on the peripheral taste receptor system. Therefore, the gustatory system evolved as a rapidly renewing specialized epithelial system. This is in contradistinction to most other sensory systems, including olfaction, in which stimuli are detected by sensory neurons.

Structures that are involved in peripheral taste reception, in decreasing order of size, are the taste papillae, taste buds, receptor cells, and taste receptor proteins. Taste papillae are visible with the unaided eye and are located throughout the oral cavity on the tongue, palate, pharynx, and epiglottis. There are four major types of gustatory papillae: circumvallate, foliate, fungiform, and taste stripes (from the original German *geschmaksstreifen*). However, the most abundant papillae on the tongue, the filiform papillae, are nongustatory (Fig. 1). These are prone to overgrowth, staining (especially with coffee and food dyes), or excessive shedding, which impart a white, coffee-brown, or raspberry appearance, respectively.

Of the gustatory papillae, the circumvallate papillae are located in a V-shaped array in the posterior third of the tongue. In humans, there are between 3 and 13 circumvallate papillae, and their number varies in other animals: rats and mice have only one, whereas cows may have as many as 25. Symmetrically located on the lateral posterior side of the tongue are the foliate papillae, which are pocket-shaped invaginations lined with taste buds. Distributed over a large surface area, the mushroom-shaped fungiform papillae cover the anterior dorsal surface of the tongue. The number of

 Copyright 2004, Elsevier (USA). All rights reserved.

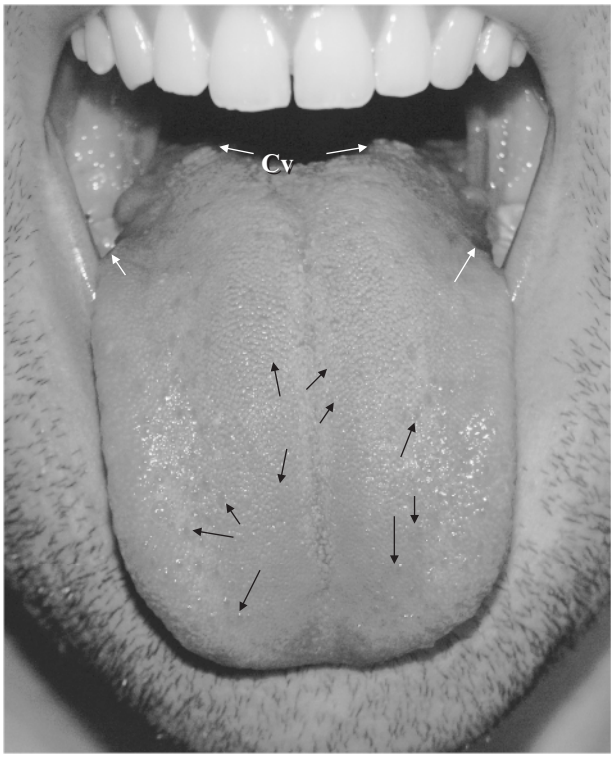


FIGURE 1 Human tongue. The dorsal surface of the tongue has four types of papillae: fungiform (Fu), nongustatory filiform (Fi), circumvallate (Cv), and foliate (Fo). Foliate papillae are not visible.

fungiform papillae in humans varies from 50 to 200. Finally, the taste stripes are located on both sides of the palatal midline at the borderline of the soft and hard palates.

The different taste papillae contain varying numbers of taste buds. For instance, in humans, the circumvallate papillae contain 100–200 taste buds, the foliate papillae have 320–2950 buds, and the fungiform papillae have 1–10 taste buds. The taste bud is the functional unit of the sense of taste. It is onion-shaped and contains 50–100 continuously maturing taste receptors and supporting taste cells (Fig. 2). Over 95% of the taste bud is shielded from the oral cavity by tight junctions, the structures that are responsible for the epithelial barrier. Only the apical portions of a few taste cells are exposed to the oral cavity through a 3- to 5-μm-wide opening, the taste pore (Fig. 3).

Unlike components of any other sensory system, taste cells have a rapid turnover rate of 10.5 days. The progenitor cells, the basal cells, are located at the base of the taste bud. As cells continuously grow and mature, they move from the basal area of the bud toward the taste pore. At any given time, the taste pore may contain the apical tips of 8–10 taste cells. The resident time of these 8–10 cells is as brief as a few hours, before they are shed into the oral cavity and washed away by saliva. This rapid turnover of cells, characteristic of many epithelial cells (e.g., certain cells lining the small intestine), means that the exposed taste receptor cells used for lunch are not the same as those used for dinner.

TASTE RECEPTORS, SIGNAL TRANSDUCTION, AND GUSTATORY PROCESSING

Generally, gustatory stimuli interact first with specific protein receptors on the apical, exposed surface of the taste receptor cells. This interaction then leads to changes in ion flux across the taste receptor cell membrane. The resulting depolarization induces release of neurotransmitter from the receptor cell to the nerve fiber innervating the cell. Changes in the firing rate of this innervating nerve are conveyed to specific regions of the central nervous system, where the taste message is decoded into a perceptual modality. During the past 5 years, a variety of taste receptor candidates—ion channels, ligand-gated channels, enzymes, and G-protein-coupled receptors (GPCRs)—have been identified for

FIGURE 2 Taste cells. (A) Transmission (phase contrast) photomicrograph of a single mouse taste cell. (B and C) Scanning electron micrographs of dissociated mouse taste cells. Bars = 10 μm. Some contaminating tissue is attached to the taste cell in B. Reproduced from Spielman *et al.* (1989) with permission from Elsevier Science.

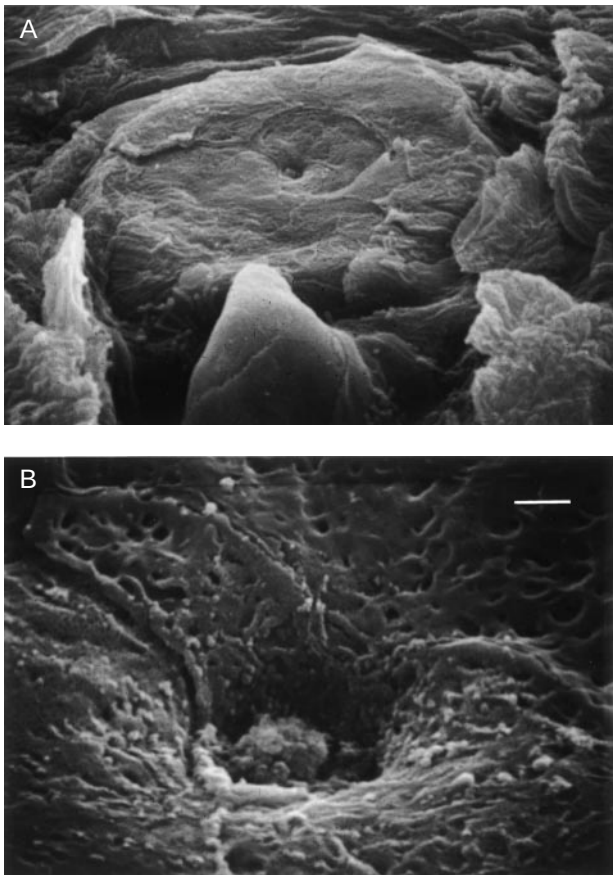


FIGURE 3 Taste pore. (A) Scanning electron micrograph of a rat fungiform papilla. The central pit represents the taste pore. (B) High-power scanning electron micrograph of the taste pore shown in A. Bar = 1 μm. Reproduced from Spielman and Brand (1997) with permission from Elsevier Science.

the five basic taste qualities: sweet, bitter, sour, salty, and savory (or umami, the amino acid taste of monosodium glutamate).

Sweet Taste

Sweet taste in humans is elicited by a variety of compounds, including sugars and sugar derivatives, D-amino acids, some of the small L-amino acids (glycine and L-alanine), and artificial sweeteners (such as cyclamate, saccharine, aspartame, sucralose, and very high-potency sweeteners). Recent evidence suggests that GPCRs are responsible for detecting sugars. A candidate sweet receptor, the T1R3, was cloned and found to be functional only as a heterodimer with a previously cloned receptor, the T1R2. Both are expressed in about 20% of the taste cells located in the posterior, lateral, and anterior taste buds of the tongue. The mechanism by which sweet taste is transduced has been previously elucidated. T1R3/T1R2 receptors couple through an increase in cGMP and activation of a cyclic-nucleotide-gated channel that leads to depolarization through calcium influx. Interestingly, artificial sweeteners use a different pathway, similar to many bitter compounds.

Bitter Taste

Detection of potentially harmful, even toxic, compounds is one of the primary roles of bitter taste. This gustatory stimulus is represented by a large and diverse array of compounds, ranging from ions (potassium) to complex artificial (denatonium) or naturally occurring compounds (caffeine, strychnine, quinine). Because of its potential survival benefit, bitter taste has the lowest detection threshold of all taste qualities, the largest known set of taste receptors, and is assumed to have the most diverse set of mechanisms of signal transduction.

Similar to sweet taste, bitter taste transduction involves primarily GPCRs as cell surface binding sites for many bitter stimulants. A family of 40–80 membrane-associated bitter taste receptors, termed T2Rs, was recently identified in rodents and humans. In humans, T2Rs are encoded in 24 genes located on three chromosomes. It is assumed, although not yet tested, that all T2Rs are bitter responsive. Bitter taste receptors apparently are coupled through gustducin, a taste tissue-enriched G protein α subunit, and associated β3 and γ13 subunits to the cyclic nucleotide and the phosphoinositide signal transduction pathways. Gustducin activates one or more phosphodiesterases, reducing the levels of cyclic nucleotides (cAMP and cGMP), leading to opening of a cyclic nucleotide-gated cation channel and depolarization. The β3/γ13 also activates phospholipase C-β2, which releases two second messengers, inositol trisphosphate (IP_3) and diacylglycerol (DAG). The former releases intracellular calcium, leading to cell depolarization. The specifics of the interplay between these two second messengers, the reduction of cyclic nucleotides, and the increase of IP_3 and DAG are not known; nor is it obvious which is the leading event in depolarization.

Sour Taste

Sour taste quality, similar to bitter taste, is a protective/warning system. It indicates the protons of acids. Protons may have a local effect on oral soft and hard tissues or a systemic effect when acidity indicates spoiled food. Several protein candidates have been implicated in sour taste transduction, including amiloride-sensitive epithelial sodium channels

(ENaCs), proton-gated channels [mammalian degenerin 1 (MDEG1), K^+ channels], hyperpolarization-activated cyclic nucleotide-gated channels (HCNs), H^+-gated ion channels, and the acid-sensing ion channels (ASICs). A variety of mechanisms may be associated with these channels, indicating the potential complexity of this taste quality. Generally, all potential mechanisms lead to an increase in intracellular positive charge that results in direct depolarization. Some of these mechanisms are supported by behavioral studies using the channel blocker, amiloride, which was shown to reduce aversion to acids in some species (the hamster, for example). The specific signal transduction mechanisms for most of these receptors remain to be elucidated.

Humans have a characteristic strong facial grimace, a "sour face," when exposed to sour stimuli. The grimace induces a strong contraction of facial muscles, which channels saliva onto the surface of the tongue. The mechanisms of salivation and tasting are tightly linked, and sour taste is the strongest salivary stimulant. With increasing salivary flow rates, higher levels of bicarbonates are secreted, which leads to buffering of the acid protons, protecting oral tissue from damage.

Salty Taste

Similar to sour taste, salty taste represents ions. Unlike sourness, however, saltiness is an essential indicator of minerals and serves as a monitor for ion homeostasis. The most important representative of this taste stimulant is sodium chloride. In rodents, an amiloride-sensitive epithelial sodium channel detects sodium chloride; the chloride appears to be mediated by a paracellular mechanism. In humans, the ENaC is less prominent and additional mechanisms not yet identified may be involved.

Umami (Amino Acid) Taste

Umami, from the Japanese word for "delicious" (*umai*), describes a taste quality specific for monosodium glutamate (MSG). This taste is synergistically enhanced in the presence of 5′ ribonucleotides, especially inosine 5′-monophosphate (IMP) and guanosine 5′-monophosphate (GMP). MSG and glutamate, the excitatory neurotransmitter, are almost identical. It was, therefore, reasonable to expect that their receptors might be related. Indeed, a truncated form of the brain glutamate receptor, mGluR4, was found in the taste system and is one of a number of candidate receptors for umami. The signal transduction mechanism for umami using the mGluR4 receptor is assumed to be similar to that in brain, a reduction in the level of cAMP leading to a closure of an unspecified cation conductance. One problem with this mechanism is that the mGluR4 receptors are generally inhibitory. Because tasting MSG likely requires an excitatory response from the taste cell, the actual role that an inhibitory receptor such as mGluR4 plays in transduction of the taste of MSG is questionable.

A completely different receptor type for MSG has been recently proposed, one that has little homology with other known glutamate receptors. This receptor is a dimer of two of the receptor proteins of the T1R family, namely, T1R1/T1R3. Note that this dimer is similar to the proposed sweet receptor, except that one monomer of the dimer pair is different. For sweet taste, the dimer is T1R2/T1R3. One interesting feature of the T1R1/T1R3 receptor for MSG is that its activity toward glutamate is enhanced by the ribonucleotides. This synergism between glutamate and the ribonucleotides is a hallmark of umami taste, and the observation that the T1R1/T1R3 dimer is enhanced by IMP lends credence to the suggestion that T1R1/T1R3 is the major receptor for umami taste. Japanese cuisine has taken advantage of appropriate combinations of foods to maximize this synergistic effect. The combination of pork, chicken, black mushrooms, sea bream, etc., which contain nucleotides, and tomatoes, cauliflower, celery, carrots, and mushrooms, which are rich in MSG, lead to an enhanced taste for glutamate via this synergistic culinary effect.

Other Tastes

In aquatic animals, other amino acids act as taste stimulants. The catfish, for instance, shows sensitivity to L-arginine, L-alanine, and glycine. The L-arginine receptor is a ligand-gated nonselective cation channel and is primarily located on the barbel, a tactile process located on the lip of the catfish.

Additional taste qualities exist. Fats have been recently found to act on taste cells, in addition to stimulating the trigeminal system. In particular, some free fatty acids activate taste cells through a potassium channel blockage. Water and metallic tastes have also been proposed as distinct, although nontraditional, taste qualities.

Signal Transduction and Processing

Taste receptor cells, similar to neurons, exhibit action potentials in response to gustatory signal transduction, leading to release of neurotransmitters. The receptor cells synapse with first-order neurons at the taste bud level. Gustatory information is carried for central processing by three cranial nerves: the VIIth,

or facial (of which the chorda tympani and greater superficial petrosal branches innervate the anterior two-thirds of the tongue and palate), the IXth, or glossopharyngeal (innervating the foliate and circumvallate papillae), and the Xth, or vagus (innervating the base of the tongue, epiglottis, and pharynx). Pain and thermal and tactile information, crucial for food detection and appreciation, are carried by the maxillary and mandibular branches of the Vth or trigeminal nerve.

Although it has been assumed for many years that specific regions of the tongue are tuned for specific taste qualities, it is now clear that all three gustatory nerves carry all taste stimuli. Even single nerve fibers may be broadly tuned to carry information about multiple types of gustatory stimuli. Gustatory information carried by the three cranial nerves is passed on to the nucleus of the solitary tract in the medulla oblongata. From there, information is sent to the ventral posteromedial thalamus and eventually to the gustatory cortex in the lower tip of the parietal lobe.

PERIPHERAL ORGANIZATION OF THE OLFACTORY SYSTEM

The sense of smell, although generally not considered as important as some of the other senses, allows human beings to detect thousands of odors in their environment. The nasal cavity is divided by the septum, and humans have three folds, or turbinates, in the dorsal part on each side. Sensory neurons are located predominantly on the superior turbinate and to a lesser extent on the middle turbinate and the septum, whereas nonsensory epithelium lines the other areas. The sensory portion of the olfactory mucosa contains several cell types. Olfactory receptor neurons (ORNs) are the cells that detect chemical stimuli. They are bipolar neurons with a dendritic process ending in an apical swelling called an olfactory knob, which is exposed to the outside world, and an axon that projects through the cribriform plate into the olfactory bulb. The olfactory knob carries either cilia or microvilli, which contain the receptor molecules that detect odors and the elements of signal transduction pathways that convert the binding of odor molecules into electrical signals. Other cells in the sensory area of the olfactory mucosa are sustentacular cells (also called supporting cells), which surround sensory neurons and produce part of the mucus that covers the epithelium and basal cells; the basal cells are immature precursor cells for ORNs and give the olfactory neuroepithelium the ability to regenerate after injury. There also is a constant turnover of ORNs, with new cells arising from dividing basal cells, then maturing into functional receptor neurons, and finally dying and being resorbed. The life span of ORNs varies but is in the range of 30 to 90 days. The nonsensory areas of the mucosa contain respiratory cells that possess motile cilia at their exposed apical ends. Ciliary movement generates a continuous retrograde flow of mucus toward the throat. Bowman glands, scattered throughout the epithelium, produce mucus. The mucus contains odorant-binding proteins, which are lipocalins, having relatively low molecular weight and high affinity for odorant molecules. Although several potential roles for odorant-binding proteins have been proposed, including transport of odorant to and/or from receptors, facilitation of odor binding to receptors, termination of receptor binding of odorant, and detoxification of the mucosa, none of these functions has been clearly established.

OLFACTORY RECEPTORS AND SIGNAL TRANSDUCTION

To detect and distinguish thousands of different odorant molecules, the nose needs a large number of different receptors. In the early 1990s, a large family of genes that encoded apparent receptor molecules was identified in olfactory tissue. These receptors, similar to sweet and most bitter receptors, possess seven transmembrane domains that contain conserved sequences and a highly variable exposed region that is believed to be the ligand-binding site. More recently, some of these proteins have been confirmed to be odorant receptors by functional expression in cell systems such as *Xenopus* oocytes and immortalized cell lines. About 1000 different members of this family of receptors have been predicted for the mouse, one of the first and best studied mammalian models of olfaction. For humans, 350 functional olfactory receptor genes have been identified and cloned. It appears that humans have a large number of pseudogenes, which have high similarities with the nucleotide sequences encoding functional receptors but contain mutations that lead to nonfunctional proteins.

The most extensively documented signal transduction mechanism in ORNs involves the second messenger molecule cAMP. Binding of an odorant to a receptor molecule leads to the activation of a GTP-binding protein (G protein); an ORN-specific G protein, G_{olf}, has been identified. Activation of G proteins causes dissociation of the α subunit from the $\beta\gamma$ subunit complex. The α subunit activates the enzyme adenylyl cyclase, which converts ATP into the second messenger cAMP. cAMP in turn activates nonselective cation channels, resulting in depolarization of the receptor

neuron and leading to generation of action potentials. Calcium ions constitute a major component of the depolarizing current through the cAMP-activated channels, and the transient rise in intracellular calcium activates a second conductance through chloride channels, which amplifies the depolarization of the cell because internal chloride is unusually high in ORNs.

Several additional messenger molecules have been implicated in signal transduction in ORNs, either as part of different transduction pathways or as modulators of the cAMP cascade. These substances include cGMP, 1,4,5-inositol trisphosphate, nitric oxide (NO), and carbon monoxide (CO), although their exact contributions to signal transduction remains controversial.

OLFACTORY BULB AND HIGHER CENTERS

Depolarization of ORNs by odors leads to the generation of action potentials, which travel along the axons that form the olfactory nerve projecting to the olfactory bulb. The axon terminals form synapses with mitral and tufted cells in the outer layer of the olfactory bulb, in discrete structures called glomeruli (Fig. 4). Mitral and tufted cells are output neurons, sending their axons to the next center in the brain. Several ORNs project onto one output neuron (convergence), and it has been shown that ORNs express only one or very few olfactory receptor molecules and that all ORNs that have the same receptors project to the same one or two glomeruli in the bulb. There are local interneurons, called periglomerular cells, that make synapses with the output neurons within a glomerulus. These dendrodendritic synapses are reciprocal, with synaptic neurotransmitter release sites on both sides of the synaptic cleft. Release of neurotransmitter from the output neuron side has an excitatory effect on the periglomerular cell, whereas release of neurotransmitter from the periglomerular cell is inhibitory to the output neurons. Activation of these synapses can result in lateral inhibition that is similar to that observed in the retina. In deeper layers of the bulb, a second population of local interneurons, the granule cells, forms the same type of dendrodendritic synapses with secondary dendrites of output neurons. The interneurons are believed to play an important role in the processing of olfactory information.

The signals are finally sent via the axons of the output neurons, which form the olfactory tracts, to higher centers of the brain, including parts of the limbic system and the orbitofrontal cortex, where the integration of olfactory information with inputs from other sensory systems, including taste, takes place.

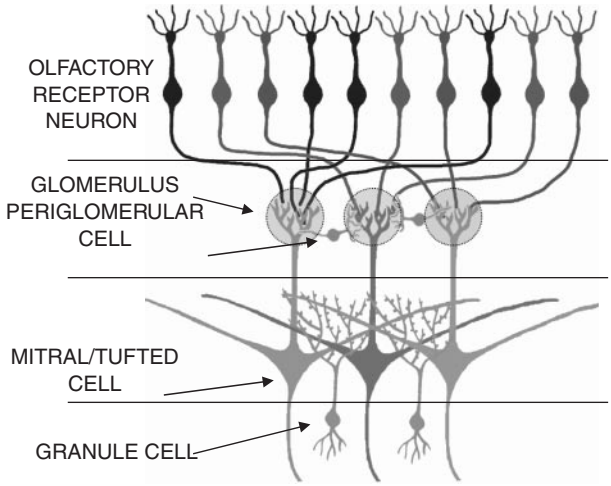


FIGURE 4 Organization of the olfactory bulb. Olfactory receptor neurons form synapses with dendrites of mitral and tufted cells in glomeruli located in the outer layer of the olfactory bulb. Local interneurons, called periglomerular cells, make dendrodendritic synapses with mitral/tufted cells in adjacent glomeruli. Another set of interneurons, granule cells, form dendrodendritic synapses with mitral/tufted cells in the deeper layer of the bulb. The color coding illustrates the fact that receptor neurons expressing the same receptor molecules project to the same glomeruli in the olfactory bulb. Courtesy of Graeme Lowe, Monell Chemical Senses Center, Philadelphia.

THE VOMERONASAL ORGAN

The vomeronasal organ (VNO) is an important sensory system in many vertebrates, particularly in mammalian species (rodents, cats, and horses, among others). It is often described as the sensory system that detects pheromones, or, in a more general sense, "social odors," a notion that has been supported for several animals and stimuli. However, it is important to note that in some cases pheromone function can be exerted through the main olfactory organ and that some volatile stimuli that are not considered behaviorally relevant can be detected by sensory neurons of the VNO.

The presence of a functional VNO in humans and the existence of human pheromones are among the most intensely debated questions in the field of olfaction. Anatomical studies have described the presence of a VNO in almost all subjects, although the organ varies greatly in size and shape among individuals and even within the nostrils in an individual. Some groups have found bipolar receptor-like cells in the area of the human VNO and electrical activity has been recorded on stimulation with derivatives of human hormones. However, no evidence has been found for an axonal connection with the olfactory bulb or other parts of

the brain. Genes that show significant homology with receptor genes isolated from rodent VNOs have been identified and found to be expressed in the olfactory mucosa of humans. Although many are pseudogenes, some appear to code for a functional receptor molecule. These receptors have not yet been isolated and their ligands are unknown.

The literature contains reports of physiological effects (synchronization of menstrual cycle) as well as behavioral effects of human scents that suggest the existence of pheromone-like substances in humans, most likely associated with apocrine secretions of the skin. With a VNO that appears nonfunctional because of its lack of neuronal connections, such stimuli might be detected with the olfactory mucosa.

INTERACTION OF TASTE, SMELL, AND OTHER SENSORY SYSTEMS

Several sensory systems must be activated to enjoy food. Gustatory, olfactory, and somatosensory (temperature, touch, and pain) systems are activated by chemical ingredients in food, dependent in part on the quality of the food. Each sensory system contributes to provide part of the overall sensation called flavor.

The absence of only one sensory system may significantly affect the pleasure of eating. For instance, individuals with an upper respiratory infection may experience a decrease in their sense of smell, and therefore a reduction in the appreciation of flavor. This can often be confused with a loss of taste. Indeed, 9 out of 10 patients complaining of loss of taste turn out to have a smell disorder. The relative ease with which the olfactory system can become compromised can be traced back to the anatomy of the olfactory system. With inhalation, the olfactory neurons, which have a slow rate of renewal, are directly exposed to potentially toxic agents. In addition, the olfactory nerve is prone to physical damage. This combination, along with frequent obstructions of the upper respiratory pathway, is a major reason why olfactory factors are the main cause of chemosensory disorders.

Texture, temperature, and carbonation (which induces mild pain) of food also affect taste and smell. The fat content of potato chips and cream cheese, the temperature of ice cream, the carbonation and temperature of beer, and the spicy nature of certain foods all contribute to the overall enjoyment of eating. Most foods are appreciated slightly below body temperature (36°C). This maximizes the emission of volatile compounds that are sensed by the olfactory system after swallowing. Other foods, such as ice cream and beer, are best at lower temperatures (closer to 4°C), whereas tea and coffee are most appreciated slightly above body temperature.

SUMMARY

A variety of chemicals act on taste and smell. The primary binding sites for these stimuli are a large number of recently identified cell surface receptors that are coupled with diverse signal transduction mechanisms. Taste and smell interact with other sensory systems, in particular those for perceiving temperature, texture, and pain. Together, they provide an overall assessment of the chemosensory and somatosensory properties of food.

See Also the Following Articles

Digestion, Overview • Salivary Glands, Physiology

Further Reading

Adler, E., Hoon, M. A., Mueller, K. L., *et al.* (2000). A novel family of mammalian taste receptors. *Cell* **100**, 693–702.

Li, X., Straszewski, L., Xu, H., Durick, K., Zoller, M., and Adler, E. (2002). Human receptors for sweet and umami taste. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4692–4696.

Lindemann, B. (2001). Receptors and transduction in taste. *Nature* **413**, 219–225.

Meredith, M. (2001). Human vomeronasal organ function: A critical review of best and worst cases. *Chem. Senses* **26**, 433–445.

Mombaerts, P. (1999). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**, 707–711.

Mori, K., Nagao, H., and Yoshihara, Y. (1999). The olfactory bulb: Coding and processing of odor molecule information. *Science* **286**, 711–715.

Schild, D., and Restrepo, D. (1998). Transduction mechanisms in vertebrate olfactory receptor cells. *Physiol. Rev.* **78**, 429–466.

Spielman, A. I. (1999). Gustatory function and dysfunction. *Crit. Rev. Oral Biol. Med.* **9**, 267–291.

Spielman, A. I., and Brand, J. G. (1997). Tongue and taste. *In* "Encyclopedia of Human Biology" (R. Dulbecco, ed.), 2nd Ed., Vol. 8, pp. 455–466. Academic Press, San Diego.

Spielman, A. I., *et al.* (1989). A method for isolating and patch-clamping individual mammalian taste cells. *Brain Res.* **503**, 326–329.

Th1, Th2 Responses

GERI BROWN
University of Texas Southwestern Medical Center, Dallas

antigen-presenting cells Mononuclear phagocytes, B cells, and dendritic cells. All can present antigen to major histocompatibility complex class II restricted T helper cells; which cell type presents antigen depends on where the antigen first encounters cells of the immune system.

cytokines Proteins produced by T cells and other immune cells; transmit signals that are important for communication among the cells of the immune system and between the cells of the immune system and other cells.

major histocompatibility complex Proteins encoded by the genetic loci involved in rejection of foreign or nonself tissues.

signal transducers and activators of transcription Proteins (Stat 1, Stat 4, and Stat 6) involved in Th1 and Th2 differentiation.

T helper 1 (Th1) cells Produce proinflammatory cytokines such as interferon γ and interleukin-2, which are important in macrophage activation as well as inflammatory and autoimmune reactions.

T helper 2 (Th2) cells Produce interleukins (IL-4, IL-5, IL-9, IL-10, and IL-13), which are involved in controlling humoral and allergic immune responses.

Studies on T helper 1 and 2 (Th1 and Th2) cells have elucidated the processes involved in regulation and development of Th1 and Th2 cells, the cytokine production of Th1 and Th2 cells, and the effects of T cell responses in the gastrointestinal tract. Th1 and Th2 cells are characterized by the cytokines that they produce. Th1 cells are predominantly involved in cell-mediated responses and Th2 cells are predominantly involved in humoral responses.

INTRODUCTION

There are several subsets of T lymphocytes that are defined by their cell surface receptors. Cluster of differentiation-3 (CD3) T cell receptors, for example, function in close association with either cluster of differentiation-4 or -8 (CD4 or CD8) surface coreceptors. In addition, these CD4+ and CD8+ T cells have also been further characterized as belonging to T helper (Th1 and Th2) subsets. The development of Th1 and Th2 T cells is regulated by a number of different signaling pathways, including the interleukin-12 (IL-12) receptor signaling pathway for Th1 cells and the IL-4 receptor signaling pathway for Th2 cells. Th1 and Th2 cell subsets can largely be defined by production of subset-specific cytokines. The hallmark cytokine of Th1 cells is interferon γ (IFNγ), but Th1 cells also produce IL-2, tumor necrosis factor α (TNFα), and lymphotoxin α and β. The signature cytokine of Th2 cells is IL-4, but Th2 cells also produce IL-5, IL-9, IL-10, and IL-13. Each of these cytokine responses is associated with responses within target organs. Within the gastrointestinal tract, the Th1 response is predominantly proinflammatory. For example, Crohn's disease and intestinal graft-versus-host disease are associated with increased production of Th1 cytokines, IFNγ and TNFα. Th2 responses are important in parasitic infections in the intestine and in allergic responses, and may be antiinflammatory or selectively inflammatory in inflammatory bowel disease. For example, the cytokine profile in ulcerative colitis is characterized by increased production of the Th2 cytokine, IL-5.

REGULATION OF TH1 AND TH2 RESPONSES

Th1 polarization is initially signaled by the T cell receptor (TCR) CD3, after its interaction with the antigen/major histocompatibility complex (MHC) on antigen-presenting cells (APCs). The second signal is produced by a number of costimulatory molecules, typified by CD28/B7. Importantly, in addition to these initiating signals, the two critical cytokines that control Th1 and Th2 differentiation are IL-12 and IL-4, respectively, though other cytokines have been reported to play a role. These two cytokines enhance the generation of their own Th subset and simultaneously inhibit the generation of the opposing subset. Th1 cells produce IFNγ and TNF, which activate macrophages but inhibit Th2 proliferation, as illustrated in Fig. 1. Th2 cells produce IL-4 and IL-5, which induce B cell activation, and IL-10, which inhibits production of IFNγ and TNF by Th1 cells.

 Copyright 2004, Elsevier (USA). All rights reserved.

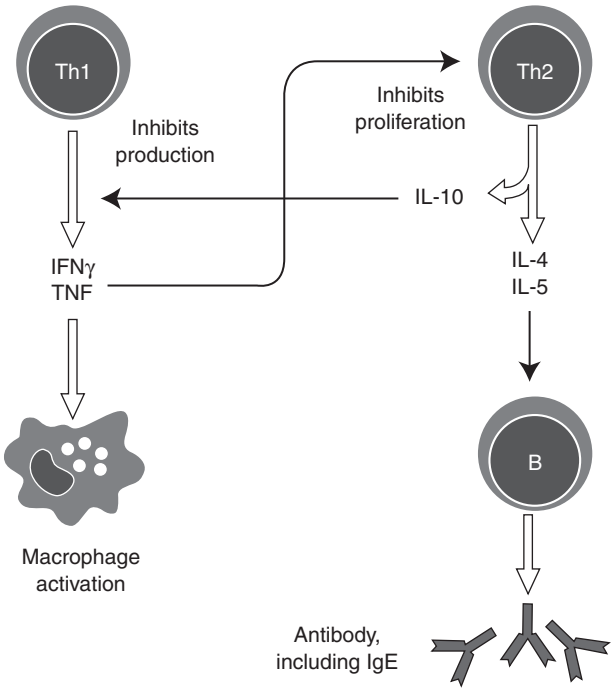


FIGURE 1 Th1 and Th2 cells produce cytokines that enhance the generation of their own Th subset and simultaneously inhibit the generation of the opposing subset. Adapted from Roitt, *et al.* (1998). "Immunology." 5th Ed, p. 125.

The critical cytokine initiating Th1 differentiation, IL-12, a heterodimeric molecule consisting of p35 and p40 subunits, is secreted predominantly by antigen-presenting cells such as dendritic cells and activated macrophages. IL-12 signaling through the IL-12 receptor (IL-12R) on the T cell induces signal transducer and activator of protein 4 (Stat 4) activation and translocation to the nucleus. Stat 4 activation induces high levels of IL-2 and IFNγ production. Importantly, not all IFNγ production by T cells is Stat 4 dependent, because Stat 6/Stat 4 double-deficient T cells produce some IFNγ, and CD8+ T cells are largely Stat 4 independent for TCR-induced IFNγ production (Fig. 2).

Another intracellular transcription factor has been recently found to be important in Th1 polarization in CD4+ T cells but not in CD8+ T cells. This factor, named T-box expressed in T cells (T-bet), was isolated using an IL-2 promoter-reporter and a cDNA library from activated Th1 cells. T-bet is a novel member of the T-box family of transcription factors. The target for T-bet is homeoprotein H1X and it seems to contribute to the capacity of T-bet to induce IFNγ production. Mechanistic studies show that overexpression of T-bet in T cells is sufficient to induce IFNγ production by direct transactivation of the IFNγ gene promoter in an IL-12-independent fashion. T-bet also up-regulates IL-12 receptor β2 chain expression on T cells (Fig. 2).

IL-1 also appears to be important in Th1 development. Differential actions of IL-1 on T cell subsets have long been recognized, and IL-1R-deficient mice display enhanced Th2 responses. IL-18, an IL-1-related factor, has been recently shown to be a selective activator of IFNγ in Th1, but not Th2, cells. Both IL-1 and IL-18 activate IL-1 receptor-associated kinase (IRAK) in Th1 cells. IRAK-deficient mice have defective IL-18-mediated Th1 type responses *in vivo*.

Importantly, IL-12 and IL-18 signaling processes appear to use distinct pathways. For example, activation of IFNγ by IL-18 has been recently found to involve activation of nuclear factor κB (NF-κB), which is the collective name for a group of transcription factors that contribute to the control of expression of many of the genes that participate in inflammation and immune responses, acting at discrete cis-acting elements in the IFNγ regulatory region. In most cells, NF-κB factors

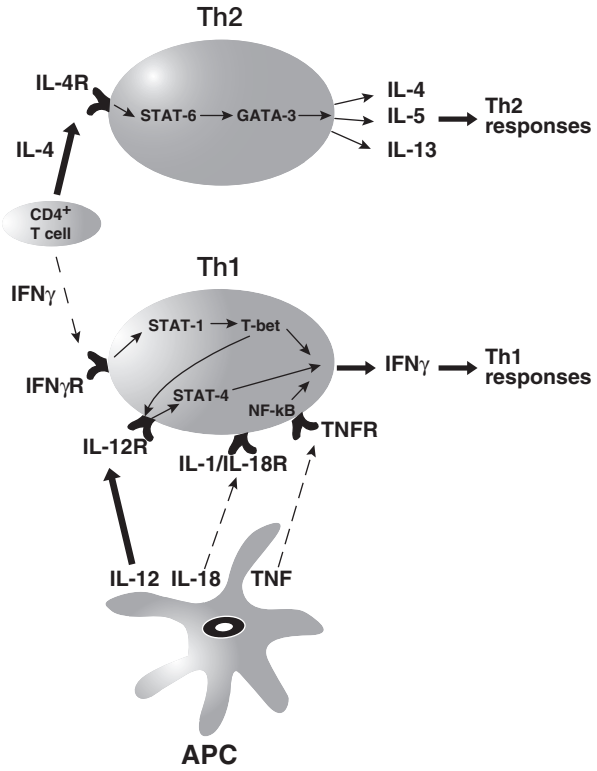


FIGURE 2 An Antigen-presenting cells (APC) and $CD4^+$ T cells secrete factors that stimulate T helper 1 and 2 (Th1, Th2) cells. IL, Interleukin; R, receptor; TNF, tumor necrosis factor; IFN, interferon; Stat, signal transducer and activator; T-bet, T-box transcription factor expressed in T cells. Adapted from Weigmann and Neurath (2002). "T-bet and mucosal Th1 responses in the gastrointestinal tract." **Gut** 51, 301–303.

normally occur in a latent form imposed by their association with inhibitory κB (I-κB) proteins, which dictate the cytoplasmic location of the proteins. Phosphorylation targets the inhibitor I-κB for proteasomal degradation, inducing NF-κB activation. TNF/TNF receptor (TNFR) interactions also involve the activation of NF-κB (Fig. 2).

IL-4, the critical cytokine inducing Th2 differentiation, signals through the IL-4 receptor (IL-4R) on the T cell and induces Stat 6 activation and translocation to the nucleus. Stat 6 activation induces high levels of IL-4 and IL-5. In addition, other transcription factors have been reported to affect Th2 cytokine production (Fig. 2).

CYTOKINES INVOLVED IN TH1 AND TH2 RESPONSES

Cytokines are proteins that transmit signals that are important for communication among and between the cells of the immune system and other cells. Each cytokine has many activities and can act on several different cell types. Some cytokines activate and regulate inflammatory cells, whereas others regulate the growth, differentiation, and activation of lymphocytes and other cell types.The cytokines TNF, IL-1, IL-18, and IL-6 are regarded as proinflammatory. Th1 cytokines (IFNγ and IL-2) are important in the activation of inflammatory cells. IFNγ is well known for antiviral properties and is produced by CD4+ and CD8+ T cells. It activates mononuclear phagocytes, increases the expression of human leukocyte antigen (HLA) class I and II molecules, promotes T and B cell differentiation, activates natural killer (NK) cells and vascular endothelial cells, and enhances the respiratory burst in neutrophils. Other cytokines produced by Th1 cells are TNFα, which causes activation of macrophages, granulocytes, and cytotoxic T lymphocytes (CTLs), influences endothelial cell adhesion, and stimulates MHC class I production. Lymphotoxin α (LTα), in combination with LTβ, is important in maintenance and development of the intestinal lymphoid system.

The Th2 cytokine IL-4 is a growth and differentiation factor for B cells but also has activity as a growth factor for CD4+ T cells and mast cells. IL-4 is known to act as an immunoglobulin E (IgE) switch factor in that it selectively stimulates B cells to switch to the production of the IgE isotype. In addition, it simulates CD23 expression on mononuclear phagocytes and B cells. IL-5 is produced by CD4+ Th2 lymphocytes and influences the inflammatory response through its ability to stimulate the growth, differentiation, and degranulation of eosinophils and to act as an eosinophil chemotactic factor. IL-5 stimulates the growth and differentiation of B cells and, in the mucosal immune system, increases IgA secretion by IgA-committed B cells. IL-10 is reported to inhibit Th1 cytokine synthesis, especially in the mucosal immune system. IL-10 has been recently reported to have some immunostimulatory activities as well. IL-9 enhances T cell survival, mast cell activation, and synergy with erythropoietin. IL-13 (similar to IL-4) is produced by activated T cells and affects B cell growth and differentiation and inhibits proinflammatory cytokine production.

GASTROINTESTINAL TH1 AND TH2 RESPONSES

Evidence that immune mechanisms are important in the initiation and the progression of ongoing tissue injury in inflammatory bowel diseases, intestinal graft-versus-host disease, food allergies, and celiac disease derives from the histologic and clinical features of these diseases as well as from laboratory studies.

Allergic Responses and Responses to Parasitic Infection

Th2 cells and their responses are involved in many food allergies, because the Th2 IL- 4 and IL-5 cytokine responses mediate immunoglobulin switching and eosinophilic activation. A variety of hypersensitivity responses to ingested food antigens (IgE-mediated responses) have been reported and are associated with food-specific IgE antibodies. These IgE antibodies bind to high-affinity Fc receptors on mast cells and basophils as well as to low-affinity Fc receptors on macrophages, monocytes, lymphocytes, and eosinophils. When food allergens penetrate the mucosal barrier and reach IgE antibodies bound to mast cells or basophils, the cells are activated and mediators are released; this induces vasodilation, smooth muscle contraction, and mucus secretion, leading to symptoms of immediate hypersensitivity. With repeated ingestion of a food allergen, mononuclear cells are stimulated to secrete histamine releasing factors, a cytokine that interacts with IgE molecules bound to the surface of basophils and perhaps mast cells. A variety of symptoms have been associated with IgE-mediated allergic reactions: shock, urticaria, angioedema, pruritic rash, vomiting, diarrhea, nasal congestion, tongue and laryngeal edema, and wheezing.

Intestinal parasitic infections are also governed by Th1 and Th2 responses. Intestinal helminths are some of the most prevalent and successful parasites in the world. Th2 cells are important in the resistance to intestinal helminths and current data also show that a

Th1 response is associated with susceptibility. Target disruptions of Th2 cytokines, their receptors, or their signaling molecules have highlighted the importance of the Th2 response in protecting the host against parasites, in addition to the importance of the IL-4-like Th2 cytokine, IL-13. There are also indications that usual mechanisms of resistance, such as eosinophils and IgE-mediated hypersensitivity reactions, may not play as important a role as previously thought.

Celiac Disease

Because both cell-mediated and humoral responses are thought to play a role in celiac disease, both Th1 and Th2 responses are considered important in the disease process. Gliadin and related cereal proteins are the undisputed triggers of celiac disease. In active disease, there is an increased expression and altered distribution of HLA class II DR molecules on small intestinal epithelial cells, with greater expression of HLA DR on epithelial cells in the crypt regions. This phenomenon is likely secondary to an increased Th1 cytokine response (IFNγ) by mucosal T cells. The isolation of T cell clones from affected intestinal mucosa that can be stimulated with gliadin peptides supports a role of gliadin interactions with T cells in the pathogenesis of the disease. Furthermore, investigators have demonstrated that mucosal T cells cultured with specific stimuli will exhibit Th1 features and release TNF, and TNF triggered intestinal fibroblasts to secrete matrix metalloproteinases (MMPs) that induced disruption of connective tissue. Increased focal expression of MMP-1 and MMP-3 mRNA in fibroblast cells isolated from the small intestinal mucosa of patients with celiac disease has been reported. Furthermore, specific intraepithelial lymphocytes that are up-regulated in patients with celiac disease modulate the antigen-specific immune response by secreting IL-4, which dampens the Th1 in favor of Th2 reactivation and protects the intestinal mucosa from chronic exposure to damaging agents such as dietary gluten.

On the humoral side, active celiac disease is accompanied by mucosal autoantibodies to reticulin, a common stimulator of the extracellular matrix. IgA antiendomysial autoantibodies (anti-EMAs) allow a screening test for biopsy-proved celiac disease. The role of the Th2 cells is suggested by their ability to regulate other T cell subsets and antibody-producing B cells.

Crohn's Disease

Crohn's disease also appears to involve both Th1- and Th2-mediated responses. Importantly, patients with Crohn's disease have been noted to benefit from the inhibition of a proinflammatory Th1 cytokine TNF, and therefore Th1-mediated responses are likely very important in this disease. The inflammatory infiltrate in Crohn's disease has characteristics of an activated phenotype, likely due to a response to the same antigens over a period of time, inasmuch as there is clonal expansion of CD4+ lymphocytes in the peripheral blood. Importantly, the Th1 cytokine IL-2 mRNA is up-regulated in active CD and the T cells show a hyperreactive response to IL-2. A clinical remission occurs in patients who have fewer IL-2-secreting T cells. Furthermore, the Th1 cytokine IFNγ has been shown to be a critical initiator and perpetuator of the disease. There is a spontaneous release of IFNγ and increased IFNγ mRNA expression by lamina propria mononuclear cells and the presence of IFNγ-secreting T cells in actively inflamed mucosa. Furthermore, there is an enhanced spontaneous production of IL-12 (the critical regulator of Th1 development) in patients with Crohn's disease. IL-18 (another regulator of Th1 development) also appears to be up-regulated in Crohn's disease.

The Th2 responses mediated through IL-4 and IL-5 have also been reported in some patients with Crohn's disease. For example, there is increased production of IgG2 in the intestine, along with a massive number of plasma cells. Other studies have demonstrated an increase in production of other immunoglobulins (IgA, IgM, and particularly IgG). Recurrent Crohn's disease is associated with eosinophilic infiltration and high IL-5 mRNA levels by *in situ* hybridization. However, IL-5 production by cultured mucosal cells is decreased in Crohn's disease.

IL-10 has also been noted to play a role in Crohn's disease. IL-10 has both anti- and proinflammatory effects, depending on local concentrations of IL-10, the types of antigens present in the microenvironment, and the activation state of the immune cells in the vicinity. In the colons of some patients with Crohn's disease, there are already higher levels of IL-10 compared with controls and the mononuclear cells isolated from the ileum of patients with Crohn's disease appear to be nonresponsive to IL-10. Importantly, low ileal IL-10 concentrations have been shown to predict relapse after ileocecal resection and there is reduced IL-10 production within the activated T cell subset in patients with Crohn's disease in remission.

Intestinal Graft-versus-Host Disease

As in many inflammatory disorders in other organs, intestinal graft-versus-host disease (GVHD) appears to

involve complex interactions among immunologic, environmental, and genetic components. T cell cytokines affect the development of GVHD. Both Th1 and Th2 cytokine-mediated occurrences of GVHD have been reported, but intestinal GVHD appears to be mediated predominantly through Th1 cytokines.

The immunopathophysiology of intestinal GVHD involves presentation of antigen to major histocompatibility complex class II disparate CD4+ T cells, which causes activation of the CD4+ T cell and induction of a Th1 cytokine profile. The proinflammatory cytokine release also causes an activation of secondary effector cells, including the macrophage, the NK cell, and the CD8+ T cell. TNF-dependent increases in membrane permeability occur early in the disease process, which likely leads to presentation of microbial products to macrophages, which further induces CD4+ Th1 cell responses in the intestine. A Th1 cytokine profile has been noted in these intestinal lymphocytes of many models of GVHD. The impetus for the Th1 polarization has been shown to involve both IL-12 and TNF. Importantly, TNF has been shown to be efficacious in the amelioration of intestinal GVHD in human trials.

Ulcerative Colitis

The Th2 cytokine IL-5 is up-regulated in ulcerative colitis (UC), thus providing evidence that UC is mediated through these responses. Investigators have reported that IL-5 protein production is increased in cultured mucosal cells from patients with UC and that the IL-5 mRNA is elevated in biopsies from patients with UC. IL-4 mRNA has also reported to be elevated in the diseased colon of UC. Furthermore, specific antibody responses mediated by IL-4 are noted in UC but are not noted in other inflammatory diseases. Antineutrophil cytoplasmic antibodies (ANCAs) have been reported in the serum of a high proportion of patients with ulcerative colitis and sclerosing cholangitis. Moreover, the perinuclear (p-ANCA) pattern of immunofluorescent staining noted in many UC patients differs from the cytoplasmic staining seen in patients with Wegener's granulomatosis. UC patients have been reported to have a serum IgG antibody that specifically reacts with their diseased colonic tissue but not with normal colon. Importantly, specific IgG1 appears to be up-regulated in active ulcerative colitis.

See Also the Following Articles

Celiac Disease • Crohn's Disease • Colitis, Ulcerative • Endomysial and Related Antibodies • Food Allergy • Mast Cells • Tumor Necrosis Factor-α (TNF-α)

Further Reading

Fiocchi, C. (1998). Inflammatory bowel disease: Etiology and pathogenesis. *Gastroenterology* **115**, 182–205.

Glimcher, L. H., and Murphy, K. M. (2000). Lineage commitment in the immune system: The T helper lymphocyte grows up. *Genes Dev.* **14**(14), 1693–1711.

Hsieh, C. S., Macatonia, S. E., Tripp, C. S., Wolf, S. F., O'Garra, A., and Murphy, K. M. (1993). Development of TH1 CD4+ T cells through IL-12 produced by *Listeria*-induced macrophages. *Science* **260**, 547–549.

Kaplan, M. H., Wurster, A. L., and Grusby, M. J. (1998). A signal transducer and activator of transcription (Stat)4-independent pathway for the development of T helper type 1 cells. *J. Exp. Med.* **188**, 1191–1196.

Maggi, E., Parronchi, P., Manetti, R., Simonelli, C., Piccinni, M. P., Rugiu, F. S., DeCarli, M., Ricci, M., and Romagnani, S. (1992). Reciprocal regulatory effects of IFN-γ and IL-4 on the *in vitro* development of human Th1 and Th2 clones. *J. Immunol.* **148**, 2142.

Parronchi, P., DeCarli, M., Manetti, R., Simonelli, C., Sampognaro, S., Piccinni, M. P., Macchia, D., Maggi, E., Del Prete, G., and Romagnani, S. (1992). IL-4 and IFN (alpha and gamma) exert opposite regulatory effects on the development of cytolytic potential by Th1 or Th2 human T cell clones. *J. Immunol.* **149**, 2977.

Powrie, F., and Coffman, R. L. (1993). Cytokine regulation of T cell function: Potential for therapeutic intervention. *Immunol. Today* **14**, 270–274.

Seder, R. A., Gazzinelli, R., Sher, A., and Paul, W. E. (1993). IL-12 acts directly on CD4+ T cells to enhance priming for IFN-γ production and diminishes IL-4 inhibition of such priming. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10188–10192.

Zhong, Z., Wen, Z., and Darnell, J. E. (1994). Stat3 and Stat4: Members of the family of signal transducers and activators of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4806–4810.

Toxic Megacolon

CHRISTOPH GASCHE
School of Medicine, University of Vienna

inducible nitric oxide synthase Key enzyme for production of nitric oxide, a potent endogenous vasodilator that acts through smooth muscle relaxation. The synthase expression is up-regulated through inflammatory processes.

polymerase chain reaction Laboratory method for amplification of genetic sequences from various sources (cells, plasma, feces, etc.).

systemic inflammatory response syndrome Systemic reaction occurring after a variety of insults, including infection, trauma, ischemia, or immune-mediated organ injury; involves one ore more of the following clinical manifestations: (a) a body temperature >38°C or 36°C; (b) a heart rate of >90 beats/min; (c) tachypnea >20 breaths/min; (d) white blood cell counts >12,000 cells/mm^3 or <4000 cells/mm^3. These physiologic changes should represent an acute alteration from the baseline in the absence of other known causes for such abnormalities.

toxic megacolon Dilatation of the colon during the course of fulminant colonic inflammation.

Toxic megacolon refers to dilatation of the colon during the course of fulminant colonic inflammation, typically characterized by severe diarrhea, abdominal distension and tenderness, and total or segmental colonic dilatation without distal obstruction. This rare but potentially lethal complication of infectious colitis or inflammatory bowel disease is associated with symptoms of the systemic inflammatory response syndrome such as fever, elevation or drop in white blood cell count, anemia, tachycardia, and hypotension. The colonic dilatation that may occur in patients with Hirschsprung's disease, chronic constipation of any cause, or intestinal pseudo-obstruction is fundamentally different from toxic megacolon, in that colitis or systemic signs of systemic inflammatory response syndrome do not accompany these disorders.

PATHOGENETIC MECHANISMS

The term "toxic" points to the pathogenetic mechanisms underlying colonic dilatation. Whereas typical colitis is limited to the mucosal layer, toxic megacolon is characterized by inflammation extending into the smooth muscle and thereby leading to motility disturbances, including atony and various degrees of myocyte degeneration. The extent of dilatation relates to the severity and depth of inflammation and to the expression of inducible nitric oxide synthase (iNOS) in the colonic muscularis propria. iNOS may locally generate excessive amounts of nitric oxide, which is known to induce smooth muscle relaxation, and may be responsible for the colonic paralysis. In animal models, high levels of iNOS expression have been reduced by bowel decontamination with oral nonabsorbable antibiotics or administration of dexamethasone, and iNOS inhibitors prevented colonic dilatation.

INCIDENCE

There is no reliable information on the frequency of toxic megacolon. Retrospective analyses done during the 1960s and 1970s reported a prevalence of 10 cases in 100 ulcerative colitis patients in respective centers. Based on this author's personal observations since 1990 in a European inflammatory bowel disease (IBD) referral center, the estimate is a prevalence of 1 case in 1000 patients. When analyzing the literature on toxic megacolon, it is seen that most of the knowledge that gathered since 1965 is extracted from case reports (Fig. 1). Only 30 original publications are identifiable, most of which are retrospective analyses of small diagnostic or interventional series. No prospective study has ever been published on toxic megacolon in humans. The publication frequency (and possibly also the incidence) has dropped further since 1990 (Fig. 2) and since that time, most series have dealt with various non-IBD-related causes of toxic megacolon. The tremendous improvements in IBD awareness, diagnosis, and clinical management may have made this life-threatening complication disappear and may have shifted the most common cause of toxic megacolon to infectious causes such as *Clostridium difficile*.

ETIOLOGY

Toxic megacolon is not a specific complication of ulcerative colitis but may appear in the course of any

 Copyright 2004, Elsevier (USA). All rights reserved.

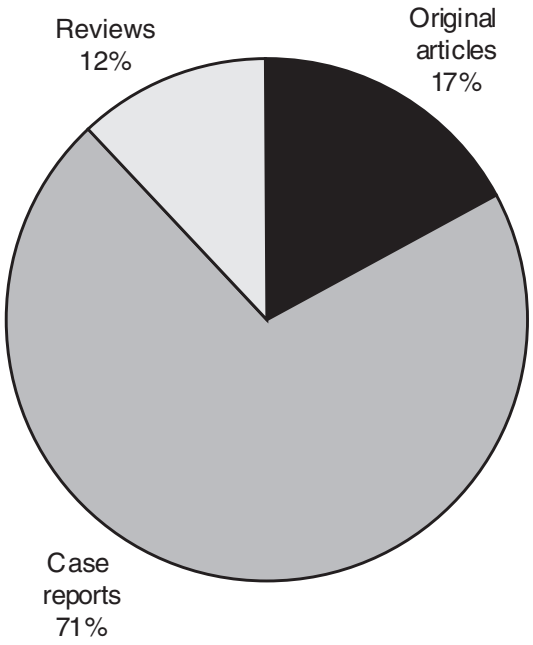


FIGURE 1 Publications on toxic megacolon between 1965 and 2001. Between 1965 and 2001, 251 articles were retrieved form the PubMed database (www.ncbi.nlm.nih.gov) using "toxic megacolon" or "toxic <and> dilatation <and> (colon <or> colitis)" as search terms. Articles with English language publication were further analyzed. Of 174 English language articles, 30 (17%) were identified as original publications, 123 (71%) were case reports, and 21 (12%) were reviews. The 30 original contributions were of a retrospective nature.

inflammatory lesion of the colonic wall, including Crohn's disease, indeterminate colitis, Behçet's disease with colonic involvement, and colitis due to ischemia (Table I). It has been learned in recent years that

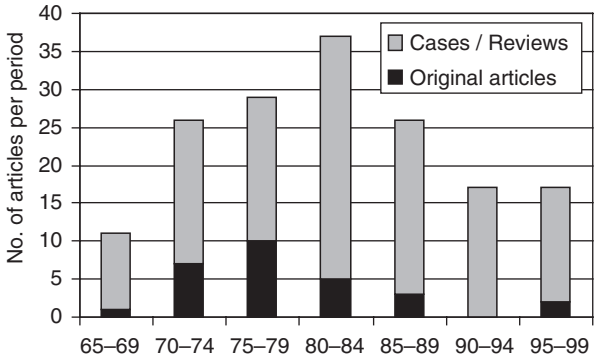


FIGURE 2 English language literature on toxic megacolon over 5-year-periods. Most of the reports on inflammatory bowel disease-associated toxic megacolon were published in the 1970s and 1980s. Notably fewer reports have been published since 1990, and most of them have dealt with various infectious causes of toxic megacolon rather than inflammatory bowel disease.

TABLE I Etiology of Toxic Megacolon

Infectious	Noninfectious
Viral	Inflammatory bowel disease
Cytomegalovirus	Ulcerative colitis,
AIDS related (includes	Crohn's colitis,
Kaposi's sarcoma)	indeterminate colitis
Bacteria	Behçet's disease
Clostridium	Ischemic colitis
difficile-associated	Drug induced
pseudomembranous	Chemotherapy,
colitis	methotrexate,
Campylobacter, *Salmonella*,	loperamide, overdose of
Shigella, *Yersinia*	tricyclic antidepressants
Parasitic	
Entamoeba histolytica,	
Cryptosporidium	

pseudomembranous colitis due to *C. difficile* infection is a major cause of toxic megacolon. In a retrospective evaluation from the Harvard Medical School, 21 of 710 patients with pseudomembranous colitis (3%) had required intensive care submission or even had died from this complication. Among patients with HIV infection or AIDS, cytomegalovirus colitis is the leading cause of toxic megacolon. Evidence of cytomegalovirus infection of the colon has also been found also in resected colons from patients with ulcerative colitis and toxic megacolon, perhaps implying that cytomegalovirus superinfection may be the actual cause of this complication of ulcerative colitis. Other immune-compromised patients are also at risk. AIDS-related toxic megacolon may arise from Kaposi's sarcoma of the colon or cryptosporidiosis.

DIAGNOSTIC PROCEDURES

Patients have distension of the colon as measured on plain abdominal films (Fig. 3) and some signs of systemic inflammatory response syndrome (SIRS) (e.g., fever, tachycardia, tachypnea, or white blood cell count alteration) associated with hypotension, electrolyte imbalances, or a decreased level of consciousness. Besides plain abdominal films, computer tomography (CT) scans are appropriate for early detection of life-threatening intraabdominal complications such as colonic perforation or septic thrombosis of the portal vein. Factors thought to increase the risk of complications include procedures that increase colon trauma, such as barium enema and colonoscopy, medications that decrease gastrointestinal motility, and electrolyte imbalances. Colonoscopy, however, has been proposed for

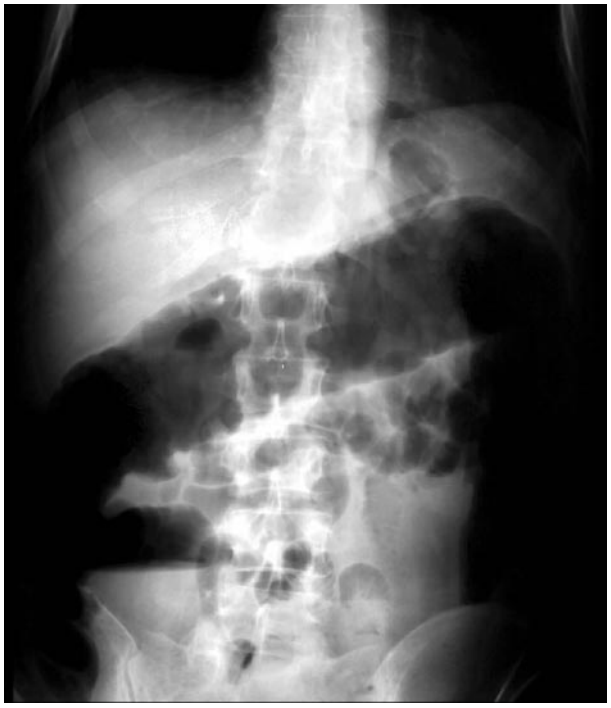


FIGURE 3 Toxic megacolon. Colonic dilatation (9 cm in the cecum, or 8 cm in the transverse colon, or 7 cm in the descending colon, or 6.5 cm in the sigmoid colon) with intraluminal air and/or fluid is typically associated with a distorted colonic contour. Courtesy of P. Pokieser, University of Vienna.

diagnostic purposes and for therapeutic decompression and may be used with caution.

The differential diagnosis of toxic megacolon is based on the patient's history (with specific consideration of inflammatory bowel diseases, AIDS, or immunosuppression) and the results of microbiological stool tests (including the detection of *C. difficile* toxin). Diagnosis of cytomegalovirus (CMV) infection may include serology, phosphoprotein-65 detection in leukocytes, and cytomegalovirus polymerase chain reaction (PCR) from plasma and feces. Timely colonoscopy, though associated with increased risk of perforation, may be crucial in patients with negative diagnostic results.

TREATMENT

Medical management of toxic megacolon depends on the appropriate etiological diagnosis (see Table 1). Even if the diagnosis is obscure, intravenous antibiotics such as metronidazole and ciprofloxacin are safe to be initiated right after the primary diagnostic procedures. In case of *C. difficile* colitis, the offending antibiotic must be stopped and vancomycin should be added via the nasogastric tube. Antiviral therapies are indicated in cytomegalovirus colitis. Intravenous steroids are still the first-line treatment of severe inflammatory bowel disease. In severe ulcerative colitis, high-dose cyclosporine has proved helpful in avoiding emergency colectomy and should therefore be added to steroids. Though studies are lacking, infliximab may have advanced to second-line therapy in severe colonic Crohn's disease.

Patients should be monitored in the intensive care unit for fluid volume, electrolyte replacement, hemoglobin levels, and bowel distension (plain abdominal X rays every 12–24 hours). Further supportive care includes parenteral nutrition, bowel rest with nasogastric tube, and colonic decompression either by frequent "rolling" of the patient to the prone position or to the knee–elbow position or by endoscopy. Medication that slows motility should not be used.

Surgical intervention is necessary if there are signs of progressive dilatation, systemic deterioration, perforation, or hemorrhage. In the emergency setting, subtotal colectomy with Brooke ileostomy and Hartmann closure of the rectum is the procedure of choice. Surgery after colonic perforation is associated with higher mortality. It is regarded an art to find the optimal timing for such a surgical procedure. Frequent cooperative assessments of the patient's status between medical and surgical teams are the key to reduce mortality in the long run.

See Also the Following Articles

Behçet's Disease • Colectomy • Colitis, Indeterminate • Colitis, Ulcerative • Colonic Ischemia • Crohn's Disease • Cytomegalovirus

Further Reading

Beaugerie, L., Ngo, Y., Goujard, F., Gharakhanian, S., Carbonnel, F., Luboinski, J., Malafosse, M., Rozenbaum, W., and Le Quintrec, Y. (1994). Etiology and management of toxic megacolon in patients with human immunodeficiency virus infection. *Gastroenterology* **107**, 858–863.

Greenstein, A. J., Sachar, D. B., Gibas, A., Schrag, D., Heimann, T., Janowitz, H. D., and Aufses, A. H., Jr. (1985). Outcome of toxic dilatation in ulcerative and Crohn's colitis. *J. Clin. Gastroenterol.* **7**, 137–143.

Imbriaco, M., and Balthazar, E. J. (2001). Toxic megacolon: Role of CT in evaluation and detection of complications. *Clin. Imaging* **25**, 349–354.

Mourelle, M., Casellas, F., Guarner, F., Salas, A., Riveros-Moreno, V., Moncada, S., and Malagelada, J. R. (1995). Induction of nitric oxide synthase in colonic smooth muscle from patients with toxic megacolon. *Gastroenterology* **109**, 1497–1502.

Rubin, M. S., Bodenstein, L. E., and Kent, K. C. (1995). Severe *Clostridium difficile* colitis. *Dis. Colon Rectum* **38**, 350–354.

Sheth, S. G., and LaMont, J. T. (1998). Toxic megacolon. *Lancet* **351**, 509–513.

Trace Minerals: Metabolism and Deficiency (Copper, Zinc, Selenium, Manganese)

ROBERT B. RUCKER AND LOUISE LANOUE
University of California, Davis

antiport Transport that is dependent on the countertransport of another substrate across a membrane barrier.
cis-acting/trans-acting elements Factors that regulate gene expression in eukaryotic cells, by the interaction of specific proteins (trans-acting elements) with specific short nucleotide sequences (cis-element motifs), usually in the promoter region of genes.
egress Cellular export; from the Latin *egressus*.
initiation/elongation factors Proteins involved in the initiation and synthesis of polypeptide chains at the level of translation.
natural resistance-associated macrophage proteins Novel family of proteins (Nramp1, Nramp2, and yeast proteins Smf1 and Smf2); functionally related transporters, defined by a conserved hydrophobic core of 10 transmembrane domains, that target Fe, Mn, and Zn.
P-ATPases Family of adenosine triphosphate-dependent metal transporters that have specificity for divalent cations.
reactive oxidant species Compounds derived from oxygen-mediated reactions.
transmembrane proteins Portions of the polypeptide chain, usually lipophilic in nature; capable of traversing the cell membrane lipid bilayer, often multiple times.
zeta-interacting proteins Family of metal transporters with a high avidity for zinc.
zinc finger motifs Structural domains associated with specific DNA–protein transcription factor interactions that are dependent on zinc for structural integrity.

Regulatory features are important to copper (Cu), zinc (Zn), manganese (Mn), and selenium (Se) metabolism and their metabolic functions. From a nutritional perspective, normal growth, development, and many aspects of disease prevention require sustained intakes of Cu, Zn, Mn, and Se in amounts sufficient to meet the recommended daily requirements and allowances for each of these elements. The sophistication of regulatory mechanisms to maintain cellular homeostasis for Cu, Zn, Mn, and Se is in keeping with the unique and broad range of functions they perform as cofactors.

COPPER

Metabolic Roles and Dietary Deficiency

Copper is essential as a redox cofactor in enzymes that usually function as oxidases. The signs and symptoms of Cu deficiency reflect perturbations in the activity of Cu-containing oxidases and range from poor energy utilization to impaired protection from oxidative free-radical damage. Specific examples include reduced cytochrome oxidase activity, defects in extracellular matrix formation, abnormal production of brain neurotransmitters, and poor iron utilization. Although Cu deficiency in humans is uncommon, experimentally induced deficiencies of Cu in animals result in anemia, defective bone development, and vascular accidents and aneurysms.

Cu deficiency is important to all aspects of growth and development. In the fetus and neonate, Cu deficiency can result in inappropriate patterns of cell death, alterations in the migration of neural crest cells, and changes in the expression of key patterning genes. These defects can be attributed to both morphological abnormalities and related epigenetic or developmental changes in DNA, e.g., in methylation patterns.

For humans, the estimated safe and adequate intake for Cu is 1.5–3.0 mg/day. This is in keeping with the requirement for optimal growth in most animals of 4–8 mg/kg of dry food or 0.5–1.0 mg per 1000 kcal (4.2 MJ). Although nutritional surveys indicate that many individuals consume 1.0 mg or less of copper per day, it is not difficult to meet the recommended level of intake given the amounts of Cu in many foods; for example, nuts, shellfish, organ meats, and legumes contain from 0.3 to 1 mg per typical serving. Condiments (various spices) and chocolate also have appreciable levels of Cu.

Cellular Transport and Regulation

Cu uptake occurs through both high- and low-affinity transport systems. Environmental factors can influence the response to transporters. Most important are factors that influence solubility and redox state. Cu exists in two different valence states; the cupric ion (Cu^{+}) is the primary substrate for the transport systems that take Cu across plasma membranes. Reduction ($Cu^{2+} \rightarrow Cu^{+}$) is catalyzed by plasma membrane reductases. However, cuprous ion (Cu^{2+}) in the intestinal lumen is more soluble than cupric ion (Cu^{+}).

 Copyright 2004, Elsevier (USA). All rights reserved.

Chemical reduction of luminal contents (e.g., by reducing agents such as ascorbic acid) can decrease the amount of bioavailable Cu that may be potentially delivered to the surface of intestinal cells.

From a conceptual perspective, recent studies in yeast have shed light on proteins involved in the process of Cu transport. For example, in *Saccharomyces cerevisiae*, high-affinity Cu ion uptake has been characterized as temperature and ATP dependent. Cu ion uptake appears to be coupled with K^+ efflux with a 1 : 2 stoichiometry, suggesting that the process may take place via a $Cu^+/2K^+$ antiport mechanism. In yeast, the gene for Cu reductase activity, designated *FRE1*, is regulated by the Mac1 transcription factor in response to cellular Cu levels. The entry of Cu into cells is orchestrated by the action of Fre1, the Cu reductase, when Cu contacts high-affinity Cu transporters, currently designated as Ctr1 and Ctr3 (Ctr2 is a low-affinity transporter). Ctr1, Ctr2, and Ctr3 are products of *CTR* genes. High-affinity Cu uptake is facilitated by Ctr1 and Ctr3and is saturable; with a K_m of 1−4 μmol/liter. Under Cu-limiting conditions, there is evidence that the transporters and proteins involved in Cu redox are up-regulated, whereas under Cu-replete conditions, they are down-regulated.

In addition to the transporters, cellular chaperones specific for Cu deliver Cu to specific cellular proteins. Other important features of Cu regulation include the role of metallothionein, a metal-binding protein for Cu, Zn, and Cd that acts to buffer abnormal shifts in the cellular concentrations of Cu, and the proteins and transporters involved in the egress of Cu from cells. Cu egress, or transport out of cells, is controlled by P-ATPase Cu transporters that are located on the surface of vesicles that arise from Golgi processing. A change in Cu status does not appear to alter Cu-transporting P-ATPase gene expression, but it does affect Cu movement to and from that outer cell membrane. Cu homeostasis must be coordinated, because the release of free Cu ions causes damage to cellular components by catalyzing the generation of reactive oxidant species (ROS).

Systemic Regulation of Cu

From the intestine, a case can be made for the transport of Cu on albumin and in the form of low-molecular-weight complexes (e.g., histidine) to target tissues, particularly the liver. Exactly how albumin and histidine relinquish Cu to organs and tissues is currently unclear. From the liver, ceruloplasmin transports Cu to other tissues. Ceruloplasmin, the predominant Cu-containing protein in mammalian serum, is a glycosylated multi-Cu ferroxidase that carries >95% of total serum Cu. Although ceruloplasmin may function in Cu transport, the absence of ceruloplasmin has not been shown to alter Cu levels in the peripheral tissues. Such observations come from what is known about individuals and animal models that are aceruloplasminemic, a genetic disorder of ceruloplasmin deficiency, Moreover, analbuminemic rats do not have significantly impaired Cu metabolism.

The average level of Cu stored in the body ranges from 50 to 120 mg. Cu is found in all organs and tissues of the human body. In cells, Cu is always bound to proteins or to organic compounds and is not found as free Cu ions. There are higher concentrations of Cu in very young compared to adolescent or mature animals (e.g., four- to fivefold greater). The high concentration of Cu in the fetal liver is impressive; in humans, fetal liver contains 20−50 μg per gram of liver, compared to 4−5 μg per gram of liver in adults. The high levels of liver Cu in children persists for 3−5 years, followed by a decline to adult levels.

With regard to overall systemic control, under normal situations, little Cu is excreted via the kidney. Rather, Cu is excreted primarily via the bile and is released into the gastrointestinal tract with limited reabsorption. The uptake of Cu by the intestine and elimination through the bile allows Cu to be conserved and tightly regulated from a systemic perspective.

Genetic Conditions and Cu Metabolism

The understanding of two genetic conditions, Menkes kinky-hair and Wilson disease, has contributed to the understanding of general Cu transport processes. In Menkes kinky-hair disease, there is a problem with Cu absorption and Cu transport in mesenchymal cells. In Wilson disease, there is an increased liver Cu content, leading to severe hepatic damage, followed by increased brain Cu levels and neurological lesions. Menkes disease results in pathology resembling Cu deficiency, as opposed to the pathology of Wilson disease, which resembles Cu toxicity.

Both the Wilson and the Menkes genes code for one of the P-type ATPases involved in Cu egress. In Menkes disease, the mutation in the *P-ATP-7A* gene prevents Cu transport across the basal lamina of the intestine. J. H. Menkes first described this Cu transport disorder in 1962 in a family of English−Irish descent. It was recognized immediately as an X-linked recessive disorder, characterized by retardation, impaired growth, peculiar hair, and focal cerebral and cerebellar degeneration. The condition is often lethal, with death occurring in the first or second year of life, usually from a vascular accident, i.e., aneurysm or stroke. Menkes patients also show signs of osteopenia (poor bone development) and vascular disease.

In cell culture, mesenchymal, epithelial, and neural cells from Menkes patients abnormally sequester Cu. Moreover, the ability to transfer Cu to some Cu-requiring enzymes, e.g., lysyl oxidase, is lacking or abnormal. The frequency for Menkes disease is now estimated to be about 1 in 35,000–40,000 live births among those of English–Irish descent. In contrast, Wilson disease is an inherited, autosomal recessive disorder of Cu accumulation and toxicity that occurs in about 1 of every 40,000 people. The responsible gene (*P-ATP-7B*) also codes for a vesicular membrane-bound, Cu-binding protein, but, unlike the Menkes gene, is expressed primarily in the liver. P-ATPase-7B has considerable homology to P-ATP-7A. As in Menkes disease, there are mutations in *P-ATP-7B* that account for symptoms associated with Wilson disease. Owing to its location in liver, when *P-ATP-7B* is altered by mutations, biliary excretion of Cu is impaired. An important detail is that the vesicles to which P-ATPase-7B is localized also appear to transport ceruloplasmin from cells. In cells adjacent to biliary canaliculi, some of the vesicular movement is to the cellular membrane that is exposed to the biliary canaliculus, whereas in other cells, the movement is to the cell membrane exposed to sinusoids and distensible vascular channels. It has therefore been postulated that the liver "packages" Cu for excretion into the bile by binding Cu to ceruloplasmin for release into bile or plasma. This accounts for the observations that defects in P-ATPase-7B activity often result in low levels of Cu bound to ceruloplasmin in blood and eventually failure of whole-body Cu regulation, because of hepatic accumulation of Cu.

In summary, several complex strategies are used to maintain Cu homeostasis at the cellular and organismal levels. The complexities are in part related to maintaining Cu in an appropriate redox state and the need to accommodate a diverse array of enzymatic functions. Fortunately, Cu deficiency is probably a rare occurrence, but genetic polymorphisms involving Cu transporters can occur, which mimic the signs of Cu deficiency and toxicity observed in animal models.

ZINC

Metabolic Roles and Dietary Deficiency

Zinc functions at the active site of many enzymes by facilitating strong, but readily exchangeable, substrate or ligand binding. Zn is not capable of redox, thus can be used biologically in novel ways at the functional sites of proteins without causing oxidative changes. Zn also plays important structural roles in proteins. One example is the zinc finger motif, the most common recurring motif in proteins that serve as transcription factors.

The importance of Zn in humans from a physiological perspective did not begin to emerge until the 1950s. Zn deficiency in animals as a cause of parakeratosis was first established. With respect to human health and disease, Zn received considerable attention when it was shown that Zn deficiency was an etiological factor in the syndrome of "adolescent nutritional dwarfism." Further work emphasized that it was particularly important to provide children with an adequate dietary Zn intake. Clinical observations in patients fed intravenously also indicated that Zn deficiency might be important to iatrogenic failure. When Zn deficiency does occur, the range of responses includes dermal lesions, poor wound healing, malabsorption, diarrhea, and immunologic defects (especially compromised T cell function).

Specific biochemical changes associated with the clinical features of Zn deficiency are not easy to identify. As a general rule, epithelial cells and cells involved in immune function are most affected by Zn deprivation. The principal biochemical lesion centers on the noncoordination of events important to the normal differentiation of cells; perhaps related to the important function that Zn plays in transcription factor integrity and structure. Of interest, there are greater changes in immune responsiveness than in changes in the activities of Zn-requiring enzymes. Zn can also have a significant impact on the hormonal regulation of cell division, specifically, the pituitary growth hormone (GH) and insulin-like growth factor-1 (IGF-1) axis. Changes in the concentrations of GH are observed in Zn deficiency, and circulating IGF-1 concentrations are consistently decreased. Other evidence suggests that reduced Zn availability affects membrane signaling systems and intracellular second messengers that coordinate cell proliferation, e.g., in response to IGF-1. Measurements of Zn concentrations in plasma are useful in identifying children who are more likely to have a growth response to Zn supplements. Regrettably, precise functional tests for Zn status are not available.

The dietary intake of Zn is around 10–15 mg/day. Sources of Zn include meat and protein-enriched foods. Factors known to influence absorption include the amount of Zn present in the intestinal lumen, and the presence of dietary promoters (e.g., human milk, animal proteins) or inhibitors (e.g., phytate, other minerals). Calcium, especially in the presence of phytate, interferes with Zn absorption.

Cellular and Systemic Regulation of Zn

The primary site of absorption of exogenous Zn in the human is the proximal small bowel—either the distal

duodenum or proximal jejunum. Absorption studies in animal models indicate an inverse relationship between percentage of Zn absorbed and dietary Zn intake. The physiological state also affects absorption. Pregnancy and lactation can enhance absorption, which can vary from 25 to 50% depending on the dose and physiological state.

In humans, the total plasma Zn concentration is 12–25 μmol/liter, with over 90% associated with albumin, about 10% associated with α2-macroglobulin, and less than 1%, complexed to other low-molecular-weight species. Zn homeostasis is achieved largely by enterohepatic recirculation, although both the magnitude and the process differ from those for Cu. For example, a higher percentage of Zn is reabsorbed from endogenous excretion, i.e., about the same as that absorbed from dietary sources. A major source of Zn in the intestinal lumen is from pancreatic secretions, because of the importance of Zn as a cofactor for pancreatic peptidases, various hydrolases, and proteinases.

As is the case for other essential metals, several transporter systems have been identified based on corresponding homologues in yeasts. Similar to the situation for Cu, there are both high-affinity and low-affinity receptors and transporters for Zn. Further, two distinct families of zinc transporters are known: the zeta-interacting protein (ZIP) family, which imports zinc, and the ZnT family, which functions in releasing zinc or sequestering zinc internally. The ZIP transporters are found in the duodenum in the crypts and lower villi and appear available for the uptake of several metal ions, including Zn. Uptake assays demonstrate that Cu^{+} and Fe^{2+} can be potential substrates, because they inhibit Zn^{2+} uptake, whereas Co^{2+}, Mn^{2+}, Mg^{2+}, and Ni^{2+} have no effect on Zn^{2+} uptake. The ZIP transporters are under transcriptional control based on the observation that one of the family of ZIP transporters, ZRT1, is inversely expressed relative to cellular Zn^{2+} levels; Zn^{2+}-depleted cells have 10-fold more ZRT1 mRNA than do Zn^{2+}-repleted cells. There is also evidence for posttranslational regulation. When cellular Zn is elevated, there is degradation of the transporters by vacuolar proteases.

Genetic and Other Conditions Influencing Zn Metabolism

Genetic disorders of Zn metabolism are rare. A condition known as acrodermatitis enteropathica, which responds dramatically to oral Zn supplementation, occurs in children. Acrodermatitis enteropathica is an inborn error of zinc metabolism, which is autosomal recessive. Characteristic symptoms in infancy include periorificial (oral, anal, and genital) and acral dermatitis, diarrhea, behavioral and mental changes, neurological disturbances, and secondary bacterial and fungal infections. Disorders of Zn metabolism secondary to the primary disease have also been reported in alcoholic cirrhosis, inflammatory bowel diseases, diabetes, and renal disease, although the significance of these observations in relation to Zn deprivation remains unclear.

MANGANESE

Metabolic Roles and Dietary Deficiency

Manganese is an essential trace element that is required for the activity of enzymes with transferase or hydrolase functions. The mitochondrial form of superoxide dismutase also requires Mn. Further, Mn deficiency is associated with perturbations in glucose metabolism, insulin function, and cholesterol regulation.

The estimated safe and adequate daily dietary intake for Mn in humans is 2–5 mg. This estimate is based primarily on dietary intake data and inferences from animal requirement studies. Regrettably, well-controlled Mn balance and excretion data are limited or have been difficult to obtain. In studies involving young adult male subjects fed conventional foods, the minimum requirement for Mn was estimated to be between 1 and 2 mg/day, an amount easily obtained in most diets. Variations in Mn intake reflect variations in food choice and the amount in the water supply (4 mg/day is defined as the lowest observable adverse-effect level, or LOAEL). Of indirect measures, serum Mn concentrations in combination with lymphocyte Mn-dependent superoxide dismutase (MnSOD) activity and blood arginase activity may have utility as indirect measures of Mn status.

Regarding excessive Mn exposure, Mn toxicity is a health risk to miners and other workers exposed to Mn-enriched dust. Excessive Mn exposure can cause impaired neurological/movement disorders. Serum Mn concentrations in combination with brain magnetic resonance imaging (MRI) scans, and neurofunctional tests seem to be the best way to monitor excessive exposure to Mn.

Cellular and Systemic Regulation of Mn

In intestinal cell models, the uptake and transport of Mn appear controlled by saturation-type kinetics. The transport characteristics under steady-state conditions in the intestine can exhibit two components that probably reflect transcellular (carrier-mediated) and paracellular (diffusional) pathways. Calcium, calcium antagonists, ATP synthesis inhibitors, and high levels of iron decrease Mn fluxes. From the exsorption side,

the Mn flux is approximately 20-fold less than in the absorptive direction. In humans and rodent models, the absorption of Mn is decreased by phytic acid, when compared with corresponding dephytinized food components, and by high intakes, corresponding to four to five times the normal requirements. Reducing agents, such as ascorbic acid do not influence uptake.

The transport of Mn most likely involves homologues of the natural resistance-associated macrophage protein (Nramp) transporters found in yeast. In *Saccharomyces cerevisiae*, the expression of three Nramp-family transporters (Smf1p, Smf2p, and Smf3p) responds to changes in cellular Mn. In particular, Smf1p and Smf2p appear to function in Mn uptake and trafficking. Analogous to the regulation of Zn and Cu transporters, the differential regulation of Nramp transporters in yeast by Mn occurs at the level of protein stability and protein trafficking through the various secretory pathways in addition to transcriptional control. For example, Smf1p and Smf2p are degraded in cells with sufficient Mn. In contrast, Smf1p and Smf2p accumulate to high levels in Mn-deficient cells. Although the evidence suggesting Nramp transporters is inferential, much of it is very strong. As an example, animal models with iron transporter defects also have impaired Mn transport. Homozygous Belgrade rats, which have hypochromic anemia due to impaired iron transport, also have abnormalities in Mn metabolism

At present, there are few cases of Mn deficiency in the medical literature. Given the heterogeneity of the North American food supply, it is difficult to make a strong case for deficiency without involving other factors or etiologies. There may be reasons, however, to be concerned about Mn toxicity, given that intakes do exceed what are thought to be the requirements for Mn.

SELENIUM

Metabolic Roles and Dietary Deficiency

Selenium is of fundamental importance to human health. It is an essential component of several major metabolic pathways, including thyroid hormone metabolism, antioxidant defense, and immune function. Se is incorporated as selenocysteine at the active site of a wide range of selenoproteins. In the past two decades, over 30 new selenoproteins have been identified, of which 15 have been characterized. A number of provocative clinical studies suggest putative roles for Se in cancer protection, ROS protection, and even relationships involving viral exposure.

Plant foods are the major dietary sources of Se in most countries throughout the world. The amount of Se in soil, which varies by region, may determine the amount of Se in the food chain, wherein Se is found as selenomethionine and selenocysteine. Se is one of the few mineral elements in which the soil concentration can influence the relative amounts found in food. In the United States, the high plains of northern Nebraska and the Dakotas have the high levels of soil Se. People living in those regions generally have the highest Se intakes in the United States. Soils in central China and parts of Russia have low amounts of Se, and dietary Se deficiency is often reported in those regions. Se is also found in some meats and in seafood. Animals that eat grains or plants that were grown in Se-rich soil have higher levels of Se in muscle tissue. Some nuts, in particular Brazil nuts and walnuts, are also very good sources of Se.

The Recommended Daily Allowance (RDA) for Se is 50–100 μg in the United States and 10–75 μg in Europe. Keshan disease, also known as Se deficiency, has been observed in low-Se areas of China, where dietary intake is $<10-20$ μg/day, an intake that is significantly lower than the RDA for Se. Se deficiency has also been observed in people who rely on total parenteral nutrition (TPN) as their sole source of nutrition. The signs of Keshan disease are cardiomyopathy, increased susceptibility to infection, cataracts, and muscular lesions resulting from impaired ROS protection. There can also be loss of pigmentation in skin and hair, and, when severe, growth retardation.

Although rare in the United States, there is also a health risk of too much Se, e.g., selenosis. Symptoms include gastrointestinal upsets, hair loss, white blotchy nails, and nerve damage. The few reported cases have been associated with industrial accidents or manufacturing errors that have led to excessively high amounts of Se in supplements. The Institute of Medicine has set a tolerable upper intake level for Se at 400 μg/day.

Cellular and Systemic Regulation of Se

Both organic and inorganic forms of Se can be utilized in the body. The order of uptake in Caco-2 cells is $SeO_3^{2-} \leq$ selenocysteine $<$ selenomethionine $<SeO_4^{2-}$. Both amino acid-related and anion transporters are involved in Se transport. Many of the details, however, have yet to be resolved. For example, the transport of selenomethionine is inhibited by its sulfur analogue, methionine, whereas inhibition of the transport of selenocysteine by cysteine is not observed. The transport of SeO_4^{2-} is inhibited by thiosulfate, but not sulfate. A Na^+,K^+-ATPase is probably responsible for energizing the brush border transport of selenate, where the ileum is the site of absorption. A number of intestinal inflammatory diseases and short-bowel syndrome can lead to Se deficiency.

In contrast with intestinal cells, red cells in particular demonstrate selective uptake of organic forms of selenium. Of the most common Se compounds (selenate, selenite, selenomethionine, and selenocysteine), selenite injected intravenously is taken up rapidly and selectively through an anion-exchange carrier or transporter. The rapid and selective uptake of selenite by red blood cells is explained by the selective and efficient uptake through the anion-exchange carrier, followed by reduction by glutathione-requiring steps. Selenite uptake is inhibited by chromate, which promotes glutathione depletion.

Another unique aspect of Se regulation is that its insertion into protein occurs posttranslationally in the form of the amino acid, selenocysteine (SeCys). Such occurrence of this element in protein is widespread in all forms of life. Elucidating how Se is incorporated as SeCys has modified our understanding of the genetic code. The formation of SeCys with its novel codon expands to 21 the codon used for naturally occurring amino acids. Although it was recognized in the mid-1960s that the codon AUG had a dual role of initiating protein synthesis and inserting methionine at protein translation start sites, the possibility that a second codon also had two functions was not considered at the time. It is now known that UGA serves both as a termination and a SeCys codon.

SeCys can be attached to $tRNA^{Cys}$ by cysteinyl-tRNA synthetase and can be incorporated nonspecifically into protein in response to Cys codons, which is the reason why many proteins contain Se at sites other than the Se-containing active sites in selenoenzymes. However, SeCys is synthesized by a novel process. The translation of selenoprotein mRNAs requires both cis-acting and trans-acting factors. SeCys is inserted into nascent selenopeptides in mammals by means of a unique amino acid insertion system. Specific 3′ untranslated (UTR) mRNA structures, designated SECIS elements, function in recruiting SBP2, a SeCys-specific elongation factor, and selenocysteinyl-$tRNA^{Ser,SeCys}$ into the SeCys insertion complex, the selenosome. SeCys $tRNA^{Ser,SeCys}$ is used as the site for SeCys biosynthesis and for its incorporation into the active site of specific selenoproteins. Further work also suggests considerable complexity in the regulation of specific selenoenzymes, e.g., transcriptional as well as translation regulatory controls exist for glutathionine peroxidase.

CONCLUDING COMMENTS

Although it is difficult to make a case for frank nutritional deficiencies of Cu, Zn, Mn, or Se in humans, there is a good likelihood of identifying polymorphisms that may alter an individual's need for a given element. Although perhaps obvious from a biochemical perspective, it is also important to underscore that each mineral contributes unique chemical properties that are essential to a diverse array of enzymatic and related functions. The complexities of these functions underscore the need in turn for equally complex regulatory and homeostatic controls.

See Also the Following Articles

Dietary Reference Intakes (DRI): Concepts and Implementation • Digestion, Overview • Malnutrition • Wilson's Disease

Further Reading

Bull, P. C., and Cox, D. W. (1994). Wilson disease and Menkes disease: New handles on heavy-metal transport. *Trends Genet.* **10**, 246–252.

Burk, R. F. (2002). Selenium, an antioxidant nutrient. *Nutr. Clin. Care* **5**, 75–79.

Burk, R. F., and Hill, K. E. (1993). Regulation of selenoproteins. *Annu. Rev. Nutr.* **13**, 65–81.

DiDonato, M., and Sarkar, B. (1999). Genetic disorders of metal transport and metabolism. *In* "The Molecular Biology and Toxicology of Metals" (R. K. Zalups and D. J. Koropatnick, eds.), Chap. 6, pp. 186–208. Taylor and Francis, London.

Driscoll, D. M., and Copeland, P. R. (2003). Mechanism and regulation of selenoprotein synthesis. *Annu. Rev. Nutr.* **23**, 17–40.

Eide, D. J. (2000). Metal ion transport in eukaryotic microorganisms: Insights from *Saccharomyces cerevisiae*. *Adv. Microb. Physiol.* **43**, 1–38.

Eide, D. J. (1998). The molecular biology of metal ion transport in *Saccharomyces cerevisiae*. *Annu. Rev. Nutr.* **18**, 441–469.

Freeland-Graves, J. H., Behmardi, F., Bales, C. W., Dougherty, V., Lin, P. H., Crosby, J. B., and Trickett, P. C. (1988). Metabolic balance of manganese in young men consuming diets containing five levels of dietary manganese. *J. Nutr.* **118**, 764–773.

Greger, J. L. (1998). Dietary standards for manganese: Overlap between nutritional and toxicological studies. *J. Nutr.* **128**, 368S–371S.

Guerinot, M. L. (2000). The ZIP family of metal transporters. *Biochim. Biophys. Acta* **1465**, 190–198.

Harris, E. D. (2000). Cellular copper transport and metabolism. *Annu. Rev. Nutr.* **20**, 291–310.

Hellman, N. E., and Gitlin, J. D. (2002). Ceruloplasmin metabolism and function. *Annu. Rev. Nutr.* **22**, 439–458.

Keen, C. L., Ensunsa, J. L., and Clegg, M. S. (2000). Manganese metabolism in animals and humans including the toxicity of manganese. *Met. Ions Biol. Syst.* **37**, 89–121.

Menkes, J. H. (1999). Menkes disease and Wilson disease: Two sides of the same copper coin. Parts I and II: *Eur. J. Paediatr. Neurol.* **3**, 147–158; **3**, 245–253.

O'Halloran, T. V., and Culotta, V. C. (2000). Metallochaperones, an intracellular shuttle service for metal ions. *J. Biol. Chem.* **18275**(33), 25057–25060.

Pena, M. M., Lee, J., and Thiele, D. J. (1999). A delicate balance: Homeostatic control of copper uptake and distribution. *J. Nutr.* **129**, 1251–1260.

Simpkins, C. O. (2000). Metallothionein in human disease. *Cell. Mol. Biol.* **46**, 465–488.

Transforming Growth Factor-β (TGF-β)

JOHN BARNARD
Columbus Children's Hospital and Ohio State University

growth factor A protein, usually secreted from cells, that exerts its biological activity by binding to high-affinity, cell surface receptors at low concentrations; most growth factors have diverse biological activities including actions independent of cell growth regulation.

Transforming growth factor-β (TGF-β) is a dimeric polypeptide growth factor belonging to a superfamily of related proteins. All superfamily members are distinguished structurally by a cluster of conserved cysteine residues held together by disulfide bonds. These proteins display a remarkable diversity of biological activities, including cell growth regulation, cell fate determination, immune function, and morphogenesis. TGF-β-related peptides and TGF-β receptors are synthesized by virtually all normal cells including epithelial, smooth muscle, and hematopoietic cells in the gastrointestinal tract.

THE TRANSFORMING GROWTH FACTOR-β GENES AND PROTEINS

The mammalian transforming growth factor-β (TGF-β) family consists of three proteins with similar biological activities. These three "isoforms" are designated TGF-β_1, β_2, and β_3 and are localized to human chromosomes 19, 1, and 14, respectively. All three TGF-β proteins are first synthesized and secreted in biologically inactive, latent precursor complexes that are incapable of binding to cell surface TGF-β receptors. "Activation" of TGF-β occurs when the mature, 25 kDa biologically active form of TGF-β is liberated from the precursor complex. Activation of latent TGF-β is believed to be a major step in the regulation of TGF-β activity. Mature, biologically active TGF-β is released from the latent complex by extremes of pH, proteolysis by plasmin, and transglutaminase-dependent interaction with the extracellular matrix. A great deal remains to be learned about the normal process of TGF-β activation and how it is altered under pathophysiological conditions.

TRANSFORMING GROWTH FACTOR-β RECEPTORS

The TGF-β receptors are designated TGF-βRI (~53 kDa), TGF-βRII (~70 kDa), and TGF-βRIII (~300 kDa). Each has an extracellular ligand-binding domain, a transmembrane sequence, and a cytoplasmic sequence. In general, all three isoforms of TGF-β bind to receptor types I, II, and III, albeit with modest differences in affinity. Binding of TGF-β to TGF-βRII recruits TGF-βRI into a tetrameric receptor complex, resulting in the transphosphorylation and activation of a serine/threonine kinase in the cytoplasmic tail of TGF-βRI. It is convenient to think of TGF-βRII as a "sensor" receptor and TGF-βRI as a "transducer" receptor. TGF-βRIIIs are abundant, but have no known signaling capability and have mostly been considered "reservoir" receptors. Regulation of receptor signaling is remarkably complex because of a large number of receptor-interacting proteins that modify signaling. These include FKBP12, the α-subunit of farnesyl transferase, TRIP-1, STRAP, and TRAP-1, among others.

INTRACELLULAR SIGNALING BY TRANSFORMING GROWTH FACTOR-β

The main signaling proteins activated by TGF-β superfamily members are called Smads (Fig. 1), although a variety of other traditional intracellular signaling pathways are also modulated. The Smads most relevant to TGF-β signaling in the gastrointestinal tract are Smad2 and Smad3 (receptor-regulated or R-Smads), Smad4 (a common-partner or Co-Smads), and Smad6 and Smad7 (inhibitory Smads). Phosphorylation of Smad2 and Smad3 by serine/threonine kinase activity in TGF-βRI is necessary for activation of TGF-β signaling. Phosphorylation of Smad2 and Smad3 permits interaction with Smad4. This is the final configuration required for translocation of the TGF-β signal into the nucleus, where modification of gene expression occurs by binding to unique Smad-binding elements in TGF-β-responsive genes. Smad7, an inhibitory Smad, binds to activated TGF-βRI and inhibits the phosphorylation of Smad2, the assembly of Smad complexes, and the nuclear translocation of Smads.

An increasingly diverse and complex array of transcriptional activators and repressors interact with the Smad complex to modify gene transcription in the cell nucleus. These include transcriptional activators such

 Copyright 2004, Elsevier (USA). All rights reserved.

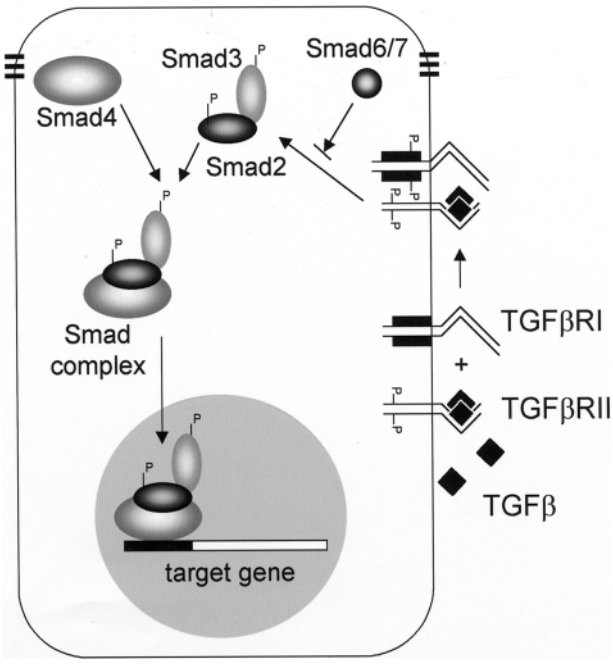


FIGURE 1 TGF-β signaling: TGF-β binds to a dimerized TGF-β type II receptor, initiating an interaction with type I receptors and phosphorylation of the cytoplasmic tail of TGF-β type I receptor, as well as Smad2 and Smad3. In concert with Smad4, the Smad complex translocates to the nucleus to modify gene expression and alter cell behavior.

as CREB-binding protein (where CREB denotes Ca^{2+}/cAMP-response element-binding protein)/p300 and transcriptional repressors such as SnoN, c-Ski, and TGIF. Many of these proteins secondarily interact with classical transcriptional repressors, such as histone deactylases.

BIOLOGICAL ACTIONS OF TRANSFORMING GROWTH FACTOR-β IN THE HUMAN GASTROINTESTINAL TRACT

TGF-β has diverse biological actions in the gastrointestinal tract, including inhibition of epithelial cell growth, stimulation of extracellular matrix (ECM) synthesis, stimulation of angiogenesis, stimulation of cell motility, and suppression of the immune system. The growth inhibitory actions of TGF-β are especially important, as their loss may result in gastrointestinal cancer. TGF-β normally arrests cell growth by causing a reversible block in the G1 phase of the cell cycle. This occurs as a result of inhibition of c-myc proto-oncogene expression and blockade of cyclin-dependent kinase activities, both of which are critical preparatory steps for DNA synthesis and cell replication. Another major biological effect of TGF-β is the modulation of genes regulating ECM deposition, which results in extracellular matrix formation. This occurs by increased synthesis of ECM proteins (collagen, fibronectin) and proteinase inhibitors (plasminogen activator inhibitor), decreased synthesis of matrix-degrading activities (plasminogen activator, collagenase, elastase), and modulation of matrix receptors and binding proteins (integrins). The effect of TGF-β on extracellular matrix synthesis appears to have major clinical significance, as a large number of fibrotic diseases, such as cirrhosis, pulmonary fibrosis, stricture, and glomerulonephritis, are associated with increased TGF-β activity. Because of its effects on the extracellular matrix, TGF-β increases wound healing, including incisional wounds in the gastrointestinal tract. The effects of TGF-β on the immune system are complex, but TGF-β is predominantly immunosuppressive. The marked inflammation observed in the gastrointestinal mucosa of TGF-β-deficient "knockout" mice underscores the importance of TGF-β in regulating the gastrointestinal immune system. Blockade of TGF-β signaling in immune cells, with resultant increased immune activity, has been proposed as an important factor in inflammatory bowel disease and autoimmune hepatitis.

A ROLE FOR TRANSFORMING GROWTH FACTOR-β IN GASTROINTESTINAL CANCER

More than 75% of human colon and pancreatic cancers are resistant to growth arrest by TGF-β, an observation believed to be key in the causation of these tumors. For example, an inactivating mutation in the TGF-βRII gene occurs in persons with defects in DNA mismatch repair, leading to hereditary nonpolyposis colon cancer, a familial cancer syndrome with an increased incidence of gastric, endometrial, and colon cancers. Overall, these types of mutations account for approximately 15% of all colorectal cancers. An additional 6–15% of sporadic colon cancers have mutations in TGF-βRII. Inactivating mutations in Smad2 and Smad4 occur in approximately 6 and 20% of sporadic colorectal carcinomas, respectively. Smad4 mutations are found in approximately 40% of kindreds with a rare autosomal dominant disorder called familial juvenile polyposis coli. Approximately 50% of pancreatic adenocarcinomas harbor inactivating Smad4 mutations. In addition to these genetic abnormalities, resistance to TGF-β growth arrest occurs in a significant percentage of colon and pancreas carcinomas because of the presence of mutant, activated

K-Ras, an oncogene that blocks Smad signaling. Restoration of TGF-β signaling may someday be a viable option in the treatment of gastrointestinal cancer.

See Also the Following Articles

Cancer, Overview • Epithelium, Proliferation of • Growth Factors

Further Reading

Barnard, J. A. (2003). Peptide growth factors. *In* "Gastrointestinal Cancers" (A. K. Rustgi and J. M. Crawford, eds.), 1st Ed., pp. 33–54. W. B. Saunders, Philadelphia, PA.

Blobe, G. C., Schiemann, W. P., and Lodish, H. F. (2000). Mechanisms of disease: Role of transforming growth factor beta in human disease. *N. Engl. J. Med.* **342**, 1350–1359.

Kim, S. J., Im, Y. H., Markowitz, S., and Bang, Y. J. (2000). Molecular mechanisms of inactivation of TGF-beta receptors during carcinogenesis. *Cytokine Growth Factor Rev.* **11**, 159–168.

Shu, H. J., and Burgess, A. W. (2001). Regulation of transforming growth factor β signaling. *Mol. Cell. Biol. Res. Commun.* **6**, 321–330.

Ten Dijke, P., Goumans, M. J., Itoh, F., and Itoh, S. (2002). Regulation of cell proliferation by Smad proteins. *J. Cell Physiol.* **191**, 1–16.

Transplantation Immunology

NICHOLAS L. TILNEY
Harvard University and Brigham and Women's Hospital

alloantigen Antigenic differences between members of the same species.
cytokines Peptide products of activated leukocytes.
HLA The symbol for the major histocompatibility complex in humans.
ligand A structure on one cell recognized by a receptor on another cell.
major histocompatibility complex A closely linked series of genes that account for the strong transplantation barrier between individuals.

Transplantation immunology is a subject that examines (1) the biology of tissue and organ grafts transferred from donor sources to genetically dissimilar recipients and (2) the activity and mechanisms of the multifaceted host defenses that react against the foreign stimulus. If the resultant panoply of antigen-specific cellular and humoral events is not modified or inhibited by X-radiation or chemical or biological means of immunosuppression, they will inexorably and irreversibly destroy the graft. Understanding of the intricacies of these responses has accelerated rapidly during recent decades, with elucidation of the functions, interplay, and effects of the cascade of adhesion molecules and other inflammatory mediators, leukocyte populations, their relationship to antigens presented to them, and the influence of a variety of effector products. Indeed, understanding of the process of "rejection" of foreign tissues has evolved from the simplistic concept that specifically sensitized host lymphocytes infiltrate and directly kill the transplant, to the more refined appreciation of a sequential series of cellular and molecular events of astounding complexity.

BACKGROUND

The basic doctrine of transplantation is that tissue transferred between different sites on the same individual (autografts), or between genetically identical twins (isografts), will heal and function normally. In contrast, transplants from members of the same species (allografts) or from different species (xenografts) will be rejected promptly by the host. The phenomenon of acute rejection was defined during World War II by gross and histological examination of skin grafts in humans and in animals. The involvement of immune mechanisms in the process was ascertained by the

Copyright 2004, Elsevier (USA). All rights reserved.

K-Ras, an oncogene that blocks Smad signaling. Restoration of TGF-β signaling may someday be a viable option in the treatment of gastrointestinal cancer.

See Also the Following Articles

Cancer, Overview • Epithelium, Proliferation of • Growth Factors

Further Reading

Barnard, J. A. (2003). Peptide growth factors. *In* "Gastrointestinal Cancers" (A. K. Rustgi and J. M. Crawford, eds.), 1st Ed., pp. 33–54. W. B. Saunders, Philadelphia, PA.

Blobe, G. C., Schiemann, W. P., and Lodish, H. F. (2000). Mechanisms of disease: Role of transforming growth factor beta in human disease. *N. Engl. J. Med.* **342**, 1350–1359.

Kim, S. J., Im, Y. H., Markowitz, S., and Bang, Y. J. (2000). Molecular mechanisms of inactivation of TGF-beta receptors during carcinogenesis. *Cytokine Growth Factor Rev.* **11**, 159–168.

Shu, H. J., and Burgess, A. W. (2001). Regulation of transforming growth factor β signaling. *Mol. Cell. Biol. Res. Commun.* **6**, 321–330.

Ten Dijke, P., Goumans, M. J., Itoh, F., and Itoh, S. (2002). Regulation of cell proliferation by Smad proteins. *J. Cell Physiol.* **191**, 1–16.

Transplantation Immunology

NICHOLAS L. TILNEY
Harvard University and Brigham and Women's Hospital

alloantigen Antigenic differences between members of the same species.
cytokines Peptide products of activated leukocytes.
HLA The symbol for the major histocompatibility complex in humans.
ligand A structure on one cell recognized by a receptor on another cell.
major histocompatibility complex A closely linked series of genes that account for the strong transplantation barrier between individuals.

Transplantation immunology is a subject that examines (1) the biology of tissue and organ grafts transferred from donor sources to genetically dissimilar recipients and (2) the activity and mechanisms of the multifaceted host defenses that react against the foreign stimulus. If the resultant panoply of antigen-specific cellular and humoral events is not modified or inhibited by X-radiation or chemical or biological means of immunosuppression, they will inexorably and irreversibly destroy the graft. Understanding of the intricacies of these responses has accelerated rapidly during recent decades, with elucidation of the functions, interplay, and effects of the cascade of adhesion molecules and other inflammatory mediators, leukocyte populations, their relationship to antigens presented to them, and the influence of a variety of effector products. Indeed, understanding of the process of "rejection" of foreign tissues has evolved from the simplistic concept that specifically sensitized host lymphocytes infiltrate and directly kill the transplant, to the more refined appreciation of a sequential series of cellular and molecular events of astounding complexity.

BACKGROUND

The basic doctrine of transplantation is that tissue transferred between different sites on the same individual (autografts), or between genetically identical twins (isografts), will heal and function normally. In contrast, transplants from members of the same species (allografts) or from different species (xenografts) will be rejected promptly by the host. The phenomenon of acute rejection was defined during World War II by gross and histological examination of skin grafts in humans and in animals. The involvement of immune mechanisms in the process was ascertained by the

Copyright 2004, Elsevier (USA). All rights reserved.

observation that a "second set" of grafts from the same donor was rejected at a faster tempo than the first via some type of "memory" intrinsic in the recipient. For the next two decades, however, the most that was known about these events was that increasing numbers of host leukocytes collecting in and around the allografts led inexplicably to their inevitable and irreversible destruction. Circulating lymphocytes were believed to be particularly critical, as evidence mounted that they were important "immunologically competent" components of host reactivity. The continuous recirculation of a substantial lymphocyte population among the blood, lymph, and somatic tissues allowed the cells to contact and respond to any antigenic stimulus in any location in the body.

It also became apparent by differential ablation of "central" lymphoid organs that there were two major classes of lymphocytes normally present in the body. The effects of removal of the thymus gland were first examined in neonatal mice. The majority of animals lost weight, wasted, and died of laboratory infections. Numbers of circulating lymphocytes profoundly decreased and peripheral lymphoid tissues, including spleen and lymph nodes, atrophied. Most significantly, skin allografts healed and grew hair normally in those that survived. Humoral activity was unaffected. As thymectomy of adult mice with already mature lymphoid tissues produces no such effects, it was concluded that the thymus was the central lymphoid organ critical in the early "schooling" of immature lymphocytes from the bone marrow. These cells become the primary mediators of cellular immunity in the body. Unaffected by thymectomy, the population responsible for antibody production was next defined. The bursa of Fabricius, a lymphocyte-rich blind pouch connected to the cloaca of birds, was found to have immunological function. Following its surgical ablation, young chicks but not adult chickens were unable to form antibody against antigens. However, as their cellular immune mechanisms remained intact, they could still reject skin grafts. Subsequent search for a "bursa equivalent" organ in mammals initially implicated the tonsil and appendix but finally settled on diffuse lymphoid aggregations in the gut and bone marrow.

Stimulated by these findings in the 1960s and 1970s, immunology, transplantation biology, and related sciences exploded. Increasing knowledge about cellular and humoral activity unleashed a torrent of investigations. Effector lymphocytes were found to require the influence of other subpopulations before they could attain full function, as illustrated by the phenomenon of "help" given by thymus-derived (T) lymphocytes to bone marrow- or bursa-derived (B) lymphocytes to produce antibody, and the effects of one T-cell subpopulation on the behavior of others via elaboration of specific products. Identification of histocompatibility antigens and definition of their critical role in the presentation of antigen at both the cellular and molecular levels unfolded. The introduction of hybridoma technology led to the creation of monoclonal antibodies that have allowed characterization of the inflammatory/immunological cascade involved in a variety of host defenses. Elaboration of receptor physiology, intracellular machinery, and T-cell interaction with the major histocompatibility complex (MHC) of foreign cells and tissues has opened entire fields in molecular biology, with further unraveling of the intricacies of host immuno-responsiveness.

ACUTE REJECTION

In overresponsive recipients or those in whom immunosuppression is relatively ineffectual, acute rejection of an organ may be a dramatic and frightening event. After several days of satisfactory function of a kidney transplant, for instance, levels of serum creatinine begin to rise. The patient often develops fevers and malaise. The graft, easily palpable through the lower abdominal wall, may enlarge and become tender. Occasionally, the tense and swollen organ ruptures, with significant bleeding into the retroperitoneal space.

This immunological destruction of an allograft involves a cascade of host-directed, lymphocyte-mediated events. Small numbers of naive lymphocytes become specifically sensitized by antigen-presenting cells in the graft. Some of these stay at the site; others circulate through host peripheral lymphoid tissues where they disseminate the antigenic message and attract newly activated cell populations to the foreign tissues. Progressive infiltration by host mononuclear cells is characteristic of acute rejection of all allogeneic organs. Within a few hours of revascularization, increasing numbers of T and B lymphocytes enter perivascular areas and then invade the graft substance itself over the next days or weeks. The primary responsibility of T cells in the acute destructive process has been emphasized *in vitro* by their ability to lyse donor target cells directly and *in vivo* by the inability of T-cell-depleted animals to reject allografted tissues. Allostimulated B lymphocytes differentiate into antibody-producing immunoblasts and plasma cells in graft and host lymphoid tissues within a few days of transplantation. Although these secrete both antigen-specific and nonspecific antibodies, their precise role in the acute process remains enigmatic. As inflammation proceeds, the increasingly obvious presence of macrophages is associated with progressive

destruction of pericapillary tissues, interstitial inflammation, and eventual tissue necrosis. Macrophages have several roles in graft destruction. They act as antigen-presenting cells to activate lymphocyte populations. They are a principal site of control by MHC immune response genes, mediated particularly by receptors on their plasma membranes for some histocompatibility antigens. They also elaborate cytokines when activated, particularly interleukin 1 (IL-1) and IL-6, which have specific effects on cells of the host and tissues of the graft.

HISTOCOMPATIBILITY ANTIGENS

Antigens on allogeneic cell surfaces allow the graft recipient to recognize that the transplanted tissue is not "self." It has long been realized that the immune responses between genetically dissimilar humans are directed against a single cluster of alloantigens, designated by the World Health Organization as HLA and encoded by MHC genes found on chromosome 6. There are two MHC antigen groups: class I (HLA-A and B) and class II (HLA-DR, -DQ, and -DP). HLA, -B, and -DR are the most important targets for host alloreactivity. These are expressed on various tissues, may be up- or down-regulated by host cells and their products, and may differentially influence graft rejection. Class I antigens are relatively ubiquitous and are constitutively expressed on all somatic cells. They interact exclusively with the cytotoxic/suppressor $CD8^+$ T-lymphocyte subpopulation. Class II antigens activate selectively the helper/inducer $CD4^+$ T-lymphocyte subpopulation. They are distributed more selectively throughout lymphoid tissues and are present on various antigen-presenting cells, circulating B lymphocytes, and monocytes. Up-regulation of these antigens on the vascular endothelium of transplanted organs by early injury such as ischemia/reperfusion become particularly important as the vascular endothelial cells are exposed continuously to circulating effector cells and their products. Particular leukocyte-derived factors can also stimulate MHC antigens differentially by inducing class II gene products or promoting class I antigens over class II. Interferon-γ (IFN-γ), for instance, up-regulates class II but not class I antigens on several cell types, increasing antigen presentation and amplifying graft immunogenicity.

ANTIGEN RECOGNITION

Recognition of graft "foreignness" by T lymphocytes requires both MHC molecules and alloantigen. Alloantigen is recognized by the T-cell receptor either "directly" as intact allo-MHC on the surface of donor cells or "indirectly" as processed peptides derived from donor MHC and presented by recipient MHC molecules in their peptide-binding groove on the surface of the antigen-presenting cells (Fig. 1). The recognition of a foreign peptide bound to a MHC molecule is the physiological pathway by which all microbial invaders are recognized and eliminated by T cells. Genetic rearrangements of specific molecules covalently linked by disulfide bonds produce combining sites on lymphocyte surfaces that can recognize virtually any antigen. Acute rejection is thought to be mediated primarily by direct allorecognition, whereas the indirect pathway appears to play a major role in chronic rejection.

Antigen recognition by T cells alone, however, is not sufficient to trigger host events. A two-signal process of lymphocyte activation involves binding of transplantation antigens to T-cell surface receptors (signal 1) coincidentally receiving a co-stimulatory signal 2 from adjacent accessory molecules (Fig. 2). These include a series of receptor−ligand interactions such as CD2−LFA-3 (cluster determinates 2−late functioning adhesion molecule) and LFA-1−intercellular adhesion molecule. Another particularly well-characterized combination is the T-cell surface molecule CD28 that binds to either of its ligands, B7-1 or B7-2, themselves expressed on the surface of bone marrow-derived antigen-presenting cells. The interaction between the T-cell receptor and the MHC alloantigen in the presence of appropriate "co-stimulatory" signals forming a multimolecular patch is recognized as the central event that initiates lymphocyte activation. This triggers further events that lead ultimately to immunological rejection.

CELL PRODUCTS

With understanding of the cell populations involved in host alloresponsiveness has come appreciation of the critical importance of their products, the cytokines and chemokines (Fig. 3). These molecules may activate or inhibit the function of other cell populations. Antigen-activated macrophages elaborate IL-1, for instance, a monokine that stimulates $CD4^+$ cells to elaborate a series of associated factors, and IL-6, a fibrosis-inducing factor associated with chronic rejection. Other cytokines released by macrophages and important in graft destruction include tumor necrosis factor α (TNFα) and IFN-γ. T-cell- and macrophage-associated chemokines include a series of small molecules such as RANTES (regulated upon activation, normal T-cell expressed and secreted) and monocyte chemotactic protein-1. These are chemoattractants that bring uncommitted macrophages to the site of inflammation.

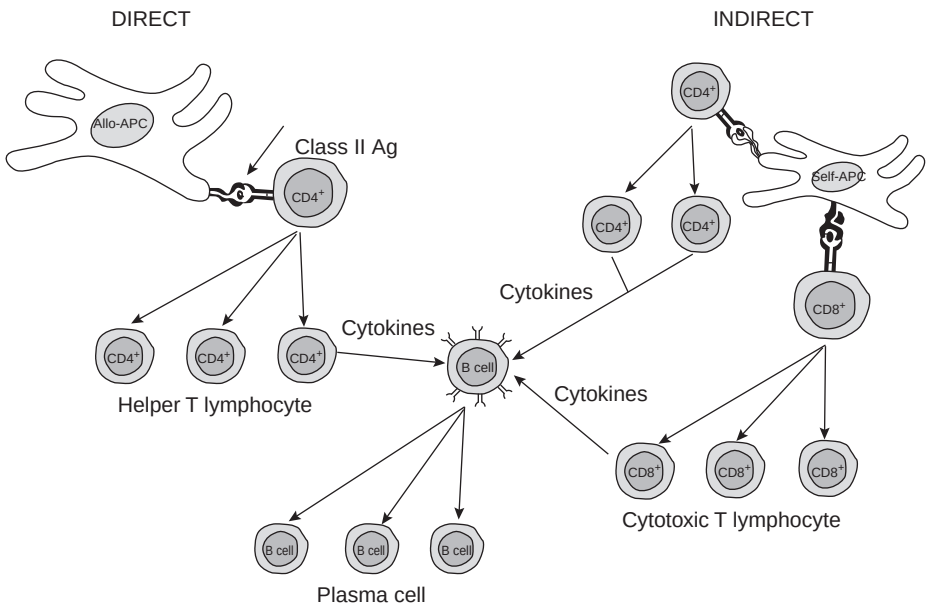


FIGURE 1 Mechanisms of recognition of alloantigen by the T-cell receptor are shown. "Direct" recognition involves intact allo-MHC on the surface of donor cells. "Indirect" involves processed donor MHC peptides presented in the antigen-binding site of recipient antigen-presenting cells.

T cells can be divided into two subclasses, T_H1 and T_H2, each of which elaborates distinct cytokines with their own properties and actions. One of the most important, T_H1-derived IL-2, stimulates activated T and B lymphocytes to differentiate and proliferate in both an autocrine and a paracrine fashion. Other cytokines encourage B-cell maturation or affect cell populations in the bone marrow. T_H1-derived IFN-γ has several effector roles including augmentation of the alloaggressiveness of macrophages previously uncommitted toward the foreign tissue, induction and amplification of MHC antigen expression on graft cells, and stimulation of B cells to increase antibody production. This factor may also increase lymphocyte adhesiveness to an antigen-presenting site by enhancing expression of surface LFA-1. Factors produced by T_H2 cells, primarily IL-4 and IL-10, in contrast, are inhibitory proteins to T_H1 cells and act to balance or modulate the activity of effector molecules. However, T_H2 cells can also increase the expression of B cells and immunoglobulin G production.

Antigen stimulation causes T lymphocytes to upregulate receptor expression on their surfaces that are specific for transferrin, insulin, IL-1, IL-2, and other cell products. The development of high-avidity surface receptors for IL-2 (IL-2R) on most activated $CD4^+$ and $CD8^+$ T lymphocytes, some B cells, dendritic cells, and macrophages, is particularly important in the rejection cascade. Binding of this cytokine to its receptor is followed by internalization of the entire complex,

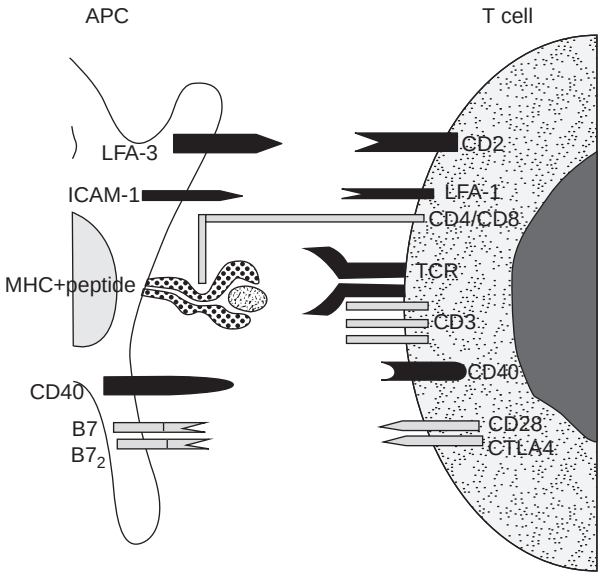


FIGURE 2 The two-signal process of lymphocyte activation is noted. Signal 1 involves the MHC antigen–T-cell receptor interaction. The co-stimulatory signal 2 includes a series of receptor–ligand interactions.

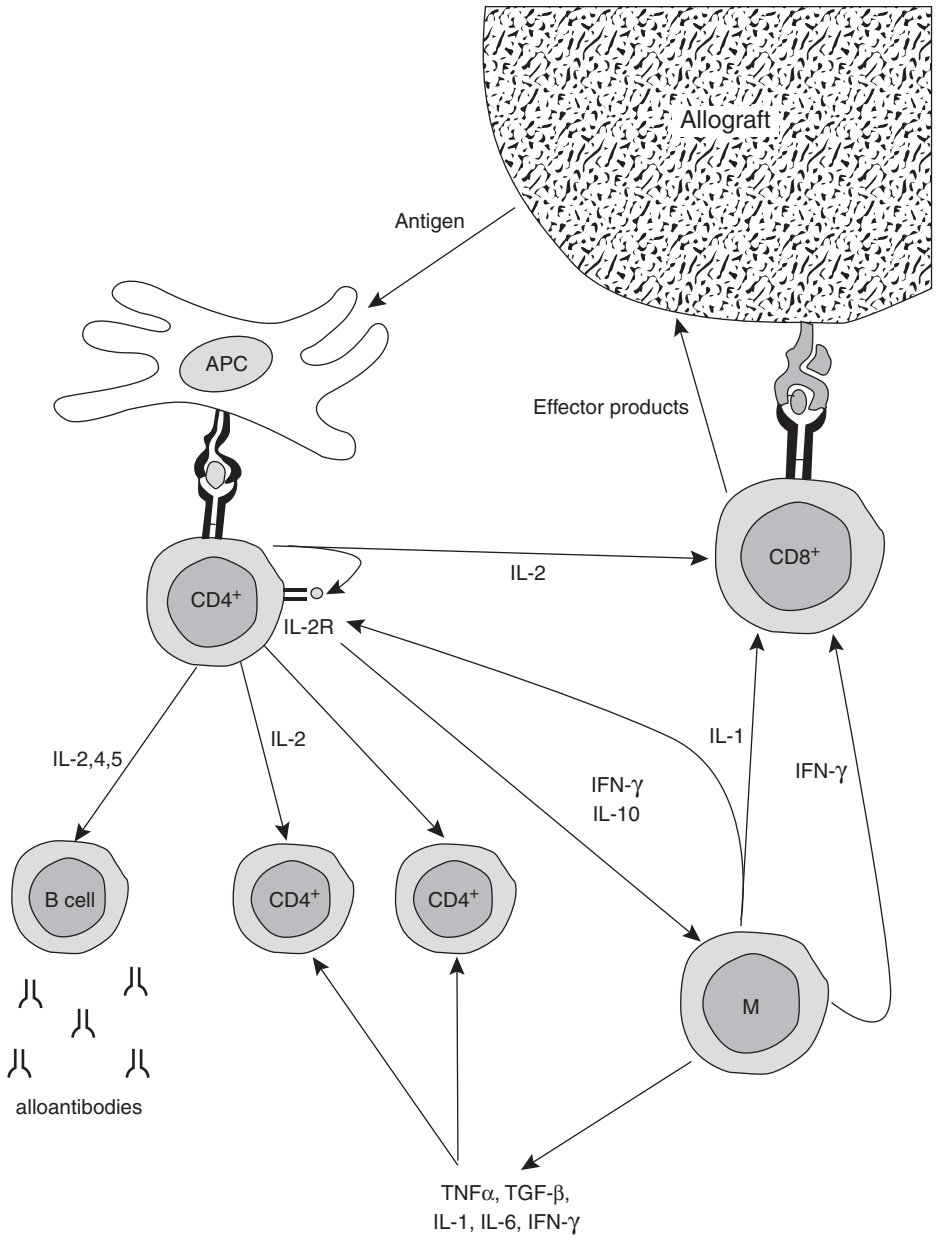


FIGURE 3 Activated lymphocytes and macrophages elaborate a series of cytokines and chemokines that influence the function of other cell populations.

transducing the signal for proliferation and clonal expansion of the antigen-activated cells and driving the entire rejection event forward. This important triggering population is relatively small; only approximately 15% of infiltrating cells in acutely rejecting rat cardiac grafts, for instance, are IL-2 receptor positive. The actual killing of foreign cells may occur via specific T-cell products, particularly granzyme B, a serine esterase protein, and perforin, a pore-forming lytic protein. These plus other effector molecules, in addition to actual cell to cell interactions, produce actual graft destruction.

CHRONIC REJECTION

The phenomenon of chronic rejection is characterized by gradual functional deterioration of a grafted organ associated with characteristic histopathological changes. Chronic rejection affects all solid organ transplants;

for instance, cadaver kidney transplants have a half-life of less than a decade despite progressive improvements in early results. The clinical manifestations of chronic kidney transplant rejection include the development and worsening of proteinuria and decline in other functional parameters. Morphologically, glomeruli become increasingly sclerotic, tubules become atrophic, and interstitial fibrosis increases progressively. The prognosis for heart and lungs is worse. Graft arteriosclerosis occurring in heart transplants and broncheolitis obliterans developing in lung transplants are characteristic chronic changes in those organs.

Several risk factors have been implicated. Alloantigen-dependent factors include histocompatibility differences between donor and recipient, the occurrence of early acute-rejection episodes, and continuing host alloresponsiveness not fully inhibited by routine immunosuppression. Alloantigen-independent factors include donor brain death, ischemia/reperfusion injury, and donor-associated factors such as age, diabetes, and hypertension. These early injuries or conditions predispose the organ to immunological injury and progressive later dysfunction.

The pathophysiology of the chronic process is multifactorial. A critical feature of the phenomenon, the compromise of vessels or other hollow structures, is conceptualized as stemming from recurrent injury to endothelial cells, leading to their persistent activation. Initial acute damage, regardless of etiology, may lead to up-regulation of adhesion molecules, antibody, and antigen–antibody complexes. A sequential pattern of adhesion molecule and cytokine expression has been associated experimentally with macrophage infiltration. This population may release cell mediators including endothelin, a potent vasopressor agent, and IL-1, IL-6, TNFα and transforming growth factor-β, all of which are fibrosis-inducing cytokines. Growth factors and other pro-inflammatory mediators increase the deposition of mesangial matrix. The resultant intimal proliferation, hypertrophy, and repair produce gradual lumenal obliteration.

The deleterious effect of chronic rejection on graft survival continues to reduce substantially much of the promise that transplantation can offer patients with organ failure.

Acknowledgments

This work was supported by U. S. Public Health Service Grants 5 RO1 1DK46190–28 and 1 PO1 AI140152–05.

See Also the Following Articles

Liver Transplantation • Lymphocytes • Pancreatic Transplantation • Short Bowel Syndrome and Intestinal Transplantation, Pediatric • Small Bowel Transplantation

Further Reading

Chertow, G. M., Milford, E. L., Mackenzie, H. S., *et al.* (1997). Antigen-independent determinants of cadaver kidney transplant failure. *J. Am. Med. Assoc.* **276**, 1732.

Häyry, P., von Willebrand, E., Parthenais, N., *et al.* (1984). The inflammatory mechanisms of allograft rejection. *Immunol. Rev.* **77**, 85–142.

Krensky, A. M., Weiss, A., Crabtree, G., *et al.* (1990). T-lymphocyte antigen interactions in transplant rejection. *N. Engl. J. Med.* **322**, 510.

Mason, D. W., and Morris, P. J. (1986). Effector mechanisms in allograft rejection. *Annu. Rev. Immunol.* **4**, 119.

Sayegh, M. H., and Turka, L. A. (1990). The role of T-cell costimulatory activation pathways in transplant rejection. *N. Engl. J. Med.* **338**, 1813.

Springer, T. (1994). Traffic signals for lymphocyte recirculation and leukocyte emigration. *Cell* **76**, 301.

Tilney, N. L., Whitley, W. D., Diamond, J. R., *et al.* (1991). Chronic rejection: An undefined conundrum. *Transplant* **52**, 389–398.

Trauma, Overview

DOUGLAS W. WILMORE
Harvard University and Brigham and Women's Hospital

blunt injury A mechanism of injury that deforms tissue on impact and causes tissue disruption when the elastic forces of the tissue are exceeded.
hypotension Low blood pressure.
multiorgan failure Malfunction progressing to complete failure of vital organs that occurs after shock and infection in a trauma patient.
penetrating injuries Caused primarily by stab wounds and bullets, these injuries are the result of cutting or lacerating tissue.
peritoneal lavage A technique to access hemorrhage into the abdominal cavity whereby a small tube is inserted into the space and fluid is instilled into the cavity and then removed to determine whether free blood is present.
shock A marked decrease in tissue perfusion usually associated with a fall in blood pressure that deprives vital tissues of oxygen.

Trauma refers to injuries that cause structural alterations and physiological imbalances in the body. Physical or chemical agents can damage tissue. In general terms, trauma refers to accidental injuries, although there are occasions, such as wars or suicide, when trauma is intentionally inflicted on an individual. Moreover, the surgeon who performs elective surgical procedures causes injury to tissue with the initial skin incision, yet the circumstances are such that the skin is free of bacteria, the scalpel is sharp, and blood loss is controlled, thus minimizing tissue injury and its subsequent effects. In contrast, accidental injury disrupts tissues and is associated with bacterial contamination and uncontrolled blood loss. The body responds to such tissue disruption by clotting the damaged blood vessels to control the hemorrhage and then initiating an inflammatory response to clear the dead tissue and kill bacteria that have contaminated the wound. Finally, major systemic changes, such as fever, tachycardia, and increased protein breakdown, occur to support this regional inflammatory process and to enhance wound healing.

SCOPE OF THE PROBLEM

Unintentional deaths due to trauma are a major cause of mortality and disability in the United States, occurring primarily in young active adults and children. In the year 2000, 97,300 individuals died due to accidents; a fatal injury occurred every 5 min and a disabling injury occurred every 1.5 s. The National Safety Council estimates that the cost of these injuries, including lost wages, medical expenses, property damage, employer costs, fire and other loss, was $512.4 billion in 2000, which equaled approximately $5000 for each U.S. household. The leading causes of unintentional-injury deaths in the United States are shown in Table I.

The major cause of accidental injuries in the United States is motor vehicle crashes. There is a death every 12 min and a disabling injury every 14 s due to automobile accidents. The age group most affected ranges from 15 to 24 years but the very old (>75 years) are also greatly affected. Motor vehicle accidents are now the major cause of death for teenagers. More than one-third of the accidents involved an intoxicated or alcohol-impaired driver or nonmotorist. The increased use of recreational drugs also contributes to this high vehicle-related accident rate.

There were over 5000 deaths in the workplace due to unintentional injuries and almost 4 million workers suffered disabling accidents. This accounts for a cost to Americans of $131.2 billion, which exceeds the combined profits of the top 13 companies in the Fortune 500.

Approximately 30,000 fatalities and over 7,000,000 disabling injuries occurred in the home. There is a fatal injury every 18 min in the home and a disabling injury every 4 s. The leading causes of these fatal events are poisonings, falls, fires and burns, and suffocation.

There were approximately 22,000 fatalities in the community caused by falls, drowning, and transportation accidents. More than one-half of the 750 deaths

TABLE I Leading Causes of Unintentional-Injury Deaths in the United States, 1998

Motor vehicle accidents	43,501
Falls	16,274
Poisoning	10,255
Drowning	4,406
Choking	3,515

 Copyright 2004, Elsevier (USA). All rights reserved.

resulting from boating accidents could have been prevented if the victim had been wearing a life jacket.

MECHANISMS AND PATTERNS OF INJURY

There are two main types of mechanical injury: blunt and penetrating. The mechanisms and the patterns of injury that are sustained by the body are unique to the type of injury and determine in large part the treatment and outcome.

Blunt injury occurs in automobile accidents, falls, or assaults with blunt objects. The injury caused is a deforming injury related to the force at impact that alters tissue shape by compression. With such force, body tissues stretch and when this elastic capacity is exceeded there is tissue disruption. Blunt injuries damage the skin (abrasions, contusions, and lacerations), cause bone fractures, and result in visceral or vascular rupture. The specific body part injured depends on the magnitude of the force, but also the fixation of the specific tissue and its strength. For example, bones such as the first two ribs, scapula, pelvis, or femur are extremely durable and these areas are fractured only when extreme forces are exerted on these specific areas. Internal organs rupture at points of fixation such the falciform ligament of the liver, the ligament of Trietz that holds the small bowel, or the retroperitoneal cecum. Internal injuries also occur because of close proximity of organs to rigid external structures, the best example being the close proximity of the brain to the skull.

The four common injuries observed in unrestrained drivers are fracture of the femur, fracture of the pelvis, injury to the thoracic cage, and fracture of facial bones. The common pattern of injury seen in a pedestrian struck by a motor vehicle is fracture of the upper third of the fibula and tibia. Injuries to the ribs and abdominal viscera occur if the individual is thrown onto the car hood. As the individual falls to the ground, head and facial injuries may occur.

Penetrating injuries are primarily caused by stab wounds and bullets. The energy involved with a stab wound is relatively low and hence these wounds are lethal only if a vital organ, such as the heart, is penetrated. Missile wounds penetrate by cutting or lacerating tissue. In addition, depending on the bullet type and the force exerted, there may be extensive tissue damage due to cavitation. The injury pattern in patients with penetrating injuries is much more random and depends on the location of the penetration and the cause of the injury.

PREVENTION

Clearly, the success to the reduction of accidental injuries is prevention. Because automobile accidents account for approximately one-half of the deaths due to trauma, there is increasing pressure to utilize preventative measures to reduce vehicle accidents. In fact, deaths from this cause are decreasing. The public health measures contributing to this success include improved design of automobiles and equipment, the mandatory use of restraint devices, reduction in highway speed, and campaigns against individuals who drink and drive.

In the area of penetrating injuries, prevention has a role by controlling gun use and gun availability. This, however, is an extremely controversial issue, but it is quite clear that countries that have gun restrictions have a much lower injury rate related to missiles than the United States.

Prevention plays a major role in injury reduction in the workplace. The high cost of such accidents more than justifies preventative programs and individuals skilled in this area are frequently employed by large companies to participate in program planning and education. Annual reports of accidents are often mandated by governmental agencies.

Accident prevention in the home is more difficult to control and occasionally is related to the manufacture of unsafe products used in the home and the sale of toys that are potentially dangerous. Consumer pressures work to improve product design in this area.

CONSEQUENCES OF INJURY

Shock

One of the first problems associated with injury is blood loss or hemorrhage. This may be visible, associated with loss of or damage to skin or a fracture of a long bone that is compounded and protruding through skin and soft tissue. However, much bleeding is internal with loss accumulating into cavities of the chest or thorax or into soft tissue, causing marked swelling.

The volume of blood in an adult represents approximately 7–8% of the body's weight and blood pressure is usually well maintained with losses up to 25% of this amount. However, with additional bleeding, hypotension, e.g., a low blood pressure, occurs.

Shock refers to this situation when blood flow and pressure are inadequate to sufficiently oxygenate essential tissues of the body. As a result, the normal process that produce high-energy compounds, such as ATP, which serve as fuel for cells, is impaired and cell function is depressed. Acid accumulates within the body

and this acidemic environment further limits cardiac, renal, pulmonary, and brain function. The hemorrhage must be controlled and the blood volume restored. The latter is accomplished by starting an intravenous infusion of a salt-containing solution that temporarily expands the blood volume and raises the blood pressure. With ongoing loss of blood, the transfusion of red cells is required, for these are the cells that provided for the oxygen-carrying capacity of blood. The type of shock associated with blood loss is referred to as hypovolemic shock.

There are several other causes of shock in the injured patient. Cardiac contusion or cardiac tamponade may result in a falling blood pressure because the heart has limited ability to pump blood throughout the body. This situation is referred to as cardiogenic shock. There are occasions when an injury to the nervous system occurs and the blood pressure falls. The usual injury is associated with spinal contusion or transection of the cord and these situations result in dilation of blood vessels that have been interrupted from their nervous connections.

Systemic Injury Responses

Trauma produces a fairly predictable and reproducible set of systemic responses. Initially, injury responses were said to follow an ebb and flow course; that is, metabolic and physiologic processes were generally slowed during the shock period and accelerated with fluid resuscitation and restoration of blood volume. With debridement and repair of the injured site, the body increases its responses, presumably in an effort to accelerate wound healing. The clinical manifestations of this response are fever, tachycardia, and increased ventilation. The body's metabolism is altered and there is increased oxidation of fat and the accelerated breakdown of protein, presumably to provide fuel and amino acids for the healing wound. In short, the patient's own tissue is broken down to ensure tissue repair. With mild to moderate injuries, these catabolic responses cause minimal debility, but with more extensive injuries these responses cause marked weight loss and erosion of muscle tissue. This can be counterbalanced in part by aggressive nutritional support and exercise. With wound healing and/or resolution of the injury, the processes abate and the individual starts to regain weight and strength.

Infection and Organ Failure

Skin is an important barrier to microorganisms and its integrity is paramount to defense against infection. With trauma to this important barrier, the individual becomes susceptible to invasive infections. This situation is compounded by the fact that many injuries occur in environments where the wounds are contaminated by dirt and other debris. With stabilization of the patient, the wound must be washed and cleaned in order to reduce the bacterial load and achieve a situation where host defenses can handle the number of bacteria that are present. Antibiotics are administered intravenously and tetanus toxoid is also given to enhance antibody production to the organisms that are responsible for this potentially lethal infection. Other infections may be expected, particularly in the abdomen if rupture of an abdominal viscera has occurred. If the patient requires long-term mechanical ventilation, associated pneumonia is also a possibility.

With the occurrence of shock and associated infection, there is a high probability that associated organ failure may occur. This syndrome is manifest as failure of one or more organs including the kidneys, liver, coagulation system, lungs, and central nervous system. The cause of this sequence of organ failure in not known, but it is hypothesized that products of the inflammatory response overwhelm the body and cause organ damage. Sequential organ failure is now the most common cause of late deaths in the patient who sustains traumatic injuries and the only available therapy is supportive care.

Site-Specific Injuries and Their Consequences

Head Injuries

Injuries to the head are frequent, disabling, and potentially life-threatening because of the delicate nature of the brain. Although emergency measures can be initiated in the field where the injury occurred, careful rapid evacuation of the patient to a hospital's emergency facility is essential for appropriate diagnosis and treatment.

Physicians assess the extent of the patients brain injury using a scoring system based on physical signs and symptoms. This scoring system aids prediction of the mortality and morbidity related to the injury. For example, 10% of individuals with a score of 14–15 will have some chronic disability related to their injury and almost all those with a score of 8 or lower will have disabilities and approximately 30% will die.

Once the patient has stabilized, a computed tomography (CT) scan is indicated, and if necessary, an emergency operation is performed. These procedures aim to control hemorrhage and evacuate blood clots within the brain. In addition, by opening the bones of the skull, the pressure within the cranium is reduced. Because the skull is a closed container, brain swelling

results in increased intracranial pressure, which at some point causes damage to the central nervous system. This pressure is carefully monitored and can be controlled by withdrawing fluid from compartments within the brain or by controlling blood pressure and the amount of intravenous fluid administered to a patient. This type of care is best performed in an intensive care unit, particularly those with individuals trained in neurosurgical care.

Extremity Injuries

These injuries involve damage to the skin and soft tissue, injuries to blood vessels and fractures to the long bones. All of these injuries may occur in the same patient and be apparent in the same extremity, and such situations are extremely challenging to the surgical team. With the development of refined techniques in vascular and peripheral nerve surgery, transected limbs or digits can be reimplanted with a reasonably good chance of rehabilitation.

Injuries to the Abdomen

Injuries to the abdomen vary greatly, depending on the mechanism of injury and the force involved. For example, approximately 30% of stab wounds to the abdomen fail to fully penetrate the subcutaneous adipose tissue and do not enter the abdominal cavity. The initial role of medical providers following abdominal injury is to assess the extent of intra-abdominal hemorrhage. This is accomplished by performing a procedure referred to as peritoneal lavage. This involves placing a small tube in the peritoneal cavity, instilling liquid through the catheter, and then withdrawing the fluid. The sample can then be examined to assess the amount of blood present and a decision can be made as to whether an operation is indicated to control the hemorrhage. An alternate diagnostic approach is to use ultrasound to detect the presence of intra-abdominal fluid. A CT scan is then performed to further localize the site of injury.

Extensive injuries to the abdomen often involve destruction of portions of the liver or the spleen and associated damage to other organs. These traumatic injuries are devastating, for blood loss is massive and tissue damage is extensive and highly lethal. Surgeons often perform several sequential operations in an attempt to salvage these patients. First, the bleeding is stopped and bacterial contamination controlled. Six to 12 h after the initial operation, when the patient has returned to a more normal physiological state, the individual is taken back to the operating room and more permanent repair of the damaged tissue is undertaken. Several additional operations may be necessary before restoration of the damaged organs is satisfactory. This approach is referred to as damage control surgery.

Burn Injury

Fires and explosions cause cutaneous burns, which are one of the most extensive injuries that can be sustained. Burns are graded by their depth; a second-degree burn will regenerate skin, whereas a third-degree injury will require skin grafting. The extent of burn over the body is expressed as a percentage of body surface area. A 60% total body surface burn means that 60% of the entire body surface is injured. Both morbidity and mortality are associated with burn size. However, the injury may be more severe with involvement of the hands, face, and genitalia and the association of inhalation of smoke, which causes injuries to the lungs.

Burn patients are cared for in several sequential phases. First, the individual must be resuscitated with intravenous fluid to replace the fluid that leaks out through the area of burn. After several days when the patient has stabilized, the burned skin can be surgically removed and skin grafts applied to these areas. This is performed by taking very thin sections of skin from uninjured areas. The harvested skin is then placed on the site where the burn occurred. The donor site may be reharvested in approximately 2 weeks, for the surface area regenerates in 7–10 days. In patients with large burns covering 70–90% of their body, this procedure of grafting is quite slow and may require 1–3 months of hospitalization. The final stage is rehabilitation, which is required to aid in ambulation, eating, and other aspects of self-care.

See Also the Following Articles

Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Occult Gastrointestinal Bleeding • Upper Gastrointestinal Bleeding • Variceal Bleeding

Further Reading

American College of Surgeons Committee on Trauma (1997). "Advanced Trauma Life Support for Doctors," 6th Ed. American College of Surgeons, Chicago.

Block, E. F. J. (1999). Diagnostic modalities in acute trauma. *New Horizons* **7**, 10–25.

Feliciano, D. V. (1991). Diagnostic modalities in abdominal trauma: Peritoneal lavage, ultrasonography, computed tomography scanning, and arteriography. *Surg. Clin. North Am.* **71**, 241–256.

National Safety Council (2001). "Report on Injuries in America." Available at http://www.nsc.org/library/rept2000.htm.

Velmahos, G. C., Tatevossian, R., and Demetriades, D. (1999). The "seat belt mark" sign: A call for increased vigilance among physicians treating victims of motor vehicle accidents. *Am. J. Surg.* **65**, 181–185.

Traveler's Diarrhea

HERBERT L. DUPONT
St. Luke's Episcopal Hospital, University of Texas, and Baylor College of Medicine, Houston

diarrheagenic *Escherichia coli* A group of *Escherichia coli* strains that cause diarrhea; two strains are important causes of diarrhea in international travelers to tropical and semitropical areas: enterotoxigenic *E. coli* and enteroaggregative *E. coli*.

rifaximin An unlicensed and poorly absorbed rifamycin antimicrobial agent with activity against enteric bacterial pathogens and effective in reducing the duration of diarrhea when used therapeutically.

traveler's diarrhea Diarrhea that occurs when a person leaves his or her home region, consuming meals at foreign eating establishments, away from local medical care.

When persons leave their home region they may experience diarrhea, with rates depending on the type of food and beverages consumed, the place where they were consumed, and the level of hygiene in the country of origin and in the destination. Most cases of diarrhea that occur in high-risk regions of tropical and semitropical countries are due to bacterial enteropathogens found in food and, to a lesser degree, in beverages. Treatment is directed to fluid and salt replacement, diet to facilitate intestinal repair, drugs to improve symptoms, and antimicrobial therapy to eliminate the causative agent and cure the illness. Vaccines are being developed to prevent infection with the most common pathogen, enterotoxigenic *Escherichia coli*.

INTRODUCTION

Diarrhea has historically been important in migrating groups when moving from one region of the world to another. It has played a major role in the outcome of many wars up to the 19th century. It has been known for some time that rates of diarrheal illness were reduced as populations remained in a region of risk, suggesting the occurrence of natural immunity. With the growing importance of international travel for pleasure and business, the occurrence and severity of traveler's diarrhea have become a major public health issue. In the past 20 years, a great deal has been learned about the incidence, sources, prevention, and therapy of this illness. A commonly used definition of traveler's diarrhea is the passage of three or more unformed stools per day, together with one or more additional symptoms or signs of enteric infection including nausea, vomiting, fever, abdominal pain or cramps, tenesmus, or fecal urgency, after leaving a home region to travel to a distant location. Illness that occurs shortly after returning home from a trip qualifies when the expected incubation period of enteric infections of 1 to 7 days is considered.

INCIDENCE

Each year more than 600 million persons cross international boundaries. Of these, more than 50 million travelers venture into tropical and semitropical regions with reduced hygiene levels and put themselves at risk for the development of diarrhea. The world can be divided into three regions based on the risk of acquiring illness by international visitors. Low-risk areas include the United States, Canada, northwestern Europe, Japan, Australia, and New Zealand, with diarrhea in international travelers visiting low-risk regions occurring at a rate of 2–4%. The high-risk areas include Latin America, southern Asia, and most parts of Africa. The risk of diarrhea when visiting these areas from a low-risk country is approximately 40%. There are areas within these high-risk regions where the rate of illness is lower among travelers than would be expected, such as Costa Rica and southern Africa. Countries in the northern Mediterranean, Caribbean countries, Russia, and China represent the third category of intermediate-risk regions where illness rates for diarrhea among international visitors range from 10 to 20%. Again, a homogenous risk is not seen for travelers to these regions. Within the Caribbean, rates are higher on certain islands (e.g., Jamaica) than others (British and U. S. Virgin Islands). The background rate of traveler's diarrhea of 2–4%, which is seen in all travelers, may relate to the frequency of meals eaten out of the home and possibly to psychic stress and increased alcohol consumption seen in many travelers. This 2–4% rate of illness is also seen among persons from high-risk regions traveling to low-risk countries (e.g., Mexicans traveling to the United States).

 Copyright 2004, Elsevier (USA). All rights reserved.

RESERVOIRS AND ETIOLOGY

In high-risk and moderate-risk areas, food- and public-hygiene standards are low, explaining the common exposure to diarrhea-producing microbes. The major source of enteric infection in high-risk urban areas is contaminated food. Growing crops in pathogen-enriched soil, failure to wash food items after harvesting, improper handling and cross-contamination within kitchens, and keeping foods at room temperature between meal services all contribute to widespread microbial contamination of food. Beverages may become contaminated because of inadequate water treatment and by groundwater contamination of water supplies during periods of heavy rainfall.

The major pathogens of traveler's diarrhea are bacterial pathogens, which explain up to 80% of illness. The most important bacterial pathogens occurring worldwide among international travelers are enterotoxigenic *Escherichia coli* (ETEC) and enteroaggregative *E. coli* (EAEC), together explaining approximately half of the illnesses seen. ETEC produces a secretory diarrhea as a result of elaboration of toxins and EAEC produces diarrhea secondary to an inflammatory, cytokine-mediated process. Other bacterial pathogens include *Campylobacter jejuni*, which is particularly important in travelers to Asia, *Shigella* spp., *Salmonella* spp., *Aeromonas hydrophila*, *Plesiomonas shigelloides*, and noncholera *Vibrio* spp. Viruses, including rotavirus and Norwalk-like calciviruses, explain approximately 10% of illness of travelers, and parasitic agents (*Giardia* spp., *Cryptosporidium* spp., *Cyclospora cayetanensis*, *Microsporidium* spp., and *Entamoeba histolytica*) are responsible for approximately 2% of illness in international visitors to high-risk areas. Parasitic pathogens are important in certain regions of the world. *Giardia* and *Cryptosporidium* are commonly seen in travelers to Russia (particularly St. Petersburg). *Giardia* and *Cyclospora* are commonly found in travelers to Nepal.

CLINICAL FEATURES

Most patients with traveler's diarrhea (~75%) experience the passage of watery stools, without significant fever, with untreated illness lasting between 5 and 7 days. Abdominal cramps and pain can be disabling. All the enteric pathogens mentioned above may produce this syndrome. In approximately 3% of travelers with diarrhea, febrile dysentery (passage of bloody stools) is the clinical expression of illness and in these cases, strains of *Shigella*, *Campylobacter*, *Salmonella*, noncholera *Vibrio*, and *Aeromonas* may be responsible. In approximately 10% of cases, vomiting is the major clinical expression of the illness where viruses (Norwalk-like calciviruses or rotaviruses) or preformed toxins of *Staphylococcus* or *Bacillus cereus* may explain the illness. One-fifth of patients with traveler's diarrhea will be confined to bed, incapacitated with their illness.

In approximately 2% of travelers with diarrhea, symptoms will persist for longer than 1 month, with a good percentage of these persons remaining ill for 6 months or longer. These patients with persistent and chronic diarrhea may develop an illness compatible with irritable bowel syndrome or Brainerd diarrhea and, in a small percentage, inflammatory bowel syndrome might be precipitated or unmasked. *C. jejuni* has been identified as one cause of chronic diarrhea in travelers, with symptoms resembling those of irritable bowel syndrome.

PREVENTION

Food and Beverage Selection

It may be possible to reduce rates of diarrhea by selection of safer items to eat and drink. Food and beverage contamination by microbes is reduced or eliminated by heating the item to 59°C. Practically, this is seen for foods that are too hot to touch (steaming hot foods, coffee, tea, and soup). It is a mistake to assume that cooked foods are safe. Hamburgers, which are invariably cooked, often are unsafe because the cooking has taken place at an earlier time before consumption and recontamination has occurred. Foods that are dry (bread), those that have been peeled (fruits), and items that have high sugar content (syrup and jelly) are generally safe. Bottled drinks can be considered safe, particularly if carbonated.

Chemoprophylaxis

Bismuth subsalicylate can be taken to prevent traveler's diarrhea. Taking two tablets with meals and at bedtime (eight tablets per day) during periods of risk will prevent 65% of the illness that would otherwise occur. Antimicrobial agents may be more effective, with rates of protection reaching 90%. The quinolones are currently the most used outside of areas where quinolone-resistant *Campylobacter* is common, such as Thailand, where one tablet (400 mg norfloxacin, 500 mg ciprofloxacin, or 500 mg levofloxacin) is taken daily during the period of risk.

Immunoprophylaxis

Vaccines that would prevent infection with the most important cause of traveler's diarrhea, enterotoxigenic *E. coli*, are being developed. An ETEC

whole-cell-binding subunit of cholera toxin, related to ETEC heat-labile enterotoxin (LT), given as an oral vaccine on two occasions, has been found to be ~80% effective in preventing ETEC diarrhea, provided the strains to which the patient is subsequently exposed produce LT or colonization factor antigens contained in the vaccine. Research groups are working on a multivalent vaccine to prevent traveler's diarrhea, considering the multitude of causative agents involved.

TREATMENT

Fluids and Electrolytes

As is the case in all forms of diarrhea, fluid and electrolyte therapy is fundamental to treatment. In most cases, encouraging travelers with diarrhea to take fluids (soft drinks and soups) and salt (saltine crackers and soups) is all that is needed to ensure hydration in these otherwise healthy persons. Special attention to oral rehydration should occur with infants, the elderly, and the infirm when diarrhea occurs during travel.

Symptomatic Treatment

Drugs that treat symptoms may be used to keep travelers healthy enough to make excursions and to function while out of town. Of this group of agents, loperamide is the most active and will reduce the symptoms of diarrhea by 60%. Newer antisecretory agents, which act more physiologically to reduce the loss of fluids and salt, appear to have a bright future in this area. Calmodulin-inhibiting drugs (e.g., zaldaride), a chloride channel-blocking drug (SP 303), and an enkephalinase inhibitor (racecadotril) have all been shown to reduce the passage of unformed stools in patients with acute diarrhea. They may not be quite as effective as loperamide, reducing diarrhea by only approximately 40%, but their more physiologic effect supports their development and general use.

Antimicrobial Agents

Since most traveler's diarrhea cases are due to infection by bacterial agents, it is not surprising that antibacterial drugs remain a mainstay of therapy. Resistance to trimethoprim–sulfamethoxazole, ampicillin, and doxycycline is currently high, rendering these agents of limited value. The quinolones are the current treatment of choice for traveler's diarrhea in adults in most regions of the world. The doses are norfloxacin 400 mg bid, ciprofloxacin 500 mg bid, or levofloxacin 500 mg qd, for 1 to 3 days, depending on the clinical response. In Thailand, quinolone-resistant *Campylobacter* is a major pathogen. In this locale and for children with severe traveler's diarrhea, azithromycin is probably the preferred drug. The dose of azithromycin for adults is 500 mg once a day for 1 to 3 days, depending on response. For children, 5–10 mg/kg/day for 1 to 3 days for more severe cases of traveler's diarrhea is suggested. A newly developed treatment for this illness is rifaximin, an unlicensed and poorly absorbed rifamycin derivative, which has been shown to be as effective as ciprofloxacin. The dose of rifaximin recommended is 200 mg tid or 400 mg bid for 3 days. For children 2–6 years of age, rifaximin may be given in an oral suspension at a dose of 100 to 200 mg bid and for those 6–12 years of age the recommended dose is 200 to 400 mg bid.

See Also the Following Articles

Anti-Diarrheal Drugs • Bacterial Toxins • *Campylobacter* • Diarrhea • Diarrhea, Infectious • Foodborne Diseases • Food Poisoning • Food Safety • Parasitic Disease, Overview

Further Reading

DiCesare, D., DuPont, H. L., Mathewson, J. J., *et al.* (2002). A double-blind, randomized, placebo-controlled study of SP 303 (Provir) in the symptomatic treatment of acute diarrhea among travelers to Jamaica and Mexico. *Am. J. Gastroenterol.* **97**, 2585–2588.

DuPont, H. L. (2002). Traveler's diarrhea. *In* "Infections of the Gastrointestinal Tract" (M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant, eds.), 2nd Ed., Chap. 19, pp. 253–265. Lippincott Williams & Wilkins, Philadelphia, PA.

DuPont, H. L., and Ericsson, C. D. (1993). Prevention and treatment of traveler's diarrhea. *N. Engl. J. Med.* **328**, 1821–1827.

DuPont, H. L., and Steffen, R. (2001). "Textbook of Travel Medicine and Health," 2nd Ed., B. C. Decker, Hamilton, Ontario, Canada.

DuPont, H. L., Jiang, Z.-D., Ericsson, C. D., *et al.* (2001). Rifaximin versus ciprofloxacin for the treatment of traveler's diarrhea: A randomized, double-blind clinical trial. *Clin. Infect. Dis.* **33**, 1807–1815.

Ericsson, C. D., DuPont, H. L., and Steffen, R. (2003). "Textbook on Travelers' Diarrhea." B. C. Decker, Hamilton, Ontario, Canada.

Gorbach, S. L., and Edelman, R. (eds.) (1986). Travelers' diarrhea: National Institutes of Health Consensus Development Conference. *Rev. Infect. Dis.* **8**, S109–S233.

Kean, B. H. (1963). The diarrhea of travelers to Mexico: Summary of five-year study. *Ann. Intern. Med.* **59**, 605–614.

Steffen, R., van der Linde, F., Gyr, K., and Schar, M. (1983). Epidemiology of diarrhea of travelers. *J. Am. Med. Assoc.* **249**, 1176–1180.

Trematodes

BRIAN F. JONES AND MICHAEL CAPPELLO
Yale University School of Medicine

cercaria Juvenile trematode stage that is shed from an infected snail.
definitive host Species in which the adult parasite develops and produces eggs.
digenean Having two hosts.
metacercaria Encysted trematode stage; life cycle continues following consumption by definitive host.
miracidium Ciliated larval trematode stage that is released from the egg.
monogenean Having one host.
operculum Hinged portion of a trematode egg through which the larva hatches.

Parasitic trematodes (flukes) consist of monogenean and digenean orders, which require one or more intermediate hosts, respectively, to complete their life cycle. Only the digenean parasites are of medical importance to humans. The major parasitic trematodes of humans include bloodborne, intestinal, and tissue-dwelling forms. Gastrointestinal symptoms, including abdominal pain, diarrhea, and vomiting, may be accompanied by systemic symptoms.

TREMATODE LIFE CYCLE AND BIOLOGY

The digenean (two-host) life cycle of a parasitic trematode typically consists of a vertebrate primary host, in which sexual reproduction of the parasite occurs, and an intermediate host, typically an aquatic snail, in which the parasite reproduces asexually. This asexual reproduction produces cercariae, which are motile forms of the worm. After being shed by the intermediate host, cercariae infect the definitive host by percutaneous penetration, or encyst on aquatic plants until consumption by the vertebrate host allows further development. The trematode life cycle is completed when eggs shed by adult worms are excreted in host feces and hatch to release ciliated miracidia, which then infect a suitable intermediate host.

Structurally, trematodes are flat and elongated worms whose outer surface (tegument) contains microvilli that both protect the worm and act as a nutrient absorptive surface. Adult worms possess anterior and ventral suckers, which are useful in maintaining attachment to host tissue. Trematodes posses a blind intestine that originates from the anterior sucker and provides additional absorptive capacity. Insoluble intestinal contents are regurgitated through the apical sucker, whereas liquid waste may be expelled through specialized excretory cells (flame cells).

Diagnosis of trematode infection is commonly accomplished by identification of eggs in feces or urine. The eggs of most species are structurally distinct and the diagnosis can be made by standard light microscopy. With the exception of fascioliasis, the drug of choice for trematode infections is praziquantel.

BLOOD TREMATODES (SCHISTOSOMES)

Schistosomes are major parasites of humans; an estimated 200–300 million people worldwide are currently infected. The predominant species of *Schistosoma* that infect humans include *Schistosoma mansoni*, found in South America, Africa, and Caribbean islands; *Schistosoma japonicum*, which occurs in Southeast Asia and Japan; and *Schistosoma haematobium*, found in Africa, India, and the Middle East. Infection occurs when tailed cercariae, aquatic motile forms of the parasite shed from an infected snail, contact and penetrate the skin of a suitable mammalian host. After shedding their tail, the cercariae transform into the first parasitic stage, the schistosomula. The schistosomula enter the bloodstream and migrate via the circulatory system through the heart and lungs, ultimately reaching the mesenteric veins (*S. mansoni* and *S. japonicum*) or the veins draining the bladder (*S. haematobium*). The males and females mate, producing eggs that are released into the bloodstream and the intestine or bladder, depending on the infecting species.

Acute schistosome infection may be associated with a variety of systemic symptoms, including fever, chills, cough, abdominal pain, vomiting, and urticaria. These symptoms generally occur 4–8 weeks after infection, and coincide with the emergence of antischistosomal antibodies. When associated with *S. japonicum* infection, this acute syndrome is referred to as Katayama

 Copyright 2004, Elsevier (USA). All rights reserved.

fever. Chronic schistosomiasis results primarily from the host immune response directed against parasite eggs. The eggs of *S. mansoni* and *S. japonicum* are shed into the mesenteric venules; some of the eggs migrate to the small intestine, where they may cause severe intestinal irritation. Many eggs are carried by the mesenteric circulation to the liver, where they elicit an intense, immunoglobulin E (IgE)-mediated inflammatory response. Chronic intrahepatic inflammation leads to granuloma formation and ultimately scarring and fibrosis. These changes eventually cause cirrhosis of the liver, with portal hypertension, ascites, formation of varices, and splenomegaly. Evidence also suggests that the risk of hepatocellular carcinoma is increased with chronic schistosomiasis. Because *S. haematobium*, unlike *S. mansoni* and *S. japonicum*, resides within the venous plexus surrounding the urinary bladder, eggs accumulate in bladder tissue and are excreted in the urine. Pain and hematuria are common symptoms of infection with *S. haematobium*, with chronic infection associated with an increased risk of squamous cell carcinoma of the bladder.

The diagnosis of schistosomiasis is made by identifying parasite eggs in the stool (*S. mansoni*, *S. japonicum*) or urine (*S. haematobium*) (Fig. 1). Because egg excretion may be sporadic, concentrating techniques may be required. Serologic testing is available through the United States Centers for Disease Control and Prevention, although the tests are not routinely utilized in clinical practice.

The treatment of *S. mansoni* and *S. haematobium* infection is praziquantel (40 mg/kg orally in two divided doses); *S. japonicum* infection is also resolved with praziquantel (60 mg/kg in three divided doses).

INTESTINAL TREMATODES (*FASCIOLOPSIS BUSKI*, *HETEROPHYES HETEROPHYES*, AND *METAGONIMUS YOKOGAWAI*)

The life cycle of *Fasciolopsis buski* requires a single snail intermediate host, which becomes infected when eggs excreted in the feces of an infected human hatch to release a miracidium. After developing in the snail, cercariae are released and encyst on water plants. These metacercaria are consumed by the definitive host; the released larvae then migrate to the host intestine and develop into adults. Adult worms in the intestine digest mucosal epithelial cells, which leads to inflammation and microabscess formation. The symptoms of infection with *F. buski* include abdominal or epigastric pain, nausea, diarrhea, and vomiting. Heavy infection may be associated with intestinal obstruction or protein-losing enteropathy. The diagnosis is made by identifying the characteristically large eggs, which measure approximately 130 by 80 μm in size, in the feces of an infected individual. Treatment is praziquantel (75 mg/kg), given in three divided doses for 1 day.

Infection with *Heterophyes heterophyes* and *Metagonimus yokogawai* occurs after ingesting undercooked fish containing metacercariae. The immature adults then migrate to the host small intestine, where they mate and attach to the intestinal mucosa. Eggs excreted in the feces hatch to release miracidia, which then infect an intermediate snail host. The snail sheds the cercariae, which infect a second intermediate host, usually a freshwater fish. Symptoms in the human host are usually mild, although heavily infected individuals

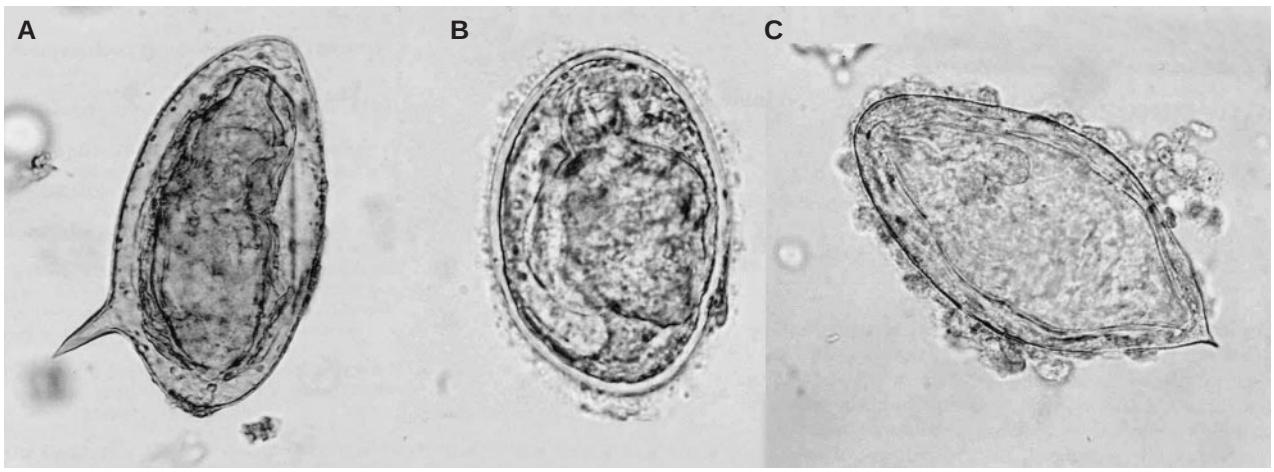


FIGURE 1 Eggs from the three major schistosome species that infect humans. *Schistosoma mansoni* (A), *Schistosoma japonicum* (B), and *Schistosoma haematobium* (C). Reproduced with permission from Ash and Orihel (1997). Copyright © 1997 by the American Society of Clinical Pathologists.

may experience abdominal pain, nausea, and diarrhea. Rarely the parasites will elicit an intestinal inflammatory response at the mucosal surface that may lead to granuloma formation. Treatment of *Heterophyes* and *Metagonimus* infections is praziquantel, 75 mg/kg administered orally in three divided doses for 1 day.

TISSUE TREMATODES (*FASCIOLA* SPP., *CLONORCHIS SINENSIS*, AND *PARAGONIMUS WESTERMANI*)

Two species of *Fasciola* cause zoonotic disease in humans. *Fasciola hepatica* (sheep liver fluke) is found in more temperate climates, including Europe, the Americas, and the Middle East, whereas *Fasciola gigantica* (cattle liver fluke) is localized to tropical regions of Asia and Africa. Adult worms, which reside in the bile ducts, release eggs that migrate to the intestine and are excreted in feces. After hatching in fresh water, the miracidia invade a snail intermediate host, in which they multiply and eventually are shed as cercariae. The cercariae attach to aquatic vegetation, and when ingested by a mammalian host, excyst and migrate to the bile ducts. Adults mate for years, and the females release eggs into the biliary tree to complete the life cycle.

Acute infection is characterized by a vigorous host immune response directed at tissue-migrating parasites, and may be associated with fever, abdominal pain, weight loss, and night sweats. Chronic infection is often asymptomatic, but may be associated with signs of biliary obstruction, including jaundice, abdominal pain, and fatty food intolerance. The diagnosis should be considered in individuals with eosinophilia and evidence of biliary obstruction or chronic liver disease. Unfortunately, fecal egg excretion is not present during the acute stage of infection and may be sporadic in chronically infected individuals. Immunodiagnostic tests have been developed, but are not routinely available. The treatment of fascioliasis is bithionol, given orally at a dose of 10 mg/kg for 10 days.

Infection with the liver fluke *Clonorchis sinensis* occurs primarily in countries of Asia, including Japan, Korea, China, Hong Kong, Taiwan, and Vietnam. Adult worms, measuring 10–20 mm in length and 3–4 mm wide, reside in the bile duct of humans. Two intermediate hosts are required for completion of the life cycle. Snails are infected with miracidia that hatch from eggs excreted in human feces. The cercariae are shed from the snail host and infect a cyprinoid fish, which, when ingested by humans, releases the encysted cercariae (metacercaria). Eating raw or undercooked fish is, therefore, the major risk factor for acquiring clonorchiasis. The larvae migrate from the small intestine to the bile duct, where they develop into adults. Within 3–4 weeks following infection, eggs appear in the feces. Symptoms occur as a direct result of the attachment of the worms to the bile duct and the host response to worm antigens. Chronic inflammation of the bile duct may lead to fibrosis and occasionally cirrhosis, as well as to secondary cholangitis, obstructive jaundice, and pancreatitis. In endemic areas, there appears to be an association between clonorchiasis and early-onset adenocarcinoma of the gallbladder. The diagnosis is made by identifying characteristic eggs in the feces of an infected individual. Eosinophilia, leukocytosis, and hyperbilirubinemia are suggestive but nonspecific laboratory findings. Treatment consists of praziquantel, 75 mg/kg divided in three doses for 1 day.

The related fluke *Paragonimus westermani*, like *C. sinensis*, requires two intermediate hosts to complete its life cycle. The first host is a snail, which is infected with the miracidium stage released from eggs in the feces of the primary host. The miracidia enter the second intermediate host, usually a crab, and develop into metacercariae. Humans become infected with the parasite by consuming undercooked crabmeat. The worms hatch in the small intestine, invade the mucosal epithelium, and eventually migrate to the peritoneal cavity. The parasites then traverse the diaphragm, where they reach the lungs and produce eggs. The eggs enter the airways and are coughed up, swallowed, and eventually excreted in feces. Adult worms in the pleural cavity may elicit an intense inflammatory response that triggers abscess formation and fibrosis. Symptoms of paragonimiasis include pleuritic chest pain and cough productive of reddish-brown sputum. Hemoptysis is rare and may signal rupture of a pulmonary blood vessel. Occasionally, adult worms may be found in the heart, mediastinum, peritoneal cavity, and even the central nervous system. The diagnosis is confirmed by identifying adult worms in tissue specimens, or eggs in sputum. Eosinophilia may be present, although this is a nonspecific laboratory finding. The treatment of paragonimiasis is praziquantel, 75 mg/kg divided in three doses for 1 day.

See Also the Following Articles

Cestodes • Helminth Infections • Nematodes • Parasitic Diseases, Overview

Further Reading

Ash, L., and Orihel, T. (1997). "Atlas of Human Parasitology." 4th Ed. ACSP Press, Chicago.

Haswell-Elkins, M. R. and Elkins, D. B. (1996). Food-borne trematodes. *In* "Manson's Tropical Diseases" (G. C. Cook, ed.), 20th Ed., pp. 1457–1477. W. B. Saunders, London.

Madacsy, L., Molnar, T., Nagy, I., Tiszlavicz, L., and Lonovics, J. (2003). Recurrent nonvariceal upper gastrointestinal bleeding in a patient with gastroduodenal schistosomiasis. *Endoscopy* **35**, 230–233.

Ross, A. G., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y., Williams, G. M., and McManus, D. P. (2002). Schistosomiasis. *N. Engl. J. Med.* **346**, 1212–1220.

Watanapa, P., and Watanapa, W. B. (2002). Liver fluke-associated cholangiocarcinoma. *Br. J. Surg.* **89**, 962–970.

Trichinella

GILBERT A. CASTRO* AND STEPHEN M. COLLINS†

**The University of Texas Health Science Center, Houston, and †McMaster University, Canada*

cytokines Proteins produced by T cells and other immune cells; transmit signals that are important for communication among the cells of the immune system and between the cells of the immune system and other cells.

T helper 1 (Th1) cells Produce proinflammatory cytokines such as interferon γ and interleukin-2, which are important in macrophage activation as well as in inflammatory and autoimmune reactions.

T helper 2 (Th2) cells Produce interleukins (IL-4, IL-5, IL-9, IL-10, and IL-13) that are involved in controlling humoral and allergic immune responses.

***Trichinella spiralis* is a nematode parasite that infects humans and a broad range of mammalian carnivores and omnivores. The resulting disease, trichinellosis, is transmitted through the ingestion of infected meat. Most humans acquire infection by eating infected pork. Experimental infections in mice, rats, guinea pigs, rabbits, and dogs have revealed that all stages of the *T. spiralis* life cycle occur in a single host. The life cycle involves an initial intestinal phase, lasting for about 3 to 4 weeks, followed by a muscle phase in which the parasite encysts and remains in the host for life. The extensiveness of its host range, its ease of maintenance in the laboratory, and its capacity to induce a consistent and multifaceted inflammatory response make *T. spiralis* a unique model for studying the interface between the immune and physiological systems of the gastrointestinal tract.**

LIFE CYCLE

The life cycle of *Trichinella spiralis* has five stages, of which the first four are larval and the fifth is the adult. Transformation of one stage to another involves molting by the worm. Infection is initiated when a host ingests the flesh of an animal that contains the first-stage larva. Muscle tissue of the ingested flesh and the surrounding cyst walls of the larvae are digested in the host stomach, releasing the larval form. On passage from the stomach to the intestine, larvae immediately invade enterocytes, where they develop into adult male and female worms within 30 hours. The bisexual worms then copulate and 3–4 days later the female worms deposit, ovoviviparously, F_2 generation larvae in the intestinal mucosa. These "newborn" larvae access the lymphatics and, through hematogenous spread, reach the capillaries of skeletal muscle, invade skeletal muscle fibers, and become encapsulated, or "encysted." The intestinal phase of trichinellosis lasts about 3–4 weeks, and the muscle phase lasts for the life of the host. Within the host, the intestinal worm population is relatively stable during early infection and is gradually depleted as the host develops immunity. Encysted muscle larvae are not eliminated by the host immune response, but may eventually become calcified.

PATHOGENESIS OF INFECTION

Signs or symptoms of disease, such as diarrhea, anorexia, peripheral blood changes, and muscle pain, are associated with physiological malfunctions and closely track inflammatory changes caused by the parasite. Inflammation may be induced by nonspecific stimuli due to traumatic damage to tissues or may be mediated through specific immunological reactions involving specific cell types and neural pathways linked

 Copyright 2004, Elsevier (USA). All rights reserved.

Haswell-Elkins, M. R. and Elkins, D. B. (1996). Food-borne trematodes. *In* "Manson's Tropical Diseases" (G. C. Cook, ed.), 20th Ed., pp. 1457–1477. W. B. Saunders, London.

Madacsy, L., Molnar, T., Nagy, I., Tiszlavicz, L., and Lonovics, J. (2003). Recurrent nonvariceal upper gastrointestinal bleeding in a patient with gastroduodenal schistosomiasis. *Endoscopy* **35**, 230–233.

Ross, A. G., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y., Williams, G. M., and McManus, D. P. (2002). Schistosomiasis. *N. Engl. J. Med.* **346**, 1212–1220.

Watanapa, P., and Watanapa, W. B. (2002). Liver fluke-associated cholangiocarcinoma. *Br. J. Surg.* **89**, 962–970.

Trichinella

GILBERT A. CASTRO* AND STEPHEN M. COLLINS†

**The University of Texas Health Science Center, Houston, and †McMaster University, Canada*

cytokines Proteins produced by T cells and other immune cells; transmit signals that are important for communication among the cells of the immune system and between the cells of the immune system and other cells.

T helper 1 (Th1) cells Produce proinflammatory cytokines such as interferon γ and interleukin-2, which are important in macrophage activation as well as in inflammatory and autoimmune reactions.

T helper 2 (Th2) cells Produce interleukins (IL-4, IL-5, IL-9, IL-10, and IL-13) that are involved in controlling humoral and allergic immune responses.

***Trichinella spiralis* is a nematode parasite that infects humans and a broad range of mammalian carnivores and omnivores. The resulting disease, trichinellosis, is transmitted through the ingestion of infected meat. Most humans acquire infection by eating infected pork. Experimental infections in mice, rats, guinea pigs, rabbits, and dogs have revealed that all stages of the *T. spiralis* life cycle occur in a single host. The life cycle involves an initial intestinal phase, lasting for about 3 to 4 weeks, followed by a muscle phase in which the parasite encysts and remains in the host for life. The extensiveness of its host range, its ease of maintenance in the laboratory, and its capacity to induce a consistent and multifaceted inflammatory response make *T. spiralis* a unique model for studying the interface between the immune and physiological systems of the gastrointestinal tract.**

LIFE CYCLE

The life cycle of *Trichinella spiralis* has five stages, of which the first four are larval and the fifth is the adult. Transformation of one stage to another involves molting by the worm. Infection is initiated when a host ingests the flesh of an animal that contains the first-stage larva. Muscle tissue of the ingested flesh and the surrounding cyst walls of the larvae are digested in the host stomach, releasing the larval form. On passage from the stomach to the intestine, larvae immediately invade enterocytes, where they develop into adult male and female worms within 30 hours. The bisexual worms then copulate and 3–4 days later the female worms deposit, ovoviviparously, F_2 generation larvae in the intestinal mucosa. These "newborn" larvae access the lymphatics and, through hematogenous spread, reach the capillaries of skeletal muscle, invade skeletal muscle fibers, and become encapsulated, or "encysted." The intestinal phase of trichinellosis lasts about 3–4 weeks, and the muscle phase lasts for the life of the host. Within the host, the intestinal worm population is relatively stable during early infection and is gradually depleted as the host develops immunity. Encysted muscle larvae are not eliminated by the host immune response, but may eventually become calcified.

PATHOGENESIS OF INFECTION

Signs or symptoms of disease, such as diarrhea, anorexia, peripheral blood changes, and muscle pain, are associated with physiological malfunctions and closely track inflammatory changes caused by the parasite. Inflammation may be induced by nonspecific stimuli due to traumatic damage to tissues or may be mediated through specific immunological reactions involving specific cell types and neural pathways linked

 Copyright 2004, Elsevier (USA). All rights reserved.

through cytokine intermediaries. Mast cell, eosinophil, and goblet cell hyperplasia characterize the mucosal tissues in infected hosts and are accompanied by physiological changes in digestion, absorption, secretion, and motility. These changes usually diminish when the inflammation subsides and the worms are eliminated.

INFLAMMATION, PATHOLOGY, AND INTESTINAL IMMUNITY

Immunity can be demonstrated by reinfecting a host that has been previously infected or has been "vaccinated"; vaccination is accomplished by enteral or parenteral injection with antigens from somatic tissues of *T. spiralis* or from excretions or secretions of the nematode. Immunity is evident in that fewer worms from the challenge infection are able to infect the host, and those that do are eliminated at a much faster rate, compared with a primary infection. Various serological or biochemical correlates are also used to measure the immune response.

IMMUNOLOGICAL REACTIONS AND THEIR INTERPLAY

Infection produces an inflammatory response in the gastrointestinal tract and consists of innate and adaptive immune responses. The innate response involves infiltration of the tissue by neutrophils and macrophages and is associated with changes in secretion, mucus production, and motility. These changes are mediated by cytokines such as interleukin-1 (IL-1). A specific immune response follows and involves the differentiation of T helper (Th) lymphocytes into subsets that are characterized by specific cytokine profiles; these are referred to as Th1 or Th2 responses. *Trichinella* is a potent stimulus of the Th2 response, which is characterized by the secretion of IL-4, IL-5, IL-9, IL-10, and IL-13. These cytokines in turn influence intestinal physiology and hence the ability of the host to evict the parasite. Thus, *Trichinella* infection serves as a robust model in which not only to study the modulation of intestinal physiology by the mucosal immune system, but also to place these interactions in a context of host defense.

Subsequent exposure of the host to *Trichinella* results in rapid expulsion of the parasite due to the acquisition of natural immunity. This is an antibody-mediated type I hypersensitivity response. Cytokines from Th2 cells stimulate B lymphocyte growth, the production of immunoglobulin E (IgE), and the activation of mast cells. Mast cells that become armed with IgE on their surface degranulate when reexposed to parasite-specific antigens. In the process, 5-hydroxytryptamine (5-HT), histamine, and prostaglandin are released from mast cells and act directly on epithelial cells and stimulate secretion of chloride ions and water. Histamine and 5-HT stimulate epithelial cell secretion through another pathway by acting indirectly through enteric cholinergic nerves that regulate chloride secretion. Additionally, mast cell-derived 5-HT and histamine act on intestinal smooth muscle to cause contraction.

The complexity of the mucosal immune response has captured the interest of gastroenterological researchers and has generated insights into human disease states as well as into new treatment strategies. The critical role of the Th2 response in host defense has generated new knowledge regarding the ability of Th cytokines to modulate secretion, mucus production, and gut motility. This has led to an understanding of how diseases characterized by immune activation, such as inflammatory bowel disease (IBD), disrupt intestinal physiology to generate symptoms such as diarrhea and pain. Studies using the *T. spiralis* model in certain mouse strains have prompted the notion that transient infection and immune activation may lead to physiological disruption that persists long after infection and inflammation are resolved. This has provided insights into a common chronic human condition known as irritable bowel syndrome (IBS), which may develop following acute infective gastroenteritis. Finally, the robust nature of the Th2 response during *T. spiralis* infection has provided a model to test the concept of immunological distraction. Prior exposure of the host to a Th1 stimulus distracts the immune system away from a protective Th2 response, and results in chronic infection. Conversely, prior infection with *T. spiralis* polarizes the immune response toward Th2 and away from the deleterious Th1 response associated with experimental colitis, with consequent amelioration of colonic inflammation. This strategy is now being harnessed to treat human IBD and is an example of the utilization of nematode infection to explore mechanisms and treatment of human disease.

See Also the Following Articles

Foodborne Diseases • Food Poisoning • Food Safety • Helminth Infections • Nematodes • Parasitic Diseases, Overview • TH1, TH2 Responses

Further Reading

Barbara, G., De Giorgio, R., Deng, Y., Vallance, B., Blennerhassett, P., and Collins, S. M. (2001). Role of immunologic factors and

cyclooxygenase 2 in persistent postinfective enteric muscle dysfunction in mice. *Gastroenterology* **120**, 1729–1736.
Barbara, G., Vallance, B. A., and Collins, S. M. (1997). Persistent intestinal neuromuscular dysfunction after acute nematode infection in mice. *Gastroenterology* **113**, 1224–1232.
Castro, G. A. (1989). Parasite Infections and Gastrointestinal Motility. *In* "Handbook of Physiology—Gastrointestinal Physiology" (J. D. Wood, ed.), pp. 1133–1152. American Physiological Society, Bethesda.
Castro, G. A. (1991). A Physiological perspective: Physiological adaptations in host-parasite systems. *In* "Parasite–Host Associations—Coexistence or Conflict?" (C. A. Toft., A. Aeschlimann, and L. Bolis, eds.), pp. 356–374. Oxford University Press, New York.
Castro, G. A., and Arntzen, C. J. (1993). Immunophysiology of the gut: A research frontier for integrative studies of the common mucosal immune system. *Am. J. Physiol.* (*Gastrointest. Liver Physiol.*) **28**, G599–G610.
Castro, G. A., and Powell, D. W. (1994). The physiology of the mucosal immune system and immune-mediated responses in the gastrointestinal tract. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Vol. 1, pp. 709–750. Raven Press, New York.
Elliott, D. E., Urban, J. F., Jr., Argo, C. K., and Weinstock, J. V. (2000). Does the failure to acquire helminthic parasites predispose to Crohn's disease? *FASEB J.* **14**, 1848–1855.
Khan, W. I., Blennerhassett, P. A., Deng, Y., Gauldie, J., Vallance, B. A., and Collins, S. M. (2001). IL-12 gene transfer alters gut physiology and host immunity in nematode-infected mice. *Am. J. Physiol.* (*Gastointest. Liver Physiol.*) **281**, G102–G110.
Khan, W. I., Blennerhassett, P. A., Varghese, A. K., Chowdhury, S. K., Omsted, P., Deng, Y., and Collins, S. M. (2002). Intestinal nematode infection ameliorates experimental colitis in mice. *Infect. Immun.* **70**, 5931–5937.

Tropical Sprue

GUILLERMO ROBLES-DÍAZ
Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City

acclimatization Process of adaptation to a new and different climate, environment, or situation.
bacterial overgrowth Abnormal bacterial proliferation; within the small intestinal lumen, this results in mucosal alterations and changes in bile salt metabolism.
malabsorption Impaired or incomplete absorption of nutrients by the intestine; can result from any abnormality in the process of digestion and/or absorption of nutrients.
osmotic diarrhea High-output diarrhea resulting from ingested, unabsorbed nutrients, which act as osmotic agents, drawing free water into the intestinal lumen.
sprue A word anglicized by Patrick Manson in 1880 from the Dutch term *sprouw*, meaning "trush," used in seventeenth century Europe to describe a diarrheal disorder with oral aphthous ulcerations.

Tropical sprue is a malabsorption syndrome of unknown etiology that occurs in residents of and visitors to certain tropical and subtropical areas. It is characterized by abnormalities in the small bowel structure and function that lead to chronic diarrhea and malabsorption with megaloblastic anemia, for which no specific causes can be identified. Tropical sprue may be the ultimate manifestation of multiple diseases with different etiologies.

INTRODUCTION

Descriptions of malabsorptive illnesses in the tropics, or tropical sprue, can be found in ancient medical literature, but little progress has been made in the understanding the exact nature of the disease. Two factors, the typical geographic distribution of sprue within the tropics and, in modern times, the response of patients with tropical sprue to sulfonamides and folic acid, have allowed differentiation of tropical sprue from other similar syndromes previously described as nontropical sprue, such as celiac disease, which can be controlled with a gluten-free diet. However, epidemiologic and etiopathogenetic controversies have limited the establishment of a unanimously accepted definition of tropical sprue, and the syndrome remains an enigmatic and commonly overlooked disorder. The absence of a

Copyright 2004, Elsevier (USA). All rights reserved.

cyclooxygenase 2 in persistent postinfective enteric muscle dysfunction in mice. *Gastroenterology* **120**, 1729–1736.
Barbara, G., Vallance, B. A., and Collins, S. M. (1997). Persistent intestinal neuromuscular dysfunction after acute nematode infection in mice. *Gastroenterology* **113**, 1224–1232.
Castro, G. A. (1989). Parasite Infections and Gastrointestinal Motility. *In* "Handbook of Physiology—Gastrointestinal Physiology" (J. D. Wood, ed.), pp. 1133–1152. American Physiological Society, Bethesda.
Castro, G. A. (1991). A Physiological perspective: Physiological adaptations in host-parasite systems. *In* "Parasite–Host Associations—Coexistence or Conflict?" (C. A. Toft., A. Aeschlimann, and L. Bolis, eds.), pp. 356–374. Oxford University Press, New York.
Castro, G. A., and Arntzen, C. J. (1993). Immunophysiology of the gut: A research frontier for integrative studies of the common mucosal immune system. *Am. J. Physiol.* (*Gastrointest. Liver Physiol.*) **28**, G599–G610.
Castro, G. A., and Powell, D. W. (1994). The physiology of the mucosal immune system and immune-mediated responses in the gastrointestinal tract. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), Vol. 1, pp. 709–750. Raven Press, New York.
Elliott, D. E., Urban, J. F., Jr., Argo, C. K., and Weinstock, J. V. (2000). Does the failure to acquire helminthic parasites predispose to Crohn's disease? *FASEB J.* **14**, 1848–1855.
Khan, W. I., Blennerhassett, P. A., Deng, Y., Gauldie, J., Vallance, B. A., and Collins, S. M. (2001). IL-12 gene transfer alters gut physiology and host immunity in nematode-infected mice. *Am. J. Physiol.* (*Gastointest. Liver Physiol.*) **281**, G102–G110.
Khan, W. I., Blennerhassett, P. A., Varghese, A. K., Chowdhury, S. K., Omsted, P., Deng, Y., and Collins, S. M. (2002). Intestinal nematode infection ameliorates experimental colitis in mice. *Infect. Immun.* **70**, 5931–5937.

Tropical Sprue

GUILLERMO ROBLES-DÍAZ
Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City

acclimatization Process of adaptation to a new and different climate, environment, or situation.
bacterial overgrowth Abnormal bacterial proliferation; within the small intestinal lumen, this results in mucosal alterations and changes in bile salt metabolism.
malabsorption Impaired or incomplete absorption of nutrients by the intestine; can result from any abnormality in the process of digestion and/or absorption of nutrients.
osmotic diarrhea High-output diarrhea resulting from ingested, unabsorbed nutrients, which act as osmotic agents, drawing free water into the intestinal lumen.
sprue A word anglicized by Patrick Manson in 1880 from the Dutch term *sprouw*, meaning "trush," used in seventeenth century Europe to describe a diarrheal disorder with oral aphthous ulcerations.

Tropical sprue is a malabsorption syndrome of unknown etiology that occurs in residents of and visitors to certain tropical and subtropical areas. It is characterized by abnormalities in the small bowel structure and function that lead to chronic diarrhea and malabsorption with megaloblastic anemia, for which no specific causes can be identified. Tropical sprue may be the ultimate manifestation of multiple diseases with different etiologies.

INTRODUCTION

Descriptions of malabsorptive illnesses in the tropics, or tropical sprue, can be found in ancient medical literature, but little progress has been made in the understanding the exact nature of the disease. Two factors, the typical geographic distribution of sprue within the tropics and, in modern times, the response of patients with tropical sprue to sulfonamides and folic acid, have allowed differentiation of tropical sprue from other similar syndromes previously described as nontropical sprue, such as celiac disease, which can be controlled with a gluten-free diet. However, epidemiologic and etiopathogenetic controversies have limited the establishment of a unanimously accepted definition of tropical sprue, and the syndrome remains an enigmatic and commonly overlooked disorder. The absence of a

Copyright 2004, Elsevier (USA). All rights reserved.

universal endemicity in tropical and subtropical regions strongly suggests that the etiologic or predisposing conditions are geographically restricted. Multiple factors seem to play a role in the genesis and persistence of the clinical manifestations. The diagnosis can be assumed when other causes of malabsorption have been excluded.

ETIOLOGY AND GEOGRAPHIC DISTRIBUTION

The etiology and distribution of tropical sprue still defy coherent explanation. Different theories have been put forward, including climatic and sanitary conditions, infectious agents, local dietary factors, and/or immune responses, which would explain the occurrence of the disease in some cases. However, environmental as well as hypothetical genetic factors have eluded discovery so far, and instead of a single disease, tropical sprue may represent a syndrome caused by different etiologies, each with its own distribution.

Climatic

Lack of acclimatization was one of the initial explanations for the higher incidence of tropical sprue among Europeans living in tropical colonies. Natives were considered resistant to the disease, and adverse climatic conditions were thought to disturb the metabolic balance of unacclimatized people. Some weather conditions, such as a combination of humidity and heat, seemed to correlate with a seasonal variation of the disease. However, a climatic explanation was eventually rejected because there was enough evidence to suggest that tropical sprue was by no means absent in native populations, but rather was overlooked. Furthermore, the disease is not present in all tropical areas of similar latitude and similar atmospheric conditions, and it has been described in climatically diverse regions.

Infectious

Tropical sprue has been considered an infectious disease because of the endemic and epidemic presentation, the abrupt onset, similar to episodes of traveler's diarrhea seen in some cases, the isolation of different bacteria in jejunal and stool cultures, and the remarkable response of patients to antibiotic therapy.

Tropical sprue occurs both sporadically and in well-defined epidemics, with a patchy distribution in some regions of Asia and Caribbean Islands and at a much lower frequency in Mexico, Central and South America, and Africa. Attack rates are higher in adults than in children, who are normally more frequently affected by most infective diarrheal diseases. The prevalence of endemic tropical sprue has not been clearly established; it tends to be underestimated but may be declining because of the widespread and early use of antibiotics for acute traveler's diarrhea.

Bacteria, viruses, protozoa, and fungi or yeasts have been considered as possible causative agents, but numerous attempts have failed to identify an agent capable of producing all the features of tropical sprue. Enterotoxigenic coliform bacteria (mainly *Klebsiella pneumoniae, Enterobacter cloacae*, and *Escherichia coli*) in jejunal aspirates and greater numbers of coliform bacteria in stool cultures have been found in tropical sprue patients in Puerto Rico and India, but studies done in South Africa have shown no difference in the number of coliform organisms, leading to the suggestion of a different pathophysiology. In visitors to the tropics, the disease may be secondary to an impairment in the expulsion of enterotoxigenic coliform bacteria from the gut after an episode of acute diarrhea. Other observations have prompted the conclusion that one or more protozoan parasites, such as *Cryptosporidium parvum, Isospora belli, Cyclospora cayetanensis*, or *Blastocystis hominis*, may play a causative role. It is also thought that perhaps a yet unrecognized pathogen causes persistent alimentary tract infection or that different infectious agents specific to geographic areas are involved. It is also possible that all microorganisms could represent secondary infections rather than being the primary cause of tropical sprue, as was found to be the case with a superimposed intestinal mycosis that was originally implicated in the etiology of tropical sprue.

Nutritional

Nutritional deficiencies and unsaturated fat consumption have been proposed as the primary causes of tropical sprue. The former, however, have been eliminated as causative in light of the frequency of tropical sprue in patients on a good diet and the absence of tropical sprue in tropical areas where the population exists on a marginal diet, as well as the recognition that vitamin deficiency can be a consequence of rather than the cause of malabsorption; however, subclinical folate deficiency, being very common in the tropics, may be a contributory factor.

The disease is more prevalent in areas where food is usually fried with cooking oils rich in unsaturated long-chain fatty acids, mainly oleic and linoleic acids (i.e., from pork lard, sesame seeds, and soybeans), whereas its absence in Africa and the English-speaking

Caribbean Islands may be related to the use of more saturated oils (i.e., coconut and palm oils) and the consumption of food that is usually boiled rather than fried. This difference may be important because linoleic acid exerts an important influence in diminishing intestinal *Lactobacillus acidophilus* and thus producing an increase in coliform bacteria. Moreover, unsaturated oils are thought to be susceptible to oxidative rancidity after prolonged storage and exposure to sunlight and high temperature, and rancid dietary fats may induce changes in intestinal functions.

The seasonal incidence of tropical sprue may be related to increased fat consumption during national holidays and festivals, as occurs in Puerto Rico during Christmas season, or to increased fat oxidation during the hotter summer months, as has been observed in the Philippines. These findings provide some explanation for disease distribution but fail to accommodate other features of the disease, such as epidemics. The current trend toward globalization of diets and foods may eventually modify the role of diet in the development of tropical sprue, especially in expatriates.

Tropical Malabsorption

Some mild abnormalities in intestinal anatomy and physiology are widely prevalent in children and adults in different parts of the tropics. In this subclinical "tropical enteropathy," T cell activation in the lamina propria, may be involved in mucosal abnormalities; this enteropathy is acquired and is usually reversible when a patient leaves the tropics. Its association with tropical sprue is unknown, but tropical enteropathy and tropical sprue seem to be two unrelated conditions with two different, yet unidentified, etiologies, rather than two extremes of the same disease, responding with differential severity to unidentified environmental factors. Tropical enteropathy may, however, render a patient susceptible, and subsequent exposure to an unknown agent may trigger the development of tropical sprue.

Genetic

Associations of tropical sprue with human leukocyte antigens (HLA) Aw-19 and Aw-31 in Puerto Rico and B-8 in India have been reported. It is possible that these associations represent potential for development of disease in these populations when they are exposed to different specific but as yet unknown agents. Another possibility is that other genes associated with tropical sprue may be in linkage disequilibrium with different HLA genes in different populations.

PATHOPHYSIOLOGY

There is not universal agreement on the events involved in the pathophysiology of tropical sprue. However, there have been consistent observations of different factors that may interact and lead to development of the clinical manifestations of tropical sprue.

Basic Lesion

The initial lesion seems to be an acute or subclinical intestinal infection, possibly differing by geography and/or dietary fat intake but leading to persistent and repetitive enterocyte injury. Nutrient malabsorption, mainly impaired fat and folic acid absorption, is secondary to mucosal involvement and leads to osmotic diarrhea and malnutrition, as well as to other alterations that worsen intestinal structure, function, and recovery. Major pathophysiologic disturbances occur in the small bowel but colonic dysfunction may also be found.

Impairment of Intestinal Structure and Function

Both mucosal damage and increased luminal fat, possibly via gut hormones such as enteroglucagon or peptide YY, lead to small intestinal stasis, which results in bacterial overgrowth and colonization by enterotoxin-producing coliforms. The enterotoxins cause the mucosal abnormalities seen in tropical sprue and weaken the mucosal resistance to bacterial adhesion. At the same time, there is an inhibition of normal gram-positive flora. Administration of antibiotics breaks this vicious cycle by controlling the abnormal flora.

The unabsorbed intestinal fat, mainly an excess of free unsaturated fatty acids, is thought to cause inhibition of Na^+,K^+-ATPase and Mg^{2+}-ATPase (yielding an increased mucosal pH, which decreases the folate absorption), and inhibition of water and electrolyte intestinal absorption; the latter also results from direct damage to colonocytes by the unknown etiologic factor of tropical sprue. Patients suffer from net fluid excretion instead of the normal net absorption, which adds a secretory component to the diarrhea.

Micronutrient Deficiencies

Low serum folate levels are one of the most common and important findings of tropical sprue. Vitamin B_{12} deficiency occurs as the damage progresses. Both micronutrients play an important role in tissues with the

fastest cell turnover, namely the small intestine and the bone marrow, which characterizes and perpetuates the main clinical manifestations. Although tropical sprue is characterized by a deficiency of all nutrients, treatment with only folic acid hastens enterocyte recovery and usually secures clinical remission.

Digestive Abnormalities

Impairment of the proximal gut mucosa may lead to a reduction in meal-stimulated exocrine pancreatic secretions. This hyposecretion is associated with the degree of duodenal atrophy and it seems to be secondary to diminished production and release of pancreatic secretagogue hormones, such as secretin and cholecystokinin. Also, most patients exhibit lactose intolerance due to uniformly reduced levels of disaccharidase activity in the atrophic jejunal mucosa. These anatomic abnormalities result in a worsening of the osmotic diarrhea because of deficient nutrient digestion. Both conditions improve when patients receive the specific treatment for tropical sprue.

CLINICAL FEATURES

No single clinical manifestation or laboratory abnormality is diagnostic of tropical sprue. The outstanding classical symptom is chronic osmotic diarrhea and the clinical spectrum is related to the duration of the disease and the nutritional background of the individual.

The early phase of the disease is characterized by diarrhea of variable intensity, with pale, fetid, and greasy stools; abdominal distension; prominent bowel sounds; anorexia; and fatigue. Tropical sprue usually begins after years in the tropics, but may appear within a short period of arrival or may occur even months or years after leaving a tropical location. Also, it may develop as an acute attack of watery diarrhea associated with fever and malaise followed by chronic manifestations. This latter development usually occurs during epidemics and/or in foreign individuals living in endemic areas. In general, chronic diarrhea with lactose intolerance, progressive weight loss, and megaloblastic anemia are commonly described. Subsequently, clinical findings of nutritional and vitamin deficiencies become prominent as a consequence of malabsorption and reduced dietary intake secondary to anorexia. It is important to note that glossitis, stomatitis, edema, or other data related to nutritional deficiencies, although common, are not consistently present and predominate in undernourished indigenous populations.

DIAGNOSIS

Tropical sprue should be suspected in any patient who lives, has lived, or visited the tropics and presents with chronic diarrhea with or without clinical evidence of intestinal malabsorption. The first step in diagnosis is the exclusion of a specific cause of chronic diarrhea and then the assessment of intestinal absorption and morphology. Thus, a complete and specialized workup must be done to exclude other diseases with similar manifestations, including bacterial overgrowth, celiac disease, and, particularly, parasitic infection, which may resemble or coexist with tropical sprue, exacerbating its clinical presentation, mainly in endemic areas.

Malabsorption Assessment

The malabsorption syndrome can be established with fecal fat quantification, D-xylose, and Schilling tests. Fat malabsorption is not consistently present. Serum beta-carotene may be used as a first-line screening test for fat absorptive capacity. Xylose absorption is almost uniformly decreased, making it a sensitive test, although with low specificity. Malabsorption of vitamin B_{12} and folic acid is present in nearly all cases, rendering serum levels of these nutrients low. Hypoalbuminemia can be secondary to impaired protein absorption, but also to inadequate diet, excessive intestinal loss, and reduced synthesis by the liver.

Morphological Alterations

The jejunal mucosa seems to be the main target in tropical sprue, but the entire small bowel is affected in advanced stages. The degree of mucosal abnormalities correlates with the severity of the malabsorption. Morphological alterations can be identified radiologically or endoscopically. One of the main values of both types of examination is the exclusion of other specific disorders.

Abnormalities in the barium small bowel follow-through examination tend to be somewhat diffuse and nonspecific. There is a slow transit of the contrast material through the gut, an increase in the caliber of the small intestine, and thickening of the folds. The endoscopic evaluation can reveal scalloping of the valvulae conniventes and a mosaic pattern of the mucosa similar to that seen in patients with celiac disease. It also enables a tissue biopsy, which is necessary to exclude other diseases.

Intestinal specimens tend to exhibit a variable loss of villous pattern, with increased crypt depth and cell hyperplasia, apparently resulting from increased cell production and migration to replenish cell loss.

These features tend to decrease in severe lesions. There are lymphocytic and plasmacytic infiltrates in the lamina propria and increasing plasma cell infiltrates with chronicity. Also, lipid droplets stained with oil red O have been consistently found in a thickened basement membrane subjacent to the intestinal epithelium. This finding sharply contrasts with lipid deposits in other conditions but it is not pathognomonic.

TREATMENT

The priorities in the initial management of tropical sprue are the restoration of water and electrolyte balance and replacement of nutritional deficiencies, mainly folic acid and vitamin B_{12}. Symptomatic treatment with antidiarrheal drugs may be required. Pharmacological doses of folic acid (5 mg, by mouth, one time daily) lead to a great improvement in clinical and hematologic manifestations within 10 days, including diarrhea, even if treatment fails to completely correct gut abnormalities. However, the response to this single treatment depends on the chronicity of the intestinal lesions. On the other hand, short-term antibiotic therapy with tetracycline (250–500 mg, by mouth, four times daily) or succinylsufathiazole (4 g, by mouth, one time daily) improves intestinal morphology and gastrointestinal symptoms significantly. Controversy exists regarding the role of broad-spectrum antibiotics and the duration of treatment. A therapeutic approach with both folic acid and antibiotics, usually given over a period of 1 to more than 6 months, is recommended. The rate and time of recovery may vary in different populations but a complete and permanent cure is expected, mainly in expatriates leaving the tropics. Recurrence of symptoms years after discontinuing treatment has been reported in some native patients.

If there is uncertainty in the diagnosis of tropical sprue in endemic areas, treatment based on folic acid and antibiotics may be wise because there is a rapid clinical recovery in these cases, and the approach can be used as an empiric trial. The efficacy of broad-spectrum antibiotics, the duration of treatment, and evaluations of long-term followup of patients needs to be assessed. Importantly, identification of a clear etiology would allow a more specific therapeutic approach.

See Also the Following Articles

Bacterial Overgrowth • Celiac Disease • Diarrhea • Malabsorptiuon

Further Reading

Cook, G. C. (1996). Tropical sprue: Some early investigation favored an infective cause, but was a coccidian protozoan involved? *Gut* **39**, 428–429.

Fantry, G. T., Fantry, L. E., and James, S. P. (1999). Chronic infections of the small intestine. *In* "Textbook of Gastroenterology" (T. Yamada, ed.), 3rd Ed., Vol. 2, pp. 1641–1659. Lippincott Williams & Wilkins, Philadelphia.

Farthing, M. J. G. (2002). Tropical malabsorption and tropical diarrhea. *In* "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., Vol. 2, pp. 1842–1854. Saunders Elsevier Science, Philadelphia.

Glynn, J. (1986). Tropical sprue—Its aetiology and pathogenesis. *J. R. Soc. Med.* **79**, 599–606.

Haghighi, P., and Wolf, P. L. (1997). Tropical sprue and sub-clinical enteropathy. A vision for the nineties. *Crit. Rev. Clin. Lab. Sci.* **34**, 313–341.

Klipstein, F. A., Holdeman, L. V., Corcino, J. J., and Moore, W. E. C. (1973). Enterotoxigenic intestinal bacteria in tropical sprue. *Ann. Intern. Med.* **79**, 632–641.

Menendez-Corrada, R., Nettleship, E., and Santiago-Delpin, E. A. (1986). HLA and tropical sprue. *Lancet* **2**, 1183–1185.

Mery, C. M., and Robles-Dúaz, G. (2000). Esprue tropical: ¿Una enfermedad inexistente o inadvertida? *Rev. Invest. Clin.* **52**, 377–379.

Morales, M., Galvan, E., Mery, C. M., Castro, G., Uscanga, L. F., and Robles-Diaz, G. (2001). Exocrine pancreatic insufficiency in tropical sprue. *Digestion* **63**, 30–34.

Westergaard, H. (1995). Southwestern internal medicine conference: The sprue syndromes. *Am. J. Med. Sci.* **290**, 249–262.

Trypsin

JOHN A. WILLIAMS
University of Michigan, Ann Arbor

enterokinase The enzyme present on the brush border of small intestinal enterocytes that cleaves trypsinogen, producing trypsin.
serine proteases Family of proteolytic enzymes with a common mechanism of action involving a serine residue at the active site.
trypsin activation peptide The amino-terminal portion of trypsinogen that is cleaved to release trypsin.
trypsinogen The inactive protein precursor of trypsin normally activated in the intestinal lumen.

Trypsin is a pancreatic protease that plays a key role in the digestion of nutrients in the small intestine. Its precursor, trypsinogen, constitutes approximately 20% of the protein in pancreatic juice. In humans, there are three forms of pancreatic trypsinogen, each coded for by a unique gene. The two major forms, anionic trypsinogen, also known as trypsinogen-1, and the more abundant cationic trypsinogen, trypsinogen-3, have very similar molecular masses, being approximately 25,000 Da with 89% homology at the amino acid level, but can be differentiated immunologically. The cationic form has more basic lysine residues and a more basic isoelectric point. A third variant, mesotrypsinogen or trypsinogen-2, is less abundant and has an intermediate isoelectric point. All three forms can be separated by two-dimensional gel electrophoresis. In the rat, four forms of trypsinogen have been identified at the protein level, with the anionic and cationic forms being most abundant. Although trypsinogen was originally thought to be expressed only in the pancreas, recently two variants have been identified in ovarian and colon cancer. A trypsin-like protein has also been identified in sperm.

SYNTHESIS AND ACTIVATION OF TRYPSIN

All trypsins are synthesized as an inactive precursor, trypsinogen, presumably as a protective mechanism to prevent premature activity in the pancreas, which could lead to the inflammatory disease pancreatitis. As with other pancreatic digestive enzymes, trypsinogens are processed through the Golgi apparatus and packaged into zymogen granules to await secretion by exocytosis. This packaging can also be considered a protective mechanism. Trypsin is produced by cleavage of an N-terminal peptide, Ala-Pro-Phe-Asp-Asp-Asp-Asp-Lys, and the tetra-aspartyl group is present in trypsinogens of most species. This peptide, known as the trypsin activation peptide, can also be assayed and used as a measure of trypsinogen activation. Cleavage of trypsinogen can be brought about by trypsin itself at a low rate and much more efficiently by the intestinal protease, enterokinase, a large protein present on the lumenal surface of cells lining the upper intestine. This activation occurs in the intestinal lumen and begins an activation cascade that then activates all the other pancreatic proteolytic enzymes and much of the lipase activity, the latter through activation of colipase (Fig. 1). Enterokinase activity is enhanced by low levels of calcium and bile salts. Trypsin is far more efficient at activating the other enzymes than it is at activating itself. Moreover, the pancreas also synthesizes small amounts of a pancreatic secretory trypsin inhibitor (PSTI), which is packaged with trypsinogen in zymogen granules and is capable of inhibiting small amounts of active trypsin. Trypsinogen can also be activated by certain cathepsins,

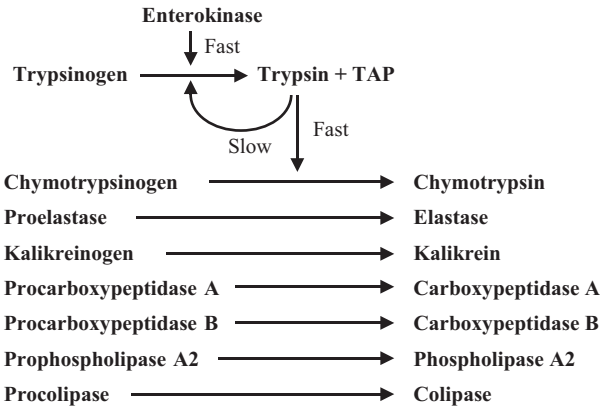


FIGURE 1 The pancreatic enzyme activation cascade. Trypsinogen(s) and other pancreatic secretory enzymes are secreted into the intestinal lumen as part of pancreatic juice. Trypsinogen is activated by enterokinase, which cleaves an amino-terminal activation peptide (TAP). Active trypsin then cleaves and activates all of the other pancreatic proteases, a phospholipase, and colipase, which is necessary for the physiological action of pancreatic triglyceride lipase.

 Copyright 2004, Elsevier (USA). All rights reserved.

lysosomal proteases that are active at low pH, but these are normally kept in a separate compartment in the cell. Activation of trypsinogen by cathepsins is thought to play a role in the genesis of acute pancreatitis.

ENZYMATIC ACTION OF TRYPSIN

This enzymatic activity of trypsin is directed at the carboxyl moiety of the basic amino acid residues lysine and arginine in proteins. Physiologically, trypsin is an endopeptidase, breaking down proteins and large peptides into small peptides. It is a serine protease with an active site that contains a Ser, His, Asp catalytic triad that is shared with other serine proteases. Its specificity is the result of a unique binding pocket that contains two glycine residues. There are a number of natural and synthetic inhibitors that will inactivate trypsin either reversibly or irreversibly. Trypsins are active over a wide pH range from 6 to 9 and are most active at pH 7.5–8.5. Calcium ion enhances trypsin activity but is not essential. Trypsins can be assayed with low-molecular-weight substrates possessing an acylarginine ester or an amide structure where activity is measured as a result of a colored or fluorescent product. Trypsinogen in plasma can also be measured by radioimmunoassay or enzyme-linked immunosorbent assay, and a urinary dipstick sensitive to trypsinogen-2 is being evaluated for use in diagnosis of acute pancreatitis.

Recently, hereditary pancreatitis has been shown to be associated with specific mutations in cationic trypsinogen or PSTI. These mutations result in reduced inactivation of trypsin and are believed to lead to premature activation of trypsin in the pancreas when associated with other provocactive events.

See Also the Following Articles

Pancreatic Digestive Enzymes • Pancreatic Enzyme Secretion (Physiology) • Pancreatitis, Hereditary

Further Reading

Lowe, M. E. (1993). The structure and function of pancreatic enzymes. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Raven Press, New York.

Rinderknect, H. (1993). Pancreatic secretory enzymes. *In* "The Pancreas: Biology, Pathobiology and Disease" (V. L. W. Go, *et al.*, eds.), pp. 219–251. Raven Press, New York.

Tumor Necrosis Factor-α (TNF-α)

LAURENCE JOHN EGAN
Mayo Clinic

caspases Proteases that operate in a cascade to initiate programmed cell death.

chemokine A chemotactic cytokine, i.e., a secreted polypeptide hormone that promotes the migration of cells along a concentration gradient.

cytokine A secreted polypeptide hormone that interacts with cell surface receptors to activate immune and other mechanisms.

ectodomain The extracellular portion of a molecule that spans the plasma membrane.

mitogen-activated protein kinases Enzymes involved in the transmission of signals from cytokine and other receptors to downstream transcription factors.

transcription factor An intracellular protein that regulates gene transcription.

Tumor necrosis factor-α (TNF-α) is one member of a superfamily of related cytokines. This cytokine is synthesized by cells and released into the extracellular space. Through the interaction with specific cell surface receptors on target cells, TNF-α initiates cascades of intracellular events that affect cellular metabolism, viability, and gene expression. The net result of these effects is highly variable, depending on the cell type and on the influence of other factors. In general, TNF-α induces innate immune mechanisms and thereby promotes inflammation, but it is also a critical regulator of cell survival. Elevated concentrations of TNF-α are characteristic of many chronic inflammatory diseases. Strategies to inhibit the biological functions of TNF-α have proven

Copyright 2004, Elsevier (USA). All rights reserved.

lysosomal proteases that are active at low pH, but these are normally kept in a separate compartment in the cell. Activation of trypsinogen by cathepsins is thought to play a role in the genesis of acute pancreatitis.

ENZYMATIC ACTION OF TRYPSIN

This enzymatic activity of trypsin is directed at the carboxyl moiety of the basic amino acid residues lysine and arginine in proteins. Physiologically, trypsin is an endopeptidase, breaking down proteins and large peptides into small peptides. It is a serine protease with an active site that contains a Ser, His, Asp catalytic triad that is shared with other serine proteases. Its specificity is the result of a unique binding pocket that contains two glycine residues. There are a number of natural and synthetic inhibitors that will inactivate trypsin either reversibly or irreversibly. Trypsins are active over a wide pH range from 6 to 9 and are most active at pH 7.5–8.5. Calcium ion enhances trypsin activity but is not essential. Trypsins can be assayed with low-molecular-weight substrates possessing an acylarginine ester or an amide structure where activity is measured as a result of a colored or fluorescent product. Trypsinogen in plasma can also be measured by radioimmunoassay or enzyme-linked immunosorbent assay, and a urinary dipstick sensitive to trypsinogen-2 is being evaluated for use in diagnosis of acute pancreatitis.

Recently, hereditary pancreatitis has been shown to be associated with specific mutations in cationic trypsinogen or PSTI. These mutations result in reduced inactivation of trypsin and are believed to lead to premature activation of trypsin in the pancreas when associated with other provocactive events.

See Also the Following Articles

Pancreatic Digestive Enzymes • Pancreatic Enzyme Secretion (Physiology) • Pancreatitis, Hereditary

Further Reading

Lowe, M. E. (1993). The structure and function of pancreatic enzymes. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Raven Press, New York.

Rinderknect, H. (1993). Pancreatic secretory enzymes. *In* "The Pancreas: Biology, Pathobiology and Disease" (V. L. W. Go, *et al.*, eds.), pp. 219–251. Raven Press, New York.

Tumor Necrosis Factor-α (TNF-α)

LAURENCE JOHN EGAN
Mayo Clinic

caspases Proteases that operate in a cascade to initiate programmed cell death.

chemokine A chemotactic cytokine, i.e., a secreted polypeptide hormone that promotes the migration of cells along a concentration gradient.

cytokine A secreted polypeptide hormone that interacts with cell surface receptors to activate immune and other mechanisms.

ectodomain The extracellular portion of a molecule that spans the plasma membrane.

mitogen-activated protein kinases Enzymes involved in the transmission of signals from cytokine and other receptors to downstream transcription factors.

transcription factor An intracellular protein that regulates gene transcription.

Tumor necrosis factor-α (TNF-α) is one member of a superfamily of related cytokines. This cytokine is synthesized by cells and released into the extracellular space. Through the interaction with specific cell surface receptors on target cells, TNF-α initiates cascades of intracellular events that affect cellular metabolism, viability, and gene expression. The net result of these effects is highly variable, depending on the cell type and on the influence of other factors. In general, TNF-α induces innate immune mechanisms and thereby promotes inflammation, but it is also a critical regulator of cell survival. Elevated concentrations of TNF-α are characteristic of many chronic inflammatory diseases. Strategies to inhibit the biological functions of TNF-α have proven

Copyright 2004, Elsevier (USA). All rights reserved.

effective in the control of certain of these diseases, notably Crohn's disease and rheumatoid arthritis.

INTRODUCTION

Tumor necrosis factor α (TNF-α) was first identified as a soluble factor that induced death in transformed cells. Independently, a molecule known as cachectin, which induced weight loss when overexpressed in animals, was described. Cloning of the genes encoding TNF-α and cachectin revealed that these are in fact the same molecule. Subsequent studies over the past 20–25 years have revealed diverse and potent effects of TNF-α on cellular functions that result in critical roles for this cytokine in mammalian physiology and pathophysiology. The cloning of genes closely related to TNF-α has revealed that this cytokine is part of a larger superfamily of related molecules. Two receptors that mediate the biological functions of TNF-α have been identified. This article examines TNF-α with particular emphasis on the factors that regulate its release, its actions on cells, and its role in chronic inflammatory diseases.

Synthesis and Secretion of TNF-α

Many cell types, including immune cells such as monocytes, macrophages, and lymphocytes, but also epithelial cells, endothelial cells, and connective tissue cells, can produce TNF-α. Under basal physiologic conditions, cells secrete very little TNF-α, but many stimuli can induce cells to release TNF-α. For example, stimulation by the pro-inflammatory cytokines interleukin-1, interferon γ, or TNF-α, infection by invasive microbes such as *Salmonella*, exposure to lipopolysaccharide from gram-negative bacteria, or engagement of the T-cell receptor can result in dramatic up-regulation of the release of TNF-α from various cell types. Numerous exogenous agents, such as glucocorticoids or cyclosporin, and endogenous proteins, including transforming growth factor-β and interleukin-10, can decrease the release of TNF-α from cells. Cells tightly control the release of TNF-α by the regulation of at least three distinct steps in the synthesis and secretion of this cytokine (Fig. 1).

Regulation of Transcription of the TNF-α Gene

A promoter region of approximately 600 bp upstream of the TNF-α gene contains binding sites for at least five distinct transcription factors. Mutational analyses of this promoter have revealed important roles for activating protein-1, nuclear factor of activated

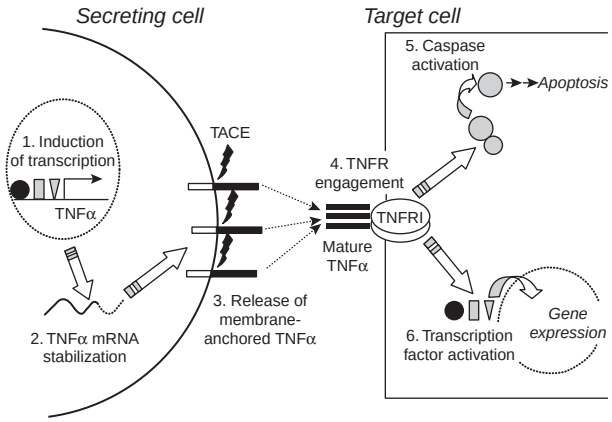


FIGURE 1 Model of TNF-α synthesis, secretion, and target cell stimulation. Several cellular mechanisms are involved in controlling the synthesis and secretion of TNF-α. A variety of transcription factors bind to the *TNF* promoter, leading to transcription of TNF-α mRNA (1). This mRNA is inherently unstable due to the presence of an AU-rich element in the 3′-untranslated region (dashed line). Stimulus-induced stabilization of TNF-α mRNA leads to greatly enhanced protein translation (2). When cells are stimulated to secrete TNF-α, the mature peptide is released by proteolytic cleavage of membrane-anchored TNF-α, by the action of TNF-α converting enzyme (TACE) (3). Released TNF-α forms homotrimers and binds to TNF receptors (TNFR) (4). Receptor engagement leads to clustering of the receptors, which facilitates the recruitment of numerous adapter and signaling molecules to the cytoplasmic tail of TNF receptors. Important downstream effects include the activation of caspase cascades (5), leading to apoptosis, and the activation of specific transcription factors (6), leading to target gene expression.

T cells, and nuclear factor κB (NF-κB) in the transcriptional regulation of TNF-α. Current concepts suggest that factors that induce TNF-α synthesis do so in part by stimulating the activity of these transcription factors, leading to increased rates of TNF-α transcription.

Regulation of TNF-α Messenger RNA Stability

The 3′-untranslated region of the TNF-α messenger RNA (mRNA), like that of many other cytokines, contains an AU-rich element that destabilizes the transcripts. As a result, mechanisms within the cell degrade TNF-α mRNA unless it is specifically stabilized. As a result, very little template is available for ribosomal translation of TNF-α protein under basal conditions. Recent evidence suggests that the activation of mitogen-activated protein kinase pathways is involved in the stabilization of TNF-α mRNA, but the exact mechanisms of this process are currently unclear. In this model, efficient translation of TNF-α requires the stabilization of the mRNA by stimulus-induced mitogen-activated protein kinase activation. The importance

of this mechanism in the regulation of TNF-α is illustrated by the development of severe multisystem inflammation in mice genetically engineered to lack the 3′ AU-rich element and thus have stable, long-lived TNF-α mRNA.

Inducible Ectodomain Shedding of Membrane-Bound TNF-α

TNF-α is synthesized as a 233-amino-acid protein that exists as a transmembrane molecule in which the biologically active portion is extracellular. In certain circumstances, such as the activation of T cells by antigen-presenting cells, this membrane-anchored form of TNF-α on the antigen-presenting cell can directly exert its biological activities by engaging TNF receptors on the T cell. More commonly, however, TNF-α is believed to act as a soluble hormone that diffuses to target cells after release from the cell of origin. Release of TNF-α requires proteolytic cleavage of the membrane-anchored form by TNF-α converting enzyme, leading to release from the cell of the mature 157-amino-acid soluble form of the cytokine. TNF-α molecules homotrimerize in the extracellular fluid and bind to their receptors on target cells.

CELLULAR RESPONSES TO TNF-α

The biological effects of TNF-α are mediated by the binding of the cytokine to specific high-affinity cell surface receptors, TNF receptor-1 and TNF receptor-2. Both receptors are widely expressed. The majority of the biological effects of TNF-α (see below) are believed to be mediated by TNF receptor-1, except in immune cells where TNF receptor-2 is also important. Cytokine trimer binding to TNF receptors induces receptor clustering that is essential for activation of downstream signaling events (Fig. 1). After activation of TNF receptor-1 by TNF-α, a number of adapter and signaling molecules are recruited to the cytoplasmic domains of the receptor. A complex cascade of downstream signaling events ensues, resulting in two major phenomena: initiation of specific target gene expression by the activation of transcription factors and the cleavage of death proteins through the activation of caspases.

Target genes of TNF-α are diverse in function, but prominently include those that promote inflammation and inhibit apoptosis. Caspase activation results in the cleavage of death substrates, which leads to apoptosis. Thus, depending on the cellular context and on the effects of numerous additional factors, TNF-α stimulation of cells can lead to the activation of innate immune mechanisms and inflammation or can lead to apoptosis. NF-κB is central to the integration of cellular responses to TNF-α. In the presence of robust NF-κB activation by TNF-α, inflammatory responses are prominent. However, if NF-κB-dependent cell survival signals cannot be activated after TNF-α stimulation, the pro-apoptotic effects of TNF-α predominate.

TNF-α IN PHYSIOLOGY AND DEVELOPMENT

In vitro and animal studies have defined a wide range of biological effects of TNF-α. At the molecular level, TNF-α induces the secretion of many cytokines, chemokines, growth factors, adhesion molecules, and enzymes that in general function to activate innate immunity. At the cellular and tissue levels, the effects of TNF-α depend not only on the cell type and the effects of other factors, but also on the local concentration of the cytokine. TNF-α induces leukocyte extravasation and activation of T and B lymphocytes and at high concentrations can induce apoptosis in some cells. Low concentrations may promote cell proliferation and the repair of tissue after injury. Experimental administration of TNF-α to animals induces fever, anorexia, and hypotension, mimicking endotoxic or septic shock.

The physiological functions of TNF-α have been explored by genetically engineering mice to lack TNF-α. These mice appear healthy and fertile, indicating that TNF-α is not essential for development or for normal physiological functions. However, TNF-α null mice lack mature B cell follicles and follicular dendritic cell networks in secondary lymphoid organs. In contrast, T-cell-rich areas are normal, suggesting an important role for TNF-α specifically in humoral immunity. Indeed, TNF-α null mice are highly susceptible to overwhelming infection with *Listeria monocytogenes*. Furthermore, these mice are somewhat resistant to developing endotoxic shock. These findings demonstrate that TNF-α is important in humoral immunity, in bacterial host defense, and in the responses to severe sepsis.

TNF-α IN HUMAN DISEASE

Characterization of some of the *in vivo* biologic effects of TNF-α led to the examination of potential roles for this cytokine in human disease. The pathophysiology of many human diseases involves chronic inflammation in the diseased tissues and in many of these diseases, the circulating or tissue concentrations of TNF-α are elevated above control values. Elevated circulating concentrations of TNF-α have been observed in patients with severe sepsis syndromes, suggesting a potential

pathogenic role for this cytokine in septic shock. However, therapeutic attempts to neutralize the biologic activity of TNF-α by administration of a neutralizing antibody to patients with severe sepsis were not beneficial (Table I). TNF-α concentrations are increased in certain chronic inflammatory diseases, notably inflammatory bowel disease and inflammatory arthritides. Since TNF-α is such an important activator of innate immune mechanisms, therapeutic inhibition of the biological activity of TNF-α has therefore emerged as an attractive target in these chronic inflammatory diseases (Table I). The management of severe, active Crohn's disease has been revolutionized by the use of infliximab, a chimeric anti-TNF-α antibody. This antibody produces therapeutic benefit in up to 70% of Crohn's disease patients who were previously refractory to conventional therapies. Infliximab and etanercept, a TNF receptor-1:Fc fusion protein, are also highly effective in inflammatory arthritides. Inhibition of the biologic activity of TNF-α comes at a price, however. Patients treated with potent anti-TNF-α agents are susceptible to opportunistic infections, especially reactivation of latent tuberculosis, and certain viral infections, such as herpes zoster. Furthermore, case reports of patients developing a multiple sclerosis-like central nervous system demyelinating disease after receiving potent anti-TNF-α therapies have generated much concern. It is possible that in certain circumstances, or at low concentrations, TNF-α may have protective effects against some disease processes.

TABLE I Efficacy of Potent Anti-TNFα Therapies in Human Diseases

Yes (proven efficacious in randomized placebo-controlled trials)
Crohn's disease
Rheumatoid arthritis
Ankylosing spondylitis
Psoriatic arthritis
Psoriasis
Maybe (positive data from uncontrolled trials)
Ulcerative colitis
Pouchitis
Behçet's disease
Pyoderma gangrenosum
Vasculitis
No (negative randomized placebo-controlled trials)
Septic shock

See Also the Following Articles

Crohn's Disease • Crohn's Disease, Pediatric • Hepatotoxins • TH1, TH2 Responses

Further Reading

Kollias, G. (1998). TNF-α. *In* "The Gene Knockout Facts Book" (T. W. Mak, ed.), Vol. 2, pp. 1033–1034. Academic Press, San Diego, CA.

Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F., and Kollias, G. (1999). Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: Implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398.

McKusick, V. A. (2002). Tumor necrosis factor. Online Mendelian Inheritance in Man. Available at http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id = 191160.

Podolsky, D. K., and Fiocchi, C. (2000). Cytokines, chemokines, growth factors, eicosanoids, and other bioactive molecules in inflammatory bowel disease. *In* "Inflammatory Bowel Disease" (J. B. Kirsner, ed.), 5th Ed., pp. 191–207. W. B. Saunders, Philadelphia, PA.

Present, D. H., Rutgeerts, P., Targan, S., Hanauer, S. B., Mayer, L., van Hogezand, R. A., Podolsky, D. K., Sands, B. E., Braakman, T., DeWoody, K. L., Schaible, T. F., and van Deventer, S. J. (1999). Infliximab for the treatment of fistulas in patients with Crohn's disease. *N. Engl. J. Med.* **340**, 1398–1405.

Targan, S. R., Hanauer, S. B., van Deventer, S. J., Mayer, L., Present, D. H., Braakman, T., DeWoody, K. L., Schaible, T. F., and Rutgeerts, P. J. (1997). A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor for Crohn's disease. *N. Engl. J. Med.* **337**, 1029–1035.

Ware, C. F., Santee, S., and Glass, A. (1998). Tumor necrosis factor-related ligands and receptors. *In* "The Cytokine Handbook" (A. Thompson, ed.), 3rd Ed., pp. 549–592. Academic Press, San Diego, CA.

Zhang, M., and Tracey, K. J. (1998). Tumor necrosis factor. *In* "The Cytokine Handbook" (A. Thompson, ed.), 3rd Ed., pp. 515–548. Academic Press, San Diego, CA.

Tyrosinemia

DAVID A. RUDNICK AND DAWN R. EBACH
Washington University, St. Louis, Missouri

acute intermittent porphyria One of a group of rare inherited disorders resulting from disturbance of the metabolism of the breakdown products of the red blood cell pigment (porphyrin). A prominent feature is intermittent attacks of abdominal pain.

NTBC Originally developed as an insecticide, this compound [2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione] was hypothesized, and then shown, to have clinical utility in the treatment of hereditary tyrosinemia type I.

Hereditary tyrosinemia type I is an inborn error of tyrosine metabolism, resulting from mutations in the fumaryl acetoacetate hydrolase gene, and affects the liver, kidneys, and peripheral nerves. A variety of interventions developed over the past several decades have improved clinical outcome, including liver transplantation and, more recently, pharmacotherapy with 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione.

INTRODUCTION

Hereditary tyrosinemia type I (referred to as tyrosinemia throughout the remainder of this article) is an autosomal recessive inborn error of metabolism that affects the liver, kidneys, and peripheral nerves. The disease results from deficiency in the enzyme fumaryl acetoacetate hydrolase (FAH), which is involved in the metabolism of tyrosine (Fig. 1). Tyrosinemia patients were first reported in the medical literature in the 1950s. Early outcomes were generally poor, with most patients succumbing in infancy or childhood. Over the past several decades, a variety of interventions have been developed, including dietary modification, liver transplantation, and prenatal screening. More recently, pharmacotherapy with the drug NTBC [2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione] has resulted in significant improvement in the natural history of this disease. This article reviews the epidemiology, genetics, clinical presentation, pathogenesis, diagnosis, and management of tyrosinemia.

EPIDEMIOLOGY AND GENETICS

Tyrosinemia is caused by deficiency of FAH (EC 3.7.1.2), the last enzyme in tyrosine catabolism. Heterozygotes

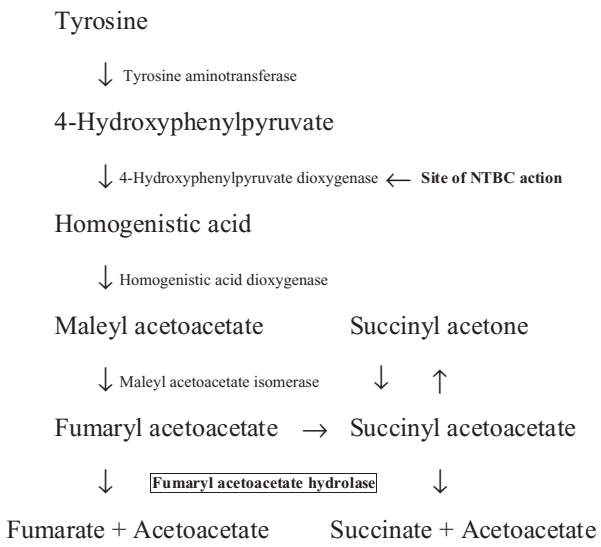


FIGURE 1 Tyrosine degradation pathway. The last step in tyrosine metabolism is catalyzed by the enzyme fumaryl acetoacetate hydrolase, which is mutated in hereditary tyrosinemia type I. This leads to increased production of metabolites including maleyl acetoacetate, fumaryl acetoacetate, succinyl acetoacetate, and succinyl acetone.

are clinically and biochemically normal. The gene encoding FAH (GenBank Accession No. NM_000137) maps to chromosome 15q23−q25 and encodes a 419-amino-acid protein, present as a homodimer within the cytosol of hepatocytes.

Epidemiology

The estimated worldwide incidence of tyrosinemia is 1 in 100,000 to 120,000. However, a much higher incidence has been recognized in northern Europe and Quebec, Canada. In the Saguenay-Lac-Saint-Jean area of Quebec, the carrier frequency of tyrosinemia is 1 in 20 and the frequency of disease in live births is 1 in 2000. Overall, the frequency of affected live births in Quebec is ~1 in 20,000. Because of this, Quebec has instituted routine prenatal testing. Tyrosinemia also occurs in ethnic groups other than French Canadian and Scandinavian.

 Copyright 2004, Elsevier (USA). All rights reserved.

Mutations in FAH

At least seven different mutations in FAH have been identified in patients with tyrosinemia. Most mutant alleles in the Saguenay-Lac-Saint-Jean population contain a single common splice mutation (designated IVS12+5g-a). Although this mutation can also occur in patients from other ethnic backgrounds, different mutations may be more prevalent in certain backgrounds (e.g., W262X in Finns and Q64H in Pakistanis). Genotype–phenotype correlations have suggested that severe mutations resulting in complete absence of FAH mRNA and protein are associated with early-onset acute disease; however, very different clinical courses have been reported within affected members of the same family, suggesting that epigenetic and environmental factors must also play a role in determining outcome.

CLINICAL PRESENTATION

Tyrosinemia should be considered in the differential diagnosis of any infant or child presenting with evidence of acute hepatocellular necrosis, conjugated hyperbilirubinemia, or decreased hepatic synthetic function of unknown etiology. Isolated or profound coagulopathy should also suggest the diagnosis. Children may also present with evidence of chronic disease with cirrhosis, nutritional rickets, renal tubular dysfunction, neurological crises, or failure to thrive. The differential diagnosis is that of liver failure in the infant and cirrhosis in the child or young adult. The natural history of tyrosinemia has been significantly modified by the development of prenatal screening for disease and pharmacological therapy with NTBC.

Acute Presentation

In the absence of prenatal testing or prior to, patients usually come to medical attention with acute evidence of liver disease before 2 years of age. This presentation may be preceded by a catabolic stress such as viral illness or bacterial infection. Symptoms are nonspecific, e.g., growth failure, irritability, vomiting, hepato- or nephromegaly, ascites, or edema. A boiled cabbage odor, related to elevated serum methionine levels, may be noted. Occasionally, presentation was associated with a bleeding diathesis manifesting as bruising or epistaxis. Untreated, the disorder leads to liver failure within 1 year.

Chronic Presentation

With chronic disease, liver dysfunction, renal tubular disease, and nutritional rickets may be present. A hallmark of the disease is the development of neurological crises that resemble acute intermittent porphyria. These are sudden-onset episodes of peripheral neuropathy with painful paresthesias, autonomic disturbance, extensor hypertonia, vomiting or paralytic ileus, and progressive weakness. Sometimes respiratory failure occurs, necessitating mechanical ventilation. The acute episode is followed by a variable recuperative period associated with weakness and paralysis. Cognition is not impaired and coma is not a feature of isolated neurological crisis but may suggest evolving hepatic encephalopathy. Tyrosinemia is also associated with a markedly increased risk of hepatocellular carcinoma even within the first 5 years of life with cases reported in patients less than 2 years of age. Renal disease is almost always present in patients with the chronic form of the disease, though its severity can be variable. Both tubular dysfunction and glomerular involvement occur. Rickets is the most common manifestation and is sometimes the primary medical problem. Tyrosinemia has also been associated with hypertrophic cardiomyopathy, as well as hypoglycemia associated with hyperinsulinism and pancreatic islet hypertrophy.

Laboratory Biochemical, Imaging, and Histopathological Abnormalities

Laboratory findings typical of acute presentation may be notable for disproportionate coagulopathy without other findings of liver disease. Transaminases and conjugated bilirubin may be near normal or mildly elevated. Plasma tyrosine and methionine levels are often markedly elevated but this is not a specific finding. Serum α-fetoprotein levels can be extremely elevated, even in excess of 100,000 ng/ml. Urinary abnormalities are consistent with proximal renal tubular dysfunction with aminoaciduria and elevated levels of urinary succinyl acetone are diagnostic. Neurological crises are generally not associated with perturbations of liver biochemistries or increased levels of succinyl acetone from baseline. Intra-abdominal imaging with ultrasound or computed tomography may show hepatomegaly with increased hepatic echogenicity as well as nephromegaly and nephrocalcinosis. Liver biopsy specimens show hepatocellular inflammation and necrosis, steatosis, and lobular distortion with pseudoacinar formation and marked nodular regeneration.

PATHOPHYSIOLOGY

Tyrosinemia results from deficiency of FAH, which catalyzes the last step in tyrosine degradation (Fig. 1). The molecular mechanisms responsible for the clinical

manifestations remain speculative. Hepatocellular and renal injury are thought to result from toxicity associated with *in vivo* accumulation of metabolites proximal to the enzyme defect, including fumaryl acetoacetate and maleyl acetoacetate and their by-products, succinyl acetoacetate and succinyl acetone. Fumaryl acetoacetate and maleyl acetoacetate are reactive unstable compounds that could theoretically consume the cellular machinery involved in protection from oxidative stress, thereby leading to tissue injury. Succinyl acetone is thought to be the most toxic intermediate. It can inhibit renal tubular transport, perhaps accounting for the proximal tubular dysfunction seen in this disease. Succinyl acetone has also been implicated as pathogenic in neurological crisis because it is a potent inhibitor of the enzyme δ-aminolevulinic acid dehydratase. This enzyme is involved in porphyrin biosynthesis, and its inhibition, which leads to the accumulation of δ-aminolevulinic acid, has been associated with neurotoxicity in acute intermittent porphyria, lead poisoning, and hereditary deficiency of the dehydratase.

Spontaneous reversion of the genetic defect in tyrosinemia has been shown to occur in nodular proliferating regions of tyrosinemic liver. In such regions, there is correction of the mutant FAH gene to wild type and detectable FAH enzyme activity. Fumaryl acetoacetate is known to be mutagenic, which may contribute to the increased frequency of genetic reversion in the diseased liver. The presence of revertant hepatocytes within proliferative nodules suggests that tyrosinemia induces a potent hepatic regenerative stimulus and that wild-type cells have a selective proliferative advantage over FAH-deficient cells.

Mouse models of FAH deficiency recapitulate much of the hepatic and renal pathophysiology that characterizes the human disease. In such mice, the disease phenotype can be reversed by re-introduction of a normal FAH transgene into the deficient background.

DIAGNOSIS

Elevated succinyl acetone in blood, plasma, or urine is pathognomonic for the disease. The diagnosis can also be made by demonstration of absent or decreased FAH activity on enzymatic assay of liver tissue or skin fibroblasts. Molecular screening for common alleles can be performed in at-risk populations, though negative results do not exclude rare mutant alleles. Prenatal diagnosis is possible by analysis of succinyl acetone in amniotic fluid and by FAH assay of cultured amniocytes or chorionic villus cells. Neonatal screening based on analysis of succinyl acetone from dried blood on filter paper is performed in Quebec.

TREATMENT

NTBC

NTBC treatment has greatly improved survival, eliminated neurological crises, and reduced the need for liver transplantation during early childhood for patients with acute tyrosinemia. It is now the cornerstone of treatment. Originally developed as an insecticide, it was later found to be a potent inhibitor of 4-hydroxyphenylpyruvate dioxygenase (Fig. 1). Thus, the hypothesis supporting its initial use in the treatment of tyrosinemia was that inhibition of this upstream enzyme in tyrosine metabolism would prevent accumulation of downstream, presumably toxic, metabolites, such as succinyl acetone. In 1992, Linstedt and colleagues tested this hypothesis by treating five tyrosinemia patients with NTBC. Treatment led to marked clinical, histological, and biochemical improvement and caused no neurological, ocular, or cutaneous complications. Therapy markedly reduced α-fetoprotein levels, suggesting significantly decreased nodular regeneration in the diseased liver. The results from treatment of larger numbers of patients over 9 years have shown that the vast majority of patients do very well. For patients in whom treatment with NTBC was started early in life, only two cases (1%) of hepatocellular carcinoma during the first year of life have been reported. If NTBC is started later in the course of disease, there appears to be increased risk for the development of hepatic malignancy and management decisions must be made on an individual case-by-case basis.

Dietary Restriction and Liver Transplantation

Prior to the development of NTBC, therapeutic intervention in tyrosinemia primarily involved dietary restriction of tyrosine and phenylalanine and liver transplantation. Hematin was used by some to treat neurological crises based on its ability to inhibit δ-aminolevulinic acid synthase.

Dietary restriction is still generally instituted at the time of diagnosis and continued throughout treatment with NTBC. The diet is stringent and based on a specialized formula that provides sufficient essential amino acids for growth but restricts tyrosine and phenylalanine. If the restriction is too great, growth failure, anorexia, and lethargy can result and so monitoring of growth and nutritional status while on the diet is important. Liver transplantation cures the metabolic liver disease and recurrent neurological crises and improves renal function. Prior to NTBC therapy, defining the optimal timing of transplantation was often challenging

owing to the high risk of development of hepatocellular carcinoma. Some would advocate transplantation early in infancy, whereas others adopted a more conservative approach. With the availability and success of NTBC in the management of tyrosinemia, liver transplantation is now likely to be restricted only to rare patients who present late or whose disease is severe and irreversible at the time of presentation. Nevertheless, NTBC therapy does not obviate the need for careful follow-up of affected patients and surveillance for hepatocellular carcinoma.

ANALYSES IN MOUSE MODELS

Analyses in FAH-deficient murine models of tyrosinemia also show that NTBC has a significant effect on the course of disease, though the correction is not complete as mice continue to show accumulation of intermediate levels of succinyl acetone and some develop hepatocellular carcinoma. Whether human patients with tyrosinemia on NTBC are at continued risk for the development of hepatocellular carcinoma is unknown. Studies in the FAH-deficient mouse model have also shown that hepatocellular transplantation can correct the disease state by repopulation of diseased liver with normal hepatocytes. Repopulation occurs only in the diseased liver and not in the livers of NTBC-treated mice. These data suggest that hepatocellular transplantation may one day be useful in the treatment of tyrosinemia in people and, perhaps, the treatment of other cell autonomous metabolic liver diseases as well.

See Also the Following Articles

Carbohydrate and Lactose Malabsorption • Food Intolerance • Galactosemia • Glycogen Storage Disease • Hereditary Fructose Intolerance • Liver Transplantation

Further Reading

Berry, G. T. (2000). Genetic and metabolic hepatic disorders. 2. Disorders of amino acid metabolism. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (W. A. Walker, P. R. Durie, J. R. Hamilton, J. A. Walker-Smith, and J. B. Watkins, eds.), 3rd Ed., pp. 1072–1075. B. C. Decker, Lewiston, NY.

Gishan, F. K. (2003). Inborn errors of metabolism that lead to permanent hepatic injury. *In* "Hepatology: A Textbook of Liver Disease" (D. Zakim and T. Boyer, eds.), 4th Ed., pp. 1425–1430. W. B. Saunders, Philadelphia, PA.

Grompe, M. (2001). The pathophysiology and treatment of hereditary tyrosinemia type 1. *Semin. Liver Dis.* **21**, 563–571.

Holme, E., and Lindstedt, S. (2000). Nontransplant treatment of tyrosinemia. *Clin. Liver Dis.* **4**, 805–814.

Mitchell, G. A., Russo, P. A., Dubois, J., and Alvarez, F. (2000). Tyrosinemia. *In* "Liver Disease in Children" (F. J. Suchy, R, J. Sokol, and W. F. Balistreri, eds.), 2nd Ed., pp. 667–685. Lippincott Williams & Wilkins, Philadelphia, PA.

Teckman, J. H., and Perlmutter, D. H. (1999). Metabolic disorders of the liver. *In* "Pediatric Gastrointestinal Disease: Pathophysiology, Diagnosis, Management" (R. Wyllie and J. S. Hyams, eds.), 2nd Ed., pp. 590–591. W. B. Saunders, Philadelphia, PA.

Ultrasonography

DAVID COSGROVE
Imperial College of Science, Technology, and Medicine and Hammersmith Hospital, London

adjuvant chemotherapy Chemotherapeutic treatment of malignancy administered after surgery or radiotherapy to reduce the risk of relapse.
bolus Intravenously injected dose of a contrast agent, administered rapidly to produce a short contrast effect.
color Doppler Colored display of blood flow.
Doppler Frequency shift caused by blood flow.
echogenic Returning a high level of echoes.
elasticity imaging Process in which the stiffness of a tissue under compression is turned into an image.
endoscopy Scans taken from within a body cavity, typically the esophagus or stomach, using specially designed small transducers.
hydrocolosinography Examination of the colon after it has been filled with fluid.
microbubbles Gas bubbles ranging in size from 1 to 10 μm that are used as contrast agents for ultrasound.
spectral Doppler Display of the constituent components of the Doppler signal over time.
transjugular intrahepatic portosystemic shunt Passage positioned so as to allow portal blood access to the hepatic veins or inferior vena cava, bypassing the liver in portal hypertension.

Ultrasound has undergone very rapid developments in recent years; in addition to improved image quality and Doppler sensitivity, the introduction of contrast agents in the form of injectable microbubbles has expanded the horizons of ultrasound applications. Many of these advances have particular significance for gastroenterology.

GASTROINTESTINAL TRACT

Ultrasound is useful in all parts of the gastrointestinal tract, allowing imaging of the five-layered structure the wall throughout the entire tract. For the esophagus and stomach, endoscopic ultrasound (EUS) probes have been developed; EUS is an excellent compliment to endoscopy because it "sees" beyond the mucosa. EUS achieves a 75% accuracy in the local and nodal staging of carcinoma of the esophagus and is very reliable in differentiating gastric leiomyomas from carcinomas because the former have a characteristic appearance as well demarcated echo-poor masses. Staging of carcinoma in the stomach is less successful than in the esophagus.

Endoscopic systems are also useful for studying the anorectal region, especially for sphincter injuries. Magnetic resonance imaging (MRI) with intracavitary coils gives similar information in this important problem and is superior for detecting fistulas. For the remainder of the gut, transabdominal scanning with high-frequency transducers is used. Thickening of the colonic wall >3 mm is readily demonstrated and allows cancers to be detected as localized masses, in contradistinction to the extensive wall thickening in Crohn's disease, in which fistulous tracts and paracolonic abscesses can also be delineated. In ulcerative colitis, only the mucosal layer is thickened; the terminal ileum is selectively thickened in terminal ileitis, a common cause of right iliac fossa pain in young people in whom the associated regional lymphadenopathy may also be demonstrable. Diverticuli are seen as gas-containing structures lying within, or adjacent to, the colon wall. Intussusception produces a characteristic multilayered mass corresponding to the triplicated walls of the gut layers (Fig. 1). Ultrasound (US) can also be used

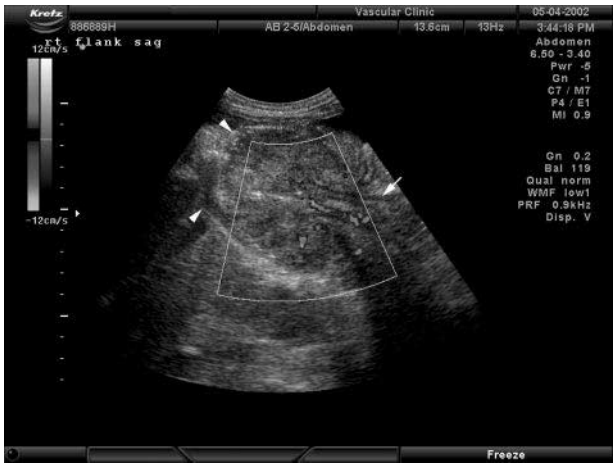


FIGURE 1 Intussusception. This patient presented with severe diffuse abdominal pain; ultrasound revealed a mass in the right flank into which the vascularized terminal ileum passes (arrow) Stretched around the hypovascular mass is a thin rim of stretched ascending colon (arrowheads). The mass proved to be cecal carcinoma, which had formed the lead of the intussusception.

 Copyright 2004, Elsevier (USA). All rights reserved.

to monitor hydrostatic reduction, thus limiting exposure to ionizing radiation.

Pyloric stenosis can often be diagnosed on clinical grounds, but in doubtful cases the thickened pyloric wall is easily detectable with US (Fig. 2). This, coupled with the demonstration of active gastric peristalsis that peters out as it reaches the pylorus, makes ultrasound a very reliable diagnostic tool. Another pediatric problem that is easily solved with US is malrotation of the small bowel. Using Doppler to determine the relative positions of the superior mesenteric artery and vein (SMA and SMV) by depicting their flow direction, the abnormal relationship of the artery and vein can be demonstrated.

The normal appendix can be demonstrated in some two-thirds of patients by using high-resolution transducers with graded compression to displace the gas that otherwise prevents acoustic access. The technique involves first localizing the cecal pole by tracing the ascending colon down into the right iliac fossa. Then, using a linear or curved array, gentle pressure is applied over several minutes to allow access to the appendix, which is seen as an elongated layered structure less than 7 mm in diameter. The inflamed appendix is thickened and the fecolith that is often present can be demonstrated as an echogenic nodule that casts an acoustic shadow. The inflammatory reaction produces an increase in Doppler signals (Fig. 3). Thickened omentum may be demonstrated as an amorphous mass of medium

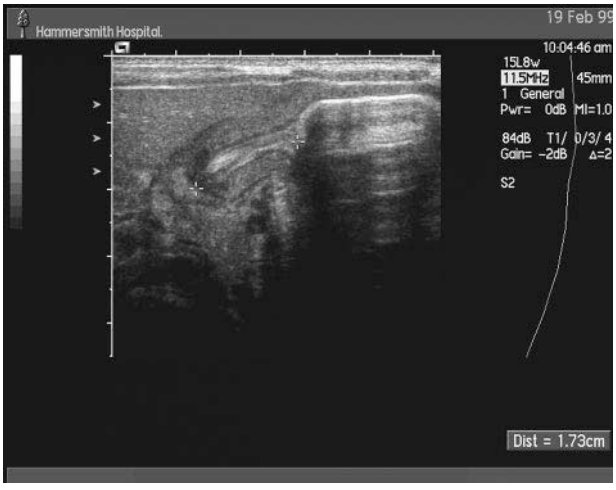


FIGURE 2 Pyloric stenosis. In a 6-week-old child who presents with projectile vomiting, the demonstration of hypertrophy of the pyloric canal is definitive. Here the pylorus measures 17 mm in length (normal, <15 mm). In addition, on the real-time study, peristaltic waves could be seen advancing along the antrum but failing to move contents into the duodenum.

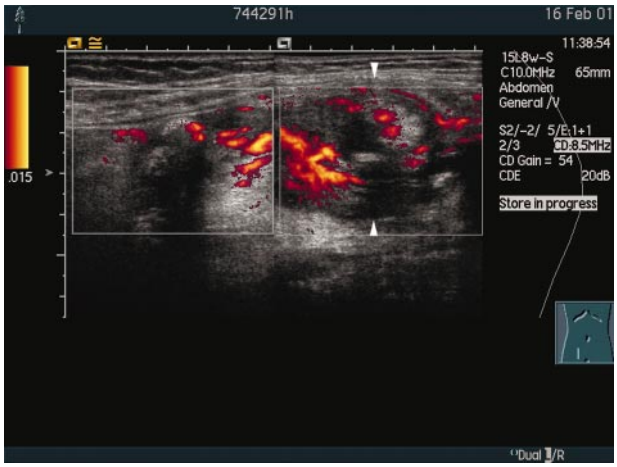


FIGURE 3 Appendicitis. In this patient with right iliac fossa pain, the hyperemic mass on ultrasound (arrowheads) corresponded with the point of maximum tenderness. The mass is the dilated inflamed appendix lying immediately inferior to the gas-containing cecum. There is no periappendicular collection to suggest an abscess. An uncomplicated acutely inflamed appendix was removed laparosopically.

echoes close to the appendix, whereas a perforation produces an adjacent fluid space. In right iliac fossa (RIF) pain, US can be used to diagnose terminal ileitis and distinguish it from gynecological causes.

Colonic polyps cannot be demonstrated by conventional ultrasound, but if the colon is first filled with water as a contrast agent, then exquisite images that clearly depict polyps down to 5 mm in diameter can be obtained using high-frequency transducers. However, the moderately invasive nature of "hydrocolosonography" and the fact that sessile or plaque-like tumors are not reliably visualized have limited the acceptance of the technique. In mesenteric angina, the tight stenosis or occlusion of the celiac axis and/or SMA produce localized fast flow on spectral Doppler or absent signals, respectively (Fig. 4).

THE LIVER

The normal liver has a uniform texture of mid-gray reflectivity, interrupted by its blood vessels (Fig. 5). Diffuse liver diseases usually cause hepatomegaly; an exception is found in cases of advanced cirrhosis, for which shrinkage is typical, although judging liver size with ultrasound is very difficult. The commonest change in diffuse liver diseases is an increase in the echo level, providing an example of "bright liver," which is recognized by comparison with echoes from the renal cortex. Normally, the two are approximately

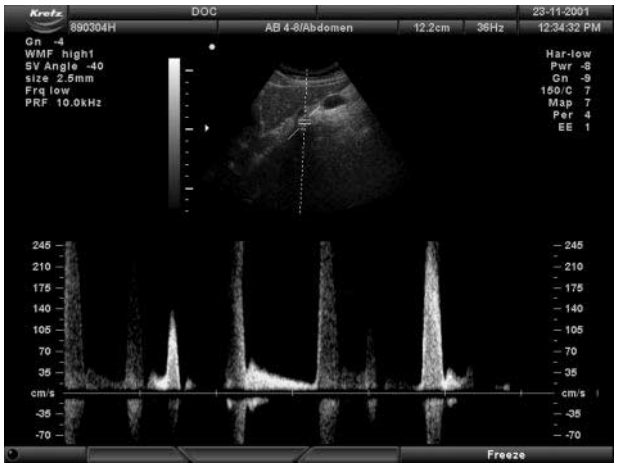


FIGURE 4 Celiac axis stenosis. Celiac axis stenosis in isolation is rarely symptomatic because of the generous anastomoses in the gut's blood supply, but, because it usually occurs in arteriopaths, two or more of the arteries are often affected. The acceleration of the blood crossing the narrowing produces fast flow on the Doppler study (here rising to more than 4 m/second, shown in the lower panel) together with flow disturbance in the form of eddies that give slower flow signals as well as the very high-velocity signals. These result in a tracing in which all velocities are more or less equally represented, so that it appears as solid white from baseline to maximum velocity.

equally reflective, but in the bright liver, the renal cortex appears to be relatively dark. The bright liver as seen in a wide range of liver diseases, the commonest being fatty infiltration, but the list includes granulomatous and fibrotic conditions as well as cirrhosis and infiltrations by amyloid. The opposite pattern, "the dark liver," is less common, and is seen in conditions in which the amount of fluid in the liver is increased. Examples are acute hepatitis and congestive heart failure.

Cirrhosis is a progressive disease and the ultrasound findings reflect this; early on, there may be no demonstrable abnormalities or there may simply be an increase in reflectivity. In more advanced cases, the liver shrinks and the caudate lobe (segment 1) enlarges. The micro- or macronodularity of cirrhosis can be visualized in ascites, when the liver surface assumes a scalloped configuration (Fig. 6). Occasionally, regenerating nodules can be seen as ill-defined, patchy alterations in reflectivity, but usually the liver texture is merely slightly heterogeneous. The major changes demonstrable on ultrasound concern the blood flow, for which Doppler is well suited. The arterialization of the liver's supply makes the hepatic artery prominent, whereas flow in the portal vein may be normal or slow. In extreme cases, this is reversed, and in an intermediate stage, hepatic vein flow is balanced, being hepatofugal in one stage of the respiratory cycle and hepatopetal in another. Obviously, this situation of zero net flow is likely to lead to portal vein thrombosis and this may be demonstrated as echogenic material filling the vein together with an absence of Doppler signals. Subsequent recannalization or opening up of collateral vessels

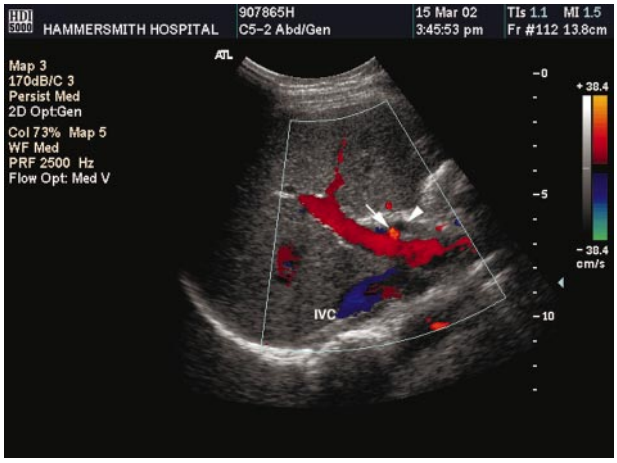


FIGURE 5 Normal liver—porta hepatis. In a longitudinal section through the porta hepatis, the uniform texture of the liver parenchyma can be seen. In the color Doppler scan, the portal vein is depicted in red, indicating flow toward the transducer. The hepatic artery (arrow), also in red, can be discerned anterior to it, and the dark circle (arrowhead) is the right hepatic duct seen in cross-section. The position of the duct represents a normal variant: usually it lies between the vein and the artery. IVC, Inferior vena cava.

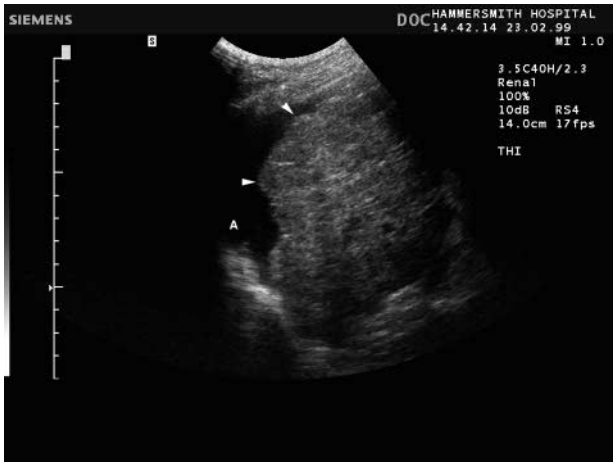


FIGURE 6 Cirrhosis. In late cirrhosis, the irregular liver surface can be visualized against the echo-free ascites (A). In this case, the outline of macronodules (arrowheads) gives a scalloped border to the liver. Its echo texture is heterogeneous (compare with the normal, even texture of the normal liver in Fig. 5).

produces "cavernous transformation" of the portal vein, which gives a spectacular color Doppler picture of numerous small tortuous vessels in the porta.

The expected slowing of portal vein flow does not always occur because intrahepatic portosystemic shunts may form, and thus measurement of portal vein flow has not proved valuable in diagnosis or grading of cirrhosis. The most dramatic shunt is recannalization of the umbilical vein beyond the free margin of the liver (Fig. 7). Other portosystemic shunts may also be demonstrated, including retroperitoneal shunts and esophageal and spleno-esophageal varices. Because ultrasound demonstrates serosal varices whereas those seen on endoscopy are submucosal, the findings on these two modalities do not always correspond. Flow in transjugular intrahepatic portosystemic shunts (TIPSs) can usually be detected, but the shadowing from the metal of the stent and slow flow within the shunt may challenge the performance of even the best ultrasound scanner (Fig. 8). Here, administration of an ultrasound contrast agent can rescue an otherwise failed study.

Of the effects of portal hypertension, spleen size and ascites are very easily assessed by ultrasound. Supervening hepatocellular carcinoma may be obvious but small and multicentric lesions are difficult to demonstrate and some regenerating and hyperplastic nodules can produce masses that mimic hepatocellular carcinomas (HCCs). However, these benign lesions do not generally show Doppler signals, so hypervascular lesions are suspicious (Fig. 9). Sensitivity to hepatocellular

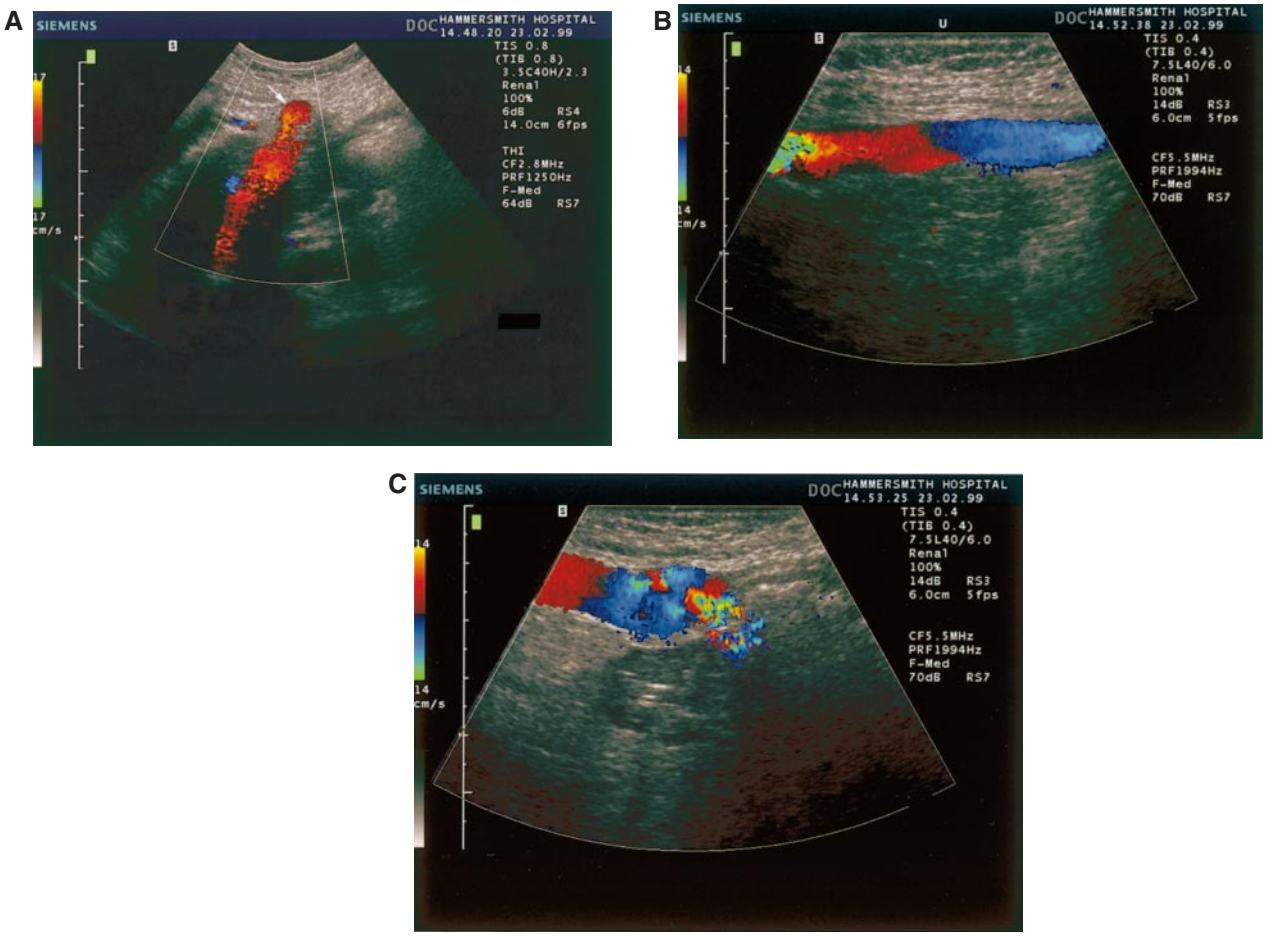


FIGURE 7 Recannalized umbilical vein. Recannalization of the umbilical vein produces this spectacular appearance on color Doppler, with signals in the position of the ligamentum teres (arrow in A) flowing toward the umbilicus (U in B) and sometimes, as in this patient, on down into the pelvis [the change from red to blue codes for a relative change in flow direction, which is toward the probe superiorly (A), then parallel (B), and finally away from it (C)]. Often this spontaneous portosystemic shunt depressurizes the system and there are no other features of portal hypertension, such as ascites.

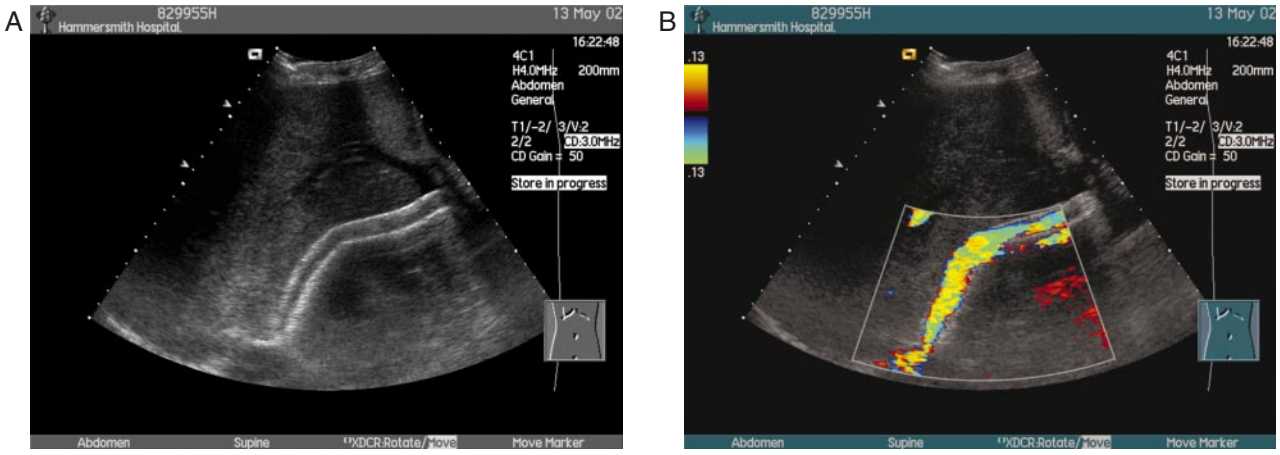


FIGURE 8 This long transjugular intrahepatic portosystemic shunt is well visualized (A) and the flow is clearly seen to be toward the hepatic vein on the Doppler study (B).

carcinomas is improved by microbubble contrast agents, which demonstrate the HCC vascularity in the arterial phase. In common with other malignancies, HCCs have a lower vascular volume compared to the liver and so they appear as defects in the late (sinusoidal or liver-specific) phase (Fig. 10). In schistosomiasis, a characteristic finding is prominence of the portal vein walls, corresponding to the per portal fibrosis of pipe stem cirrhosis. In the Budd–Chiari syndrome, lack of Doppler signals in the main hepatic veins is diagnostic. There is often reversed flow in the portal vein with ascites. If the thrombosed vessels recannalize, a thready meshwork replaces the normal pattern of three main veins (Fig. 11). Its microvascular equivalent, veno-occlusive disease, cannot be diagnosed with current ultrasound techniques.

FOCAL LESIONS

A wide variety of focal liver lesions can be diagnosed by ultrasound, notably cysts, for which US is the most specific and sensitive test. Hydatid cysts have a variety of appearances, depending on the condition of their contents, but consistently have a prominent capsule that is lacking around simple cysts. In endemic countries, ultrasound-guided aspiration (perhaps with injection of a sclerosant) is widely used for symptomatic control with good safety.

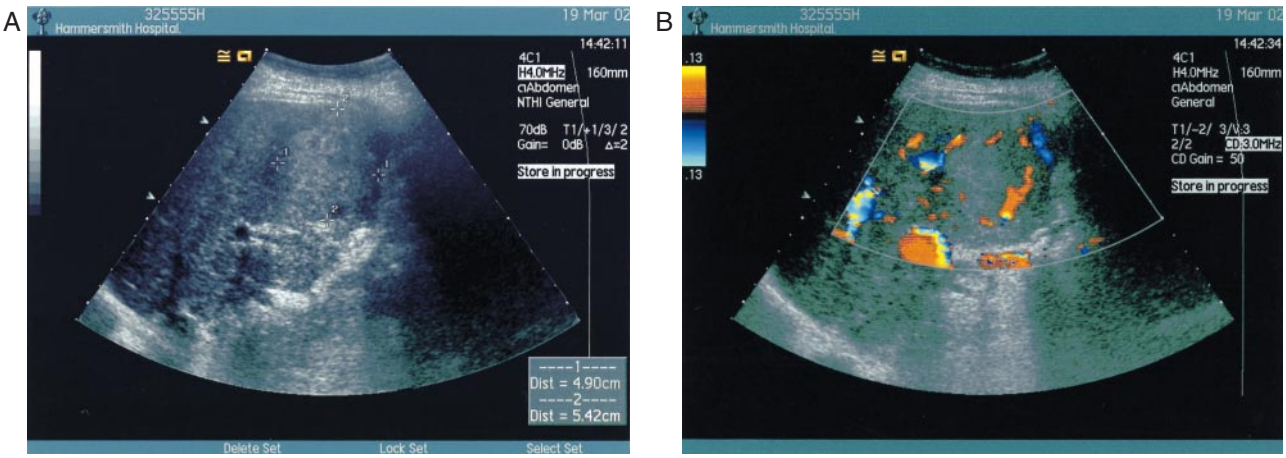


FIGURE 9 Hepatocellular carcinoma. This irregular 5-cm mass in segment V in a cirrhotic patient (A) is suggestive of an hepatocellular carcinoma, and this is reinforced by the demonstration of numerous tortuous blood vessels within it on color Doppler (B).

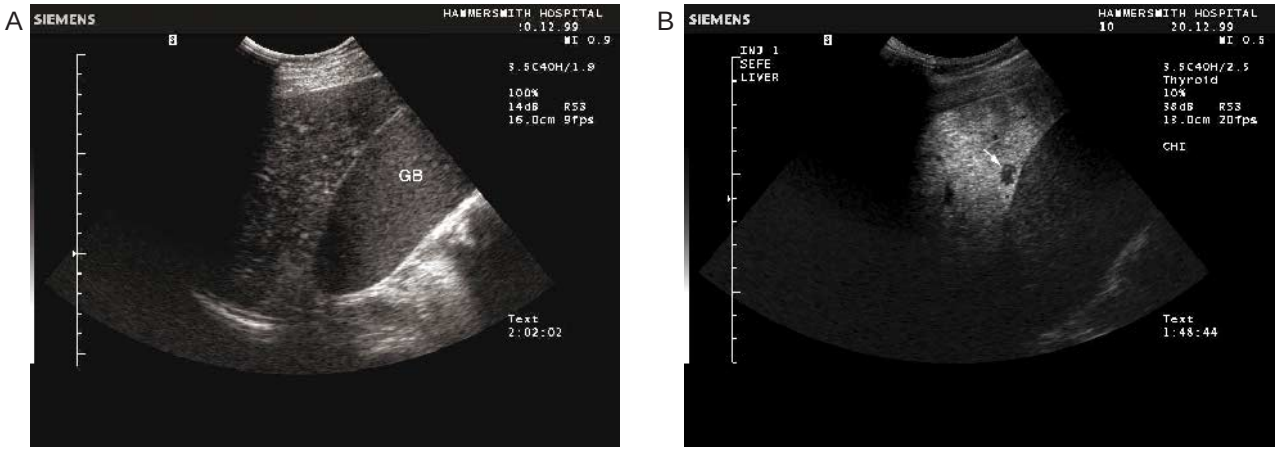


FIGURE 10 Occult hepatocellular carcinoma—contrast study. In this patient with cirrhosis, followup scans were performed to monitor the development of hepatocellular carcinoma. Only minor irregularity of the liver's texture was seen on the scan (A). He was entered into a phase 3 study of a new contrast agent, Sonazoid (Amersham, UK), and in the liver-specific phase a clear-cut lesion was revealed (arrow in B). Though the α-fetoprotein level was normal, an enhanced MRI study was performed and was read as normal. Three months later, the lesion became visible on unenhanced ultrasound and was confirmed on MRI and by biopsy to be hepatocellular carcinoma. Radiofrequency ablation was performed. GB, gallbladder.

Abscesses are typically seen as shaggy-walled cavities that are often multiple (Fig. 12). On Doppler, the hyperemia of the surrounding tissue is often obvious. However, in the initial stages of abscess formation, before pus has collected, a solid mass is found and the lesion can be very subtle.

Hemangiomas are the commonest benign liver tumor and typically appear as echogenic masses. On Doppler, they are hypovascular, only occasionally showing weak venous signals. In the late vascular phase after contrast injection, they typically show the same peripheral clumping that is characteristic on CT, with slow and often incomplete fill-in from the periphery over several minutes (Fig. 13). Focal nodular hyperplasia contains normal liver elements in an abnormal arrangement. Some have a vascularized central scar seen as an echo-poor streak with Doppler signals that typically radiate outward in a spoke-wheel fashion, and these arterial features are elegantly demonstrated after

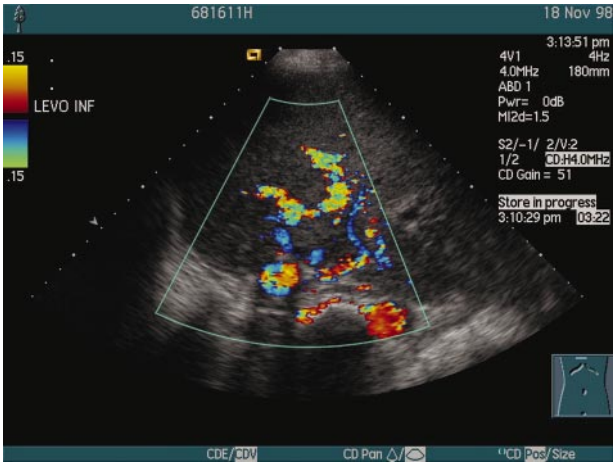


FIGURE 11 Budd–Chiari syndrome. Recannalization of the thrombosed hepatic veins leads to a tangled pattern of vessels in the superior part of the liver.

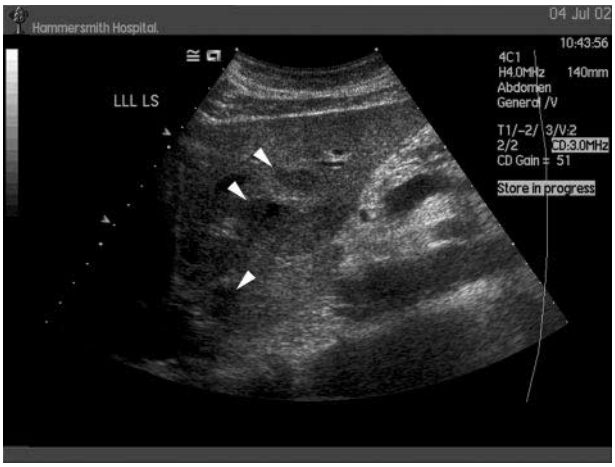


FIGURE 12 Liver abscess. A region of altered reflectivity (arrowheads) is seen in the left liver in this patient presenting with a fever. The appearance is typical of an abscess in the early (preliquefactive) phase of the development of an abscess. The ultrasound appearance is indistinguishable from metastatic disease, but the history usually provides the necessary clues and ultrasound-guided aspiration should be performed.

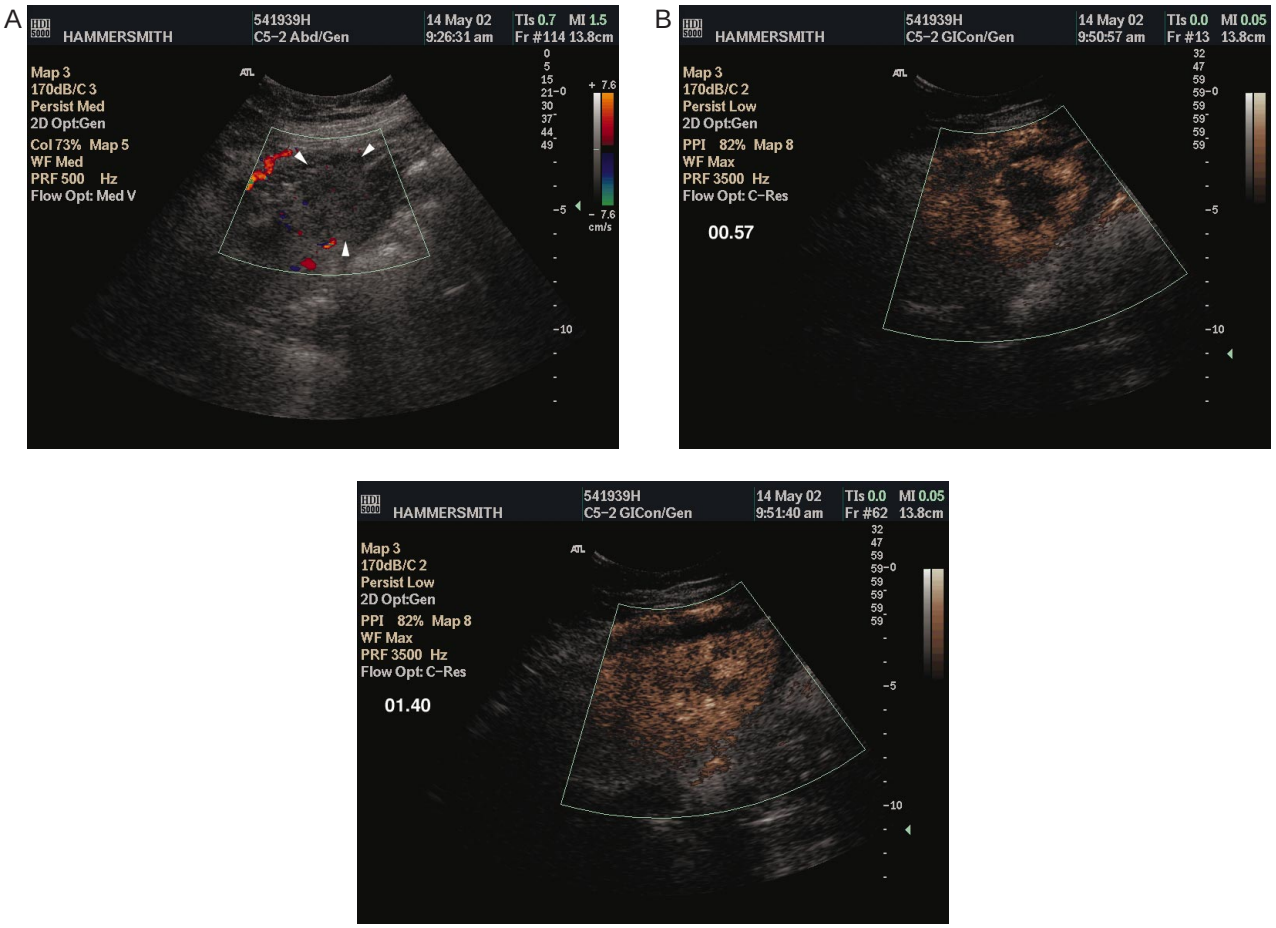


FIGURE 13 Hemangioma contrast study. (A) This lesion in the inferior portion of segment V (arrowheads) in a patient with a known malignancy is suspicious for a metastasis because it is echo poor. (B) Soon after injection of a microbubble contrast agent (SonoVue, Bracco, Italy), clumpy peripheral filling is observed (contrast shown in red at 50 seconds after injection; and thereafter there is progressive centripetal enhancement, until (C) at 1 minute, 40 seconds, when the lesion is almost completely enhanced and has become much less conspicuous. This dynamic pattern is typical of a hemangioma.

contrast enhancement. In the liver-specific phase of microbubble enhancement, their behavior is pathognomomic: they take up contrast strongly and so blend with the normal liver or occasionally even stand out with more intense signals (Fig. 14).

The variety of appearances of metastases on ultrasound is bewildering and remains largely unexplained (Fig. 15). Though there are trends (for example, echogenic metastases are typically gastrointestinal or urogenital in origin), it is not possible to tie particular patterns with the primary site. The success of ultrasound in liver staging has been exaggerated in the past, quoted figures of accuracies around 80% being calculated on a per patient basis rather than for individual lesions. These limitations have been highlighted by the improvement in detectability when ultrasound contrast agents are used. In several prospective studies, use of microbubble-specific modes allow the detection of lesions down to a diameter of 3 mm (Fig. 16). Though not yet perfected (a depth limitation of around 10 cm is an important limitation), this approach reveals more lesions than three-phase helical computed tomography (CT) and comes close to the sensitivity of MRI. Specificity is also addressed because malignancies, with their low blood volume, show as defects in the sinusoidal or liver-specific phases, and in some the feeding artery is obvious in the arterial phase (Fig. 17).

A particularly useful application of contrast agents is in evaluating the completeness of ultrasound-guided interstitial therapy: when all tumor appears to have

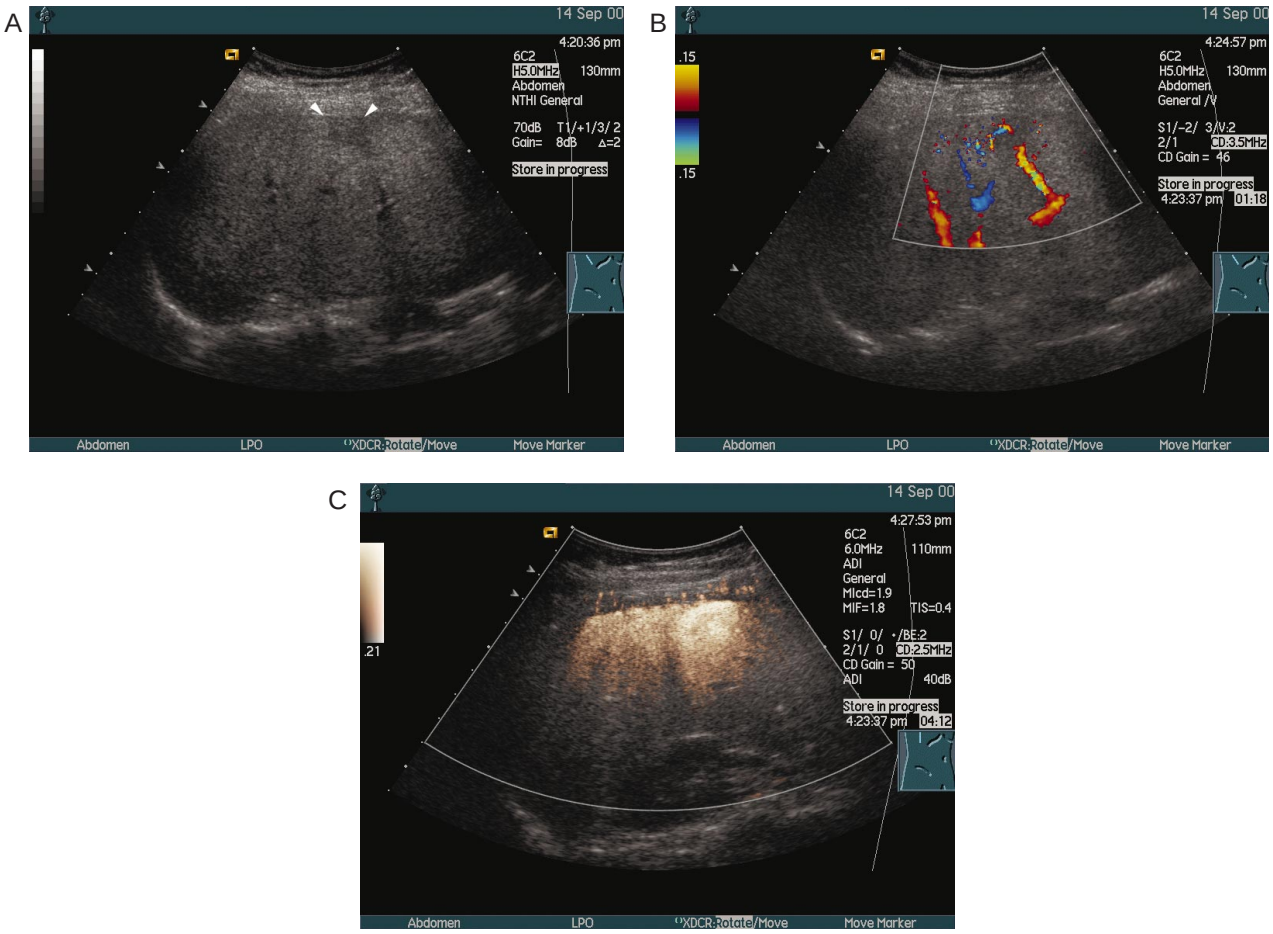


FIGURE 14 Focal nodular hyperplasia. (A) This focal lesion in the anterior part of segment 5 (arrowheads) was an incidental finding. (B) Its ill-defined margins and vascularity (color Doppler) raised the suspicion of malignancy. However, following administration of the liver-specific contrast agent Levovist, nonlinear imaging shows that it takes up the agent to the same extent as the liver and it has disappeared (C). This makes the diagnosis of focal nodular hyperplasia most likely. The lesion remained unchanged a year later.

been destroyed, microbubbles often reveal residual portions of perfused tumor that can be ablated immediately, so that the patient does not have to be moved to CT. Microbubbles can also be used as tracers and in the liver the time taken for them to cross into the hepatic veins (normally 30 seconds or more, owing to the slow flow in the liver sinusoids) is shortened when there is arteriovenous shunting, as occurs in metastases and cirrhosis. This simple test seems to be able to detect occult metastases (for example, in colorectal cancer), and should prove useful in selecting those patients who would benefit from adjuvant chemotherapy. It is also very promising in differentiating uncomplicated chronic hepatitis from cirrhosis and this might avoid the frequent biopsies that are currently required to guide antiviral treatment in patients with hepatitis C (Fig. 18). In liver trauma (and that of other solid abdominal organs), conventional ultrasound is limited to detecting intraperitoneal bleeding. The advent of contrast agents that allow the microvasculature to be demonstrated promises to change that, and to provide useful imaging by the patient's bedside, without interrupting observations and supportive care (Fig. 19).

BILIARY TREE

The exquisite sensitivity of ultrasound to liquid spaces has made it the primary imaging modality for the biliary tree and the standard test for gallstones (Fig. 20). Stone composition does not affect detectability. Occasional false positives occur when polyps are mistaken for

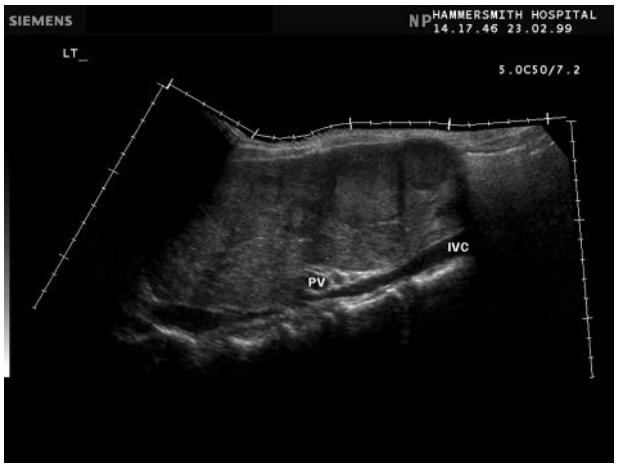


FIGURE 15 Liver metastases. This extended field-of-view (Sciescape, Siemens) scan of an enlarged left liver exhibits numerous focal lesions of varying appearances, some echo poor, others echogenic. The patient had a carcinoma of the colon (minor scale divisions equal 1 cm). IVC, Inferior vena cava; PV, portal vein.

stones; their attachment to the gallbladder wall provides a clue.

In acute cholecystitis, major signs such as splitting of the thickened wall, tenderness localized to the gallbladder (the ultrasound Murphy's sign), and the presence of stones, coupled with minor signs such as dilatation of the gallbladder, are useful (Fig. 21). Ultrasound may also demonstrate alternative causes for right upper quadrant pain. Acalculus cholecystitis remains a difficult diagnosis but the demonstration of vascularity within the wall on Doppler can be useful. In critical cases, ultrasound can be used to direct a draining cannula as a holding measure.

Carcinoma of the gallbladder is seen as a mass, unfortunately often large, with invasion of the adjacent liver surface (Fig. 22). Hilar cholangiocarcinomas (Klatskin tumors) have infiltrating margins that are ill defined on imaging so it may be very difficult to determine the extent of the tumor. Here, microbubble contrast agents have contributed substantially because they outline the limits of the functioning liver, against which the tumor is clearly seen as a signal void, as with other liver malignancies.

Ultrasound is the primary imaging technique in obstructive jaundice because US sensitivity to fluid spaces is exploited. The diameter of the common duct is easily measured. Dilatation is evidence of obstruction and, provided it is appreciated that this anatomical feature may persist after relief of obstruction, is a highly reliable sign in the appropriate clinical setting (Fig. 23). Similar considerations apply to dilatation of the intrahepatic biliary tree, which produces the classic "parallel channel" and "double-barreled shot-gun" signs when the pairs of tubular structures are cut along or across, respectively (Fig. 24). Ultrasound is very sensitive for "surgical" jaundice, provided the history is taken into account, because the biliary tree may remain dilated indefinitely, even after relief of an obstruction. For the same reason, ultrasound cannot be used to tell whether a biliary stent has reobstructed, though the demonstration of aerobilia provides useful indirect

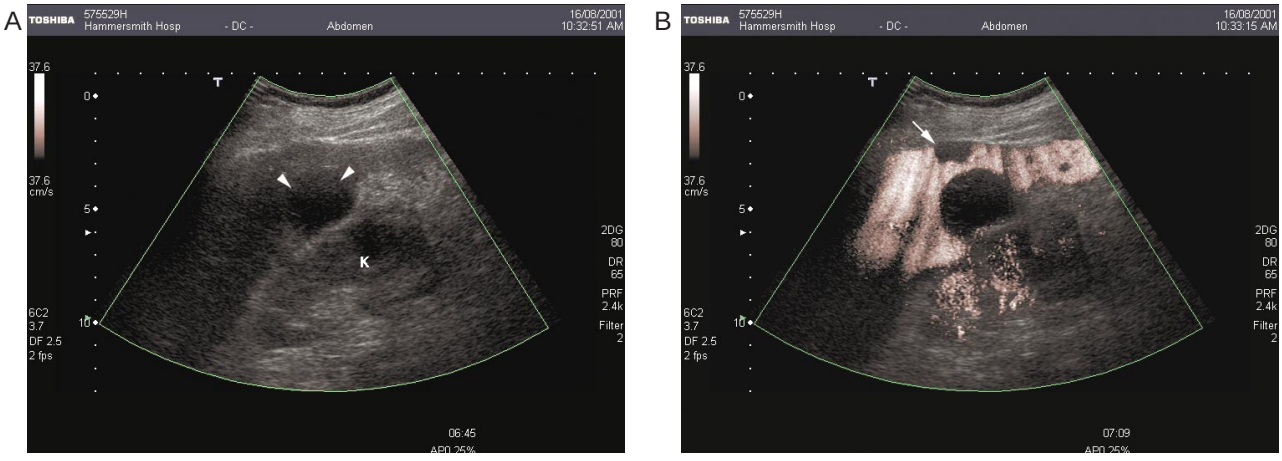


FIGURE 16 Liver metastasis contrast study. (A) In this patient with a known colonic cancer, an obvious lesion compatible with a metastasis is seen in the inferior part of segment 5 (arrowheads); K, kidney. (B) In a registered image taken in the liver-specific phase after injection of the contrast agent Levovist (Schering, Berlin), using a liver-specific mode (Agent Detection Imaging) that shows the presence of contrast in color, this lesion is more clearly delineated, but an additional lesion that was not apparent before is also obvious (arrow).

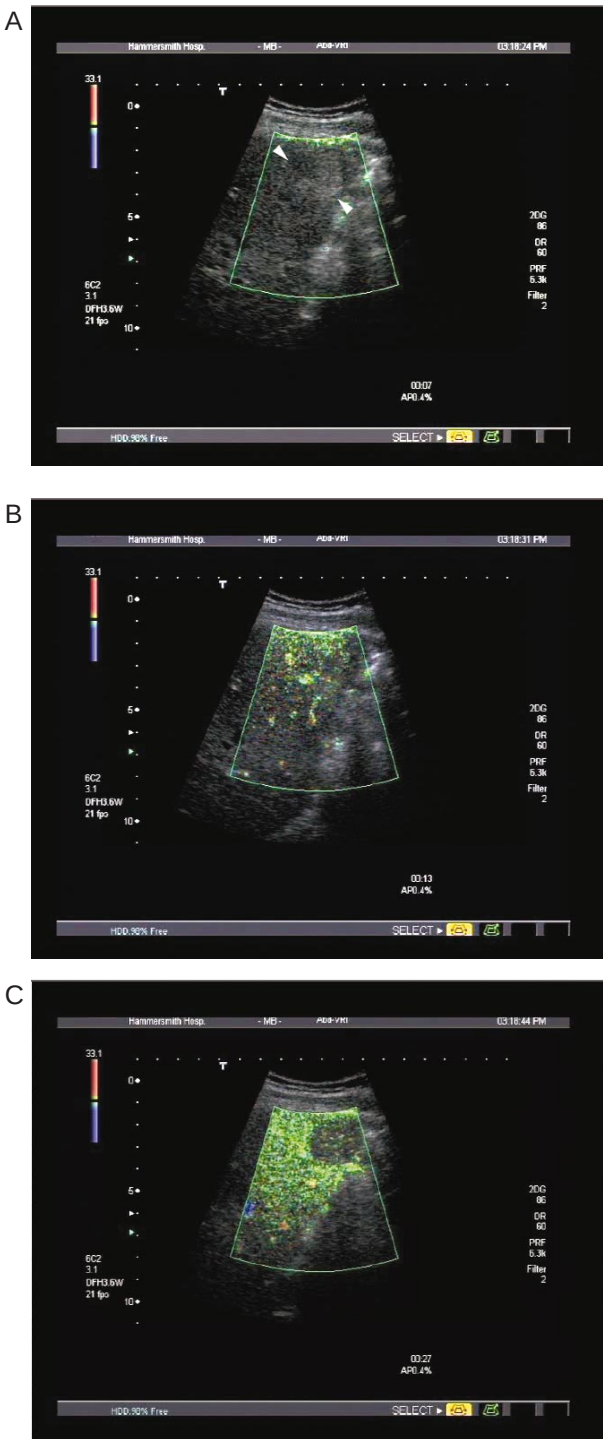


FIGURE 17 Liver metastasis contrast study. (A) A metastasis from a colonic carcinoma is seen as an echogenic mass (arrowheads) at the beginning of an injection of a microbubble contrast agent, SonoVue (7 seconds postinjection). (B) In the next image from the real-time sequence at 13 seconds postinjection (arterial phase), numerous small vessels enter the tumor from its periphery. In this microbubble-specific mode (Vascular Recognition Imaging, Toshiba, Tokyo), the contrast is depicted in green when it is stationary; the red and blue tints show moving microbubbles. (C) In the third frame, at 27 seconds, contrast is filling the liver parenchyma but the metastasis, with its low vascular volume, is highlighted as a signal void.

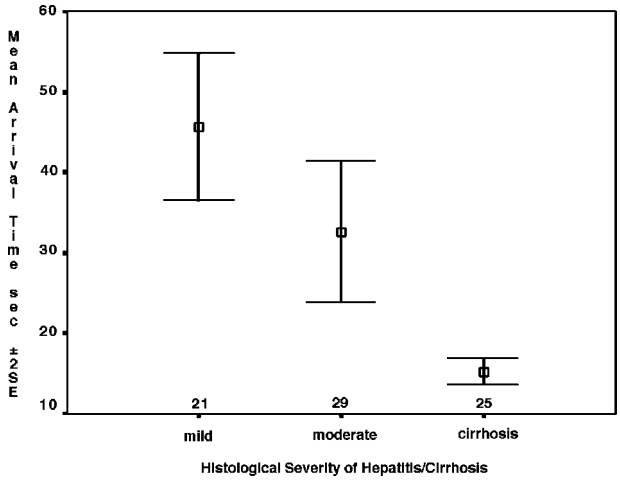


FIGURE 18 Hepatic vein arrival time. By noting the arrival of a bolus of a contrast agent in the hepatic veins, the delay from the moment of injection can be measured. When arteriovenous shunts develop in the liver, the arrival time is shortened from its normal of around 45 seconds. This test can detect the onset of cirrhosis in patients with chronic hepatitis showing mild and moderate changes in necroinflammatory scores on histology. The numbers above the *X* axis refer to the numbers of cases. Courtesy of Dr. Adrian Lim.

evidence of patency. Ultrasound can usually indicate the level of an obstruction by demonstrating the lowest point of dilatation, but is less useful in determining the cause, mainly because strictures cannot be detected, but also because many stones simulate soft tissue masses with no shadowing.

PANCREAS

Long considered a challenge for ultrasound, the normal pancreas can usually be imaged, at least in part, by a combination of modern equipment and graded compression, using the probe to displace bowel gas (Fig. 25). Oral contrast agents may further improve access in the epigastrium, whereas the pancreatic tail can be imaged using the spleen as a window. The pancreas is relatively echogenic in the adult, though the uncinate process commonly retains the lower levels of echo intensity typical of the child's pancreas. The

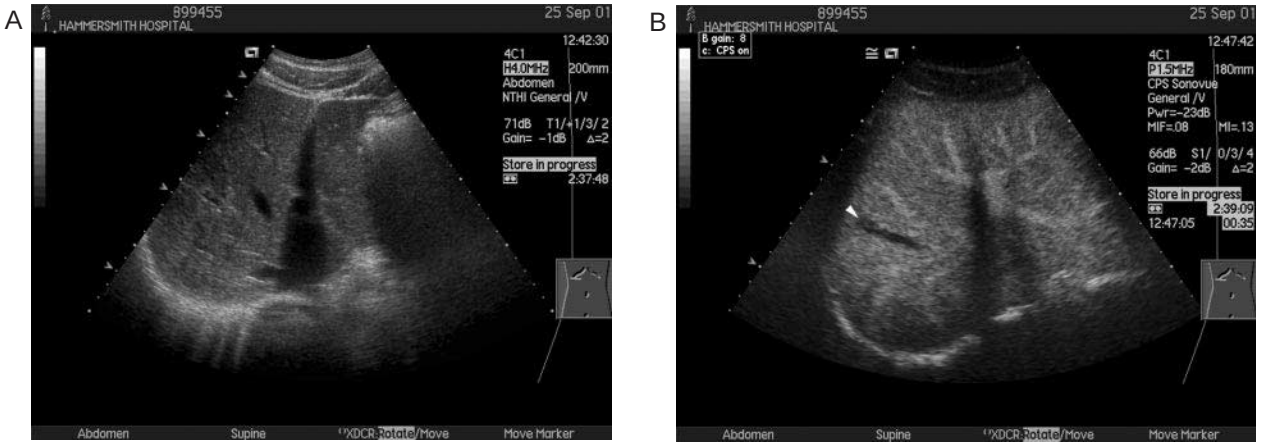


FIGURE 19 Stab wound of the liver. This patient suffered a penetrating injury to the right liver; despite knowing exactly where to look, it could not be demonstrated on grayscale ultrasound (A). Following administration of a contrast agent (SonoVue, Bracco, Milan), the lesion was clearly shown (arrowhead in B).

main pancreatic duct is seen as "tram lines" with a maximum caliber of 3 mm.

In acute pancreatitis, the gland swells, losing its normal shape, which has a waist at its neck, and becomes echo poor, but in practice, acute pancreatitis is a clinical diagnosis. An ultrasound scan should be obtained early in the course of the disease to look for gallstones. In chronic pancreatitis, there may be no changes on ultrasound until fibrosis and calcification supervene (Fig. 26). Dilatation and tortuosity of the main duct can be demonstrated.

Carcinoma of the pancreas produces a mass of variable echogenicity, often accompanied by duct dilatation and also by dilatation of the bile duct (the "double-duct sign") when, as is commonest, the tumor arises at the head of the pancreas (Fig. 27). Sometimes the diagnosis is obvious, with the mass perhaps involving the portal vein and lymphadenopathy, but often the features cannot be distinguished from focal pancreatitis. Endoscopic ultrasound offers some advantages but is, of course, more invasive.

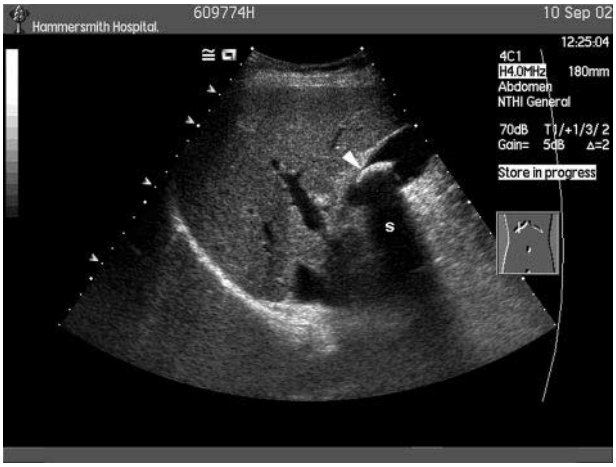


FIGURE 20 Gallstone. The strongly echogenic structure (arrowhead) in the gallbladder and its tell-tale shadowing (S) are diagnostic of a gallstone.

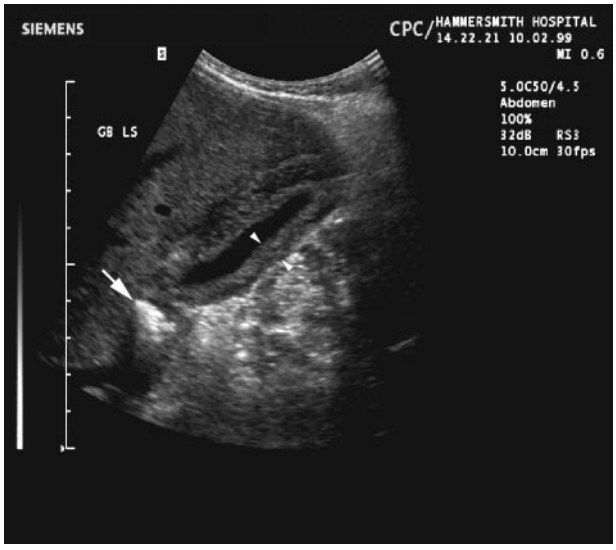


FIGURE 21 Acute cholecystitis. Wall thickening (arrowheads), often with splitting from edema, is typical in acute cholecystitis. There is usually local tenderness under the probe, which may be obvious in real time. The offending stone (arrow) is visible.

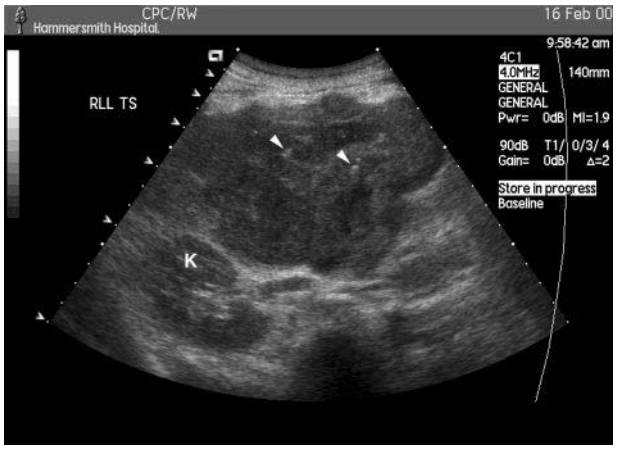


FIGURE 22 Carcinoma of the gallbladder. A large mass occupies the gallbladder fossa in this patient, who presented with jaundice. Embedded within the mass are several calculi (arrowheads). K, Kidney.

Of the rarer forms of pancreatic tumor, cystic carcinomas may produce a multicystic mass, though in the microcystic type the individual cysts are usually too small to be resolved and instead their numerous walls produce an echogenic mass. Papillary tumors are almost never visualized with ultrasound because they present at too small a size. Endocrine tumors are seen as echo-poor masses, but conventional scanning usually does not demonstrate hormonally active tumors because of their small size. However, the superior resolution of intraoperative scanning can reveal them and this is very useful to locate impalpable tumors (Fig. 28).

NEW METHODS

The clinical impact of the advent of contrast agents for ultrasound has been described. Contrast agents also offer both anatomical information for delineating and characterizing focal lesions and functional information, derived by their use as tracers in tracking the transit of a bolus across a region of interest. In this role, they are supported by the small volumes required (0.5–5 ml is typical) and by the fact that new scanning modes are specific to the microbubbles, so that they can be almost completely separated from tissue signals and displayed as perfectly registered fusion images. In addition, they have a relatively short persistence after injection (2–10 minutes), so that repeat studies are feasible, and they are nontoxic. The fact that they can be selectively destroyed by acoustic pressures toward the top of the acceptable diagnostic range offers the unique opportunity for negative bolus studies that display the reperfusion of a tissue to give true hemodynamic information.

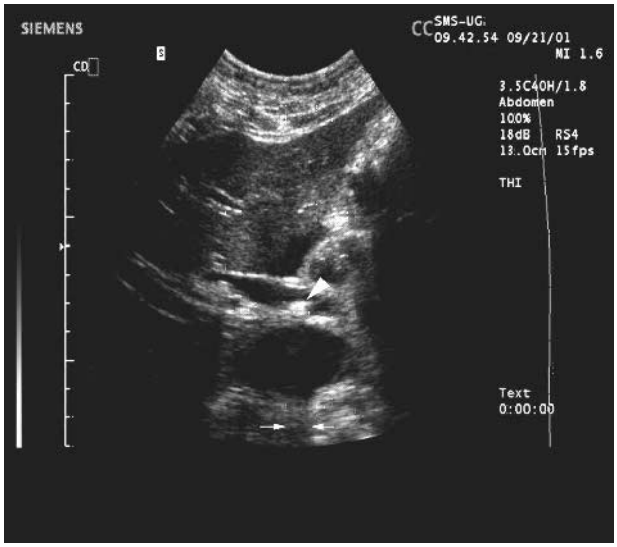


FIGURE 23 Stone in a dilated common bile duct. The echogenic structure within the dilated common bile duct (arrowhead) is identified as a stone by its shadowing. Dilatation is easily recognized on ultrasound, but stones are often much more elusive and endoscopic retrograde pancreatography may be required.

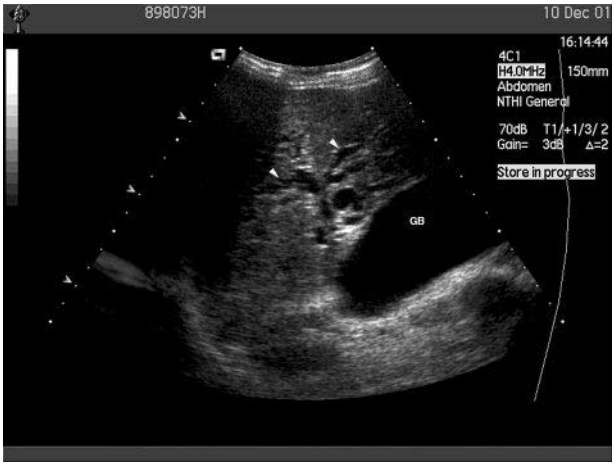


FIGURE 24 Dilated intrahepatic ducts. Normal ducts can just be resolved with ultrasound, but, when dilated, they are readily identified as tubes running alongside the vascular structures, producing a sort of duplicated anatomical pattern seen as "too many tubes" or the "parallel channel," signs of the ducts converging on the porta hepatis. GB, Gallbladder.

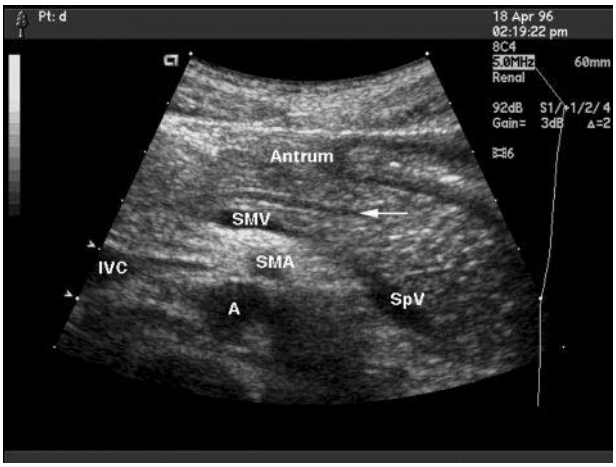


FIGURE 25 Pancreas. The proximal pancreas with its duct (arrow) lying across the superior mesenteric vessels is seen in this epigastric transverse section. A, Aorta; antrum, antrum of stomach; IVC, inferior vena cava; SMA, superior mesenteric artery; SMV, superior mesenteric vein; SpV, splenic vein.

A new way to use ultrasound (and other imaging) data that could prove to be important is elasticity imaging. For this, a series of images is collected while the tissue under study is distorted by applying gentle pressure, in order to move the tissue by about 5 mm. The images of the response to such a stress are used to form a strain image, which relates to the elastic (Young's) modulus; because this has several orders of magnitude greater range in human tissues than does the bulk modulus that is used for conventional ultrasound, the contrast between normal and pathological tissue is increased. In experimental situations, elastography can be implemented in real time and, for the liver, cardiac motion can be used as the stress. Although not ready for clinical use, this approach, which essentially images what the palpating hand discerns, is of great interest. Ultrasound can be implemented in three dimensions and this has proved useful when complex anatomy, such as the fetal face, needs to be depicted, to look for developmental abnormalities. In gastroenterology, this is probably of less importance, but the fact that three-dimensional images can be obtained in real time (so-called four-dimensional imaging) could be useful for interventional procedures in which it is important to visualize the needle path interactively in three dimensions.

CONCLUSIONS

Developments in ultrasound have afforded it a unique role in the field of gastroenterology. It is the choice technique for the gallbladder and biliary tree, whereas Doppler gives unique information in portal hypertension and other vascular disorders of the liver. New techniques, especially with microbubble contrast agents, have made US comparable to helical CT for detecting and differentiating liver tumors and in abdominal trauma. The role of US is further extending to hollow abdominal organs, thanks to new technologies and systematic scanning approaches.

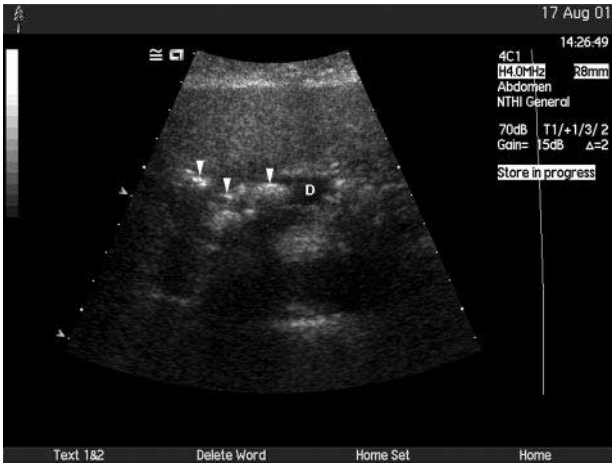


FIGURE 26 Chronic pancreatitis. Ultrasound is not sensitive to the changes of chronic pancreatitis, but when there is calcification, within the duct in this case (arrowheads), the changes are easily demonstrated. D, Pancreatic duct.

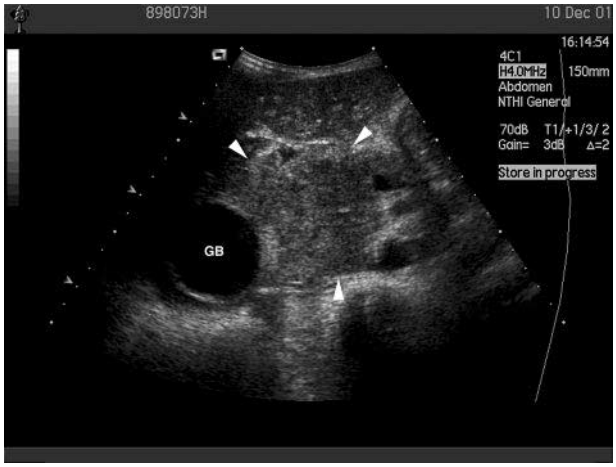


FIGURE 27 Carcinoma of the head of the pancreas. This large mass in the position of the head of the pancreas (arrowheads) proved to be a cancer, but the distinction from a chronic pancreatic mass is often impossible. Note the dilated gallbladder (GB).

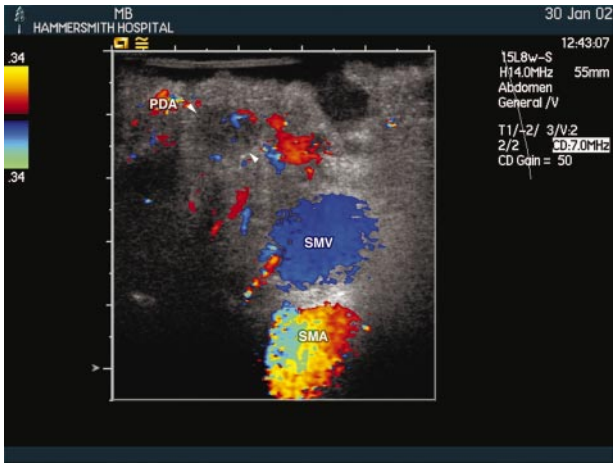


FIGURE 28 Insulinoma intraoperative scan. This 7-mm insulinoma in the head of the pancreas could not be located in theater. This high-resolution scan obtained directly via the exposed pancreas clearly shows the lesion (arrowheads); the associated blood vessels are highlighted on color Doppler. PDA, Pancreatico-duodenal arcade; SMA, superior mesenteric artery; SMV, superior mesenteric vein.

See Also the Following Articles

Computed Tomography (CT) • Endoscopic Ultrasonography • Endoscopy, Complications of • Picture Archiving and Communication Systems (PACS) • Upper Gastrointestinal Endoscopy

Further Reading

Albrecht, T., and Blomley, M. J. (2001). Characteristics of hepatic hemangiomas at contrast-enhanced harmonic US. *Radiology* **220**, 269–270.

Bartolozzi, C., and Lencioni, R. (2001). Contrast-specific ultrasound imaging of focal liver lesions. Prologue to a promising future. *Eur. Radiol.* **11**(Suppl. 3), E13–E14.

Bartram, C. (2001). Radiologic evaluation of anorectal disorders. *Gastroenterol. Clin. North Am.* **30**, 55–75.

Bennett, G. L., and Hann, L. E. (2001). Pancreatic ultrasonography. *Surg. Clin. North Am.* **81**, 259–281.

Brand, R. (2001). The diagnosis of pancreatic cancer. *Cancer J.* **7**, 287–297.

Damore, L. J., 2nd, Cook, C. H., *et al.* (2001). Ultrasonography incorrectly diagnoses gallbladder polyps. *Surg. Laparosc. Endosc. Percutan. Tech.* **11**, 88–91.

Freeny, P. C. (2001). Pancreatic carcinoma: Imaging update 2001. *Dig. Dis.* **19**, 37–46.

Gaiani, S., Volpe, L., *et al.* (2001). Vascularity of liver tumours and recent advances in doppler ultrasound. *J. Hepatol.* **34**, 474–482.

Giorgio, A., Tarantino, L., *et al.* (2001). Hydatid liver cyst: An 11-year experience of treatment with percutaneous aspiration and ethanol injection. *J. Ultrasound. Med.* **20**, 729–738.

Hunt, G. C., and Faigel, D. O. (2002). Assessment of EUS for diagnosing, staging, and determining resectability of pancreatic cancer: A review. *Gastrointest. Endosc.* **55**, 232–237.

Kalimi, R., Gecelter, G. R., *et al.* (2001). Diagnosis of acute cholecystitis: Sensitivity of sonography, cholescintigraphy, and combined sonography–cholescintigraphy. *J. Am. Coll. Surg.* **193**, 609–613.

Li, P. C., and Wu, C. L. (2001). Strain compounding: Spatial resolution and performance on human images. *Ultrasound Med. Biol.* **27**, 1535–1541.

Martinez-Noguera, A., Montserrat, E., *et al.* (2001). Ultrasound of the pancreas: Update and controversies. *Eur. Radiol.* **11**, 1594–1606.

Mirilas, P., Koumanidou, C., *et al.* (2001). Sonographic features indicative of hydrostatic reducibility of intestinal intussusception in infancy and early childhood. *Eur. Radiol.* **11**, 2576–2580.

Mortele, K. J., and Ros, P. R. (2001). Imaging of diffuse liver disease. *Semin. Liver Dis.* **21**, 195–212.

Roberts, S. A. (2002). Laparoscopic ultrasound staging of upper gastrointestinal tract tumors: Is it really necessary? *Abdom. Imaging* **27**, 1–5.

Shorvon, P. (2002). Imaging of appendicitis: A cautionary note. *Br. J. Radiol.* **75**, 717–720.

Sofia, S., Casali, A., *et al.* (2001). Sonographic diagnosis of adult intussusception. *Abdom. Imaging* **26**, 483–486.

Tandon, M., and Topazian, M. (2001). Endoscopic ultrasound in idiopathic acute pancreatitis. *Am. J. Gastroenterol.* **96**, 705–709.

Wilson, S. R., and Burns, P. N. (2001). Liver mass evaluation with ultrasound: The impact of microbubble contrast agents and pulse inversion imaging. *Semin. Liver Dis.* **21**, 147–159.

Wilson, S. R., Burns, P. N., *et al.* (2000). Harmonic hepatic US with microbubble contrast agent: Initial experience showing improved characterization of hemangioma, hepatocellular carcinoma, and metastasis. *Radiology* **215**, 153–161.

Yarmenitis, S. D. (2002). Ultrasound of the gallbladder and the biliary tree. *Eur. Radiol.* **12**, 270–282.

Upper Gastrointestinal Bleeding

C. MEL WILCOX
University of Alabama, Birmingham

hematemesis Vomiting of blood.
hematochezia Passage of bright red or wine-colored stool, usually representing bleeding from the lower gastrointestinal tract, often the colon.
melena Passage of dark black stool, representing blood typically from an upper gastrointestinal source.
upper gastrointestinal bleeding Loss of blood from the gastrointestinal tract proximal to the ligament of Treitz.

Acute upper gastrointestinal bleeding is a common medical condition encountered by generalists and specialists alike. Despite the significant strides in the development of effective endoscopic techniques for hemostasis, this type of bleeding, especially in the elderly, remains a significant cause of morbidity and potential mortality. Prompt recognition of the patient with acute upper gastrointestinal bleeding will ensure timely endoscopic evaluation, which will provide not only a diagnosis but also the opportunity for endoscopic therapy and triage. Patients with low-risk lesions at endoscopy can be managed with earlier hospital discharge, assuming there is no significant comorbidity.

INTRODUCTION

Current estimates of the incidence of hospitalization for acute upper gastrointestinal bleeding (UGIB) in the United States population range from 30 to 100 per 100,000, which translates to approximately 400,000 hospital admissions yearly in the United States. Whether the widespread adoption of *Helicobacter pylori* eradication therapy for patients with dyspepsia and peptic ulcer as well as the use of safer nonsteroidal antiinflammatory drugs (NSAIDs) will reduce this number remains to be explored. These incidence data do not account for those patients who experience acute UGIB while hospitalized; given the narrower differential diagnosis and higher mortality, such patients represent a unique subset of bleeders.

ETIOLOGY

The most common cause of acute UGIB remains peptic ulcer disease (Fig. 1). Gastroesophageal varices, Mallory–Weiss tears, and acute gastric mucosal lesions (erosions) are the next most frequent etiologies. It is important to recognize that the most common cause identified in a particular study is dependent on the population examined. For example, peptic ulcer and acute gastric mucosal lesions represent the most common causes in NSAID users, whereas variceal bleeding is most prevalent in patients with portal hypertension and cirrhosis.

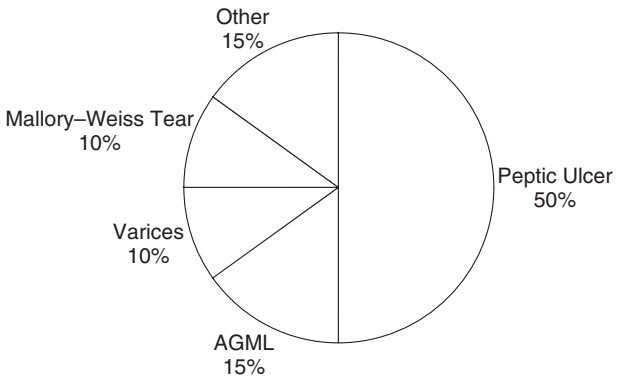


FIGURE 1 Etiology of upper GI bleeding. AGML, Acute gastric mucosal lesion.

APPROACH TO THE PATIENT

The first step in the management of the patient with UGIB is prompt diagnosis. Although most patients manifest with an alteration in stool color, usually melena associated with hematemesis, the initial presentation in some patients will be hematochezia alone, which may suggest a lower gastrointestinal (GI) bleeding source. The color of both vomitus and stool not only points to the site of bleeding (upper versus lower GI tract), but also provides important prognostic information. Patients who present with hematemesis and melena generally have more substantial blood loss as compared to patients who present with "coffee-ground" hematemesis and brown stool or melena without hematemesis. Red hematemesis also suggests more acute and significant

 Copyright 2004, Elsevier (USA). All rights reserved.

blood loss and is associated with a worse prognosis compared to a presentation with coffee-ground hematemesis alone. Generally, the more significant the blood loss, the more likely the bleeding source is arterial.

A thorough history often provides important clues to the underlying cause of bleeding. A thorough systematic review for diseases of the esophagus, stomach, and duodenum should be performed. Rarely, nasopharyngeal and oropharyngeal bleeding will masquerade as UGIB in the absence of symptoms pointing to this possibility. The presence of dysphagia or reflux symptoms may suggest esophageal cancer and gastroesophageal reflux disease, respectively. Because NSAID use is so prevalent in patients with gastric and duodenal ulcer disease and acute mucosal lesions, it is important to question specifically not only regarding prescription use but also the use of over-the-counter NSAIDs, because many patients do not consider these "medications." The dose and duration are also important to ascertain, because high-dose NSAID use of long duration is more likely to produce peptic ulcer as compared to very short-term intermittent use. A past history of peptic ulcer, particularly bleeding ulcer, is important to identify. Associated symptoms of ulcer include nausea, vomiting, and nocturnal epigastric distress. Nevertheless, the absence of abdominal pain is common in peptic ulcer regardless of NSAID use. Painless bleeding is also more common in patients with mucosal disease such as vascular ectasias. Vascular ectasias are most prevalent in the elderly and in those with underlying renal disease, typically elderly patients on dialysis. Approximately one-third of patients with Mallory–Weiss tear will not present with a classic history of repetitive retching followed by hematemesis. Portal hypertension may be suggested in the patient with established chronic hepatitis or significant alcohol use associated with stigmata of chronic liver disease. A history of aortic aneurysm or prior repair should raise the possibility of an aortoenteric fistula, which would require not only endoscopy but also computed tomography to diagnose.

Following recognition and diagnosis, initial evaluation must promptly address hemodynamic status. Persistent bleeding with subsequent volume loss, if left untreated, will lead to multiorgan failure and death caused by hypoperfusion. Initial hemodynamic assessment should use vital signs as a measure of volume status. Resting tachycardia and orthostatic hypotension represent a $>15\%$ volume loss. The degree of resuscitation with normal saline and/or blood products should be dictated by the volume status (vital signs), initial hematocrit, and evidence of ongoing blood loss as assessed clinically (recurrent/persistent hematemesis, melena, etc.). Careful administration of blood products and fluids should be provided for the elderly patients, especially those with underlying heart disease or renal dysfunction. Maintaining a hematocrit above 30% is appropriate for those with underlying coronary artery disease to maximize oxygen delivery. It is critical to remember that the hematocrit can be influenced by the degree of volume resuscitation and prior volume status. Therefore, the hematocrit must be interpreted in the context of these other factors. Determination of platelet count and prothrombin time is important, and is mandatory for those receiving medications such as coumadin. In addition, an elevation of prothrombin time and/or thrombocytopenia could point to underlying portal hypertension. Elevation of blood urea nitrogen concentration can be observed with both upper and lower GI bleeding, although this is typically higher with UGIB. Patients at risk for coronary disease should have an electrocardiogram to exclude myocardial ischemia; occult ischemia and infarction have been found in 10% or more of elderly patients. Chest radiography should be used selectively.

The site of gastrointestinal blood loss is generally apparent based on the history or witnessing of hematemesis or melena. When the diagnosis of UGIB is in doubt, nasogastric (NG) aspiration should be considered. Identification of bile in the aspirate suggests either bleeding distal to the duodenum or bleeding that has stopped. Conversely, if no bile is found, bleeding may still be present in the duodenum. Although melena almost uniformly represents bleeding proximal to the colon, in elderly patients with hematochezia in the absence of hematemesis, documentation of an upper gastrointestinal source is critical. When there is active gastrointestinal blood loss (unstable vital signs) and hematochezia, urgent upper endoscopy should be performed when bile is not found on NG aspirate, because a duodenal ulcer could be present. When the patient presents with hematemesis or gross melena with stable vitals signs, NG aspiration is not mandatory, because the site of bleeding is known. If there is concern about active bleeding, which may influence the timing of endoscopy, NG aspiration to sample for fresh blood is reasonable. Routine use of nasogastric lavage prior to upper endoscopy is unnecessary, because standard NG tubes cannot adequately clear the stomach of blood and clots; a large-bore tube (Edlich tube) is required in this setting. Generally, even with active bleeding and blood clots in the stomach, the bleeding source can be accurately identified and endoscopically treated at the time of endoscopy.

RISK ASSESSMENT

When initially evaluating the patient, outcome is intuitively assessed. For example, it is recognized that the major risk factor impacting the outcome of UGIB is uncontrolled and/or recurrent hemorrhage. Both clinical and endoscopic features help predict recurrent hemorrhage and poor outcome. Identification of these patients is how early endoscopic examination and endoscopic therapy will have the most beneficial impact.

A number of clinical factors have been examined for their predictive value for risk of recurrent hemorrhage. These include age greater than 65 years, gender, alcohol or tobacco use, preexisting liver disease, renal disease, vascular malformations, anticoagulant use, unstable vital signs (tachycardia, orthostasis, shock), ongoing bleeding (hemoglobin <9 g/dl), and blood urea nitrogen (BUN) concentration >90 mg/dl. However, when controlling for these variables, independent predictors include unstable vital signs (tachycardia, orthostasis, shock) and clinical evidence of ongoing blood loss. When combining these factors in a scoring system, more than 90% of patients who require intervention can be identified, which generally will include endoscopic therapy. Conversely, it is now recognized that patients without these clinical features are at low risk for recurrent bleeding and may be managed with less urgency.

All patients with gastrointestinal bleeding require neither intensive care unit (ICU) stay nor even hospital admission. Factors that may prompt admission for ICU monitoring include active bleeding, hypotension, coagulopathy, altered mental status, and comorbid conditions, especially in elderly patients. Nevertheless, clinical judgment should guide the need for observation in an intensive care unit. Emerging data suggest that clinical features alone may predict which patients require hospitalization for bleeding.

Most studies have combined both clinical and endoscopic features to assess risk for recurrent bleeding, poor outcome, and need for hospitalization. Rockhall and colleagues developed one of the best-studied scoring systems. This investigation found that age, vital signs, comorbidity, source of hemorrhage, and endoscopic stigmata could be used to classify low risk and thus eligibility for outpatient care. In this study, patients without these features had no mortality at short-term followup. Other investigators have used similar combinations of clinical and endoscopic features to show that early discharge or even discharge following endoscopy is both safe and cost effective. One study used well-defined criteria for low-risk discharged patients meeting these criteria after endoscopy in the emergency room. Followup demonstrated a low risk for recurrent bleeding (less than 5%) with 0% mortality.

TABLE I Prevalence and Outcome of Stigmata of Hemorrhage in Bleeding Ulcers

Endoscopic stigmata	Prevalence (%)	Rebleeding rate (%)
Spurting	10	80
Visible vessel	20	50
Clot	20	15
Flat spots	15	5
Clean base	30	5

Endoscopic features have been well established as predictors for recurrent gastrointestinal bleeding. These features have primarily been studied in peptic ulcer hemorrhage but are likely applicable to other nonvariceal causes. Although studies have suggested that large ulcers, ulcers high on the lesser curvature, and posterior duodenal ulcers have the highest risk of rebleeding, it is now recognized that endoscopic features of the lesion—so-called stigmata of hemorrhage—are the most important (Table I; Figs. 2–6). These stigmata have been repeatedly shown to predict accurately the risk of recurrent bleeding, the need for surgery, and mortality. The rebleeding rate for clots overlying an ulcer is variable among studies, and these differences are likely related to the fact that the blood clot may obscure underlying stigmata such as a visible vessel.

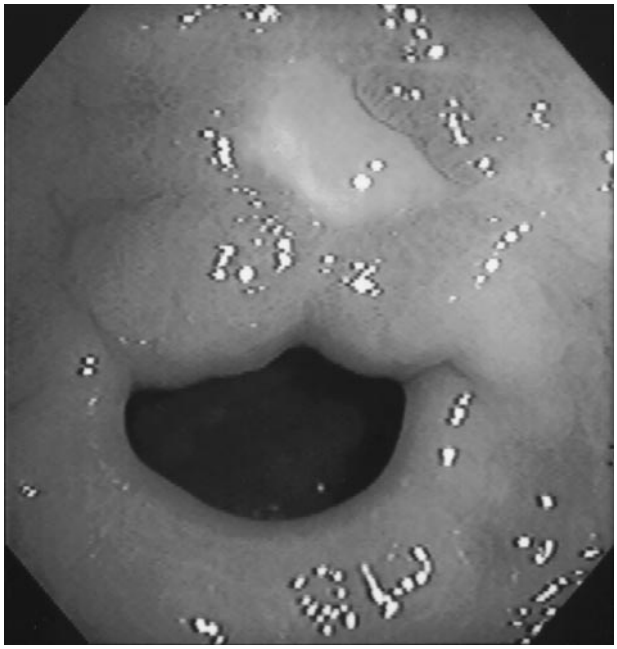


FIGURE 2 Clean-based ulcer in the peripyloric area.

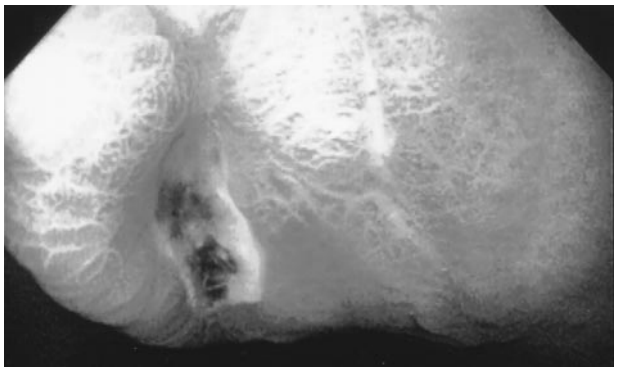


FIGURE 3 Linear ulcer in the gastric body; note flat black spots.

Over the past several decades, despite the efficacy of endoscopic therapy, the mortality rate for UGIB has remained relatively stable at approximately 5–10%. It is felt that this rate has not fallen because elderly patients, who are more likely to die, comprise a larger proportion of the bleeding population. It is recognized that age greater than 60 years increases the risk of death for UGIB by approximately 10-fold. Patients who have substantial bleeding while hospitalized have mortality rates of approximately 30%, likely due to the significant comorbidity requiring hospitalization. Regardless of the group, patients with UGIB rarely die from blood loss, but rather bleeding exacerbates any underlying comorbidities.

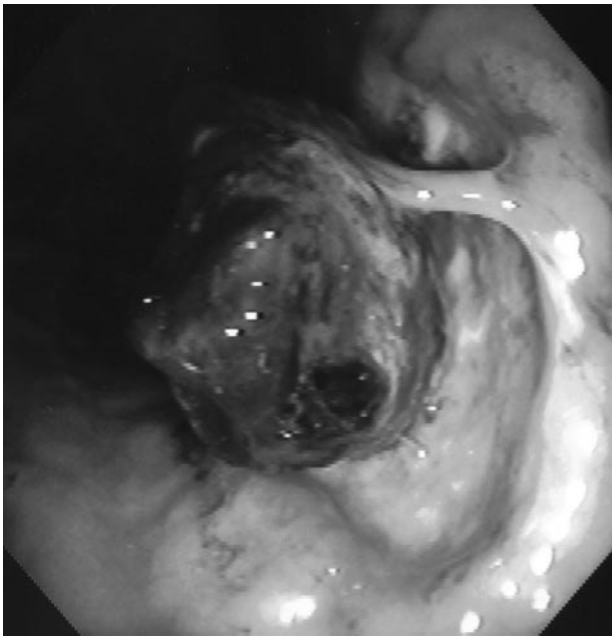


FIGURE 4 Large clot overlying a deep ulcer on the lesser curvature.

In summary, current evidence strongly supports the use of both clinical and endoscopic features to predict recurrent bleeding and outcome. When these features are combined, it is possible to determine not only if patients require more careful monitoring such as in an ICU, but also if patients require hospital admission.

ENDOSCOPIC HEMOSTASIS

A randomized trial of 100 patients with UGIB comparing routine endoscopy to upper gastrointestinal barium study in 1981 showed no difference in hospital stay, rebleeding rate, and mortality. This result is not surprising given the fact that, at that time, endoscopy represented a diagnostic tool alone. Subsequently, Laine prospectively randomized patients with bleeding peptic ulcers and high-risk stigmata to endoscopic therapy using a thermal probe or sham endoscopic therapy. This study demonstrated an improved outcome in the treated group with greater hemostasis (90% of treated patients compared to 13%), lower transfusion requirements, less urgent surgery, and a shorter hospital stay. Mortality was unchanged, perhaps because death is an

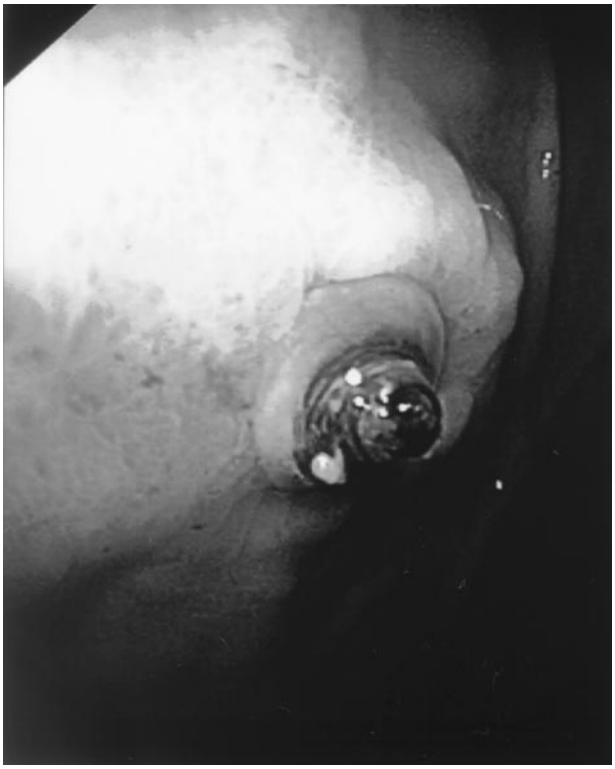


FIGURE 5 Nipplelike projection (visible vessel) from a small ulcer in the gastric body.

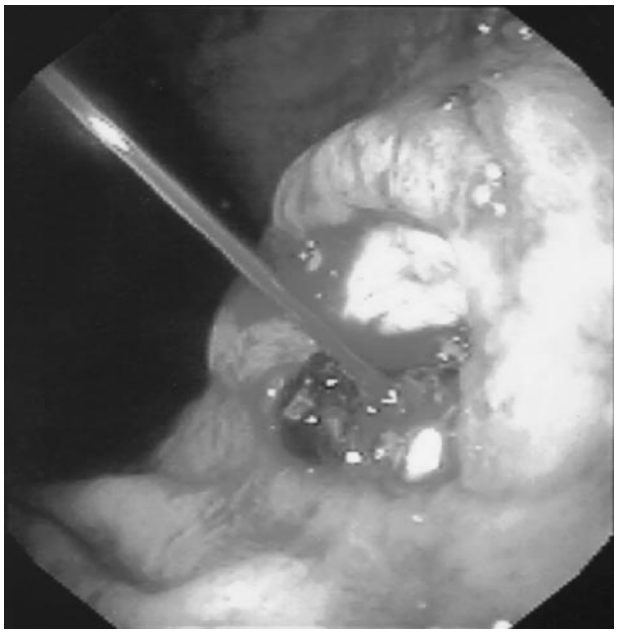


FIGURE 6 Small ulcer in the gastric body; note active arterial spurting.

uncommon event in UGIB. Since this seminal study was undertaken, numerous trials evaluating peptic ulcer, Mallory–Weiss tear, and esophageal variceal bleeding have shown the efficacy of endoscopic therapy in controlling bleeding, reducing the need for emergent surgery, and preventing rebleeding. Using meta-analysis, Cook and colleagues also showed that not only will endoscopic therapy decrease rebleeding, but also mortality in bleeding peptic ulcers.

Endoscopic Hemostatic Methods

A variety of endoscopic techniques have been developed and tested for the control of gastrointestinal bleeding, principally from peptic ulcer. Broadly, these can be categorized as thermal, injection, and mechanical techniques. The most widely tested and currently utilized techniques are the thermal methods. A 7 or 10 French probe placed through the endoscope channel delivering heat energy to the bleeding point has been found to be highly effective. One probe uses electrical energy to deliver heat at over 200°F to the tip of the probe (heater probe). The other method uses bipolar current that passes through the tip of the probe to generate heat. Thermal methods deliver heat to the bleeding vessel, which seals the defect in the arterial wall (coaptive coagulation), thus controlling active bleeding and preventing rebleeding. The depth of tissue injury for these thermal methods is shallow and thus perforation is rare. Noncontact devices include the argon and neodymium : yttrium aluminum garnet (Nd : YAG) crystal laser. Although effective, the YAG laser has fallen out of favor because of its cost, difficult and cumbersome use, and higher perforation rate. More recently, argon plasma coagulation has been utilized and is effective to ablate superficial lesions such as vascular ectasias (Fig. 7). This device works by passing Argon gas from the end of the probe, which becomes ignited, resulting in superficial tissue coagulation. Depending on the power setting, the depth of injury is also shallow and risk of perforation very small.

The technique of injection therapy is widely available, easy to perform, and efficacious. A variety of agents have been used, most frequently a dilute concentration (1:10,000) of epinephrine. It is felt that epinephrine results in both vasoconstriction and vessel occlusion (tamponade) by the fluid bolus. Injection of saline alone, however, does not appear as effective as epinephrine. Other tested agents include those causing tissue necrosis or fixation (ethanol, polidocanol, sodium tetradecylsulfate, ethanolamine) and clot formation (thrombin). Using a standard sclerotherapy needle, the technique involves injection of aliquots of 1–2 ml of dilute epinephrine (maximum 20 ml), or 0.1–0.2 ml (total volume 2 ml) of these sclerosing agents, because the sclerosing agents result in greater tissue injury and have a higher perforation rate. One large trial found thrombin to be effective and to have a low complication rate. This was given in doses of 61,000 international units (IU). Injection of cyanoacrylate glue, which is not available in the United States, has been used in

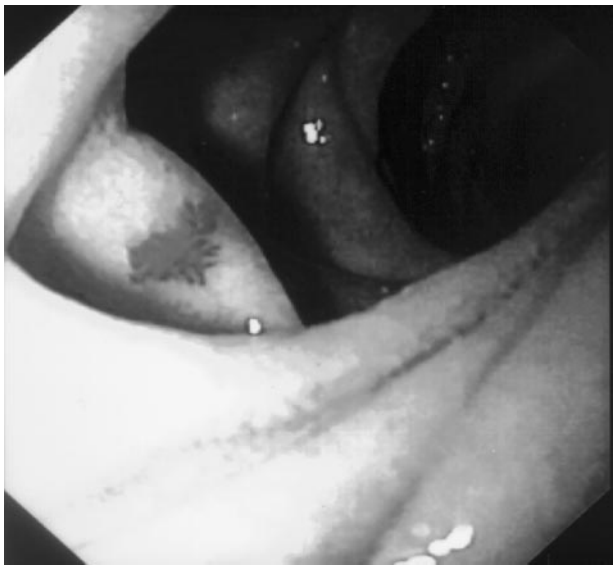


FIGURE 7 Small spiderweb-like lesion in the second duodenum, typical for a vascular ectasia.

European trials, but multiple endoscopy sessions may be required.

The most common mechanical technique employed is the endoscopic mucosal clip (hemoclip). Although best tested for use in peptic ulcer bleeding, these clips can also be used for Mallory–Weiss tears. One recent trial comparing a heater probe to a clipping device suggested superior efficacy for clips. Nevertheless, in the United States, thermal and injection therapies remain the standard of care.

EFFICACY OF ENDOSCOPY THERAPY

Numerous trials have compared each of these therapies either to placebo (medication), other endoscopic techniques, or in combination. Generally, endoscopic therapy is administered to patients with actively bleeding ulcers or with high-risk stigmata, including visible vessels. Although not as widely studied, these therapies are effective for other mucosal lesions, including esophageal ulcers and Mallory–Weiss tears. Generally, these techniques have been shown to reduce the rate of recurrent bleeding to less than 20%, although a substantial effect on mortality has rarely been shown. Several trials have used combination therapy with epinephrine injection followed by thermal therapy and suggest superiority of the combination method.

MANAGEMENT OF ACUTE VARICEAL BLEEDING

Prompt recognition of the patient at risk for portal hypertension-related bleeding is critical because of the higher morbidity and mortality of these patients, coupled with the fact that specific endoscopic therapy is required and selected pharmacological therapies are beneficial. Approximately 35–80% of patients with cirrhosis may develop varices, usually esophageal, with a much smaller percentage developing gastric varices. The patient who should be considered at risk for variceal bleeding is one who presents with stigmata of chronic liver disease (jaundice, cutaneous stigmata of chronic liver disease, ascites). Additional laboratory clues may include hypoprothrombinemia, thrombocytopenia, jaundice, and hypoalbuminemia. In the patient considered at risk by history, physical examination, and laboratory studies, pharmacologic therapy can be initiated prior to endoscopic therapy with octreotide, a synthetic long-acting analogue of somatostatin. This agent has essentially replaced the combination of vasopressin and nitrates, given their higher complications as well as superior efficacy of octreotide. Endoscopic therapy for esophageal varices includes sclerotherapy or, more recently, esophageal banding. This latter technique involves the endoscopic delivery of ringlike elastic bands that are placed directly on varices. Trials comparing sclerotherapy to variceal banding have shown improved efficacy in preventing rebleeding as well as a reduction in complication rates with banding. Active bleeding from gastric varices can be treated endoscopically, although this is not a long-term option and surgical therapy or transjugular intrahepatic portosystemic shunt (TIPS) therapy will be required. Transplantation is another potential option in the appropriate candidate. For the patient with acute variceal bleeding, antibiotic prophylaxis with a quinolone for 7 days has also been recommended because this decreases overall infection rates, including spontaneous bacterial peritonitis.

RECURRENT BLEEDING

Despite the efficacy of endoscopic therapies, bleeding still recurs in 15–20% of patients (range 5–35%) treated endoscopically with spurting vessels, and bleeding and nonbleeding visible vessels. Rebleeding from esophageal varices also remains a problem, although less so since the introduction of banding. The mortality in these groups remains substantial at 10% or greater. Patients with rebleeding can be treated with repeat endoscopic therapy, angiographic embolization, or surgery.

Repeat Endoscopic Therapy

There has been one prospective randomized trial comparing repeat endoscopic therapy to surgery for patients rebleeding from peptic ulcer. Repeat endoscopic therapy was successful in 73% of patients, and the overall complication rate was twice as high in the surgery group. Nevertheless, length of hospital stay, transfusion requirements, and mortality were no different. Given the data from this well-done study, an attempt at repeat endoscopic therapy should be performed prior to surgical therapy for patients with ulcer-related rebleeding. Although not studied, it is likely that other nonvariceal causes of rebleeding can be similarly managed, such as Mallory–Weiss tear. In the high-risk bleeder, some have performed "second look" endoscopies in the hopes of retreating endoscopically persistent high-risk lesions, thereby reducing recurrent bleeding. The data for this have been mixed and no firm recommendation can be made.

Angiographic Therapy

Angiography can be used initially in the patient with UGIB to identify the site of bleeding when endoscopy

cannot be performed or when the bleeding source cannot be identified due to massive bleeding. To demonstrate extravasation, active bleeding at approximately 1 ml/minute must be present. If active arterial bleeding is identified, selective placement of coils or foam pledgets into the artery is performed. When the site of bleeding is known but bleeding is intermittent, empiric embolization of the appropriate arterial source is an option. In the past, selective infusion of arterial vasopressin was utilized. However, the systemic effects of widespread vasoconstriction may be complicated by myocardial infarction and arrhythmias. The success of embolization is approximately 80%, with a low complication rate, including perforation.

Surgical Therapy

Many studies have shown that surgery is required for recurrent bleeding in approximately 5–10% of patients with ulcer-related bleeding. The precise timing of surgery for patients with recurrent bleeding is not well established. Although successful in halting bleeding, the complication rate with surgery is high and mortality rates of over 20% have been reported. Unfortunately, it is often the elderly patients with comorbidity who fail endoscopic therapy, and they are poor surgical candidates. The type of surgery undertaken will depend on the location of the lesion and the experience of the surgeon. Generally, in the high-risk patient, angiographic embolization can be attempted first; if unsuccessful, surgical therapy can then be undertaken.

Medical Therapy

Although antacid therapy has been routinely prescribed for UGIB for many years, large trials of over 1000 patients have failed to demonstrate efficacy of histamine-2 (H2) receptor blockers over placebo. A large meta-analysis did suggest a decrease in surgery and death, however. Despite these discouraging results with these agents, antacid therapy has some scientific rationale. *In vitro*, the combination of blood and acid at a pH less than 5 results in coagulopathy, with reduction in platelet aggregation and fibrin formation. These abnormalities were not seen when the pH was greater than 6.

With the availability of potent acid-suppressing agents (proton pump inhibitors), enthusiasm for acid reduction in the setting of bleeding has been rekindled. These agents could theoretically be effective, given their efficacy is raising gastric pH. Several studies from Europe have shown that a continuous infusion of omeprazole in bleeding peptic ulcer decreases rebleeding rates. The use of high-dose oral therapy was shown to be effective in one study, and there appears to be additional efficacy for the use of omeprazole following endoscopic therapy.

Although widely used for the therapy of variceal hemorrhage, some data suggest that intravenous somatostatin and octreotide also have benefit for ulcer-related bleeding. A recent meta-analysis suggested a reduction in continued or recurrent bleeding; however, given the efficacy of endoscopic therapies, perhaps combined with oral omeprazole, these agents should rarely be given.

POSTTHERAPY FOLLOWUP

Immediate

Assuming endoscopic therapy has been performed successfully or the patient does not require endoscopic therapy at the time of endoscopy, the intensity of subsequent monitoring for rebleeding is dependent on the likelihood of rebleeding and the overall health status of the patient. Elderly patients with comorbidity with ulcer-related bleeding requiring endoscopic hemostasis may require additional ICU observation, whereas a younger healthy patient with a Mallory–Weiss tear treated endoscopically could be managed on a ward and considered for early discharge (within 24 hours of endoscopy). Monitoring should include not only frequent vital signs and hematocrit, but also a careful recording of stool frequency and color as important clues to recurrent bleeding. Generally, ulcer-related rebleeding occurs approximately 24–72 hours after initial stabilization. For the patient considered at low risk for recurrent bleeding, the diet can be liberalized and early discharge planned, assuming there is adequate home support. If repeat endoscopy is either planned or felt likely because of high-risk stigmata, a completely liquid diet is appropriate for the first 24 hours after endoscopic therapy. Hospital discharge is appropriate 48–72 hours after endoscopic therapy in the elderly patient.

Curing the Ulcer Disease

Prior to awareness of the role of *H. pylori* in the pathogenesis of peptic ulcer, long-term strategies following ulcer hemorrhage consisted of maintenance antacid therapy to prevent recurrent ulceration and thus bleeding. In a study by Jensen and colleagues, the use of maintenance ranitidine (150 mg/day) decreased the recurrence rate of duodenal ulcer bleeding over 1 year from 36 to 9%. Given their additional potency, it is likely that long-term therapy with proton pump inhibitor therapy would be even more effective. No differences

have been shown in ulcer healing between bleeding and nonbleeding ulcers.

Determination of *H. pylori* status during the hospitalization is mandatory for patients with bleeding ulcer. Biopsy of the gastric antrum for urease testing [with a kit using the *Campylobacter*-like organism (CLO) test] can be performed at the time of endoscopy, although the sensitivity has been found to be less in this setting, perhaps due to interference from blood. Random biopsies can be obtained for gastric histology, particularly if neoplasm needs to be excluded and assessment for *H. pylori* performed. Stool antigen tests are another alternative. To ensure an accurate diagnosis, combination testing with more than one method is suggested. Numerous studies have now demonstrated that successful eradication of *H. pylori* effectively cures peptic ulcer disease. For patients with bleeding ulcer, this is translated into decreased ulcer and hemorrhage risk. Indeed, most studies suggest that the bleeding rate following successful eradication is less than 5%. Given this striking success, it is critical to verify the eradication of *H. pylori* either with a urea breath test, repeat endoscopy and biopsy, or stool antigen testing. If *H. pylori* is successfully eradicated and the patient does not require NSAID therapy, long-term acid suppression is not generally required.

ROLE OF NSAID THERAPY

For patients in whom NSAIDs are believed causative, proscribing these medications should be done if at all possible. If NSAIDs can be discontinued, standard antacid therapy will heal the lesions. If NSAIDs must be continued, antacid therapy with a proton pump inhibitor given for 12 weeks is indicated to heal the ulcer. Prospective studies have shown the superiority of omeprazole in standard doses compared to H2 receptor antagonists and misoprostol for the healing of NSAID-related ulcers despite continued NSAID use.

Patients who require long-term conventional NSAID therapy will require cotherapy. In a study randomizing over 700 patients with healed NSAID-related gastric or duodenal ulcers, but who required continued NSAIDs, to omeprazole (20 mg/day), misoprostol (200 μg BID), or placebo, the ulcer remission rates were 61, 48, and 27% at 24 weeks, respectively. Although this study clearly documented the superiority of omeprazole, it should be recognized that almost 40% of patients had ulcer relapse, which is unacceptable.

Another strategy that could be employed is the use of cyclooxygenase-2 inhibitors (COX-2). A study comparing the COX-2 agent celecoxib to diclofenac plus omeprazole in patients with NSAID-related ulcer bleeding found an equivalent rate of recurrent bleeding at 6 months of 4.9 vs 6.4%, respectively. These results suggest that in high risk patients, a COX-2 agent alone may need to be given with a proton pump inhibitor to further reduce bleeding risk. Ongoing studies should address this important question.

The use of even low-dose aspirin increases the risk of peptic ulcer bleeding. The relative risk of even low-dose aspirin (81 mg) has been shown to be greater than 2. The risk is higher with increasing aspirin dose, given the dose–response relationship. Therefore, in the patient with a bleeding ulcer who requires long-term low-dose aspirin, cotherapy with a proton pump inhibitor should be provided.

There are some data suggesting that *H. pylori* eradication may reduce the ulcer recurrence rate despite continued NSAID use. Chan and colleagues compared maintenance omeprazole to *H. pylori* eradication in patients with bleeding ulcers who required continued aspirin or naproxyn therapy. In patients requiring naproxyn, they found the incidence of recurrent ulcer was lower in patients receiving omeprazole (4%) than in those undergoing *H. pylori* eradication alone (19%). In contrast, the recurrent hemorrhage rate was equivalent for patients receiving aspirin and omeprazole and for patients receiving aspirin and *H. pylori* eradication without omeprazole.

SUMMARY

UGIB remains a common and important medical condition. Great strides have been made in the past decade in the development of endoscopic hemostatic techniques that have significantly impacted our ability to control and prevent rebleeding. The use of potent antiacid therapy, including intravenous proton pump inhibitors, may additionally improve the outcome of bleeding peptic ulcer. Prompt recognition of the patient with UGIB will ensure timely endoscopic evaluation, providing not only a diagnosis, but also the opportunity for endoscopic therapy and triage. Patients with low-risk lesions at endoscopy can be managed with earlier hospital discharge assuming there is no significant comorbidity.

See Also the Following Articles

Duodenal Ulcer Endoscopy, Complications of • Gastric Ulcer • *Helicobacter pylori* • Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Mallory–Weiss Tear • NSAID-Induced Injury • Occult Gastrointestinal Bleeding • Portal Hypertension and Esophageal Varices • Upper Gastrointestinal Endoscopy • Variceal Bleeding

Further Reading

Bornman, P. C., Theodorou, N. A., Shuttleworth, R. D., Essel, H. P., and Marks, I. N. (1985). Importance of hypovolaemic shock and endoscopic signs in predicting recurrent haemorrhage from peptic ulceration: A prospective evaluation. *BMJ* **291**, 245–257.

Chan, F. K., Hung, L. C., Suen, B. Y., Wu, J. C., Lee, K. C., Leung, V. K., Hui, A. J., To, K. F., Leung, W. K., Wong, V. W., Chung, S. C., Sung, J. J. (2002). Celecoxib versus diclofenac and omeprazole in reducing the risk of recurrent ulcer bleeding in patients with arthritis. *N. Engl. J. Med.* **26**, 2104–2110.

Chung, I. K., Kim, E. J., Lee, M. S., Park, S. H., Lee, M. H., Kim, S. J., Cho, M. S., and Hwang, K. Y. (2001). Endoscopic hemastasis on bleeding peptic ulcers. *Endoscopy* **33**, 969–975.

Chung, S. C. S., Lau, J. Y. W., and Sung, J. J. Y. (1997). Randomized comparison between adrenaline injection along and adrenaline injection plus heat probe treatment for actively bleeding ulcers. *BMJ* **314**, 1307–1311.

Cook, D. J., Guyatt, G. H., Salena, B. J., and Laine, L. A. (1992). Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: A meta-analysis. *Gastroenterology* **102**, 139–148.

Derry, S., and Loke, Y. K. (2000). Risk of gastrointestinal hemorrhage with long term use of aspirin: Meta-analysis. *BMJ* **321**, 1183–1187.

Grino, P., Pascual, S., Such, J., Casellas, J. A., Niveiro, M., Andreau, M., Saez, J., Grino, E., Palazon, J. M., Carnicer, F., and Perez-Mateo, M. (2001). Comparison of diagnostic methods of *Helicobacter pylori* infection in patients with upper gastrointestinal bleeding. *Scand. J. Gastroenterol.* **36**, 1254–1258.

Laine, L. (1987). Multipolar electrocoagulation in the treatment of active upper gastrointestinal tract hemorrhage. *N. Engl. J. Med.* **316**, 1613–1617.

Lau, J. Y., Sung, J. J., and Lam, Y. H. (1999). Endoscopic treatment compared with surgery in patients with recurrent bleeding after initial endoscopic control of bleeding ulcers. *N. Engl. J. Med.* **340**, 751–756.

Lau, J. Y. W., Sung, J. J. Y., Lee, K. K. C., Yung, M. Y., Wond, S. K. H., Wu, J. C. Y., Chan, F. K. L., Ng, E. K. W., You, J. H. S., Lee, C. W., Phil, M., Chan, A. C. W., and Chung, S. C. S. (2000). Effect of intravenous omeprazole on recurrent bleeding after endoscopic treatment of bleeding peptic ulcers. *N. Engl. J. Med.* **343**, 310–316.

Podila, P. V., Ben-Menachem, T., Batra, S. K., Oruganti, N., Posa, P., and Fogel, R. (2001). Managing patients with acute, nonvariceal gastrointestinal hemorrhage: Development and effectiveness of a clinical care pathway. *Am. J. Gastroenterol.* **96**, 208–219.

Rutgeerts, P., Rauws, E., and Wara, P. (1997). Randomised trial of single and repeated fibrin glue compared with injection of polidocanol in treatment of bleeding peptic ulcer. *Lancet* **350**, 692–696.

Udd, M., Miettinen, P., Palmu, A., Heikkinen, M., Janatuinen, E., Pasanen, P., Tarveinen, R., Kairalouma, V., Lohman, M., Mustonen, H., and Julkunen, R. (2001). Regular-dose versus high-dose omeprazole in peptic ulcer bleeding: A prospective randomized double-blind study. *Scand. J. Gastroenterol.* **36**, 1332–1338.

Wilcox, C. M., and Clark, W. C. (1999). Causes and outcome of upper and lower gastrointestinal bleeding: The Grady Hospital experience. *South Med. J.* **92**, 44–50.

Upper Gastrointestinal Endoscopy

J. BARRY O'CONNOR
Duke University Medical Center

achalasia Motility disorder of the esophagus in which (1) the lower esophageal sphincter does not adequately open during swallowing and (2) the contractions in the esophagus are not coordinated.
angularis Sharp angulation in the lesser curvature of the normal J-shaped stomach.
antrum Distal or lower portion of the stomach.
benzodiazepines Class of medications with sedating and amnestic effects.
biopsy Small tissue sample taken during endoscopy.
botulinum toxin (Botox) Neurotoxin used to relax muscle tissue.
bougie Flexible device used to dilate or stretch a narrowed segment of intestine.
cardia Upper part of the stomach just below the esophagus.
cautery Procedure that delivers heat to body tissue to destroy it.
conscious sedation Sedation to the point of diminished consciousness but not unconsciousness.
duodenum Portion of the small intestine just beyond the stomach.
dysphagia Difficulty swallowing (food becomes stuck or held up in the chest during swallowing).
endoscope Tubular flexible instrument used to look into the intestine.
epigastric Location in the central upper abdomen.
fibrosis Thickening and hardening of the tissues from chronic irritation or injury.
fluoroscopy Real-time X-ray machine used in the setting of endoscopy to monitor the position of therapeutic devices such as balloon or bougies.
fundus Proximal portion of the stomach.
guide-wire Long flexible wire used during endoscopy to guide instruments into correct position.
hemostasis Control of bleeding.
hypotension Abnormally low blood pressure.
lower esophageal sphincter Muscular valve at the lower end of the esophagus, at the junction with the stomach.
Mallory–Weiss tear Traumatic tear of the lining of the lower esophagus or upper stomach, often caused by vomiting.
mucosa Surface lining of the intestine.
odynophagia Painful sensation during swallowing.
palliation Relief of symptoms without curative intent.
pylorus Opening that leads from the stomach to the duodenum.
retroflexing Positioning of the tip of the endoscope so that it is turned around 180° to provide a backward view.
savary dilator Flexible polyvinyl dilator with a central guide-wire lumen. Used to dilate esophageal strictures.
sclerosant Agent (usually liquid) that, when injected, causes thickening and fibrosis in tissue. Used in upper gastrointestinal endoscopy to control bleeding from a vessel.
squamous Normal multilayered mucosa (seen in the esophagus).
stent Hollow tubular device deployed in a stenosed area of bowel in order to widen the lumen diameter.
stricture/stenosis Narrowed portion of the intestine.
upper esophageal sphincter Muscular valve at the upper end of the esophagus.
vasoconstriction Narrowing of a blood vessel (usually a bleeding vessel).

Upper gastrointestinal endoscopy (EGD) is the science of direct examination of the esophagus, stomach, and duodenum, using an endoscope, a flexible tube containing a light source and high-resolution optics. When connected to a monitor, an endoscope displays real-time images of the inner lining of the upper GI tract. Upper GI endoscopy may be diagnostic or therapeutic. In diagnostic endoscopy, the bowel lining is visualized but no treatment is undertaken. In therapeutic endoscopy, instruments are passed through the scope to destroy or remove tissue or deliver therapy to diseased tissue.

TYPES OF UPPER GASTROINTESTINAL ENDOSCOPES

Most modern upper gastrointestinal (GI) endoscopes are flexible, allowing passage through the normal turns in the human GI tract. Upper GI endoscopes are classified by external diameter and by the number and size of the working channels. Larger working channels allow passage of larger instruments used to treat bleeding and remove tissue. A detailed description of the components of endoscopes is beyond the scope of this article. Briefly, endoscopes are long flexible tubes with a light source, lens, and display monitor. The handle of the scope has buttons that enable air and water instillation and dial controls that direct the tip of the scope through the bowel. There are one or more

 Copyright 2004, Elsevier (USA). All rights reserved.

channels that run the length of the scope. These are used to aspirate fluid from the bowel, to push fluid through the scope to wash the mucosal surface, and to pass therapeutic instruments through.

PERFORMING UPPER GASTROINTESTINAL ENDOSCOPY

Preparation for Endoscopy

Patients must be fasting from the previous night to ensure that the upper intestine is empty. A physical examination is performed to ensure patient suitability for sedation and endoscopy. Informed consent is obtained.

Sedation for Upper Endoscopy

In the United States, most patients are sedated for gastrointestinal endoscopy procedures. Conscious sedation is usually achieved using benzodiazepines (e.g., midazolam, diazepam) and narcotic analgesics (e.g., fentanyl, meperidine) given intravenously, the dose being titrated to the desired level of sedation. Sedation is begun immediately before the endoscopic examination and additional doses of sedative medications may be given during the procedure as needed. Patient safety is an important part of satisfactory performance of upper GI endoscopy. During and after the procedure, the patient's pulse, blood pressure, electrocardiogram (EKG), and oxygen saturation are monitored continuously. After the procedure, the sedation effect gradually wears off such that most patients may be safely discharged from the endoscopy unit within 1 h after completion of their endoscopy.

Upper GI endoscopy may be performed without sedation, as is frequently the case in Europe. However, because of the relatively large diameter of the standard diagnostic upper endoscope, unsedated upper endoscopy may be somewhat uncomfortable and will not be tolerated by all patients. Recently, an ultrathin upper endoscope was introduced in the United States. It may be used to examine the upper GI tract without sedation. This scope may be introduced into the upper GI tract either via the mouth or more commonly via the nose. In most cases, the transnasal route is preferred because during transoral passage, pressure from the scope on base of the tongue induces a gag reflex, making the procedure uncomfortable. Anxious patients are less likely to tolerate an unsedated procedure. However, in properly selected patients the upper endoscopy can be completed safely and comfortably with an ultrathin scope using this technique. In rare occurrences, it is not possible to adequately sedate a patient using standard conscious sedation agents. In that case, general anesthesia can be administered by appropriate personnel.

Risks of Upper Gastrointestinal Endoscopy

Upper endoscopy is generally considered a safe procedure. Complications are very uncommon. The commonest serious risk of endoscopy relates not to the endoscopy procedure itself but rather to the conscious sedation administered for it. The most common unwanted side effects of sedation are respiratory depression and hypotension. Side effects are more likely to occur in the elderly and in patients with coexisting medical problems. However, with judicious dosing of sedatives and continuous careful monitoring of pulse blood pressure, EKG, and oxygen saturation during the procedure, upper endoscopy can be performed safely in most patients. Other risks of endoscopy are very uncommon and relate to injury from the scope itself, including perforation of the bowel and bleeding. Bleeding is usually minor and controllable by endoscopic techniques such as cautery and injection of vasoconstricting agents. Frank bowel perforation, though very uncommon, usually requires immediate surgical intervention.

Depending on the intent of the procedure, upper GI endoscopy can be categorized as diagnostic or therapeutic. In diagnostic endoscopy, the aim is to visualize the bowel and if necessary take small mucosal tissue samples (biopsies). In therapeutic endoscopy, active interventions, such as dilating a narrowed segment, placing a stent, or controlling bleeding, are conducted.

Indications for Upper Gastrointestinal Endoscopy

A complete description of the medical aspects of all diseases evaluated and treated by upper GI endoscopy is beyond the scope of this article. Broadly, the indications for EGD are to diagnose and treat esophageal, gastric, and duodenal diseases. Esophageal diseases include gastroesophageal reflux disease, achalasia, and evaluation for potential injury of the esophagus from medications or toxins. Premalignant conditions of the esophagus, such as Barrett's esophagus, may be monitored endoscopically at intervals. Epigastric pain is among the commonest indications for diagnostic upper endoscopy. Disease processes in the stomach that may cause epigastric pain include gastric ulcers and gastric cancer. Persistent vomiting and vomiting of blood also require evaluation by EGD. Similarly, ulcers are among

the commonest pathologies found in the duodenum during upper endoscopy.

The Technique of Diagnostic Upper Gastrointestinal Endoscopy

The anatomic areas visualized by upper GI endoscopy include the esophagus, the stomach, and the proximal part of the duodenum (Fig. 1). Transient views obtained during passage of the scope through the mouth and pharynx (or nose and nasopharynx in the case of transnasal endoscopy) should not be considered a substitute for a thorough evaluation by an otorhinolaryngologist.

Patient Position

The patient is placed lying on his or her left side. The nurse stands near the head of the bed, monitoring the patient and suctioning excess fluid from the mouth. The endoscopist stands at the side of the bed facing the patient. A plastic bite block is placed between the patient's upper and lower teeth to protect the teeth and prevent the patient from biting the scope.

The patient is usually sedated before the scope is introduced. The scope is introduced into the upper esophagus via the oral or nasal route. The oral route is the most common. In the case of transoral upper endoscopy, the scope passes through the bite block and between the tongue and the palate. Once the scope reaches the back of the tongue, the tip is angulated approximately 90° so that it dips down into the pharynx. Using gentle pressure, the scope is advanced close to the posterior wall of the pharynx, until the upper esophageal sphincter (UES) is reached. At this time, the tip of the scope is opposed to the mucosa, limiting visualization. Gentle pressure is applied to the scope, pressing it against the UES. This pressure is sensed by the patient as an urge to swallow. Once the patient swallows, the UES opens, allowing passage of the scope into the upper esophagus. Patience is required during passage through the UES. If too much pressure is applied, perforation is risked. In some cases, the sedated patient is slow to swallow. Techniques to encourage passage of the scope through the upper esophageal sphincter include flexing the patient's neck gently forward and verbally encouraging the patient to swallow. With patience, the scope can be passed in almost every case. Occasionally, a large-diameter scope will not pass through this area. In that case, a smaller-diameter scope will usually traverse without problem. Once the scope is in the esophagus, it is gently pushed forward under direct visualization. Examination of the upper intestinal lining is performed during scope insertion or withdrawal, or both, at the discretion of the endoscopist.

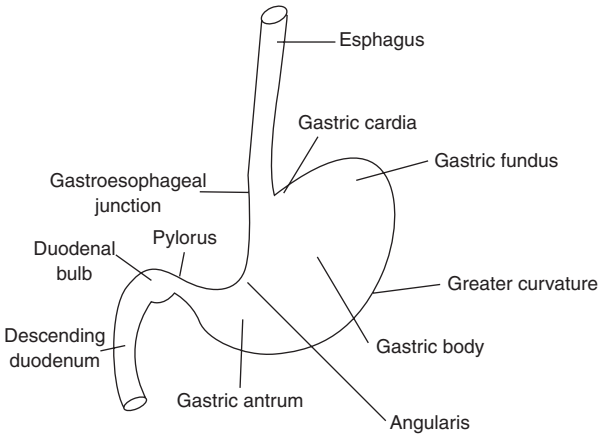


FIGURE 1 Representation of upper GI tract.

Examination of the Esophagus

The mucosa (lining) of the esophagus is examined by direct visualization. The mucosa is smooth and has a gray color (Fig. 2). The scope is gently pushed slowly down the esophagus. Contractions of the esophagus may occur during examination, causing apposition of the walls and limited visualization. In that case, the endoscopist simply waits for the contraction to pass before resuming the examination. Air is instilled into the lumen of the esophagus to expand it, thus keeping it open to improve visualization. The location of

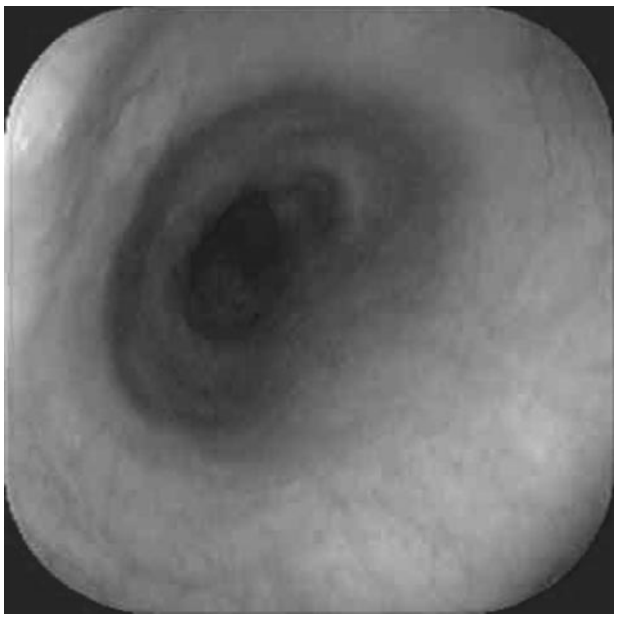


FIGURE 2 Endoscopic photograph of the normal esophagus.

anatomical landmarks within the upper intestine varies with patient size, particularly his or her height. In the average adult male, the upper esophageal sphincter is located 20 cm from the incisor teeth. The esophagus is approximately 20 cm long so the lower esophageal sphincter (LES) is located approximately 40 cm from the incisors. The muscular LES is not directly visible endoscopically, because of its location beneath the mucosa. The endoscopist will know she has reached the junction of the esophagus and the stomach when the esophageal lumen narrows and the lining changes appearance. The esophagus is lined with squamous mucosa, which has a grayish appearance. The stomach is lined by columnar mucosa, which has a salmon-colored appearance. The junction between the normal esophageal mucosa and the gastric mucosa is known as the squamocolumnar junction. In normal individuals, the squamocolumnar junction is located near the LES. There are two other important anatomical landmarks in this area. The first is the true gastroesophageal junction. This junction is determined by locating the proximal end of the normal gastric folds. In most patients, it is located in the same place as the squamocolumnar junction. Normally the esophagus passes through a hiatus in the diaphragm. At this point, the diaphragm abuts the gastroesophageal junction. In some cases, the abutment creates an impression on the bowel lumen that is visible endoscopically from inside. In patients with a hiatal hernia, this indentation is often seen in the proximal stomach.

Examination of the Stomach

Once the scope passes through the gastroesophageal (GE) junction, it enters the stomach. The stomach is divided into several parts. The most proximal part located just below the GE junction is the cardia. This is best visualized by turning the tip of the scope 180°, in a maneuver known as retroflexing (Fig. 3). The body of the stomach has a wide lumen. Gastric folds extend from the cardia down through the body and become less prominent or disappear as the lower stomach (antrum) is entered. The stomach is usually J-shaped. This shape has two implications for upper endoscopy. First, the shape causes acute angulation of the stomach on the lesser curvature, at a point known as the angularis. This location must be examined carefully as this is a "blind spot" and findings here may be missed on cursory examination. Second, as the scope is advanced through the stomach, its J-shape causes the scope to become bowed against the greater curvature of the stomach, thus stretching the stomach. Excessive stretching at this point may cause the patient discomfort, sometimes to the point of inducing undesirable retching or vomiting. To avoid this problem, the scope should be advanced slowly. The gastric antrum is recognized by its smooth, flat mucosa. In the distal portion of the antrum, there is an opening that leads into the duodenum. This is the pylorus. To pass the scope beyond the stomach into the duodenum, it is advanced gently until the scope tip is opposed to the pyloric opening. Gentle pressure is maintained, while the scope tip is maneuvered to keep it in line with the pylorus. The scope tip then passes into the duodenum.

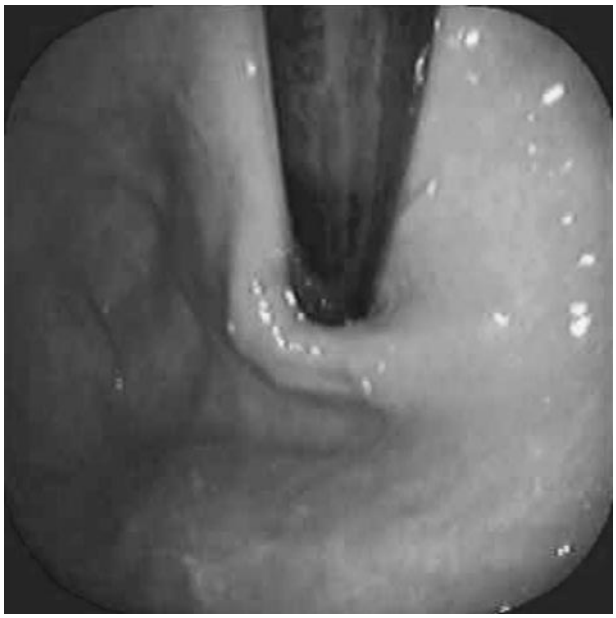


FIGURE 3 Endoscopic photograph of a retroflexed view of a normal stomach.

Examination of the Duodenum

The first part of the duodenum (bulb) is recognized by its location immediately beyond the pylorus and by its flat, smooth lining. The junction between the bulb and the second portion of duodenum is angulated, requiring maneuvering of the scope to advance further. In most cases, the second portion of the duodenum is reached by advancing the scope to the end of the bulb, torquing (twisting) the scope clockwise, and deflecting the tip of the scope upward. This maneuver creates a 90° turn clockwise and the second portion of the duodenum is entered. There is a partially blind area of mucosa here. The second portion of the duodenum has prominent folds that extend most or all of the way around the lumen (Fig. 4). The scope is advanced further into the descending portion of the duodenum. The scope may be further advanced to a variable extent. This extent varies according to how much of the scope is

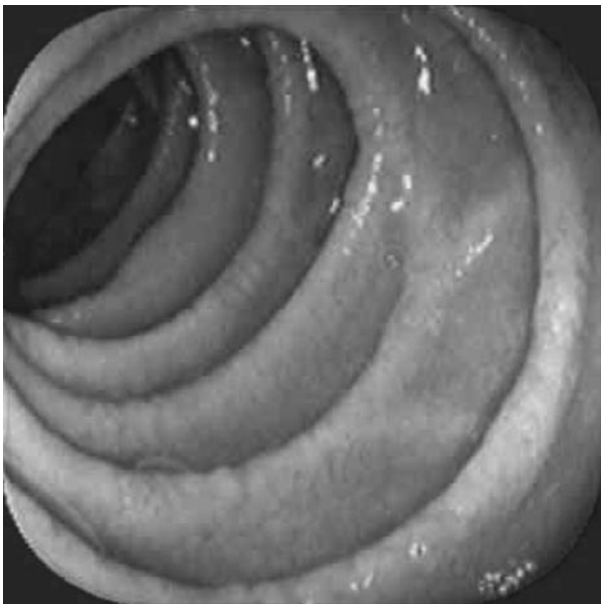


FIGURE 4 Endoscopic photograph of a normal descending duodenum.

looped or bowed in the stomach. In cases where there is much looping in the stomach, there may be insufficient scope length available to advance beyond the descending portion of the duodenum. Removing the gastric loop may help. This maneuver is known is "reducing." It is achieved by slowly pulling the scope out in small increments. As this is done, the excess scope that was previously bowed in the stomach moves away from the greater curvature and becomes available and effectively lengthens the available scope, causing the scope tip to advance despite the fact that the endoscopist is withdrawing the scope. This forward motion of the scope tip during withdrawal of the scope is known as paradoxical motion.

In most cases, detailed examination of the upper intestine is performed during withdrawal of the scope. Withdrawal from the duodenum must be done slowly, because the gastric loop will become reduced in size, causing the scope tip to rapidly pull back into the stomach. Complete visualization of the stomach requires adequate distension so that the gastric walls separate and the spaces between the gastric folds may be seen. Some portions of the stomach, especially the proximal parts, are best seen when the scope is in the retroflexed position. Air is removed by suction during withdrawal of the scope. Upon completion of the procedure, there will be some air in the intestine. The patient is encouraged to belch and to pass gas per the rectum to expel this air.

During diagnostic upper endoscopy, the endoscopist may wish to obtain a biopsy (small sample of the bowel lining) of the mucosa. In that case, a flexible wire with a small forceps at its tip is advanced through the working (biopsy) channel of the scope. When the end of the forceps is seen coming out of the scope tip, the forceps are directed to the desired part of the mucosa via a combination of advancing the forceps within the scope channel and deflecting the tip of the scope. Once the forceps are near the correct part of the mucosa, the jaws of the forceps are opened by the endoscopy assistant, by activating a lever in the handle. The open jaws of the forceps are advanced into the mucosa, pressing against it slightly so that a portion of mucosa lies between the jaws. The jaws of the forceps are then closed by the assistant, under instructions from the endoscopist. This traps a small part of the mucosa in the forceps. The tip of the forceps is quickly pulled back into the channel, thus removing a small portion of mucosa. The forceps are completely withdrawn from the biopsy channel and the tissue is retrieved from the jaws and sent for analysis.

Recovery after Endoscopy

After the scope is withdrawn, the patient is sent to the recovery area and the patient's pulse, blood pressure, EKG, and oxygen saturation values are monitored for approximately 45–60 min. By that time, most of the sedation effect will have worn off and the patient will be ready for discharge from the endoscopy unit. Discharge instructions are given and follow-up arrangements are made.

The Technique of Therapeutic Upper Endoscopy

In therapeutic upper endoscopy, the endoscopist aims to change tissue structure or anatomy by applying one or more interventions during the EGD. Choice of endoscope type is important when endoscopic therapy is planned. If an instrument channel is going to be in use for much of the therapeutic procedure, then the endoscopist will choose a scope with two working channels. One is used for the therapeutic instruments and the other is available for suctioning fluid and/or blood from the intestinal lumen or is used to flush water to wash the mucosa for better visibility. Newer endoscopes are available with very large suction channels. This allows suctioning of large pieces of solid or semisolid material from the mucosa (blood/food) to better visualize the mucosa. In therapeutic upper endoscopy, the scope is advanced into the upper GI tract

in a manner similar to that used in the diagnostic procedure.

Common Conditions Requiring Therapeutic Upper Endoscopy

Esophageal Stenosis (Stricture)

Esophageal stenoses may be benign or malignant. Different endoscopic techniques are used to dilate these two types of stenoses.

Benign strictures Benign esophageal stenoses are most often caused by fibrosis of the tissues in the lower esophagus as a result of chronic reflux of acid. Less common causes include postoperative, caustic (lye) injury, and pill esophagitis (medication-induced injury to the esophagus). These stenoses cause difficulty swallowing foods (mainly solid foods). Endoscopy demonstrates a narrowing of the esophagus, which in advanced cases may not allow passage of the endoscope beyond the stricture. There are two principal endoscopic techniques for dilating benign strictures: bougienage and balloon dilation.

Bougienage is dilation or stretching of the stricture using a long flexible device tapered at one end. To perform bougienage, the endoscope is removed and flexible dilators (bougies, French for candles) are guided into the esophageal stricture. This may be done blindly without a guiding system or by using a guide-wire and fluoroscopic monitoring. In the guide-wire technique, the scope is advanced to the distal stomach. A flexible metal guide-wire is placed through the working channel of the scope. The scope is withdrawn while the wire is maintained in place. At this point, the distal end of the wire is in the lower stomach and the proximal end is outside the patient's mouth. The bougie has a central channel so that it is fed over the wire and manually maneuvered into position. The position is monitored using X rays (fluoroscopy). The endoscopist will begin dilation using a small-diameter dilator, inserting and removing successively larger-diameter dilators through the stricture until the required amount of stretching is achieved. Dilation should be stopped if the patient experiences undue discomfort or if excessive resistance is felt during passage of the bougie.

Balloon dilation is performed with the scope tip in the esophagus. In balloon dilation, a long catheter is passed through the therapeutic channel of the scope. The distal end of the catheter has a deflated, sausage-shaped balloon on it. While the scope is kept in the esophagus, the balloon is slowly inflated under direct visualization in the region of the stricture, dilating it. To achieve the desired diameter of dilation, the endoscopist may use successively larger balloons. Variable-diameter balloons are available. In this case, the diameter of dilation is determined by how much air pressure is used to inflate the balloon. For benign esophageal strictures, the results from dilation using the balloons and bougies are approximately equal. The balloons have the theoretical advantage of maximizing the dilating forces and minimizing the shearing forces applied to the esophagus. However, bougies are reusable and may therefore have some cost advantages.

Malignant esophageal strictures Malignant strictures of the esophagus are most often caused by cancer of the lining of the esophagus itself or of the gastroesophageal junction. The patient usually presents with progressive difficulty swallowing and weight loss. When the endoscopic appearance suggests cancer, the diagnosis is confirmed with biopsies. When endoscopic palliation is deemed appropriate, the aim is to widen the esophageal lumen to allow as near normal swallowing as possible. Endoscopic treatment may include standard dilation with bougies or balloons as described above for benign strictures. However, malignant strictures restenose quickly, days or weeks after standard dilation. To prevent restenosis, a stent may be placed at the site of stricture. The most commonly used stent is the permanent self-expanding metal stent. To place this stent, the proximal and distal margins of the tumor are determined by endoscopic examination. A flexible catheter, over which a flexible stent is tightly wound, is introduced through the working channel of the scope to the narrowed area. Under direct endoscopic vision and X-ray guidance, the preloaded stent is opened. When it opens, it self-expands to widen the diameter of the esophagus at that point. Metal stents are considered permanent and once inserted are difficult or impossible to remove. In the event of recurrent narrowing, however, additional stents may be placed inside or adjacent to the original stent.

Achalasia To facilitate normal swallowing, the contractions in the body of the esophagus must be coordinated and the LES must open (relax) to allow the bolus of food to enter the stomach. Achalasia is a condition in which the LES valve does not relax adequately and the muscles of the main part of the esophagus do not contract in a coordinated fashion. The result is progressive problems swallowing both solid and liquid foods. Since the lack of peristalsis in the body of the esophagus cannot be reversed, the principal aim of therapy is to dilate the lower esophageal sphincter. This can be achieved endoscopically using two different techniques. The first is an injection of the neurotoxin Botox. The endoscopist performs a standard upper endoscopy. Then, using a needle on the tip of a catheter, 25 units of Botox is injected into each of the four quadrants at the

LES. This neurotoxin causes relaxation of the LES sufficient to help swallowing. The principal limitation of this treatment is that it lasts only approximately 9 months, thus requiring repeated treatments. Despite this limitation, it has a role in the management of older patients who have significant co-morbid disease and who might do poorly with riskier treatments.

The other endoscopic treatment for achalasia aims to rupture some of the muscle fibers in the LES, thus dilating it and thereby improving swallowing. This technique is called pneumatic balloon dilation. To begin, the upper GI tract is evaluated with a scope in the usual manner. A flexible wire is inserted via the scope as described in the therapeutic endoscopy section above. The scope is then withdrawn while the wire remains in place. A flexible catheter is placed over the wire under X-ray guidance. A balloon located near the distal end of the catheter is inflated and expands, stretching and rupturing the muscle fibers of the LES. The placement and inflation of the balloon are monitored under fluoroscopic guidance. The pressure in the balloon is monitored and controlled. This type of balloon dilation differs from dilation of benign peptic strictures in that the diameter of the dilating balloons is much larger (between 3 and 4 cm) and therefore the risk of perforation is higher.

Foreign body removal Accidentally or purposely ingested foreign bodies may be removed from the upper GI tract with the aid of endoscopy. Although small foreign bodies may pass through the GI tract without complication, larger ones that might become lodged in the small intestine are best removed endoscopically. To remove foreign bodies, upper endoscopy is performed in the usual manner. Once the foreign object is located, it is secured with a variety of different instruments, which are passed through the working channel of the scope. The instruments used to remove the foreign object include snare, forceps, net, and basket.

Control of upper gastrointestinal bleeding There are several causes of bleeding in the upper GI tract. The most common is acute bleeding from ulcers in the stomach or duodenum, Mallory-Weiss tears, and in patients with liver disease, varices (swollen veins) in the esophagus or proximal stomach. To control upper GI bleeding from ulcers, several endoscopic treatments may be applied. First, the site of bleeding is located and the area is washed to enhance visualization. Then a variety of liquids may be injected near the bleeding vessel to achieve hemostasis. The agents include epinephrine, which controls bleeding by its vasoconstricting effect, and alcohol, which compresses the bleeding vessel and also scleroses the surrounding tissues. Sclerosing agents such as tetradecyl are most often used to inject bleeding esophageal varices. One of the most common approaches to upper GI ulcer bleeding is to first inject the area of the vessel with epinephrine. This slows or stops the bleeding temporarily and allows better visualization. Then heat (cautery) therapy is applied. Most commonly, a cautery probe is placed through the working channel and out the tip of the scope. The tip of the probe has a bipolar coagulation device that compresses and delivers heat to the bleeding vessel. The probe is guided into position under direct visualization and is pressed firmly against the bleeding arteriole to oppose its walls. Then the probe is activated by a foot pedal, delivering heat energy to coagulate the vessel. The combination of injection with epinephrine and cautery is very successful in controlling bleeding. Other hemostatic control techniques include placing a clip on the bleeding vessel or, in the case of a bleeding esophageal varix, placing a rubber band to control bleeding.

See Also the Following Articles

Achalasia Endoscopic Ultrasonography • Endoscopy, Complications of • Esophageal Strictures • Foreign Bodies • Gastroesophageal Reflux Disease (GERD) • Occult Gastrointestinal Bleeding • Upper Gastrointestinal Bleeding

Further Reading

Cotton, P., and Williams, C. (1996). "Practical Gastrointestinal Endoscopy," 4th Ed. Blackwell Science Ltd., Oxford, UK.

DiMarino, A., and Benjamin, S. (2002). "Gastrointestinal Disease, an Endoscopic Approach," 2nd Ed. Slack Inc., Thorofare, NJ.

Vagus Nerve

KIRK HILLSLEY AND DAVID GRUNDY
University of Sheffield, United Kingdom

afferent fibers Nerve components that conduct sensory information from the periphery to the central nervous system. In this context, "sensory" does not necessarily imply sensation.

dorsal motor vagal nucleus Brain stem nucleus where the vagal preganglionic parasympathetic neuronal cell bodies are located.

efferent fibers Nerve components that conduct information from the brain to the periphery. In the vagus nerve, these are predominantly preganglionic parasympathetic nerve fibers. These fibers project to the viscera and make synaptic connections with postganglionic neurones within the organs they innervate, which in turn regulate motility, secretion, and blood flow. For this reason, the efferent supply is also referred to as motor innervation.

nodose, petrosal, and jugular ganglia Sensory ganglia; contain the cell bodies of afferent neurones that innervate the thoracic and abdominal viscera. These neurones project centrally to make synaptic connections in the nucleus of the tractus solitarius and peripherally to terminate in the viscera.

nucleus ambiguous Origin of motor neurones supplying the striated pharyngeal and esophageal muscles.

nucleus of the tractus solitarius Brain stem nucleus receiving vagal sensory input; the neurones project to the dorsal motor vagal nucleus to complete the brain stem circuit for vago-vagal reflexes and disseminate sensory information to higher brain regions.

The vagus nerve is the Xth cranial nerve connecting the brain to the periphery, including the gastrointestinal tract. It innervates structures from the esophagus to the colon and accessory organs such as the pancreas and liver. The vagal branches supplying these various target organs contain nerve fibers that are predominantly unmyelinated, conducting action potentials at around 1 m/sec. Classically, the vagus is referred to as a parasympathetic nerve, but this relates only to its efferent functions. However, it is also an afferent nerve and conveys a vast amount of sensory information from thoracic and abdominal viscera, which serves as a basis for reflexes that control and coordinate gut function.

INTRODUCTION

Galen (129–199 AD), who studied dissected corpses of Roman gladiators, was probably the first anatomist to describe the distribution of the vagus nerve. With the advent of modern neurobiological techniques, the cells of origin and the terminal distributions of afferent and efferent nerve fibers of the vagus have been described in terms of topography, neurochemistry, and electrophysiology. Such studies have revealed not only the anatomy of the vagus, but also the functional basis for brain–gut interactions.

MOTOR INNERVATION

The motor supply to the gastrointestinal (GI) tract is derived from two brain stem nuclei. The nucleus ambiguous provides the motor neurones that supply the striated muscle of the upper gastrointestinal tract, including the esophagus and the upper esophageal sphincter (Fig. 1). The preganglionic parasympathetic neurones that supply the smooth muscle portion of the gut originate in the dorsal motor vagal nucleus. This is a spindle-shaped nucleus running rostrocaudally through the medulla oblongata on either side of the central canal as it emerges into the 4th ventricle. It is organized into longitudinal columns of neurones that ultimately give rise to the different vagal branches that peal off the main anterior and posterior trunks to supply the various abdominal organs (Fig. 1). Within the gut wall, these preganglionic neurones synapse in the enteric nervous system, where the postganglionic neurones that supply the various gastrointestinal effectors are located. These effectors include the muscle layers and associated interstitial cells of Cajal, epithelia, and blood vessels, and regulate peristaltic contractions, secretions, and alterations in blood flow, respectively. In addition, vagal parasympathetic nerves can also influence resident immune cells, including macrophages and mast cells.

The textbook view of the parasympathetic innervation of the gut as being both exclusively cholinergic and entirely excitatory is outmoded. True, acetylcholine is the main neurotransmitter released from both motor and preganglionic vagal efferent fibers and, in the case of the latter, acts on nicotinic receptors located on the postsynaptic membrane of enteric neurones. However, it is now recognized that the enteric nervous system is not a simple relay station for the parasympathetic nerve

 Copyright 2004, Elsevier (USA). All rights reserved.

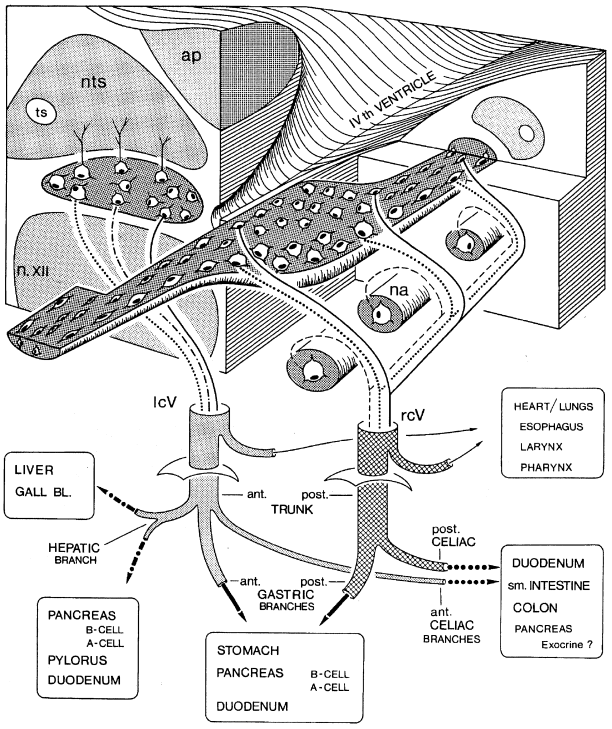


FIGURE 1 A schematic representation of the dorsal vagal complex of the medulla oblongata, viewed from a caudal and lateral (right) perspective. Also illustrated are the abdominal vagal trunks with their respective branches and a partial list of their different target tissues or organs. The dorsal motor nucleus of the vagus can be seen in a horizontal orientation on the right side of the brain stem and in a frontal orientation on the left side. The medial cell column on either side (associated, respectively, with the gastric branches) is indicated, with smaller cell bodies, and the lateral cell column on each side (associated, respectively, with the celiac branches) is indicated by larger somata. This perspective view and the different line symbols used for axons illustrate that each column of preganglionic neurons in the dorsal motor nucleus gives rise to a separate branch of the subdiaphragmatic vagus (hepatic branch neurons and axons are not shown). The nucleus ambiguous, which innervates supradiaphragmatic and striated tissues, is also illustrated on the right, where it is positioned ventrolateral and parallel to the dorsal motor nucleus of the vagus. Abbreviations: ap, area postrema; lcV, left cervical vagus; nts, nucleus of the tractus solitarius; rcV, right cervical vagus; ts, tractus solitarius; na, nucleus ambiguous. Modified by H. R. Berthoud from Ritter, S., *et al.* (1992). "Neuroanatomy and Physiology of Abdominal Vagal Afferents." CRC Press, with permission.

supply. It contains all the elements of an independent nervous system and as such can organize intrinsic reflexes to modulate contractile activity and secretion according to local demand. The enteric nervous system, in turn, is influenced by vagal preganglionic neurones and therefore provides the final pathway for both intrinsic and vagal reflexes. Acetylcholine is a transmitter in subpopulations of enteric motor neurones that innervate various effectors, and through an action on postjunctional muscarinic receptors brings about contraction or secretion. However, acetylcholine is not the only excitatory transmitter in enteric motor neurones; substance P and 5-hydroxytryptamine are examples of noncholinergic excitatory transmitters. Other enteric motor neurones contain transmitters such as adenosine triphosphate (ATP), vasoactive intestinal peptide (VIP), and nitric oxide (NO), which mediate inhibitory effects, particularly on muscle function. Consequently, vagal influence over gut function can be both excitatory and inhibitory, the latter causing relaxation of sphincters and the gastric body during accommodation of a meal or prior to emesis.

There are relatively few preganglionic vagal fibers compared to the many millions of neurones within the enteric nervous system. This fact, together with the clinical observation that gut function is relatively well maintained after vagotomy, a procedure common around the 1950s for treating peptic ulcer disease, has resulted in the importance of the vagal nerve supply being called into question. However, it is clear that the vagus nerve can have profound effects on gut function. What is intriguing, therefore, are the mechanisms that allow the enteric nervous system to maintain and restore relatively normal function in the face of a compromised vagal innervation after surgery or neuropathy.

SENSORY INNERVATION

Afferent fibers in the abdominal vagus trunks outnumber efferent fibers by about 10 to 1. The vagus nerve can be considered more a sensory than a motor nerve, and a vast amount of sensory information is conveyed to the brain stem. The sensory cell bodies are located mainly in the nodose ganglia. These are derived from the second and third epibranchial placodes during embryonic development and project centrally to the brain stem and peripherally to their terminations in the organs they innervate. The cell bodies synthesize the transmitters and enzymes that the nerve terminals utilize in the process of synaptic transmission. Glutamate is a major excitatory transmitter for gut reflexes, but other transmitter substances exist in the brain stem, including acetylcholine, catecholamines, neuropeptides (especially tachykinins), and γ-aminobutyric acid (GABA). Within the brain stem, vagal afferents terminate mainly in the nucleus tractus solitarius (nTS), although some afferents project into the dorsal motor vagal nucleus (DMVN), where they make monosynaptic connections with vagal motorneurones. Other afferent terminals project into the area postrema (AP), a region of the brain stem that integrates peripheral and central signals

involved in triggering nausea and vomiting. From the nTS, second-order neurones project a short distance into the dorsal motor vagal nucleus to complete the vago-vagal circuits through the brain stem that help control and coordinate gut function. Other second-order neurones ascend to higher regions of the brain, including the hypothalamus and limbic system, where they modulate autonomic function and behavior.

In the periphery, vagal sensory nerve terminals branch extensively in the various layers of the gut wall. Terminals in the mucosa transmit information relevant to luminal contents, although these sensory endings do not penetrate the epithelial lining. Instead, the endings form an elaborate network within the lamina propria, where they are exposed to chemicals absorbed from the lumen (nutrients and microbial antigen) or mediators released from specific cell types in the mucosal epithelium and lamina propria. Enteroendocrine cells have a specialized apical microvillus membrane that is capable of detecting the chemical constituents of the luminal content, and in response to an appropriate stimulus, release paracrine and endocrine mediators from the basolateral aspect of the cell. One such mediator is cholecystokinin (CCK), released by fat and protein digestion products in the lumen. CCK has widespread effects on gastrointestinal function, including inhibition of gastric emptying and stimulation of pancreatic secretion and bile flow, and also contributes to the satiety mechanisms that regulate food ingestion. These effects are mediated via activation of vagal afferents that are exquisitely sensitive to CCK. Other vagal afferents are activated by 5-hydroxytryptamine (5-HT) as part of a detection system for enterotoxins. Activation of these afferents can trigger diarrhea and vomiting, which dilute and expel the potential harmful material. The receptors to these and other mediators are synthesized in the cell soma and transmitted to the periphery, where they become linked to sensory signal generation. A wide range of receptors are expressed by vagal sensory neurones, including receptors to neuromodulators, cytokines, and growth factors.

Muscle mechanosensory endings transmit information concerning the level of distension or contractile activity within the gut wall. In some regions such as the stomach, elaborate sensory terminal arrays are found in the muscle layers, where they are positioned to detect changes in muscle tone. Other endings are found in association with myenteric ganglia (intraganglionic laminar endings, or IGLEs) and can detect the activity of overlying and underlying muscles that cause distortion of their endings. The threshold for activation of these vagal mechanoreceptors, unlike spinal nociceptors that respond predominantly to high levels of distension, is within the physiological range. In other words, levels of distension or contractions associated with ongoing activity are signaled to the central nervous system (CNS).

There are also paraganglia associated with the abdominal vagus that are outside of the GI tract, and these may serve associated functions. Vagal paraganglia consist of encapsulated glomus cells and occasional neurons. Paraganglia receive blood and lymph supplies, and 90% of them are innervated by ascending afferent vagal fibers that form cuplike varicosities around glomus cells. It has been proposed that vagal paraganglia may serve as a second line of chemoreceptive defense by monitoring the blood and lymph close to the GI tract and thus modulate afferent activity.

FUNCTIONAL ORGANIZATION OF VAGO-VAGAL REFLEXES

Sensory influences from both mucosal and mechanosensitive endings provide the CNS with a global view of the progress of digestion and the movement of contents through the gastrointestinal tract. Information on the luminal environment (pH, osmolarity, and antigen and nutrient content) and every contraction that occurs within the gastrointestinal tract is transmitted centrally to the brain. This information is used to set an appropriate parasympathetic outflow to the enteric nervous system. The enteric nervous system, in turn, integrates this parasympathetic input with local information provided by its own intrinsic sensory system. These reflexes help to match motor and secretor function to the digestive needs of the individual and coordinate gut regions that can be meters apart.

BEHAVIORAL CONSIDERATIONS AND SENSATION

Vagal afferent information plays an important role in regulating behavior. The role of CCK in satiety and 5-HT in nausea and vomiting are well established. More controversial is the role of vagal afferents in mediating illness behavior associated with acute-phase responses following bacterial translocation across the bowel wall. Circulating cytokines, particularly interleukin-1β (IL-1β) from intestinal macrophages, play a pivotal role, but there is also evidence that vagal afferents express receptors to IL-1β and contribute to anorexia and fever associated with acute-phase responses. Mucosal mast cells respond to antigenic signals and release mediators that can also influence vagal sensory signals. Because paraganglionic glomus cells are

immunoreactive for IL-1, it is likely that vagal paraganglia also play a role in immune responses.

The role of vagal afferents in sensation is controversial. Some vagal afferents project into the cervical spinal cord and may transmit signals to spinothalamic pathways that bring about sensations such as discomfort or pain. Current evidence suggests that vagal afferent information from the nTS projects to areas of the CNS associated with emotional aspects of visceral stimulation—for example, the limbic system—rather than cognition. There is also mounting evidence that vagal afferent traffic may influence descending spinal pathways associated with pain processing. Vagal afferents may therefore modulate pain rather than mediating pain per se.

See Also the Following Articles

Autonomic Innervation • Cholecystokinin (CCK) • Gastric Motility • Parasympathetic Innervation • Sensory Innervation

Further Reading

Berthoud, H. R., and Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Auton. Neurosci.* **85**, 1–17.

Berthoud, H. R., Patterson, L. M., Neumann, F., and Neuhuber, W. L. (1997). Distribution and structure of vagal afferent intraganglionic laminar endings (IGLEs) in the rat gastrointestinal tract. *Anat. Embryol. (Berl.).* **195**, 183–91.

Grundy, D., and Schemann, M. (2002). Motor control of the stomach. *In* "Innervation of the GI Tract" (S. Brookes and M. Costa, eds.), pp. 57–102. Taylor and Francis, New York and London.

Kirkup, A. J., Brunsden, A. M., and Grundy, D. (2001). Receptors and transmission in the brain–gut axis: Potential for novel therapies. I. Receptors on visceral afferents. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G787-G94.

Phillips, R. J., and Powley, T. L. (2000). Tension and stretch receptors in gastrointestinal smooth muscle: Re-evaluating vagal mechanoreceptor electrophysiology. *Brain Res. Rev.* **34**, 1–26.

Powley, T. L., Berthoud, H. R., Prechtl, J. C., and Fox, E. A. (1989). Fibers of the vagus nerve regulating gastrointestinal function. *In* "Brain–Gut Interactions" (Y. Tache and D. Wingate, eds.), pp. 73–82. CRC Press, Boca Raton, Florida.

Zhou, H., Ichikawa, H., and Helke, C. J. (1997). Neurochemistry of the nodose ganglion. *Prog. Neurobiol.* **52**, 79–107.

Variceal Bleeding

ALA I. SHARARA

American University of Beirut School of Medicine, Lebanon, and Duke University Medical Center

Child–Pugh classification Assignment of the severity of cirrhosis based on clinical and laboratory parameters, including encephalopathy, ascites, bilirubin, albumin, and prothrombin time; in order of increasing severity of cirrhosis, classes A, B, and C.

hepatic venous pressure gradient Defined as the gradient between the wedged, or occluded, hepatic venous pressure and free hepatic venous pressure; provides a reliable measurement of portal pressure.

primary prophylaxis Prevention of a first variceal bleed in patients with cirrhosis and varices.

sclerotherapy Injection of varices with a sclerosant such as sodium morrhuate or sodium tetradecyl sulfate.

secondary prophylaxis Prevention of recurrent variceal hemorrhage in a patient who suffered a first episode of variceal bleeding.

variceal band ligation Endoscopic placement of a rubber ligature around a varix.

Variceal bleeding, a major complication of portal hypertension resulting from cirrhosis, accounts for about 10–30% of all cases of upper gastrointestinal tract hemorrhage. More than any other cause of gastrointestinal bleeding, this complication results in considerable morbidity and mortality, prolonged hospitalization, and increased affiliated costs. Variceal hemorrhage develops in 25–35% of patients with cirrhosis and accounts for 80–90% of cases of bleeding episodes in these patients. From 10 to 30% of these episodes are fatal and as many as 70% of survivors will rebleed following an index variceal hemorrhage. Moreover, the 1-year survival is 34–80%

Copyright 2004, Elsevier (USA). All rights reserved.

immunoreactive for IL-1, it is likely that vagal paraganglia also play a role in immune responses.

The role of vagal afferents in sensation is controversial. Some vagal afferents project into the cervical spinal cord and may transmit signals to spinothalamic pathways that bring about sensations such as discomfort or pain. Current evidence suggests that vagal afferent information from the nTS projects to areas of the CNS associated with emotional aspects of visceral stimulation—for example, the limbic system—rather than cognition. There is also mounting evidence that vagal afferent traffic may influence descending spinal pathways associated with pain processing. Vagal afferents may therefore modulate pain rather than mediating pain per se.

See Also the Following Articles

Autonomic Innervation • Cholecystokinin (CCK) • Gastric Motility • Parasympathetic Innervation • Sensory Innervation

Further Reading

Berthoud, H. R., and Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Auton. Neurosci.* **85**, 1–17.

Berthoud, H. R., Patterson, L. M., Neumann, F., and Neuhuber, W. L. (1997). Distribution and structure of vagal afferent intraganglionic laminar endings (IGLEs) in the rat gastrointestinal tract. *Anat. Embryol. (Berl.).* **195**, 183–91.

Grundy, D., and Schemann, M. (2002). Motor control of the stomach. *In* "Innervation of the GI Tract" (S. Brookes and M. Costa, eds.), pp. 57–102. Taylor and Francis, New York and London.

Kirkup, A. J., Brunsden, A. M., and Grundy, D. (2001). Receptors and transmission in the brain–gut axis: Potential for novel therapies. I. Receptors on visceral afferents. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G787-G94.

Phillips, R. J., and Powley, T. L. (2000). Tension and stretch receptors in gastrointestinal smooth muscle: Re-evaluating vagal mechanoreceptor electrophysiology. *Brain Res. Rev.* **34**, 1–26.

Powley, T. L., Berthoud, H. R., Prechtl, J. C., and Fox, E. A. (1989). Fibers of the vagus nerve regulating gastrointestinal function. *In* "Brain–Gut Interactions" (Y. Tache and D. Wingate, eds.), pp. 73–82. CRC Press, Boca Raton, Florida.

Zhou, H., Ichikawa, H., and Helke, C. J. (1997). Neurochemistry of the nodose ganglion. *Prog. Neurobiol.* **52**, 79–107.

Variceal Bleeding

ALA I. SHARARA

American University of Beirut School of Medicine, Lebanon, and Duke University Medical Center

Child–Pugh classification Assignment of the severity of cirrhosis based on clinical and laboratory parameters, including encephalopathy, ascites, bilirubin, albumin, and prothrombin time; in order of increasing severity of cirrhosis, classes A, B, and C.

hepatic venous pressure gradient Defined as the gradient between the wedged, or occluded, hepatic venous pressure and free hepatic venous pressure; provides a reliable measurement of portal pressure.

primary prophylaxis Prevention of a first variceal bleed in patients with cirrhosis and varices.

sclerotherapy Injection of varices with a sclerosant such as sodium morrhuate or sodium tetradecyl sulfate.

secondary prophylaxis Prevention of recurrent variceal hemorrhage in a patient who suffered a first episode of variceal bleeding.

variceal band ligation Endoscopic placement of a rubber ligature around a varix.

Variceal bleeding, a major complication of portal hypertension resulting from cirrhosis, accounts for about 10–30% of all cases of upper gastrointestinal tract hemorrhage. More than any other cause of gastrointestinal bleeding, this complication results in considerable morbidity and mortality, prolonged hospitalization, and increased affiliated costs. Variceal hemorrhage develops in 25–35% of patients with cirrhosis and accounts for 80–90% of cases of bleeding episodes in these patients. From 10 to 30% of these episodes are fatal and as many as 70% of survivors will rebleed following an index variceal hemorrhage. Moreover, the 1-year survival is 34–80%

Copyright 2004, Elsevier (USA). All rights reserved.

and is inversely related to the severity of the underlying liver disease. Treatment of patients with gastroesophageal varices includes the prevention of the initial bleeding episode (primary prophylaxis), the control of active variceal hemorrhage, and the prevention of recurrent bleeding after a first episode (secondary prophylaxis).

PATHOGENESIS OF GASTROESOPHAGEAL VARICES

Chronic liver disease leading to cirrhosis is the most common cause of increased portal pressure, or portal hypertension. Portal pressure is a result of the relationship depicted by Ohm's law: $P = QR$, where P is the pressure along a vessel, Q is the flow in the vessel, and R is the resistance to that flow. In the majority of cases of liver disease, portal hypertension is the result of increased intrahepatic resistance and increased flow through the splanchnic system via a hyperdynamic circulation. Most forms of liver disease encompass aspects of each of these increases. The level of increased resistance to flow varies with specific forms of liver disease and may occur at presinusoidal or postsinusoidal levels, as in schistosomiasis and venoocclusive disease, respectively. Many forms of liver injury result in "sinusoidal" portal hypertension, the pathogenesis of which is complex but involves dynamic cellular factors (such as stellate cell constriction and contraction around sinusoids) and fixed elements (such as fibrosis). Recent studies suggest that an imbalance of the potent vasoconstrictor, endothelin-1, and the potent vasodilator, nitric oxide, may be important in the genesis of increased intrahepatic resistance, which is an early and critical component of most forms of portal hypertension.

Varices represent portosystemic collaterals derived from dilatation of preexisting embryonic vascular channels, including those between the coronary and short gastric veins and the intercostal, esophageal, and azygous veins. In the distal esophagus, over an area extending 2–5 cm from the gastroesophageal junction, veins are found more superficially in the lamina propria rather than the submucosa. This results in reduced support from surrounding tissues because of the predominant intraluminal location of these varices and may explain the predilection for bleeding at this site. The opening and dilation of portosystemic collaterals appears to depend on a threshold portal pressure gradient (portal minus free hepatic vein pressure) of 12 mm of mercury (Hg), below which varices do not form. This pressure gradient is necessary but not sufficient for the development of gastroesophageal varices. In other words, patients with gradients in excess of 12 mmHg do not invariably develop varices. The prevalence of esophageal varices in cirrhotic patients ranges from 24 to 90% and is partly related to the duration and severity of cirrhosis.

PREDICTION OF VARICEAL HEMORRHAGE

Not all patients with varices bleed; gastroesophageal variceal hemorrhage occurs in only 30–35% of patients with cirrhosis, most commonly within 2 years of the diagnosis of varices. It is unclear what leads to rupture and bleeding in individual patients. Local factors such as changes in lower esophageal sphincter pressure or acid reflux do not appear to play a role in the pathogenesis of variceal rupture. Rather, the main determinant of bleeding is the variceal wall tension, a function of the transmural pressure, radius, and wall thickness of the vessel (as indicated by Frank's modification of LaPlace's law). Variceal rupture is directly related to physical factors such as the thickness and elastic properties of the vessel, in addition to intravariceal and intraluminal pressure. Furthermore, the severity of bleeding is related to transmural pressure and the size of the rent in the varix, and is inversely proportional to blood viscosity.

A major focus of investigation, with important clinical implications, is understanding which patients are most likely to bleed. Clinical characteristics that denote an increased risk of first variceal hemorrhage include continued alcohol use and liver decompensation, whereas endoscopic findings that predict a higher risk of bleeding include larger size of varices and the presence of endoscopic red signs (described as cherry red spots, hemocystic spots, or red wale markings) on the variceal wall, indicating dilated intraepithelial and subepithelial superficial veins. A combination of clinical and endoscopic findings, including the Child–Pugh class, size of varices, and the presence or absence of red wale markings, was found to correlate highly with the risk of first bleeding in patients with cirrhosis. Hemodynamic parameters examined include the measurement of the hepatic venous pressure gradient (HVPG), azygous blood flow, as well as direct measurement of intravariceal pressure. HVPG calculated by the gradient of wedged and free hepatic vein pressure (normal value, 5 mmHg) is used most frequently and provides reliable measurement of portal pressure in patients with cirrhosis. The extent of elevation of HVPG may be the best indicator of risk of bleeding, severity of bleeding, and survival. A rise in pressure in a patient with known varices increases the risk of bleeding, and the extent of portal pressure elevation

appears to have an inverse relationship to prognosis after hemorrhage has occurred. In general, however, a linear relationship between the degree of portal hypertension and the risk of variceal hemorrhage or variceal formation does not exist, so this technique cannot be used routinely to identify individual patients at high risk of bleeding. Of interest, the reported circadian variation in portal pressure (HVPG is highest shortly after midnight and lowest at 7 PM) may offer a possible explanation for the predilection of variceal bleeding to occur during the late evening to early hours of the day.

PRIMARY PREVENTION OF BLEEDING FROM ESOPHAGEAL VARICES

Based on prospective studies of cirrhotic patients with varices identified at endoscopy and studies of untreated groups in randomized controlled trials, the risk of bleeding from esophageal varices has been estimated at 25−35% at 1 year. Given the poor outcome of cirrhotic patients following variceal hemorrhage, attempts have been made to identify those at high risk and prevent bleeding. In a large prospective study involving 321 patients with cirrhosis, the risk of bleeding was found to correlate with the size of varices, the presence of endoscopic red wale markings, and to the severity of liver disease. Screening endoscopy is generally recommended for patients with cirrhosis to determine the presence of large varices, but the cost-effectiveness of this approach remains controversial. The use of clinical parameters such as the presence of splenomegaly and low platelet count may help select patients likely to have large varices on endoscopy. Therapy for primary prophylaxis against variceal bleeding has evolved considerably over the past decade and is summarized in Table I.

Pharmacologic Therapy

The general objective of pharmacologic therapy for variceal bleeding is to reduce portal pressure and consequently intravariceal pressure and wall tension. Drugs that reduce portocollateral venous flow (vasoconstrictors) or intrahepatic vascular resistance (vasodilators) have been used, and include beta-blockers, nitrates, α2-adrenergic blockers, spironolactone, pentoxifylline, and molsidomine. Because varices do not bleed at an HVPG less than 12 mmHg, reduction to this level is ideal, but substantial reductions in HVPG (i.e., by >20%) are also clinically meaningful.

Beta-blockers exert their beneficial effect on portal venous pressure by diminishing splanchnic blood flow and consequently gastroesophageal collateral and azygous blood flow. The nonselective beta-blockers, such as propranolol and nadolol, are preferred because of the dual benefit of β1- and β2-receptor blockade. β1-Receptor blockade reduces cardiac output and causes splanchnic vasoconstriction by reflex activation of α-adrenergic receptors in the splanchnic circulation, whereas β2-receptor blockade results in splanchnic and peripheral vasoconstriction by eliminating β2-receptor-mediated vasodilatation, thereby allowing unopposed α-adrenergic receptor-mediated vasoconstriction. In the absence of HVPG determination, beta-blockers are titrated in order to achieve a reduction in resting heart rate to 55 beats/minute, or 25% of baseline. Propranolol is generally given as a long-acting preparation and titrated to a maximal dose of 320 mg/day. Nadolol is initiated at 80 mg daily up to a maximal daily dose of 240 mg.

The portal pressure-reducing effects of beta-blockers are, however, unpredictable and neither the resultant reductions in heart rate nor drug blood levels are good indicators of response to therapy. For example, portal venous pressure is reduced in about 60−70% of patients who receive propranolol therapy, but only 10−30% of these patients show a substantial response (i.e., >20% reduction). Additionally, approximately 20−25% of patients have no measurable decline in portal pressure despite increasing dosage of propranolol. The effectiveness of beta-blockers for primary prophylaxis against variceal bleeding has been demonstrated in several controlled trials. Additionally, meta-analyses have revealed a 40−50% reduction in bleeding and a trend toward improved survival. Further, a cost-effectiveness analysis comparing propranolol to sclerotherapy and shunt surgery found propranolol to be the only cost-effective form of primary prophylaxis.

In addition to beta-blockers, a number of vasodilators have been investigated in portal hypertensive patients and in animal models of portal hypertension; the long-acting organic nitrates, isosorbide dinitrate and isosorbide-5-mononitrate, have received the greatest attention. Isosorbide mononitrate, the active component formed by rapid denitration of dinitrate preparations in the liver, has a relatively long half-life (about 5 hours) and dose linear kinetics even in the presence of liver or kidney disease, whereas the dinitrate preparation undergoes extensive first-pass hepatic metabolism, resulting in a short half-life and unpredictable bioavailability when used in cirrhotics. The exact mechanism of action of nitrates is unclear, but is thought to be mediated primarily by reducing intrahepatic resistance—perhaps by inhibiting sinusoidal constriction caused by hepatic stellate cells or myofibroblasts—and possibly by splanchnic arterial vasoconstriction induced in response to venous pooling and vasodilation in other

TABLE I Summary of Therapy for Esophageal Varices[a]

Purpose of therapy	First-line therapy	Comments	Alternative therapy[b]	Comments
Primary prophylaxis[c]	Beta-blockers alone or in combination with isosorbide mononitrate[d]	Nitrate alone is not recommended; in advanced (Child–Pugh class C) liver disease, optimal therapy is unclear (probably band ligation); transplantation should be considered for patients in this group	Band ligation	Band ligation is indicated for patients with contraindications to or intolerance of medical therapy; the effectiveness of combined beta-blockers and band ligation is unknown; neither TIPS nor sclerotherapy is recommended for primary prophylaxis
Active variceal bleeding	Octreotide (or terlipressin) and endoscopic therapy[e]	Octreotide (or terlipressin) should be continued for a minimum of 24–28 hours; band ligation may be superior to sclerotherapy	Balloon tamponade TIPS	Tamponade is indicated primarily as a temporizing measure TIPS is reserved for those with refractory or recurrent early bleeding
		Antibiotic prophylaxis should be considered, especially in patients with ascites	Shunt surgery	Surgery is reserved for those in whom TIPS is not feasible
Secondary prophylaxis[f]	Band ligation alone or in combination with beta-blockers with or without isosorbide mononitrate	The combination of band ligation and beta-blockers with or without isosorbide mononitrate is likely to be more effective than either therapy alone	TIPS	TIPS is best used as a bridge to transplantation in patients with advanced liver disease
		Patients with advanced liver disease often have an intolerance to beta-blockers	Shunt surgery	Shunt surgery should be reserved for selected patients with Child–Pugh class A and class B cirrhosis

[a] Data from Sharara and Rockey (2001).

[b] TIPS, Transjugular intrahepatic portosystemic shunt.

[c] Variceal hemorrhage occurs in 25–30% of patients within 2 years after the documentation of varices.

[d] Beta-blockers reduce the risk of variceal hemorrhage to 15–18%, and the combination of beta-blockers and isosorbide mononitrate reduces the risk to 8.5–10%. The beta-blocker propranolol is generally given as a long-acting preparation, and the dose is titrated to a maximum of 320 mg/day. The initial dose of the beta-blocker nadolol is 20 mg/day and the dose is increased up to a maximum of 80 mg/day.

[e] Octreotide is usually given as an infusion of 25–50 μg/hour (with or without a bolus). The dosage of terlipressin is 2 mg every 4 hours for the first 24 hours, then 1 mg every 4 hours.

[f] Bleeding recurs in approximately two-thirds of patients within 1 year after the initial hemorrhage.

regional vascular beds. Monotherapy with nitrates is ineffective in primary prophylaxis and may have detrimental effects, particularly in cirrhotic patients with ascites, such as accentuation of hemodynamic and laboratory signs of hypovolemia and vasodilatation, akin to those seen in what has been called postparacentesis circulatory dysfunction.

The addition of isosorbide mononitrate to propranolol has been shown to result in an enhanced reduction in portal pressure in humans. In a randomized controlled study involving 42 patients with cirrhosis and esophageal varices, a reduction of greater than 20% in HVPG was documented in only 10% in the propranolol group, compared to 50% in the combination therapy group. In patients with Child–Pugh class A and B cirrhosis, the addition of isosorbide mononitrate to nadolol has been shown, in a randomized trial, to result in a greater than 50% additional reduction in variceal bleeding rate when compared with nadolol monotherapy (12%, compared with 29%).

Endoscopic Therapy

Over the past two decades, endoscopic therapies have assumed a prominent role in treatment of esophageal varices. Endoscopic sclerotherapy is performed by injecting a small volume of any of a number of sclerosant solutions (such as sodium tetradecyl sulfate, sodium morrhuate, ethanolamine, or absolute ethanol), usually in an intravariceal location. These sclerosing agents are all highly effective in achieving hemostasis in approximately 80–90% of patients, although their mechanisms of action are not entirely clear (e.g., local tamponade effect, or primary venous thrombosis). Multiple randomized trials have examined the effect of endoscopic sclerotherapy in the prevention of first variceal bleeding. Results are conflicting, with few trials showing significant reduction in bleeding and most showing no difference. One study was stopped prematurely because of higher mortality in the sclerotherapy group as compared to the sham-treatment group.

Band ligation of esophageal varices utilizes the same methodology employed for years in elastic band ligation of internal hemorrhoids. A suction adapter is fitted on the tip of the endoscope, the varix is suctioned into the cylinder, and a rubber band is deployed around the varix, resulting in ischemic necrosis of the mucosa and submucosa but leaving an intact muscularis propria. The resultant shallow ulcerations are replaced by granulation tissue and dense mature scar tissue and lead to variceal obliteration. The only published trial, to date, comparing ligation to propranolol in the primary prevention of variceal bleeding showed an actuarial rate of bleeding of 15% in the ligation group compared to 43% in the propranolol monotherapy group. The study was criticized, however, because of the higher than expected rate of bleeding in the propranolol group as well as the relatively low mean dose of propranolol used. Furthermore, the preliminary results of another trial published in abstract form show no advantage of ligation when compared to the combination of nitrates and beta-blockers. Nevertheless, ligation may be an acceptable option for patients at high risk of variceal bleeding and who are intolerant of, or have contraindications to, medical therapy. The optimal primary prophylaxis in patients with Child–Pugh class C cirrhosis remains unclear, and may arguably be liver transplantation.

MANAGEMENT OF ACUTE VARICEAL HEMORRHAGE

Variceal hemorrhage is usually an acute clinical event characterized by rapid gastrointestinal blood loss presenting as hematemesis (which can be massive), with or without melena or hematochezia. Hemodynamic instability (i.e., tachycardia and/or hypotension) is common. A successful outcome in all cases of upper gastrointestinal hemorrhage, including those of variceal bleeding, hinges on prompt resuscitation, hemodynamic support, and attempt at correction of any hemostatic derangement, preferably in the setting of an intensive care unit.

Although variceal bleeding is common in patients with cirrhosis presenting with acute upper gastrointestinal hemorrhage, other causes of bleeding, such as ulcer disease, must be considered in the differential diagnosis. Urgent initiation of empiric pharmacologic therapy with vasoactive agents is indicated in situations in which variceal hemorrhage is likely. Subsequently, direct examination of the upper gastrointestinal tract by an experienced endoscopist is critical to establish an accurate diagnosis and to provide the rationale for immediate and subsequent therapies. The immediate steps in the management of acute variceal bleeding include (1) volume resuscitation, (2) prevention of complications, (3) ensuring hemostasis, and (4) initiating measures to prevent early and delayed rebleeding. Optimal volume resuscitation is the most critical aspect of management of all patients with gastrointestinal hemorrhage. Patients with suspected variceal bleeding should have large-bore intravenous lines placed and should receive crystalloids and/or blood in proportion to the degree of hemodynamic instability. Coagulopathy should also be corrected. Caution should be exercised, however, in avoiding overtransfusion because of the

theoretical concern of rebleeding from the variceal rent after excessive blood volume expansion. Although largely unsupported, it is traditional in medical training and practice to use a hematocrit of 30% as an upper safety margin following transfusion.

There is no evidence that placement of a nasogastric tube contributes to variceal rupture or bleeding. Furthermore, a nasogastric tube can be helpful in providing a crude index of the activity and amount of bleeding as well as in decompressing the stomach of blood and clots, thereby reducing the risk of aspiration. Because of the risk of aspiration, elective endotracheal intubation should be considered prior to endoscopy in patients with massive bleeding, severe agitation, or altered mental status. Patients with variceal hemorrhage and ascites are at increased risk of bacterial infections, particularly spontaneous bacterial peritonitis. This risk appears to be increased in the setting of uncontrolled hemorrhage or as a result of transient bacteremia following endoscopic sclerotherapy or variceal ligation. Short-term systemic antibiotics (e.g., third-generation cephalosporins for 4 to 10 days) have been shown, in a meta-analysis, to decrease the risk of bacterial infections and to reduce mortality in cirrhotic patients with gastrointestinal bleeding. Last, patients with cirrhosis and massive gastrointestinal hemorrhage are at risk for the development (or exacerbation of preexisting) portosystemic encephalopathy. This complication should be recognized and managed accordingly. Furthermore, it has been suggested that the presence of blood in the gut may, as a result of its large protein load, lead to reflex splanchnic hyperemia, presumably via the release of vasoactive gastrointestinal peptides. Theoretically, the resultant increase in portal venous pressure may increase the risk of early rebleeding. In support of this theory is a controlled trial from France showing a reduction in blood requirements and in mortality following whole gut irrigation with isotonic mannitol in cirrhotic patients with gastrointestinal bleeding.

PHARMACOLOGIC THERAPY

Pharmacologic therapy can be administered early, requires no special technical expertise, and is thus a desirable first-line option for the management of acute variceal hemorrhage. Drugs that reduce portocollateral venous flow (vasoconstrictors) or intrahepatic vascular resistance (vasodilators) or both have been used in order to achieve this effect. Vasoconstrictors work by decreasing splanchnic arterial flow whereas vasodilators are used in combination with vasoconstrictors in order to reduce their systemic side effects but may also exert an added beneficial effect on intrahepatic resistance (Table I).

Vasopressin and Glypressin

Vasopressin is a nonselective vasoconstricting agent that causes a reduction of splanchnic blood flow and thereby a reduced portal pressure. Introduced in 1956, vasopressin was the first agent shown to reduce hemorrhage from varices. Because of its short half-life, vasopressin must be given by continuous intravenous infusion and dosage is limited because of significant systemic vasoconstriction, necessitating discontinuation of therapy in 25% of cases. Severe vascular complications such as myocardial ischemia and infarction, mesenteric and limb ischemia, and cerebrovascular accidents have been described. In addition, vasopressin prevents free water excretion from the kidneys and can result in fluid overload, hyponatremia, and worsening ascites. The concomitant use of nitroglycerin results in improved efficacy in the control of variceal hemorrhage (up to 70%) and reduction in the systemic side effects of vasopressin. The beneficial effect of nitroglycerin appears to be a result of the effect of nitric oxide on intrahepatic vascular resistance, causing a reduction in portal pressure. This favorable effect of nitroglycerin is independent of the mode of delivery (sublingual, intravenous, or transdermal). Blood pressure and electrocardiographic monitoring are recommended during therapy with vasopressin and nitroglycerin.

Triglycyllysine vasopressin (glypressin, terlipressin) is a synthetic vasopressin analogue activated *in vivo* by cleavage of its N-terminal residue. Compared with vasopressin, glypressin has fewer side effects and a longer biologic half-life, allowing its use as a bolus intravenous injection (2 mg every 4 hours for the initial 24 hours, then 1 mg every 4 hours for the next 24 to 48 hours). This advantage has led to its successful use in the "field" for cases of suspected variceal bleeding prior to or during transport to the hospital. Glypressin has been shown in multiple placebo-controlled trials to control bleeding in about 80% of cases and is the only pharmacologic therapy proven, to date, to reduce mortality from acute variceal hemorrhage. Glypressin is not currently available in the United States.

Somatostatin, Octreotide, and Vapreotide

Somatostatin, a naturally occurring peptide, and its analogues, octreotide and vapreotide, stop variceal hemorrhage in up to 80% of patients and are generally considered to be equivalent to vasopressin, terlipressin, and endoscopic therapy for the control of acute variceal bleeding. Their precise mechanism of action is unclear but may be due to an effect on the release of vasoactive

peptides (such as glucagon, vasoactive intestinal peptide, and substance P) or to reduction of postprandial hyperemia. Somatostatin is used as a continuous intravenous infusion of 250 μg/hour following a bolus injection of 250 μg. Octreotide is used as a continuous infusion of 50 μg/hour and does not require a bolus injection. Side effects are minor, including hyperglycemia and mild abdominal cramps.

A recent area of interest has been the use of octreotide or vapreotide in combination with endoscopic therapy. In two separate studies, the addition of octreotide to endoscopic sclerotherapy or banding resulted in improved control of bleeding and reduced transfusion requirements, but did not reduce overall mortality. Based on this work, a continuous infusion of octreotide of 25 μg/hour has been recommended for 5 days following emergency endoscopic therapy. The added benefit of octreotide appears, however, to be largely limited to the first 24–48 hours of use. In one study, all early rebleeding episodes occurred within 2 days of endoscopic therapy; in the other study, no statistically significant difference was noted after the first 24 hours between the octreotide or placebo groups as far as rebleeding or the mean units of blood transfused.

ENDOSCOPIC THERAPY

Endoscopic sclerotherapy stops variceal hemorrhage in 80–90% of cases. The advantages of sclerotherapy include its ability to achieve definitive control of bleeding under direct endoscopic vision, as well as the wide availability, ease of use, and low cost of the technique. Its drawbacks include a significant risk of local complications, including ulceration, bleeding, stricture, and perforation. Rare systemic complications have been reported, including bacteremia with endocarditis, splenic or brain abscess formation, and portal vein thrombosis.

Randomized trials in patients with acute variceal bleeding have shown that endoscopic variceal band ligation is essentially equivalent to sclerotherapy in achieving initial hemostasis. Because of the limited view with the fitted suction adapter, band ligation may technically be more difficult when bleeding is massive. The newer multiband devices allow the placement of up to 10 bands in one setting, obviating the need for repeated intubation or the need for an endoscopic overtube. Complications of endoscopic variceal band ligation (EVBL) include superficial ulcerations, transient chest discomfort, and, rarely, stricture formation.

Because of the deeper submucosal location of gastric varices, injection sclerotherapy and rubber band ligation are not usually effective in controlling acute bleeding from gastric varices. The injection of *N*-butyl-2-cyanoacrylate tissue glue or the use of large, detachable endoscopic minisnares has been shown, in small, uncontrolled trials, to be effective for bleeding gastric varices.

Balloon Tamponade

The use of the Sengstaken–Blakemore, or Minnesota, tube for hemostasis of variceal bleeding is based on the principle of the application of direct pressure on the bleeding varix by an inflatable (esophageal or gastric) balloon fitted on a rubber nasogastric tube. It is important to note that only physicians experienced in this technique should place these tubes. When properly applied, balloon tamponade is successful in achieving immediate hemostasis in almost all cases. However, early rebleeding following balloon decompression is high. Complications of balloon tamponade include esophageal perforation or rupture, aspiration, and asphyxiation from upper airway obstruction. Balloon tamponade is generally not recommended and should largely be reserved for rescue of cases of hemorrhage uncontrolled by pharmacologic and endoscopic methods and as a temporary bridge to more definitive therapy.

Transjugular Intrahepatic Portosystemic Shunt

Treatment with a transjugular intrahepatic portosystemic shunt (TIPS) consists of the vascular placement of an expandable metal stent across a tract created between a hepatic vein and a major intrahepatic branch of the portal system. TIPS can be successfully performed in 90–100% of patients, resulting in hemodynamic changes similar to a partially decompressive side-to-side portacaval shunt but avoiding the morbidity and mortality associated with a major surgical procedure. TIPS as been shown to be effective in the treatment of refractory, uncontrolled, acute variceal bleeding. Of note, patients with advanced liver disease and multiorgan failure at the time of TIPS have a 30-day mortality that approaches 100%.

Surgical Therapy

Surgery is generally considered in the setting of continued hemorrhage or recurrent early rebleeding (uncontrolled by repeated endoscopic or continued pharmacologic therapy) and when TIPS is not available or technically feasible. Surgical options include portosystemic shunting or esophageal staple transection alone or with esophagogastric devascularization and

splenectomy (Sugiura procedure). Devascularization procedures may be useful in patients who cannot be shunted because of splanchnic venous thrombosis. Regardless of the choice of surgical technique, morbidity is high and the 30-day mortality for emergency surgery has approached 80% in some series. Understandably, "rescue" liver transplantation is not a practical option for patients with uncontrolled variceal hemorrhage.

PREVENTION OF RECURRENT VARICEAL BLEEDING

Variceal hemorrhage recurs in approximately two-thirds of patients, most commonly within the first 6 weeks after the initial episode. This period of high risk can be subdivided into an early (first 5 days) and a late period, with the highest risk falling within the first 5 days from the initial hemorrhage. As mentioned previously, the risk of early rebleeding is reduced by the adjuvant use of octreotide or vapreotide and possibly glypressin and somatostatin, primarily in the first 24 to 48 hours, after initial endoscopic or pharmacologic control of hemorrhage. Risks for early rebleeding include clinical, endoscopic, and hemodynamic parameters, such as the severity of the initial bleed, the degree of liver decompensation, and the presence of encephalopathy and impaired renal function. Endoscopic parameters predictive of early rebleeding include the presence of active bleeding, stigmata of recent bleeding, and/or large varices. The severity of portal hypertension correlates closely with the severity and risk of rebleeding as well as actuarial probability of survival following an index episode. In a cohort of patients presenting with variceal hemorrhage, those with an initial HVPG greater than 20 mmHg had a 1-year mortality of 64% compared to 20% for patients with lesser elevations in portal pressure.

Given the high risk of recurrent hemorrhage and its associated morbidity and mortality, strategies aimed at prevention should be rapidly instituted following the index episode (Fig. 1). The choice of preventive therapy should, therefore, take into consideration the efficacy of therapy, the side effects of the selected treatment, the patient's expected survival, and overall cost. Preventative strategies include pharmacologic, endoscopic, and surgical methods.

Pharmacologic Therapy

Reducing the portal pressure by more than 20% from the baseline value pharmacologically results in a reduction in the cumulative probability of recurrent bleeding at 1, 2, and 3 years from 28% at 1 year, 39% at 2 years, and 66% at 3 years, to 4, 9, and 9%, respectively. Although adjustment of medical therapy based on portal pressure measurement would be ideal, HVPG determination may not be readily available, thus therapy must be adjusted using empiric clinical parameters.

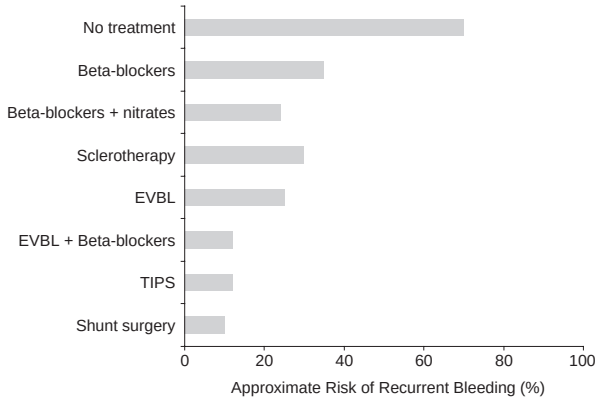


FIGURE 1 Relative effectiveness of available therapies for the prevention of recurrent variceal bleeding. The estimates shown are based on the cumulative data available in the literature (recurrent bleeding at 1 year). EVBL, Endoscopic variceal band ligation; TIPS, transjugular intrahepatic portosystemic shunt. Reproduced with permission from Sharara and Rockey (2001).

Agents with a favorable effect on portal pressure in humans with cirrhosis include beta-blockers, nitrates, α2-adrenergic blockers, spironolactone, pentoxifylline, and the venous dilator molsidomine. However, the most widely used drugs and the only ones for which there is sufficient evidence are the nonselective beta-blockers (which have been used with or without oral nitrates). Several randomized placebo-controlled trials, including a meta-analysis, have demonstrated that beta-blockers prevent rebleeding and prolong survival. The major side effects of beta-blockers are fatigue, impotence, depression, bradycardia, hypotension, and sleep disorders. Contraindications to therapy include congestive heart failure, asthma or chronic obstructive pulmonary disease, and atrioventricular heart block.

The addition of isosorbide mononitrate to beta-blockers appears to enhance the protective effect of beta-blockers alone for the prevention of recurrent variceal bleeding, but offers no survival advantage and reduces the tolerability of therapy. Compared with either sclerotherapy or endoscopic band ligation, combination medical therapy is superior in reducing the risk of recurrent bleeding in patients with esophageal variceal hemorrhage, primarily in patients with Child–Pugh class A and B cirrhosis. Notably, in patients who show a significant hemodynamic response to therapy (defined as a reduction in the hepatic venous pressure

gradient to less than 12 mmHg or more than 20% of the baseline value), the risk of recurrent bleeding and of death is significantly reduced.

Endoscopic Therapy

Endoscopic therapy has been established over the past decade as a therapeutic cornerstone for prevention of esophageal variceal rebleeding. Gastric varices, however, are not effectively treated by sclerotherapy or ligation. Patients with recurrent gastric variceal hemorrhage are best treated by *N*-butyl-2-cyanoacrylate injection or by nonendoscopic means.

Sclerotherapy reduces the risk of recurrent esophageal variceal bleeding from approximately 65% to 30–35% at 1 year, but does not appear to improve overall mortality. Sclerotherapy is performed every 10–14 days until varices are eradicated, usually after five or six sessions. A meta-analysis of nine trials found endoscopy and beta-blockers to be equivalent with respect to the risk of rebleeding and survival. Moreover, combination pharmacotherapy (beta-blockers plus isosorbide-5-mononitrate) is superior to sclerotherapy alone in patients with Child–Pugh class A or B cirrhosis.

Endoscopic variceal band ligation is highly effective at obliterating varices. Ligation is associated with a lower risk of recurrent bleeding, compared to sclerotherapy (approximately 25 vs. 30% at 1 year), fewer complications, reduced overall cost, and higher survival. Ligation should, therefore, be considered standard endoscopic therapy for secondary prophylaxis. As with sclerotherapy, ligation is performed every 10 to 14 days until complete variceal eradication, typically requiring three or four sessions.

Combination modality approaches, usually including an endoscopic and pharmacologic treatment, are attractive pathophysiologically and may be more effective than single therapy. Combined endoscopic therapy and beta-blockers reduce recurrent bleeding more than do beta-blockers alone (but provide no survival benefit). Although addition of sclerotherapy to ligation may theoretically offer greater protection against recurrent bleeding, this combination has not been shown to be advantageous.

Transjugular Intrahepatic Portosystemic Shunt

Transjugular shunting is more effective than endoscopic therapy for preventing variceal rebleeding but offers no survival benefit. The cumulative risk of rebleeding following TIPS placement is 8–18% at 1 year. The trade-off, however, is that TIPS is associated with a higher incidence of clinically significant hepatic encephalopathy (new or worsened portosystemic encephalopathy is noted in about 25% of patients after TIPS). Advanced liver disease is the main determinant of poor outcome following TIPS. Consequently, in patients with advanced liver disease, TIPS is best used as a bridge to liver transplantation.

An important concern with TIPS is the development of shunt stenosis and/or occlusion due primarily to pseudointimal hyperplasia, with reported rates of 31% at 1 year and 47% at 2 years. Doppler ultrasound examination of the stent is routinely performed at some centers to examine blood flow and estimate the shunt patency, but has a low sensitivity and specificity. Balloon dilatation of the stenosed stent or redeployment of another metallic prosthesis is done in case of significant stenosis. Management of hepatic encephalopathy and shunt stenosis may result in significant affiliated costs when TIPS is considered. A cost analysis comparing TIPS and endoscopic sclerotherapy has suggested no difference in cumulative cost despite the lower incidence of rebleeding with TIPS.

Surgical Therapy

Portosystemic shunt surgery is the most effective means by which to reduce portal pressure. Although effective at eradicating varices and preventing rebleeding, nonselective portacaval shunts are associated with a significant incidence of hepatic encephalopathy, portal vein thrombosis, and occasionally liver failure. In contrast, selective shunts decompress the portal system without endangering portal blood flow, liver function, or the feasibility of future liver transplantation. It is important to note that elective surgical therapy is largely reserved for patients with Child–Pugh class A disease and a proportion of patients with class B disease and preserved liver function. The choice of surgical therapy should be individualized and must be considered in the context of cause and severity of liver disease, patient compliance, likelihood of liver disease progression, and overall prognosis. Surgery may be preferred in patients who are not likely to be compliant with medical or repeated endoscopic therapy, or those that are not candidates for liver transplantation (e.g., active substance abuse, HIV positive).

Commonly used shunts include the distal splenorenal shunt and the low-diameter (mesocaval or portacaval) interposition shunt. Rates of recurrent bleeding are on the order of 10%, with the highest risk of bleeding occurring in the first month after surgery. Devascularization procedures (i.e., esophageal

transection and gastroesophageal devascularization) are usually considered in patients who cannot undergo shunts because of splanchnic venous thrombosis and should be performed only by experienced surgeons.

COST-COMPARISON OF AVAILABLE THERAPIES

Data examining the cost of variceal bleeding and the cost-effectiveness of commonly used therapies are limited. The treatment cost of an episode of variceal bleeding has been estimated at $10,000 to $35,000. Further, the cost-effectiveness of diagnostic methods used to guide therapy is unclear. For example, HVPG determination, which may accurately predict pharmacologic response to therapy, is an attractive, although invasive, adjunct in the management of patients with variceal bleeding, but its cost-effectiveness is an open question. Further, screening endoscopy for detection of large varices, although recommended, has not been demonstrated to be cost-effective.

Finally, there are areas in which management is controversial and not standardized. For example, given the right expertise, secondary prophylaxis with surgical shunts may be more effective than medical or endoscopic therapy in Child–Pugh class A patients. On the other hand, patients with advanced cirrhosis are often intolerant of beta-blockers—let alone in combination with nitrates—and therefore the use of combination therapy remains controversial in such patients. Arguably, the preferred treatment for such patients is early liver transplantation. Therefore, when choosing a specific treatment plan, the clinician must take into consideration the direct costs of health care utilization, as well as the efficacy and morbidity of therapy. The treatment chosen should be tailored to fit the patient's clinical condition while also taking into account the possibility that the patient's liver disease may progress and thus necessitate transplantation. Furthermore, the cost-effectiveness of various treatment modalities should factor in the cost of failed therapy (e.g., rebleeding, shunt revision) and that of treatment-related complications (encephalopathy, esophageal stricture, etc.). Therapeutic modalities used in patients with acute variceal hemorrhage, or in primary and secondary prophylaxis, are listed in Table II.

TABLE II Efficacy and Cost of Treatment for Prevention of Recurrent Variceal Bleeding in Patients with Cirrhosis[a]

Treatment	Characteristics of suitable patients[b]	Risk of bleeding at 12 months[c] (%)	Cost at 12 months[d] ($)	Comments
Medical therapy (nadolol or propranolol and isosorbide mononitrate)	Child–Pugh class A or B cirrhosis; reduction of ≥20% in HVPG with medication; high degree of compliance	4–25	3000–3700	Includes cost of HVPG determination at baseline and at 1–2 months of therapy
Endoscopic variceal band ligation	Child–Pugh class A–C cirrhosis; compliance with repeated medical therapy	20–30	8500–9500	Estimate based on a mean of four sessions until varices are obliterated followed by diagnostic esophagoscopy at 3 and 12 months
Transjugular intrahepatic portosystemic shunt	Current or future candidates for liver transplantation	8–15	12,000–15,000	Includes cost of Doppler ultrasonography of shunt every 3 months to monitor for stenosis or occlusion
Distal splenorenal shunt or low-diameter (mesocaval or portocaval) interposition shunt	Child–Pugh class A or B; good liver function	5–10	25,000–40,000	Includes preoperative venous phase arteriography and measurement of liver volume

[a] Data from Sharara and Rockey (2001).
[b] HVPG, Hepatic venous pressure gradient.
[c] The risk of bleeding varies with the severity of the liver disease.
[d] Costs represent the hospital charges, where applicable. The cost of care for bleeding episodes is not included.

SUMMARY

Gastroesophageal variceal hemorrhage is a common and devastating complication of portal hypertension and is a leading cause of morbidity and mortality in patients with cirrhosis. Because the clinical outcomes are poor once variceal bleeding has occurred, primary prophylaxis with beta-blockers is indicated but the role of endoscopic ligation deserves further evaluation. The treatment of acute variceal hemorrhage is aimed at volume resuscitation and ensuring hemostasis with pharmacologic agents and endoscopic techniques. A high risk of rebleeding after an index episode mandates the institution of preventative strategies. Wedge pressure-guided medical therapy may be the preferred mode of secondary prophylaxis in patients with Child–Pugh class A or B cirrhosis, but treatment with a combination of methods is pathophysiologically attractive. The choice of therapy should be tailored to fit the patient's clinical condition, risk factors, and prognosis.

See Also the Following Articles

Cirrhosis • Esophageal Strictures • Hemorrhage • Portal Hypertension and Esophageal Varices • Somatostatin • Upper Gastrointestinal Bleeding • Upper Gastrointestinal Endoscopy

Further Reading

Feu, F., Garcia-Pagan, J. C., Bosch, J., *et al.* (1995). Relation between portal pressure response to pharmacotherapy and risk of recurrent variceal haemorrhage in patients with cirrhosis. *Lancet* **346**, 1056–1059.

Garcia-Pagan, J. C., Feu, F., Bosch, J., and Rodes, J. (1991). Propranolol compared with propranolol plus isosorbide-5-mononitrate for portal hypertension in cirrhosis. A randomized controlled study. *Ann. Intern. Med.* **114**, 869–873.

Laine, L., and Cook, D. (1995). Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding. A meta-analysis. *Ann. Intern. Med.* **123**, 280–287.

Merkel, C., Marin, R., Sacerdoti, D., *et al.* (2000). Long-term results of a clinical trial of nadolol with or without isosorbide mononitrate for primary prophylaxis of variceal bleeding in cirrhosis. *Hepatology* **31**, 324–329.

Moitinho, E., Escorsell, A., Bandi, J. C., *et al.* (1999). Prognostic value of early measurements of portal pressure in acute variceal bleeding. *Gastroenterology* **117**, 626–631.

Poynard, T., Cales, P., Pasta, L., *et al.* (1991). Beta-adrenergic-antagonist drugs in the prevention of gastrointestinal bleeding in patients with cirrhosis and esophageal varices. An analysis of data and prognostic factors in 589 patients from four randomized clinical trials. Franco-Italian Multicenter Study Group. *N. Engl. J. Med.* **324**, 1532–1538.

Polio, J., and Groszmann, R. J. (1986). Hemodynamic factors involved in the development and rupture of esophageal varices: A pathophysiologic approach to treatment. *Semin. Liver Dis.* **6**, 318–331.

Rossle, M., Deibert, P., Haag, K., *et al.* (1997). Randomised trial of transjugular-intrahepatic-portosystemic shunt versus endoscopy plus propranolol for prevention of variceal rebleeding. *Lancet* **349**, 1043–1049.

Sharara, A. I., and Rockey, D. C. (2001). Gastroesophageal variceal hemorrhage *N. Engl. J. Med.* **345**, 669–681.

Sung, J. J., Chung, S. C., Yung, M. Y., *et al.* (1995). Prospective randomised study of effect of octreotide on rebleeding from oesophageal varices after endoscopic ligation. *Lancet* **346**, 1666–1669.

The North Italian Endoscopic Club for the Study and Treatment of Esophageal Varices (1988). Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices: a prospective multicenter study. *N. Engl. J. Med.* **319**, 983–989.

Villanueva, C., Minana, J., Ortiz, J., *et al.* (2001). Endoscopic ligation compared with combined treatment with nadolol and isosorbide mononitrate to prevent recurrent variceal bleeding. *N. Engl. J. Med.* **345**, 647–655.

Vascular Abnormalities

BRYAN J. BURNS AND LAWRENCE J. BRANDT
Montefiore Medical Center, Bronx, New York

angiodysplasia (vascular ectasia) Most common vascular abnormality of the gastrointestinal tract. Believed to be acquired with aging, represents the most frequent cause of recurrent lower intestinal bleeding in persons older than 50 years. The likely cause is intermittent, low-grade obstruction of the submucosal veins at the site where the veins penetrate the muscular layer of the colon.

angiography Radiographic technique by which blood vessels and vascular abnormalities can be visualized following injection of contrast material (dye).

Dieulafoy's lesion Rare cause of massive gastrointestinal bleeding; results from a large artery in abnormally close contact with the lining of the stomach or small intestine.

hemangioma Vascular tumor that can be found in the small intestine and colon; considered the second most common vascular lesion of the colon (after angiodysplasia). Hemangiomas can be classified into two distinct types, capillary hemangiomas and cavernous hemangiomas. They can become very large and present with gastrointestinal bleeding.

telangiectasia Vascular abnormality of small veins, arteries, and capillaries of the gastrointestinal tract. The abnormal vessels can become dilated, thin, and fragile, causing gastrointestinal bleeding. Telangiectasias are most commonly seen in the stomach and small intestine and tend to be multiple rather than single.

Vascular abnormalities are important causes of gastrointestinal bleeding. They may be single or multiple, or part of a systemic disorder that is acquired or inherited. The most commonly occurring vascular lesions in the gastrointestinal tract are vascular ectasia, gastric antral vascular ectasia, hereditary hemorrhagic telangiectasia, progressive systemic sclerosis, Dieulafoy's lesion, hemangioma, cavernous hemangioma of the rectum, blue rubber bleb nevus syndrome, and Klippel–Trenaunay–Weber syndrome.

VASCULAR ECTASIA

Vascular ectasia, also referred to as angiodysplasia, is the most common vascular abnormality of the gastrointestinal (GI) tract and likely represents the most frequent cause of recurrent major lower intestinal bleeding in persons older than 50 years. Vascular ectasias are believed to be associated with the degenerative changes of aging and are not related to other vascular abnormalities of the skin or other organs. Although these lesions are occasionally seen in the upper GI tract (stomach and small intestine), they are primarily found in the right side of the colon (cecum and ascending colon). Vascular ectasias are found equally in men and women. The lesions are usually multiple and tend to be smaller than 10 mm in diameter.

Approximately 50% of patients with bleeding ectasias have some degree of cardiac disease. Up to 25% have been reported to have stenosis of the aortic valve, although recent studies have not shown such an association. There is even some anecdotal evidence that replacement of the aortic valve in severe aortic stenosis has stopped recurrent gastrointestinal bleeding from angiodysplasias. The exact relationship between ectasias, aortic valve disease, and GI bleeding is still inconclusive and needs further research. Other diseases have been associated with vascular ectasias, including a coagulation disorder (von Willebrand's disease) and chronic renal failure requiring hemodialysis, but these associations require confirmation.

Pathology

Vascular ectasias are believed to be degenerative lesions that are acquired through aging. The likely cause is partial, intermittent, low-grade obstruction of the submucosal veins at the site where the veins penetrate the muscular layer of the colon (Fig. 1). Repeated episodes of elevated pressure during muscular contraction and distension of the right colon may ultimately result in dilatation and distortion of the submucosal vein. When resected segments of colon are injected with silicone rubber to fill the vascular lesions and then viewed under a microscope, a dilated, tortuous submucosal vein is often the earliest abnormality noted. Later, the venules and capillaries that drain into that vein may dilate and extend very close to the surface of the colon. Eventually, the sphincter regulating flow between the artery and submucosal vein becomes destroyed, allowing free flow between these vessels (arteriovenous fistula). In the most severe lesions, the innermost lining of the bowel, the mucosa, may ultimately be replaced by a maze of dilated, fragile vascular channels. Despite their complicated

 Copyright 2004, Elsevier (USA). All rights reserved.

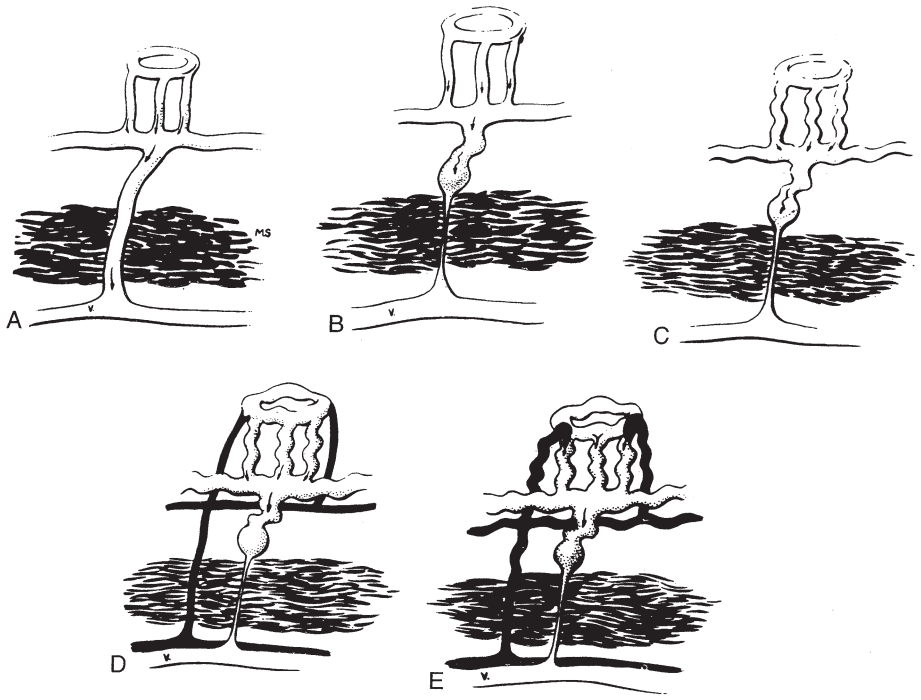


FIGURE 1 Vascular ectasia. Proposed concept of the development of vascular ectasias: (A) Normal state of vein (v.) perforating muscle layers; (B) with muscular contraction or increased intraluminal pressure, the vein is partially obstructed; (C) after repeated episodes over many years, the submucosal vein becomes dilated and tortuous; (D) later the veins and venules draining into the abnormal submucosal vein become similarly involved; (E) ultimately the capillary ring becomes dilated, the precapillary sphincter becomes incompetent, and a small arteriovenous communication is present through the ectasia. From Boley, S. J., Sammartane, S. J., Adams, A., *et al.* (1977). On the nature and etiology of vascular ectasias of the colon: Degenerative lesions of aging. *Gastroenterology* **72**, p. 650, with permission.

appearance, vascular ectasias are often tiny—smaller than a pinhead. Local factors may play a role in the development of these lesions. For example, the greater tension in the wall of the right colon and cecum compared to the rest of the colon may be responsible for the increased prevalence of ectasias in that part of the colon.

Clinical Presentation

Vascular ectasias manifest only through gastrointestinal bleeding. Incidental ectasias may be seen at colonoscopy in 3–6% of healthy, asymptomatic people, and require no treatment given the overall low risk of bleeding and the inability to predict who will and who will not bleed. Bleeding may be brisk or occult and the clinical presentation varies from overt passage of large amounts of bright red blood per rectum to iron-deficiency anemia. The majority of patients with colonic lesions present with chronic, low-grade bleeding. A quarter of patients may pass dark, tarry stools. A majority of the bleeding episodes stop spontaneously. Many patients have multiple blood transfusions and endoscopic procedures before a diagnosis of vascular ectasias is made.

Diagnosis

A majority of the ectasias are diagnosed by colonoscopy, and because most are located in the right colon, it is important that the entire colon be examined. The endoscopic appearance of colonic lesions can be variable. Most often they are small in size, have scalloped or frondlike edges, and resemble a spider or a coral reef. Endoscopic diagnosis, however, is limited by the similar appearance of other vascular lesions and of even minor trauma from the colonoscopy. Circulating blood volume, the patient's state of hydration, and sedative medications (such as meperidine) also can diminish the prominence of some vascular ectasias. Blood transfusions, intravenous fluids, or administration of naloxone (which rapidly reverses the effects of meperidine) may facilitate visualization of these vascular

abnormalities. Color images of these and other lesions can be found on-line at several web sites, including the GASTROLAB Image Gallery (http://www.gastrolab.net/pawelcom.htm.). Angiography may be used to detect angiodysplasias not seen at endoscopy or to find and treat actively bleeding lesions. Knowledge of typical angiographic signs enables a more specific diagnosis than can be made by colonoscopy.

Treatment

In most patients, bleeding can be controlled without surgery either colonoscopically or through an angiographic catheter. Endoscopic obliteration treatments include heater probe, argon plasma coagulation, and, less commonly, electrocoagulation and laser therapy. Bleeding may recur in as many as 50% of patients, prompting further therapy. A variety of medications infused though an angiographic catheter can also successfully stop bleeding, though this is rarely necessary. The most commonly used medication is vasopressin (a potent constrictor of blood vessels). As a last resort, surgical removal of the right colon is used for bleeding that does not respond to the above therapies or in circumstances in which such therapies are not available.

GASTRIC ANTRAL VASCULAR ECTASIA: "WATERMELON STOMACH"

Gastric antral vascular ectasia (GAVE) is an unusual vascular abnormality, distinct from other vascular lesions. It has been known as "watermelon stomach" based on its appearance of dilated blood vessels radiating outward from the pylorus like spokes from a wheel, and resembling the stripes on the surface of a watermelon.

Epidemiology

GAVE is seen commonly in older patients, the average age being 70 years. This disorder is more common in women. The cause is uncertain, but it has been associated with certain immunologic disorders, cirrhosis of the liver (especially when complicated by hypertension in the portal venous system of the liver), and absence of acid production by the stomach. No disease association has been strong enough to produce insight into the etiology of this disorder.

Pathology

Microscopically, the mucosa of the gastric antrum is hypertrophied and folded. The vascular channels are located primarily in the submucosa and are tortuous, dilated, and contain focal thromboses. The lesion is restricted to the gastric antrum, but may be associated with other vascular disorders of the stomach if portal hypertension is present.

Clinical Presentation and Diagnosis

Nearly all patients present with iron-deficiency anemia from slow oozing of blood, but patients may also present with overt gastrointestinal bleeding. Diagnosis is made by the characteristic endoscopy appearance of alternating red stripes with normal mucosa. Convoluted columns of ectatic vessels produce the reddish color. Angiography has not proved to be helpful in the diagnosis of these lesions.

Treatment

Medical, endoscopic, and surgical approaches have all been utilized in the treatment of GAVE. Medical therapy is mainly that of iron replacement for anemia. Estrogen–progesterone combinations have been tried with some regression of these lesions, but results have not been consistent. GAVE lesions respond to a variety of endoscopic modalities, including heater probe, electrocoagulation, lasers, or argon plasma coagulation. Surgical resection of the gastric antrum is curative but should be considered only for patients for whom endoscopic therapy has been unsuccessful.

HEREDITARY HEMORRHAGIC TELANGIECTASIA

A familial disorder associated with multisystem disease, hereditary hemorrhagic telangiectasia (HHT), or Osler–Weber–Rendu disease, is characterized by recurrent gastrointestinal bleeding from small vascular lesions (telangiectases) of the skin and mucous membranes. It is inherited in an autosomal dominant fashion. This disorder usually presents with recurrent nosebleeds during childhood, but the typical lesions on the lips and tongue may not be seen until later in life. Gastrointestinal bleeding is rare before age 30 years and has a peak incidence in the sixth decade. Bleeding typically improves during pregnancy. Gastrointestinal bleeding may be quite severe and some patients have received more than 50 blood transfusions over their lifetime.

Pathology

The major abnormalities are of small veins (venules), small arteries (arterioles), and capillaries. The ends of these vessels become dilated, tortuous, and irregular. The vessels are lined by a single layer of cells without muscle or elastic tissue, so they cannot contract.

This renders them thin and very fragile and may explain why they tend to bleed.

Clinical Presentation

Typical features of HHT include a family history in up to 80% of patients, as well as characteristic telangiectasias on the vermilion border of the lips, oral and nasal mucosae, tongue, and nailbeds. These vascular lesions may occur in the colon but are more commonly seen in the stomach and small bowel. It is at these sites where they are more likely to cause significant bleeding. Most patients will pass dark, tarry stools (melena), but vomiting of blood and the passage of red blood per rectum have also been reported. This disease is a systemic illness, and lesions are not confined to the GI tract. The telangiectasias may also be found in the lungs, brain, and liver. Pulmonary lesions may allow for shunting of pulmonary arterial blood into the cerebral circulation, which can have devastating consequences, including stroke and brain abscess.

Treatment

The management of GI bleeding in patients with HHT is difficult because of the multiplicity of lesions. Endoscopic ablation may be performed during active bleeding or between bleeding episodes. Oral estrogen–progesterone therapy has also been used, but its effectiveness is limited. Surgical resection should be reserved for bleeding lesions resistant to endoscopic therapy, especially because rebleeding from other areas of the gastrointestinal tract is common.

PROGRESSIVE SYSTEMIC SCLEROSIS (SCLERODERMA)

Telangiectactic vascular lesions are a prominent feature of progressive systemic sclerosis, especially in the variant that has as it clinical features calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias (CREST). In this disorder, the sites most frequently involved are the hands, lips, face, and tongue, but gastrointestinal bleeding has been reported from lesions in the stomach, small intestine, and colon. Bleeding is best treated by endoscopic electrocautery or by ablation using laser or argon plasma coagulation.

DIEULAFOY'S LESION

This vascular abnormality is a rare cause of massive gastrointestinal bleeding, usually from the stomach, but sometimes from the small or large intestine. It is twice as common in men as in women and presents at a mean age of 52 years. This lesion is believed to be present in 1–2% of patients with acute GI hemorrhage.

Pathology

It is generally accepted that the primary defect in a Dieulafoy's lesion is an abnormally large artery running through the submucosa, or in the some cases the mucosa, of the stomach. The vessel is normal histologically, but does not undergo the normal decrease in caliber as it penetrates from the outer to the inner layers of the stomach. As a result, the "caliber-persistent" vessel is in close contact with the stomach lining and may cause a small erosion in this lining. This can cause the vessel to become eroded, with resultant massive bleeding.

Clinical Presentation

These lesions typically present with profuse upper GI hemorrhage without symptoms of ulcer disease. The diagnosis can usually be made by endoscopy, and the site of bleeding is usually located within 6 cm of the junction of the esophagus and the stomach. The mortality rate from bleeding has been high in the past but has improved significantly with the advent of therapeutic endoscopy.

Therapy

Historically, surgery was the only effective way to manage Dieulafoy's lesions, often requiring a wedge resection of the entire area. More recently, emergency endoscopy has been used to localize and treat the actively bleeding site. A variety of techniques can be used to achieve hemostasis, including local injection of vasoconstrictive or sclerosing agents, heater probe, electrocoagulation, and laser or argon plasma coagulation. Most patients (80%) with Dieulafoy's lesions can be managed with endoscopic therapy alone. Rebleeding from these lesions may occur and, therefore, patients should be observed closely following treatment.

HEMANGIOMA

Hemangiomas are vascular growths that can be found in both the small intestine and the colon. Although rare, they are considered the second most common vascular lesion of the colon (after vascular ectasias). The lesions may be present at birth and enlarge with the normal

development of the child. They may be single or multiple and can be associated with hemangiomas on the skin or in other organs.

Pathology

Hemangiomas are often small, ranging from a few millimeters to 2 cm, but larger lesions occur, especially in the rectum (see the following discussion on cavernous hemangiomas of the rectum). Hemangiomas can be classified into two distinct histologic types, capillary hemangiomas and cavernous hemangiomas. Capillary hemangiomas are uncommon and may be found in the small intestine and appendix. These reddish-purple lesions are usually solitary and well circumscribed. They are made up of tiny vessels, closely packed together into clusters separated by very little connective tissue. In contrast, cavernous hemangiomas are most often found in the large intestine or rectum and can be quite large, extending up to 20 or 30 cm in length, with enough bulk to encroach on the bowel lumen. They also can be well-circumscribed and polypoid. Histologically, they are seen as numerous dilated, irregular, blood-filled spaces that may extend through the entire wall of the bowel.

Clinical Presentation

Hemangiomas usually present in young men and women in their second decade. Bleeding is usually slow, producing anemia or rarely black, tarry, or red stools. Cavernous hemangiomas may present with brisk bleeding, and death from exsanguination has occurred. Occasionally, large lesions can present with bowel obstruction. Diagnosis is best made by endoscopy, because X-ray studies, including angiography, frequently are normal. Endoscopically, the lesions appear as bright red spots or nodules (capillary hemangiomas) or as reddish-purple polyps or mounds (cavernous hemangiomas).

Therapy

Small, solitary hemangiomas can be ablated locally by endoscopic techniques. Most larger lesions must be approached surgically, because even a small endoscopic biopsy can result in massive bleeding or perforation. Either the hemangioma alone or the involved segment of bowel harboring the vascular growth is resected. Alternative therapies have been used in unresectable cases. These include injection of sclerosing agents, freezing methods (cryotherapy), and ablation of the arterial supply to the lesion. In most cases, these techniques have been only marginally successful.

CAVERNOUS HEMANGIOMA OF THE RECTUM

These lesions are often considered separately from other hemangiomas because of their different presentation and clinical course. They are often solitary and quite extensive in size, involving much of the rectum and occasionally the sigmoid colon as well. Because of their large vascular channels, these lesions may present with massive rectal bleeding. A soft mass may be detected on digital rectal examination. The diagnosis can also often be suggested by clues obtained from plain X rays of the abdomen. These studies may show focal calcific densities (phleboliths), which form in the dilated sinuses, or a distorted rectal air column. Endoscopically, elevated plum-red nodules are seen; ulcers and rectal inflammation may also be present. Local measures to control bleeding are only temporarily effective. Embolization and surgical ligation of major feeding vessels have also been used, but ultimately resection of the rectum may be required.

BLUE RUBBER BLEB NEVUS SYNDROME

The blue rubber bleb nevus syndrome is a rare disorder that manifests as cavernous hemangiomas on the skin and in the gastrointestinal tract. The descriptive name of this condition derives from the fact that the blue skin lesions have the appearance and feel of rubber nipples.

Pathology

The lesions are quite distinctive; they are blue and raised, ranging from 0.1 to 10 cm in size, and have a wrinkled surface. A few to several hundred of these lesions can be found on the skin, usually on the trunk, extremities, and face. They are often present at birth or appear during childhood. Microscopically, the lesions are cavernous hemangiomas composed of dilated capillaries.

Clinical Presentation

Gastrointestinal tract involvement with hemangiomas of the blue rubber bleb nevus syndrome is very common. Lesions can be found throughout the gastrointestinal tract but are most common in the small bowel. They usually present with low-grade gastrointestinal bleeding. The typical presentation is that of a young adult with chronic anemia and the characteristic skin

lesions (characteristic by appearance and by the wrinkled sac they leave on compression).

Therapy

Although barium and angiographic studies have both been used to diagnose this disorder, endoscopy is the best available modality (although limited in the length of small bowel it can visualize). Conservative treatment with iron supplementation is usually tried in these patients. If bleeding is significant and confined to one or a few segments of bowel, surgical resection of the involved areas may be needed. Endoscopic treatment with laser therapy has been used but may be dangerous if the lesion involves the full thickness of the bowel wall. Endoscopic ultrasound is a new technique that may help define the depth of bowel involvement prior to therapy.

KLIPPEL–TRENAUNAY–WEBER SYNDROME

This syndrome is characterized by three cardinal features: a large vascular lesion (hemangioma) involving a lower limb, with varicose veins, and enlargement of the bones and soft tissues of, the involved limb. Swelling of the involved leg is very common. This disorder may cause gastrointestinal bleeding from cavernous hemangiomas, usually located in the colon or rectum. Vaginal hemangiomas may also cause significant vaginal bleeding. Bleeding may be mild or heavy and is often recurrent. Surgical resection of the involved area of bowel is standard, but recently endoscopic laser therapy has been tried in several cases with good results.

See Also the Following Articles

Hemorrhage • Lower Gastrointestinal Bleeding and Severe Hematochezia • Occult Gastrointestinal Bleeding • Upper Gastrointestinal Bleeding

Further Reading

Boley, S. J., and Brandt, L. J. (1986). Vascular ectasias of the colon—1986. *Dig. Dis. Sci.* **31**, 26S.

Boley, S. J., Sammartano, R. J., Adams, A., *et al.* (1977). On the nature and etiology of vascular ectasias of the colon: Degenerative lesions of aging. *Gastroenterology* **72**, 650.

GASTROLAB Image Gallery web site. Picture archive. Retrieved at http://www.gastrolab.net/pawelcom.htm.

Guttmacher, A. E., Marchuk, D. A., and White, R. I. (1995). Hereditary hemorrhagic telangiectasia. *N. Engl. J. Med.* **333**, 918.

Jabbari, M., Cherry, R., Lough, J. O., Daly, D. S., Kinnear, D. G., and Goresky, A. (1984). Gastric antral vascular ectasia: "The watermelon stomach." *Gastroenterology* **87**, 1165.

Reinus, J. F., and Brandt, L. J. (1990). Upper and lower gastrointestinal bleeding in the elderly. *Gastroenterol. Clin. North Am.* **19**, 293.

Samuel, M., and Spitz, L. (1995). Klippel–Trenaunay syndrome: Clinical features, complications, and management in children. *Br. J. Surg.* **82**, 757.

Stark, M. E., Gostout, C. J., and Balm, R. K. (1992). Clinical features and endoscopic management of Dieulafoy's disease. *Gastrointest. Endoscop.* **38**, 545.

Vascular Abnormalities, Pediatric

XENIA B. HOM* AND STEVEN J. FISHMAN†

*Children's Hospital and University of California, San Diego and †Children's Hospital Boston and Harvard University

Blue Rubber Bleb Nevus syndrome A disorder consisting of multifocal venous malformations of the skin, soft tissues/muscles, gastrointestinal tract, or almost any organ.

hemangioma Benign endothelial tumors.

hereditary hemorrhagic telangiectasia Also known as Osler-Weber-Rendu disease, with autosomal dominant inheritance of vascular malformations in multiple organ systems, including the lungs, brain, and gastrointestinal tract.

Kaposiform hemangioendothelioma Endothelial hyperplasia, which is less discrete and more aggressive than typical hemangioma. It can be associated with very low platelet counts (known as Kasabach Merritt phenomenon).

vascular malformation A result of abnormal development of vascular structures. They are subclassified based on their predominant channel type (capillary, venous, lymphatic, arteriovenous, etc.).

Skin and soft tissue vascular anomalies are relatively common in infants and children, although visceral vascular anomalies are much less common and often asymptomatic. The nomenclature of vascular anomalies, in general, has been misleading and propagates misconceptions that often lead to incorrect diagnosis and treatment. Nevertheless, most gastrointestinal vascular anomalies have typical clinical presentations, natural histories, and endoscopic and radiologic appearances. An understanding of such patterns of presentation facilitates accurate diagnosis and appropriate management of these uncommon lesions.

CLASSIFICATION

Vascular anomalies in childhood are split into two broad categories: tumors and vascular malformations. Imprecise terminology has been inconsistently applied to a variety of vascular anomalies. A reclassification, encompassing clinical behavior, cellular kinetics, and physical appearance, was proposed by Mulliken and Glowaki in 1982 and can also be applied to visceral lesions (Table I). For practical purposes, most of these vascular tumors of childhood are hemangiomas, which are benign endothelial tumors. The descriptors "cavernous" or "capillary" hemangioma are misleading and unnecessary. Often the lesions described as cavernous hemangiomas are in fact vascular malformations rather than true hemangiomas.

TABLE I Biologic Classification of Vascular Anomalies

Tumors	Malformations
Cause: Endothelial hyperplasia	**Cause: Dysmorphogenesis with normal endothelial turnover**
Hemangioma	Capillary
Congenital hemangioma	Venous
Kaposiform hemangioendothelioma	Lymphatic
Rare forms	Arterial/arteriovenous
	Complex-combined

Vascular malformations result from errors in vascular morphogenesis and are not tumors. They are subdivided into their predominant channel type: capillary, venous, lymphatic, arterial, or a combination of these. The suffix "-oma" in modern medical terminology (e.g., lymphangioma, cystic hygroma, and angioma) often indicates a neoplastic process with up-regulated cell growth. These terms should be avoided because they misrepresent the true biology of the nonneoplastic, quiescent endothelium in vascular malformations.

HEMANGIOMAS

Epidemiology and Pathogenesis

Hemangioma is the most common tumor of infancy and childhood. It occurs in 4–10% of white infants and 23% of low-birth-weight premature infants weighing <1200 g. The incidence in dark-skinned babies is low. Most hemangiomas are single, but 20% of affected infants have multiple lesions. Multiple cutaneous tumors (more than 5) often arise in association with visceral lesions [mostly in the liver, gastrointestinal (GI) tract, lungs, and brain]. Hemangiomas occur with a female to male ratio of 3 : 1 to 5 : 1. This ratio is higher in problematic tumors and in rare cases where there are associated structural anomalies. There is also evidence suggesting that the risk of hemangioma is significantly higher in offspring of women who undergo chorionic villus sampling.

 Copyright 2004, Elsevier (USA). All rights reserved.

The triggers for hemangiogenesis are not entirely clear, although there are hypotheses suggesting a viral cause, such as human parapox virus infection or genetic alteration. In a study of monozygotic versus dizygotic twins, there was no strong evidence for Mendelian inheritance; however, there are rare kindreds that suggest familial transmission in an autosomal dominant pattern with incomplete penetrance and variable expressivity. A putative locus on 5q has been identified in three families with coexistence of hemangioma and vascular malformation. The median age for the appearance of most cutaneous hemangiomas is 2 weeks after birth (Fig. 1). However, approximately one-third to one-half are nascent at birth. Hemangiomas evolve through three stages:

1. Proliferation (usually between 0 and 1 year of age), characterized by up-regulation of two potent angiogenic peptides, vascular endothelial growth factor and basic fibroblast growth factor.
2. The involuting phase (usually between 1 and 7 years of age), characterized by a higher rate of apoptosis than mitosis, with decreased interferon-β in the overlying epidermis. The rate of regression is unrelated to the appearance, site, size of tumor, or the gender of the patient. Involution is complete in 50% of children by 5 years and in 70% by 7 years of age, with gradual improvement until 10–12 years of age.
3. The involuted phase (after 7 years of age), in which the hemangioma may be completely invisible, leave a pale fibrofatty residuum, or lead to some fine telangiectatic vessels on the overlying skin. Lesions that have ulcerated during proliferation leave a scar.

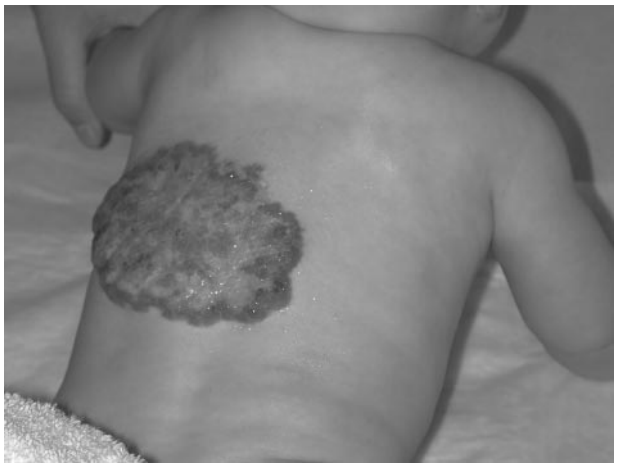


FIGURE 1 Typical cutaneous hemangioma.

Clinical Manifestations

Hemangiomatosis refers to multiple cutaneous and visceral hemangiomas. The liver is the most common visceral site involved and can range in presentation from tiny, asymptomatic tumors found incidentally, to large, single or multiple tumors, with or without cardiac complications. These typically manifest in the first 4 months of life with hepatomegaly, anemia, and high-output cardiac failure. Abdominal compartment syndrome can occur when the hepatomegaly is so marked that it limits diaphragmatic excursion and venous return to the inferior vena cava, resulting in respiratory and renal failure. Occasionally, the liver hemagioma may invade directly into the duodenum or colon or parasitize its blood supply. Symptomatic hemangiomas of the stomach, small bowel, and colon are rare. When detected, they often exhibit diffuse, patchy involvement and they are also often associated with multifocal cutaneous lesions.

Radiologic Features

Any infant with more than five cutaneous hemagiomas should be considered for screening by ultrasound and/or magnetic resonance imaging (MRI) to evaluate for possible lesions in the liver. Doppler ultrasound of the liver sometimes detects a liver hemangioma in its early proliferative phase; lesions may even be detected on antenatal ultrasound. The ultrasound can show single or multiple lesions with decreased arterial resistance, increased venous velocity, and discrete soft tissue mass. Sonography is useful in documenting tumor response to pharmacologic therapy. On MRI, hemangiomas appear to be solid tissue of intermediate intensity on T1 images and moderate hyperintensity on T2 images. Flow voids are indicators of shunting (rapid flow) between feeding arteries and dilated draining veins and are prominent around and within the tumor. One should be aware that because both intrahepatic hemangioma and arteriovenous malformations (AVMs) are fast flow, they can often be mistaken for one another. Finally, radionuclide scanning with technetium Tc-99m-tagged red blood cells can be used to document deep multiple hemangiomas in the GI tract.

Treatment

Most lesions require no treatment. However, if there is a threat to life, limb, vital function, or significant tissues, the lesions should be treated pharmacologically. The first-line agent is oral prednisone, 2–3 mg/kg/day (up to 5 mg/kg/day for refractory lesions) for 1 month, followed by a slow taper over the subsequent

5–7 months. Rebound growth after weaning the dose is a common phenomenon and the dose may need to be temporarily increased. Small focal lesions can be treated with sequential intralesional injections of triamcinolone. Approximately one-third of treated lesions undergo rapid involution, one-third stabilize, and one-third have no response. The growth of the lesions is inhibited, probably from the mild antiangiogenic effect of corticosteroids.

For steroid-resistant lesions, treatment with a more potent angiogenesis inhibitor, interferon α-2b, 2–3 million units/m^2/day, can be given subcutaneously. The response is typically slower, although most lesions respond. Side effects include a low-grade fever in the first 1–2 weeks (which can be minimized by pretreatment with acetominophen), transient transaminase elevation, neutropenia, and anemia (usually mild and not requiring discontinuation of the interferon). The last and potentially most serious side effect is spastic diplegia, which has been reported in approximately 5% of infants treated with interferon.

Gastrointestinal hemangiomas are rare and can be single, multifocal, or diffuse infiltrative tumors. Bleeding may occur and necessitate repeated transfusions. On occasion, focal lesions can be treated with local endoscopic (band ligation, argon plasma coagulation, or intralesional corticosteroid injection) or surgical techniques (enterotomy). However, one cannot always be assured that any endoscopically visualized lesion is the only bleeding site, as there may be other lesions out of reach of the endoscope or colonoscope. In the case of diffuse infiltrative hemangioma(s) with life-threatening hemorrhage, management consists of supportive care with repeated transfusions, parenteral nutrition, and antiangiogenic therapy (to accelerate involution), rather than surgical resection (as this is usually not possible).

For relatively small, asymptomatic, hepatic lesions, sequential physical exam and ultrasound are typical management. Therapy is typically reserved until there is significant hepatomegaly or symptoms such as congestive heart failure. Initial therapy is with systemic steroids and second-line therapy is interferon. If symptoms worsen or persist, hepatic arteriography may be considered. Embolization is a potential therapy, only if there are direct macrovascular shunts through the lesion(s). Chemotherapy with cyclophosphamide or vincristine has also occasionally been effective for life-endangering hepatic hemangiomas. Endoscopic evaluation is not typically needed, but may be useful in cases where there is concern about the liver hemangioma invading the bowel or its vascular supply. Endoscopy may be indicated to confirm and localize the lesion(s), but it is unlikely to be sufficient for hemostasis of gastrointestinal bleeding if it occurs. Surgical resection of liver hemangiomas is almost never indicated. Liver transplant is considered only in rare instances when the tumor is unresponsive to the aforementioned therapies and the patient is in extremis from abdominal compartment syndrome. In cases of such large lesions, it is crucial to screen for acquired hypothyroidism, as some hemangiomas have been demonstrated to express type 3 iodothyronine deiodinase, which inactivates circulating thyroid hormone. Prior screening for congenital hypothyroidism is generally normal in these individuals and a normal early thyroid-stimulating hormone does not ensure that hypothyroidism will not develop as a hemangioma enlarges. Thus, repeated testing should be performed if the lesion undergoes significant growth.

Congenital hemangioma is a rare variant in which a lesion is fully grown at birth and involutes rapidly, usually by 1 year of age. No congenital hemangioma has been reported in the gastrointestinal tract.

KAPOSIFORM HEMANGIOENDOTHELIOMAS

Kaposiform hemangioendotheliomas (KHEs) are generally much more aggressive than hemangiomas, although they follow a similar pattern of spontaneous involution. They are much less discrete and tend to blend in diffusely with the surrounding structures. Both sexes are equally affected. Although involvement in a hollow viscus has not been reported, the torso, retroperitoneal structures, and liver may be involved. Kasabach-Merritt phenomenon occurs in association with these lesions and is characterized by a severe thrombocytopenia (typically 2000 to 5000 platelets/mm^3), which places the infant at risk for hemorrhage in various sites, including the gastrointestinal tract. KHE is clearly differentiated from common hemangioma via MRI; KHEs have a poorly defined tumor margin, extend across tissues, and have relatively small vessels. Biopsy is rarely indicated for a coagulopathic tumor and it is usually too diffusely infiltrative and large to resect. In contrast to hemangiomas, interferon has been effective only in 50% of infants and no single pharmacologic therapy has proven to be consistently effective. Two caveats must be remembered in managing thrombocytopenia in these lesions: Platelet transfusion should be avoided unless there is active bleeding or a surgical procedure is indicated. Heparin should not be given, as it can stimulate tumor growth, platelet trapping, and worsened bleeding.

VASCULAR MALFORMATIONS

Capillary Malformations

Capillary malformations are composed of dilated capillary to venular sized vessels, with a paucity of normal nerve fibers around these vessels on immunohistochemical staining. In the skin, they include port wine stains and telangiectasias. It is extremely rare to have symptomatic capillary malformations of the viscera. Hereditary hemorrhagic telangiectasia (HHT), also known as Osler–Weber–Rendu disease, can present with nosebleeds, shunting through pulmonary arteriovenous malformations, cerebral lesions, and potential gastrointestinal mucosal lesions. The pulmonary complications are usually apparent before the GI manifestations, and if GI bleeding occurs, it is rare before adulthood. There are rare case reports of HHT associated with juvenile polyps, which are highly vascular and a common source of bleeding during childhood. In these cases, the child may present with gastrointestinal bleeding from the polyps rather than the vascular anomaly.

Venous Malformations

Venous malformations are the most common symptomatic vascular anomaly of the GI tract in childhood. Most patients have only single lesions, though some have multiple malformations. Most of these malformations are sporadic, although genes for several rare, autosomal dominant forms have been identified. They have a wide range in size (miniscule to massive) and shape (flat to bulky and amorphous). Grossly, they are nonpulsatile, appear deep purple, decompress with manual pressure, and refill shortly after release (Fig. 2). Microscopically, they are channels lined by endothelium but lack smooth muscle. They have also been improperly termed cavernous hemangiomas because of their soft, spongy feel and color. However, this term should not be used because these lesions are not tumors, do not involute, and do not respond to pharmacologic treatment. Venous malformations of the skin and soft tissue are often asymptomatic, other than being unsightly and sometimes disfiguring. Blood can stagnate in the nonlaminar channels and cause a local consumptive coagulopathy. Thrombosis may occur, resulting in sudden swelling, firmness, and pain. In large lesions, systemic hypofibrinogenemia and prolonged prothrombin time may result from factor consumption. Thrombocytopenia is atypical. Even without frank thrombus, pain can occur due to congestion or low-grade thrombosis. Small, palpable phleboliths can form and appear as radiopaque spherical densities on radiographic studies. If present, these

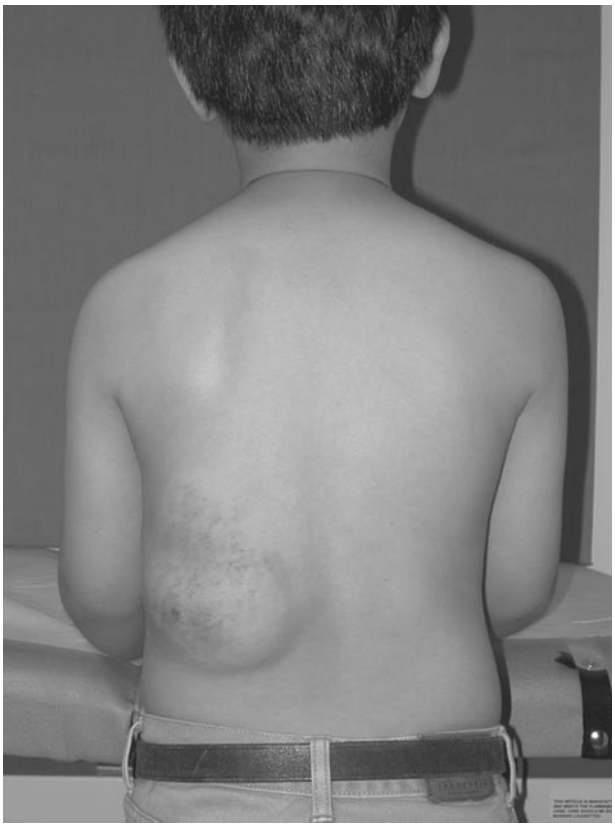


FIGURE 2 Typical cutaneous venous malformation.

serve as an easy distinguishing feature of venous malformations from hemangiomas. In the GI tract, they most commonly present with upper or lower, acute or chronic, gastrointestinal bleeding. Since GI bleeding is not common in childhood, it is often not suspected or identified until the patient is found to be profoundly anemic. This is highlighted by the fact that children independent in toileting may not report a change in bowel habits and if parents notice dark-colored stools, they may not appreciate their significance. Pain and obstruction are less common presenting symptoms of venous malformations in the gut. They are common in the liver and spleen (often improperly termed hemangiomas) and usually asymptomatic.

A single, focal venous malformation can occur anywhere in the gastrointestinal tract, vary in size (<1 cm to >20 cm diameter), and can be mucosal, mural, or transmural (Fig. 3). Endoscopic ultrasound can be used to determine the depth of mural involvement, although the entire lesion must be viewed, as there can be varying depths of involvement in different parts of the same lesion. Therapeutic endoscopic procedures should be performed only on superficial lesions and surgical backup should be available for complications that may

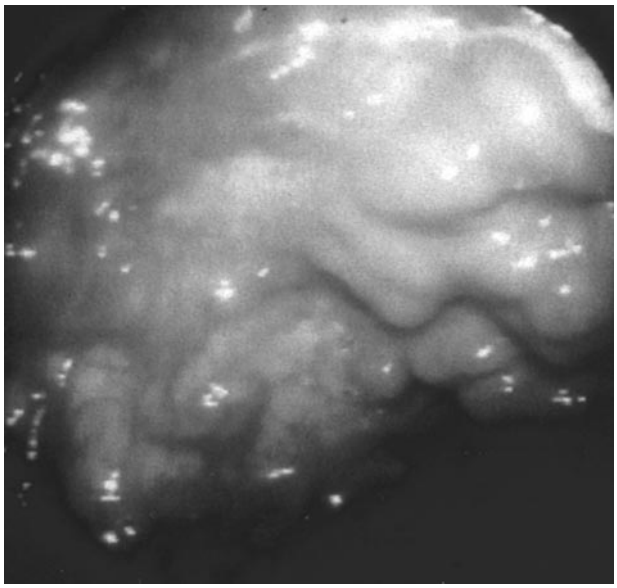


FIGURE 3 Endoscopic appearance of colonic venous malformation.

arise. Small, muscosal lesions can be eradicated with endoscopic band ligation. If the lesion is not transmural, endoscopic sclerotherapy may also be used, but is unlikely to be sufficient treatment of large lesions. Potential complications include perforation (if unsuspected transmural lesions are treated endoscopically) and intravascular migration of sclerosing material from large lesions (if there are anomalous or large vessels communicating with the lesion). In such cases, pretreatment angiography or intralesional injection of contrast under fluoroscopy may be helpful. Surgical excision is needed to eradicate large or transmural venous malformations.

Blue Rubber Bleb Nevus syndrome (BRBNS) consists of multifocal venous malformations of the skin, soft tissues/muscles, gastrointestinal tract, or almost any organ. Skin lesions vary in size (usually 3–15 mm), are deep blue or purple in color, are flat or minimally raised, may be tender to palpation, and have a predilection for the trunk, palms, and soles (Fig. 4). A total body survey with delayed imaging of Tc-99m-labeled red blood cells or scintigraphy using single-photon emission computed tomography (for a three-dimensional picture) shows pooling of the labeled cells within the lesions. Almost all patients have multiple lesions within the liver. Gastrointestinal lesions have an endoscopically pathognomonic appearance: discrete, purple berry-like protuberances, several millimeters to centimeters in size, scattered from mouth to anus (Figs. 5 and 6). Most are broad-based, but some have a narrow pedicle. Typically, there is a broad rim of normal mucosa encircling the base and extending up to the reddish blue apex of the lesion, although normal mucosa may also cover the entire mass. Affected patients present from early infancy to young adulthood, invariably with chronic GI bleeding (usually consistent black stools rather than acute hemorrhage) and anemia. Most patients require chronic transfusions, iron replacement, and erythropoietin. Some patients can also present with intussusceptions or volvulus. Single pharmacologic or endoscopic treatment studies to date have not yielded reproducible, durable efficacy. Therapies that have been evaluated include interferon-α, somatostatin, high-dose intravenous gammaglobulin, sclerotherapy, band ligation, and laser photocoagulation. Because most lesions are generally in the unexamined portions of the small intestine, endoscopic therapy alone is unlikely to decrease a patient's transfusion requirement. There is one report of endoscopic polypectomy with laparoscopic assistance to approach small bowel lesions. The authors' approach to BRBNS is to eradicate all hollow viscus venous malformations using a combined endoscopic and operative approach when a patient requires repeated transfusions or has refractory severe anemia. Since BRBNS lesions are congenital, present at birth (though they may be minute), and expand over time, operation is delayed until bleeding is significant, in hopes that all lesions are "ripe" and can all be eradicated in one procedure. Lesions are identified endoscopically in the esophagus to duodenum and rectum to colon. If lesions are superficial, endoscopic band ligation is performed. Otherwise, they are surgically excised. Small bowel lesions are identified via enteroscopy through

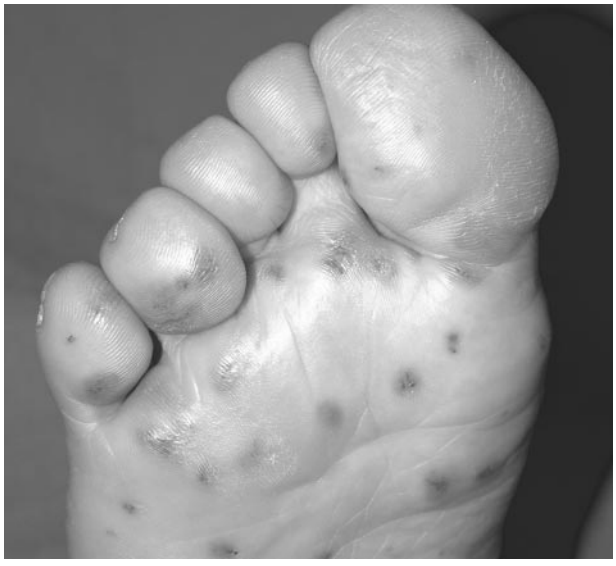


FIGURE 4 Skin lesions of the Blue Rubber Bleb Nevus syndrome.

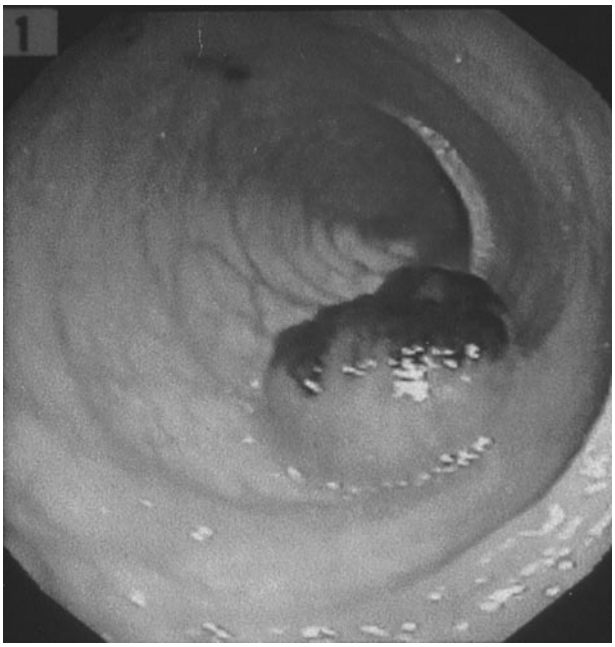


FIGURE 5 Endoscopic appearance of venous malformations of the Blue Rubber Bleb Nevus syndrome.

enterotomies or by inversion of segments of intestine by sequential intussusceptions through multiple enterotomies. Small bowel length is preserved by wedge excision or ligation of lesions rather than circumferential small bowel resections. Outcomes have been excellent.

Diffuse venous malformations involve large contiguous sections of bowel and extraintestinal structures (such as mesentery, retroperitoneum, pelvis, muscles, subcutaneous tissues, and skin). Upper visceral diffuse venous malformations may involve the mesenteric, splenic, and portal veins. If the portal vein is anomalous, one should investigate for presinusoidal hypertension. These lesions may be untreatable, although portal decompression can be helpful by decreasing intralumenal pressure of the malformation.

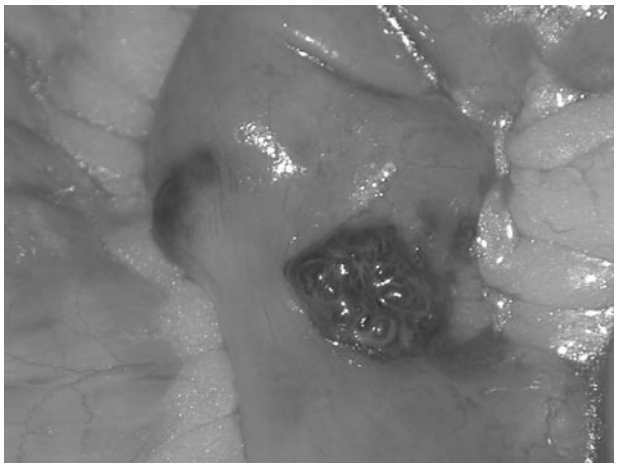


FIGURE 6 Intestinal lesions of the Blue Rubber Bleb Nevus syndrome.

Lower visceral diffuse venous malformations can extend from the anorectum proximally. They usually are transmural and extend into the pelvis. They may involve the entire colon or just the left side. Imaging studies demonstrate a markedly thickened colon and anorectum with or without phleboliths. Colonoscopy reveals a massively engorged, purple mucosa with contiguous varix-like projections, which can cause chronic bleeding requiring life-long, repeated transfusions. However, endoscopic therapy is futile in all cases and may exacerbate the bleeding. Surgery can involve partial colectomy with end colostomy or colectomy with anorectal mucosectomy and endorectal coloanal or ileoanal pull-through. Full-thickness rectal resection should be avoided because there can be uncontrollable hemorrhage from the extrarectal pelvic venous malformation.

Arteriovenous Malformations

AVMs are high-flow lesions between abnormal arteries and veins via many anomalous communications, without a normal intervening capillary bed. True AVMs of the GI tract are rare and many so-named lesions are in fact venous malformations. Endoscopically, AVMs in the gut appear pulsatile and demonstrate high flow on endoscopic ultrasound. Embolization of GI AVMs generally results in necrosis and perforation. Thus, surgical resection is the only curative therapy for gut lesions and presurgical localization may be performed by tattooing during selective arteriography.

Lymphatic Malformations

Lymphatic malformations of the GI tract are uncommon and mesenteric cysts may represent lymphatic malformations of the lacteals. Protein-losing enteropathy and ascites may be seen in association with anomalous development of the mesenteric lymphatics, cisterna chili, and thoracic duct. Endoscopy is of limited use in diagnosis and therapy.

Complex Combined Vascular Malformations

Complex combined vascular malformations result from various combinations of anomalous vessel channel types. The most common form in the GI tract is the capillary–lymphaticovenous malformation of the Klippel–Trenaunay syndrome (Fig. 7). This syndrome is characterized by an enlarged lower extremity at birth,

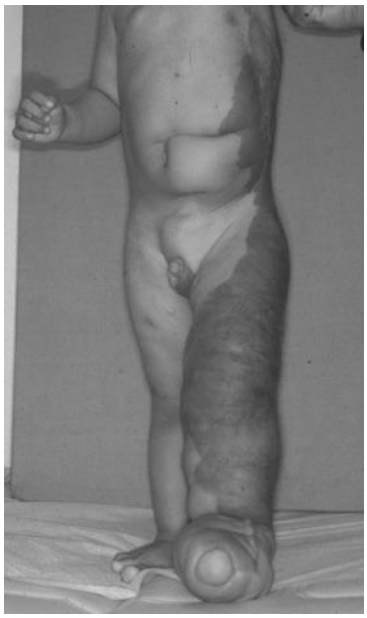


FIGURE 7 Klippel–Trenaunay syndrome.

a massive distorted foot, and axial overgrowth. Purple capillary stains are usually prominent and lymphatic vesicles protrude through the skin. Extension of the lesion into the pelvis to involve the bowel and bladder is not uncommon, though some with Klippel–Trenaunay syndrome may have no GI symptoms. On colonoscopy, the appearance of lesions is varied, sometimes with diffuse purple vascular discoloration and varied degrees of mural thickening. Therapy is indicated only for bleeding resulting in severe chronic anemia or if there is recurrent cellulitis of the buttock or thigh. These infections are often demonstrated to result from enteric organisms that translocate directly through the abnormal mucosal barrier in the malformation. Therapy may entail partial colectomy with diverting colostomy or colectomy with anorectal mucosectomy and coloanal endorectal pull-through.

An appreciation of clinical patterns, accurate nomenclature, associated visible cutaneous lesions, endoscopic appearances, and radiologic findings is important in the proper diagnosis and treatment of vascular anomalies of the gastrointestinal tract.

See Also the Following Articles

Kaposi's Sarcoma • Vascular Abnormalities

Further Reading

Fishman, S. J., Burrows, P. E., Leichtner, A. M., *et al.* (1998). Gastrointestinal manifestations of vascular anomalies in childhood: Varied etiologies require multiple therapeutic modalities. *J. Pediatr. Surg.* **33**, 1163.

Fishman, S. J., and Fox, V. L. (2001). Visceral vascular anomalies. *Gastrointest. Endosc. Clin. North Am.* **11**, 813–834.

Fishman, S. J., Shamberger, R. C., Fox, V. L., and Burrows, P. E. (2000). Endorectal pull-through abates gastrointestinal hemorrhage from colorectal venous malformations. *J. Pediatr. Surg.* **35**, 982–984.

Marler, J. J., Fishman, S. J., Upton, J., Burrows, P. E., Paltiel, H. J., Jennings, R. W., and Mulliken, J. B. (2002). Prenatal diagnosis of vascular anomalies. *J. Pediatr. Surg.* **37**, 318–326.

Mulliken, J. B., Fishman, S. J., and Burrows, P. E. (2000). Vascular anomalies. *Curr. Prob. Surg.* **37**, 517–584.

Vascular Anomalies Center Website. Available at http://www.childrenshospital.org/vascularanomalies/.

Vasculitis

SALAHUDDIN KAZI
University of Texas Southwestern Medical Center, Dallas

antineutrophil cytoplasmic antibodies Proteins directed against cytoplasmic neutrophil antigens such as proteinase 3 and myeloperoxidase; useful in the diagnosis of small-vessel vasculitis.

Janeway lesion Painless, macular lesions on the palms and soles, characteristic of infective endocarditis.

oligoarthritis Form of arthritis that involves more than one but a relatively small number of joints.

Osler node A tender, raised, cutaneous lesion, typically on the finger pads; characteristic of infective endocarditis.

Vasculitis is defined pathologically as a vessel wall lesion characterized by the presence of leukocytes, with reactive damage to mural vascular structures resulting in tissue ischemia and necrosis. Clinically, vasculitis refers to syndromes resulting from such local, regional, or systemic inflammation and necrosis of blood vessels. Vasculitis may occur in the context of a well-defined disease such as infective endocarditis (secondary vasculitis) or may be idiopathic. Secondary vasculitis is much more common than idiopathic vasculitis. Disorders that mimic clinical vasculitis include multiple cholesterol emboli syndrome and several occlusive vasculopathic disorders. Histologically, such disorders are notable for the absence of vasculitic changes. The term "vasculopathy" is often applied to such disorders.

CLASSIFICATION

The idiopathic or primary systemic vasculitides are a heterogeneous group of diseases with frequent clinical and pathologic overlap. Classification is based on vessel size, clinical features, and histology (Table I). Three main categories are recognized based on vessel size (large, medium, and small) and are further subclassified based on clinical and histologic characteristics.

EPIDEMIOLOGY

Although the causes of vasculitis are largely unknown, epidemiologic data have suggested roles for geographic, genetic, and environmental factors. Giant cell arteritis is the most common vasculitis in older individuals from North America and Europe. Takayasu's arteritis is predominantly seen in individuals from the Far East and Southeast Asia. The medium- and small-vessel vasculitides (polyarteritis nodosa and microscopic polyangiitis, Wegener's granulomatosis, and Churg–Strauss syndrome) are all relatively uncommon. Kawasaki disease primarily occurs in children under the age of 5 years. Henoch–Schönlein purpura is the most common type of vasculitis in children and is relatively infrequent in adults. Cutaneous leukocytoclastic vasculitis is relatively common and frequently seems to be precipitated by the use of medications.

ETIOLOGY AND PATHOGENESIS

The systemic vasculitides are immune-mediated inflammatory diseases for which the precise etiology is unclear. These diseases may be triggered by infection, with the release of proinflammatory cytokines and up-regulation of leukocyte adhesion molecules in an uncontrolled manner. Several factors appear to determine the size of vessel involved, the subsequent inflammatory responses, and systemic manifestations. These include the nature of the antigen involved, factors modulating endothelial cell activation, and other components of the inflammatory cascades, including cytokines, costimulatory molecules, and intracellular activation pathways. The end result is disruption of vessel integrity, producing clinical signs and symptoms of tissue ischemia.

Giant cell arteritis appears to be a T-cell-mediated disease. Macrophages produce interleukin-1, interleukin-6, and transforming growth factor-β and often become organized as multinucleated giant cells under the influence of interferon γ. Takayasu's arteritis is also a granulomatous polyarteritis in which CD8 T cells have been identified as the major cell type. Several candidate organisms have been associated with Kawasaki syndrome. It is speculated that toxins may act as superantigens. Polyarteritis nodosa is an immune complex vasculitis and has been reported in association with hepatitis B virus in some cases. Wegener's granulomatosis, microscopic polyangiitis, and Churg–Strauss syndrome have been referred to as the antineutrophil

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Names and Definitions of Vasculitides Adopted by The Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitis[a]

Name	Definition
Large-vessel vasculitis	Giant cell (temporal) arteritis Granulomatous arteritis of the aorta and its major branches with a predilection for the extracranial branches of the carotid artery; often involves the temporal artery; usually occurs in patients older than 50 years and often is associated with polymyalgia rheumatica
	Takayasu arteritis Granulomatous inflammation of the aorta and its major branches; usually in patients younger than 50 years
Medium-vessel vasculitis	Polyarteritis nodosa (classic polyarteritis nodosa) Necrotizing inflammation of medium or small arteries without glomerulonephritis or vasculitis in arterioles, capillaries, or venules
	Kawasaki disease Arteritis of large, medium, and small arteries associated with the mucocutaneous lymph node syndrome; coronary arteries are often involved and aorta and veins may be involved; usually occurs in children
Small-vessel vasculitis	Wegener's granulomatosis Granulomatous inflammation involving the respiratory tract, and necrotizing vasculitis affecting small to medium-sized vessels (e.g., capillaries, venules, arterioles, and arteries); necrotizing glomerulonephritis is common
	Churg–Strauss syndrome Eosinophil-rich and granulomatous inflammation involving the respiratory tract, and necrotizing vasculitis affecting small to medium-sized vessels; associated with asthma and eosinophilia
	Microscopic polyangiitis (microscopic polyarteritis) Necrotizing vasculitis, with few or no immune deposits, affecting small vessels (i.e., capillaries, venules, or arterioles); necrotizing arteritis involving small and medium-sized arteries may be present; necrotizing glomerulonephritis is very common and pulmonary capillaritis often occurs
	Henoch–Schönlein purpura Vasculitis, with IgA-dominant immune deposits, affecting small vessels (i.e., capillaries, venules, or arterioles); typically involves skin, gut, and glomeruli, and is associated with arthralgias or arthritis
	Essential cryoglobulinemic vasculitis Vasculitis, with cryoglobulin deposits, affecting small vessels (i.e., capillaries, venules, or arterioles), and associated with cryoglobulins in serum; skin and glomeruli are often involved
	Cutaneous leukocytoclastic vasculitis or glomerulonephritis Isolated cutaneous leukocytoclastic angiitis without systemic vasculitis

[a] Modified from Jennette *et al.* (1994).

cytoplasmic antibody (ANCA)-associated vasculitides. It is thought that the ANCAs play a direct pathogenic role in these disorders. Henoch–Schönlein purpura is an immune complex vasculitis in which immunoglobulin A (IgA) is the principal component. Leukocytoclastic vasculitis results from immune complex deposition in postcapillary venules.

CLINICAL FEATURES

General

The systemic vasculitides are inflammatory diseases that often result in simultaneous tissue ischemia of several organ systems. The associated systemic inflammatory response is often clinically evident as fever, weight loss, and malaise. Giant cell arteritis may present as a fever of unknown origin in the elderly.

Cutaneous

Leukocytoclastic vasculitis is a reaction pattern that may be seen in a variety of vasculitides. When the skin is the only organ system involved, the term "cutaneous leukocytoclastic vasculitis" is assigned. Henoch–Schönlein purpura typically manifests as palpable purpura over the lower extremities. Many other vasculitides can also produce similar lesions. In addition to leukocytoclastic vasculitis, other cutaneous patterns of injury include nodules, ulceration, and bullous lesions. Occasionally, digital gangrene occurs from medium-vessel occlusion. Secondary vasculitides (including cryoglobulinemic vasculitis and immune complex vasculitis from infections such as infective endocarditis) also have prominent cutaneous features. Cryoglobulinemic vasculitis can cause purpura and infarcts of cool peripheral structures such as the digits, ear lobes, and tip of the nose and the penis. Characteristic cutaneous lesions from infective endocarditis are referred to as Osler's nodes and Janeway lesions. Splinterlike hemorrhages in the nail bed are also frequently noted in this disorder.

Musculoskeletal

Arthritis is a relatively infrequent in the systemic vasculitides. It may involve both large and small joints. Common patterns include a symmetric small-joint synovitis of the upper extremities (as can occur in giant cell arteritis) or an asymmetric oligoarthritis of the large joints (as may occur in the ANCA-associated vasculitides). Muscle involvement may occur in polyarteritis nodosa, presenting as myalgia or muscle weakness.

Renal

Glomerulonephritis is characteristic of the ANCA-associated small-vessel vasculitides. It is nearly universal in microscopic polyangiitis, very frequent in Wegener's granulomatosis, and can occasionally occur in Churg–Strauss syndrome. It is also frequently seen in Henoch–Schönlein purpura and in cryoglobulinemic vasculitis. Clinical manifestations include hematuria, proteinuria, rapidly progressive glomerulonephritis, acute nephritic syndrome, and occasionally the nephrotic syndrome. Hematuria and mild renal insufficiency are characteristic of Henoch–Schönlein purpura, whereas rapidly progressive glomerulonephritis with moderate to severe renal insufficiency frequently occurs in Wegener's granulomatosis and microscopic polyangiitis. Polyarteritis nodosa causes a renal arteritis without glomerulonephritis. Clinically, this presents with renal insufficiency, hypertension, and proteinuria.

Gastrointestinal

Gastrointestinal involvement occurs primarily with polyarteritis nodosa and Henoch–Schönlein purpura. Polyarteritis nodosa frequently involves the mesenteric vasculature and can cause bowel infarction, ulceration, and bleeding. Involvement of the pancreatic vessels can cause acute pancreatitis. It may present as postprandial periumbilical pain. Rupture of the mesenteric microaneurysms is a life-threatening complication of polyarteritis nodosa. Localized polyarteritis nodosa is not uncommonly reported in pathologic specimens from patients with acute cholecystitis or appendicitis. Liver inflammation (hepatitis) is an uncommon feature of idiopathic systemic vasculitis and, when prominent, should suggest a secondary infectious cause of vasculitis such as hepatitis C virus, with cryoglobulinemic vasculitis, and the rickettsial diseases. In children, Henoch–Schönlein purpura frequently presents with colicky abdominal pain and gastrointestinal hemorrhage. Some patients experience nausea and vomiting. Hepatomegaly, hydrops of the gallbladder, diarrhea, and pancreatitis have all been reported in Kawasaki syndrome.

Neurologic

Vasculitis of the central nervous system can occur as an isolated disorder. It is referred to as primary angiitis of the central nervous system. In giant cell arteritis, headache from temporal artery involvement and blindness from central retinal artery ischemia are relatively common. Takayasu's arteritis may present with cerebral ischemia. Vasculitis associated with connective tissue diseases such as systemic lupus erythematosus and

Sjögren's syndrome can also involve the central nervous system. Other systemic vasculitides more commonly involve the peripheral nervous system, often presenting as mononeuritis multiplex with acute foot drop or wrist drop. This is particularly characteristic of polyarteritis nodosa and Churg–Strauss syndrome.

Pulmonary

Involvement of the upper airway is characteristic of Wegener's granulomatosis, which can involve the nasal septum, ears, larynx, and trachea. Subglottic stenosis is a common complication. Pulmonary infiltrates, nodules, and cavitary lesions are also frequently noted in Wegener's granulomatosis. Pulmonary capillaritis presenting as alveolar hemorrhage is characteristic of microscopic polyangiitis but can also occur with Wegener's granulomatosis. Churg–Strauss syndrome is characterized by asthma. Pulmonary involvement with polyarteritis nodosa is distinctly uncommon. Giant cell arteritis may present with throat pain and cough.

Other

Kawasaki disease characteristically involves the coronary arteries and, untreated, can result in coronary artery aneurysms. Epididymo-orchitis can occur in polyarteritis nodosa. Eosinophilia is universal in Churg–Strauss syndrome. Ocular manifestations in the systemic vasculitides include scleritis, uveitis, retinal vasculitis, and orbital pseudotumors. Ocular disease is particularly common in Wegener's granulomatosis and giant cell arteritis.

DIAGNOSIS

The diagnosis of systemic vasculitis is considered when patients present with the systemic inflammatory disorder associated with rapidly progressive multiorgan dysfunction. Certain clinical features are very suggestive of vasculitis. These include acute renal insufficiency with an active urinary sediment, mononeuritis multiplex, unexplained central nervous system ischemia, palpable purpura, and acute pulmonary hemorrhage. In all such cases, it is important to first consider the secondary causes of vasculitis. These include a variety of infectious diseases (such as infective endocarditis, rickettsial diseases, hepatitis C virus infection, and human immunodeficiency virus infection), lymphoproliferative disorders (especially hairy cell leukemia), left atrial myxoma, and drugs such as ergot alkaloids and cocaine. Diagnostic tests that are helpful include those that exclude the secondary causes (blood cultures, viral serology, echocardiography, and urine drug screen) and those that suggest immune complex formation (rheumatoid factor and antinuclear antibody) or the production of antineutrophil cytoplasmic antibodies. ANCA testing is particularly valuable when small-vessel vasculitis presents as a pulmonary–renal syndrome. Indirect immunofluorescence reveals two characteristic patterns: diffuse cytoplasmic immunofluorescence (c-ANCA) is moderately sensitive and specific for active Wegener's granulomatosis, and perinuclear immunofluorescence (p-ANCA) is indicative of microscopic polyangiitis. Most patients with Wegener's granulomatosis will test positive for antiproteinase 3 antibodies by enzyme-linked immunosorbent assay (ELISA) and the majority of patients with microscopic polyangiitis will test positive for antimyeloperoxidase antibodies (anti-MPOs) by ELISA. ANCA titers may fluctuate with disease activity and may predict relapse in Wegener's granulomatosis.

It is important to stress that vasculitis is a pathological diagnosis and every attempt should be made to obtain a tissue diagnosis. Angiography can be useful when obtaining tissue is not feasible (e.g., suspected mesenteric or intracranial vasculitis).

THERAPY AND PROGNOSIS

Aggressiveness of therapy of the systemic vasculitides is determined by the site, extent, and pace of organ involvement. Drug-induced leukocytoclastic vasculitis is generally a self-limited disease once the offending medication is withdrawn. Occasionally, a short course of oral corticosteroids is necessary. Similarly, Henoch–Schönlein purpura may resolve spontaneously in the absence of any specific treatment. However, if renal involvement is present, therapy is usually indicated. Some vasculitides, such as giant cell arteritis, usually respond to corticosteroids alone. Diseases such as polyarteritis nodosa and Wegener's granulomatosis typically pose a threat of catastrophic organ damage and are usually treated with simultaneous administration of corticosteroids and cytotoxic agents. Once remission is obtained, corticosteroids are the first to be withdrawn followed by the withdrawal of cytotoxic agents. In Wegener's granulomatosis, the ANCA titers fall with remission and may even revert to negative. In some cases, a rising ANCA titer predicts clinical relapse.

Early mortality in the vasculitides is usually from the disease or is the result of complications of acute immunosuppression. Late mortality usually occurs from complications of therapy. The prognosis in leukocytoclastic vasculitis and giant cell arteritis is excellent, whereas diseases such as polyarteritis nodosa and Wegener's

granulomatosis can cause serious morbidity and decreased life expectancy. The 5-year survival of polyarteritis nodosa is 60% and that of Wegener's granulomatosis is 70%.

See Also the Following Articles

Henoch–Schönlein Purpura • Hepatitis C • Kawasaki Syndrome

Further Reading

Calabrese, L. H., Duna, G. F., and Lie, J. T. (1997). Vasculitis in the central nervous system. *Arthritis Rheum.* **40**, 1189.

Fauci, A., Katz, P., Haynes, B., *et al.* (1979). Cyclophosphamide therapy of severe systemic necrotizing vasculitis. *N. Engl. J. Med.* **301**, 235.

Freeman, A. F., and Shulman, S. T. (2001) Recent developments in Kawasaki disease. *Curr. Opin. Infect. Dis.* **14**(3), 357–361.

Guillevin, L., Le Thi Hong, D., Godeau, P., *et al.* (1988). Clinical findings and prognosis of polyarteritis nodosa and Churg–Strauss angiitis: A study in 165 patients. *Br. J. Rheumatol.* **27**, 258.

Jennette, J. C., and Falk, R. J. (1997). Small-vessel vasculitis. *N. Engl. J. Med.* **337**, 1512.

Jennette, J. C., Falk, R. J., Andrassy, K., Bacon, P. A., Churg, J., Gross, W. L., Hagen, E. C., Hoffman, G. S., Hunder, G. G., Kallenberg, C. G., *et al.* (1994). Nomenclature of systemic vasculitides. Proposal of an international consensus conference. *Arthritis Rheum.* **37**(2), 187–192.

Kerr, G. (1995). Takayasu's arteritis. *Rheum. Dis. Clin. North Am.* **21**, 1041.

Lee, E., Smith, H., Miller G. I., *et al.* (1984). Ischemic pseudomembranous colitis with perforations due to polyarteritis nodosa. *Am. J. Gastroenterol.* **79**, 35.

Levine, S. M., and Hellmann, D. B. (2002). Giant cell arteritis. *Curr. Opin. Rheumatol.* **14**, 3.

Lhote, F., and Guillevin, L. (1995). Polyarteritis nodosa, microscopic polyangiitis and Churg–Strauss syndrome. *Rheum. Dis. Clin. North Am.* **21**, 911.

Lopez, L., Schocket, A., Standford, R., *et al.* (1980). Gastrointestinal involvement in leukocytoclastic vasculitis and polyarteritis nodosa. *J. Rheumatol.* **7**, 677.

Rao, J. K., Weinberger, M., Oddone, E. Z., *et al.* (1995). The role of antineutrophilic cytoplasmic antibody (C-ANCA) testing in the diagnosis of Wegener granulomatosis. A literature review and meta-analysis. *Ann. Intern. Med.* **123**, 925.

Vasoactive Intestinal Peptide (VIP)

TOAN D. NGUYEN

University of Washington and Puget Sound VA Health Care System, Seattle

nitric oxide (NO) Chemical that mediates smooth muscle relaxation through activation of guanylate cyclase and production of cyclic GMP; vasoactive intestinal peptide acts through the single-transmembrane natriuretic peptide clearance receptor C, coupled to G_{i1} and G_{i2} proteins, to activate endothelial nitric oxide synthase, with resultant production of NO.

vasoactive intestinal peptide (VIP) family VIP and peptides homologous to VIP, including secretin, glucagon, PHM (a 27-amino-acid peptide having N-terminal histidine and C-terminal methionine), pituitary adenylate cyclase-activating peptide (of either 27 or 38 amino acids in length), gastric inhibitory peptide, growth hormone-releasing factor, and helodermin.

$VPAC_1$ and $VPAC_2$ Specific, high-affinity membrane receptors for vasoactive intestinal peptide; both are G_s-coupled receptors spanning the plasma membrane seven times.

WDHA A syndrome of watery diarrhea, hypokalemia, and achlorhydria, associated with neuroendocrine vasoactive intestinal peptide-secreting tumors (VIPoma); also known as pancreatic cholera or Verner-Morrison syndrome.

In 1902, W. Bayliss and E. Starling described secretin and established the field of endocrinology. In their landmark communication to the Royal Society of London, they noted that "... secretin is associated with another body with a pronounced lowering effect on the blood pressure." This may be the first reference to vasoactive intestinal peptide (VIP), identified by S. Said and V. Mutt in 1970 as a polypeptide from the small intestine that induces systemic

Copyright 2004, Elsevier (USA). All rights reserved.

granulomatosis can cause serious morbidity and decreased life expectancy. The 5-year survival of polyarteritis nodosa is 60% and that of Wegener's granulomatosis is 70%.

See Also the Following Articles

Henoch–Schönlein Purpura • Hepatitis C • Kawasaki Syndrome

Further Reading

Calabrese, L. H., Duna, G. F., and Lie, J. T. (1997). Vasculitis in the central nervous system. *Arthritis Rheum.* **40**, 1189.

Fauci, A., Katz, P., Haynes, B., *et al.* (1979). Cyclophosphamide therapy of severe systemic necrotizing vasculitis. *N. Engl. J. Med.* **301**, 235.

Freeman, A. F., and Shulman, S. T. (2001) Recent developments in Kawasaki disease. *Curr. Opin. Infect. Dis.* **14**(3), 357–361.

Guillevin, L., Le Thi Hong, D., Godeau, P., *et al.* (1988). Clinical findings and prognosis of polyarteritis nodosa and Churg–Strauss angiitis: A study in 165 patients. *Br. J. Rheumatol.* **27**, 258.

Jennette, J. C., and Falk, R. J. (1997). Small-vessel vasculitis. *N. Engl. J. Med.* **337**, 1512.

Jennette, J. C., Falk, R. J., Andrassy, K., Bacon, P. A., Churg, J., Gross, W. L., Hagen, E. C., Hoffman, G. S., Hunder, G. G., Kallenberg, C. G., *et al.* (1994). Nomenclature of systemic vasculitides. Proposal of an international consensus conference. *Arthritis Rheum.* **37**(2), 187–192.

Kerr, G. (1995). Takayasu's arteritis. *Rheum. Dis. Clin. North Am.* **21**, 1041.

Lee, E., Smith, H., Miller G. I., *et al.* (1984). Ischemic pseudomembranous colitis with perforations due to polyarteritis nodosa. *Am. J. Gastroenterol.* **79**, 35.

Levine, S. M., and Hellmann, D. B. (2002). Giant cell arteritis. *Curr. Opin. Rheumatol.* **14**, 3.

Lhote, F., and Guillevin, L. (1995). Polyarteritis nodosa, microscopic polyangiitis and Churg–Strauss syndrome. *Rheum. Dis. Clin. North Am.* **21**, 911.

Lopez, L., Schocket, A., Standford, R., *et al.* (1980). Gastrointestinal involvement in leukocytoclastic vasculitis and polyarteritis nodosa. *J. Rheumatol.* **7**, 677.

Rao, J. K., Weinberger, M., Oddone, E. Z., *et al.* (1995). The role of antineutrophilic cytoplasmic antibody (C-ANCA) testing in the diagnosis of Wegener granulomatosis. A literature review and meta-analysis. *Ann. Intern. Med.* **123**, 925.

Vasoactive Intestinal Peptide (VIP)

TOAN D. NGUYEN

University of Washington and Puget Sound VA Health Care System, Seattle

nitric oxide (NO) Chemical that mediates smooth muscle relaxation through activation of guanylate cyclase and production of cyclic GMP; vasoactive intestinal peptide acts through the single-transmembrane natriuretic peptide clearance receptor C, coupled to G_{i1} and G_{i2} proteins, to activate endothelial nitric oxide synthase, with resultant production of NO.

vasoactive intestinal peptide (VIP) family VIP and peptides homologous to VIP, including secretin, glucagon, PHM (a 27-amino-acid peptide having N-terminal histidine and C-terminal methionine), pituitary adenylate cyclase-activating peptide (of either 27 or 38 amino acids in length), gastric inhibitory peptide, growth hormone-releasing factor, and helodermin.

$VPAC_1$ and $VPAC_2$ Specific, high-affinity membrane receptors for vasoactive intestinal peptide; both are G_s-coupled receptors spanning the plasma membrane seven times.

WDHA A syndrome of watery diarrhea, hypokalemia, and achlorhydria, associated with neuroendocrine vasoactive intestinal peptide-secreting tumors (VIPoma); also known as pancreatic cholera or Verner-Morrison syndrome.

In 1902, W. Bayliss and E. Starling described secretin and established the field of endocrinology. In their landmark communication to the Royal Society of London, they noted that "... secretin is associated with another body with a pronounced lowering effect on the blood pressure." This may be the first reference to vasoactive intestinal peptide (VIP), identified by S. Said and V. Mutt in 1970 as a polypeptide from the small intestine that induces systemic

Copyright 2004, Elsevier (USA). All rights reserved.

vasodilation, hypotension, increased cardiac output, respiratory stimulation, and hyperglycemia. VIP is now known to have protean biological, physiological, and clinical effects; a Medline search yielded ~10,000 articles published with VIP as a keyword. This article will examine the peptide structure of VIP, its receptors, cellular action, distribution, biological effects, and clinical relevance.

VASOACTIVE INTESTINAL PEPTIDE: PEPTIDE STRUCTURE AND HOMOLOGOUS PEPTIDES

Vasoactive intestinal peptide (VIP) is composed of 28 amino acid residues and belongs to a family of homologous bioactive peptides that include PHM (27 amino-acid peptide having N-terminal histidine and C-terminal methionine), pituitary adenylate cyclase-activating peptide (PACAP; of either 27 or 38 amino acids in length), secretin, glucagon, gastric inhibitory peptide (GIP), growth hormone-releasing factor (GHRF), and helodermin (Gila monster venom) (Fig. 1). PACAP is a neuropeptide and secretin, glucagon, GIP, and GHRH are hormones. PHM is co-synthesized with VIP (see below).

Human, cow, pig, goat, dog, and rat (but not guinea pig) VIPs are identical, suggesting conservation during mammalian evolution. Nonmammalian VIP (chicken, alligator, frog, trout, bowfin, dogfish, cod, and goldfish) differs from the human peptide at only four or five positions.

Peptide modeling suggests that VIP exhibits a central α-helix from Val-3 to Asn-24 with random coiled N- and C-termini. Studies of VIP analogues and alanine substitutions suggest that His-1, Val-5, Thr-11, Arg-14, Lys-15, Lys-21, Asn-28, Leu-23, and Ile-26 interact directly with VIP receptors.

Human VIP is synthesized as a precursor protein (prepro-VIP) of 170 amino acids (MW 19,169) containing, in addition to VIP, a signal peptide of ~20 amino acids and PHM. As shown in Fig. 1, human PHM exhibits marked homology to VIP (in the rat, substitution of the N-terminal methionine by isoleucine yields PHI); its effect is mediated through VIP receptors. In humans, prepro-VIP mRNA is transcribed from an 8837 bp gene that encodes seven exons and six introns and that is located on chromosome 6q24.

VIP	HSDAVFTDNY TRLRKQMAVK KYLNSILN	
PACAP	HSDGIFTDSY SRYRKQMAVK KYLAAVLGKR YKQRVKNKN	[19]
PHM	HSDGVFTSDF SKLLGQLSAK KYLESLM	[13]
Helodermin	HSDAIFTQQY SKLLAKLALQ KYLASILGSS TSPPP	[14]
Secretin	HSDGTFTSEL SRLREGARLQ RLLQGLL	[10]
GHRF	YADAIFTSDY RKVLGQLSAR KLLQDIMSQR … L (44)	[8]
Glucagon	HSQGTFTSDY SKYLDSRRAQ DVFQWLMNT	[6]
GIP	YAEGTFISDY SIAMDDIRQQ DFVNWLL	[4]

FIGURE 1 VIP peptide family. The amino acid sequences of human VIP and of homologous peptides are shown. With VIP serving as reference, the identical amino acids are underlined and the number of identical amino acids for each peptide is given in brackets.

VASOACTIVE INTESTINAL PEPTIDE RECEPTORS

This section will examine the receptors specific for VIP that mediate the different biologic effects of this peptide. Although VIP interacts with many receptors with variable degrees of affinity and specificity, only two receptors with high affinity for VIP are identified, named $VPAC_1$ and $VPAC_2$ in 1998. Of the peptides of the VIP family, PHM also acts through $VPAC_1$ and $VPAC_2$, since no receptor specific for PHM has been identified. PACAP also binds to $VPAC_1$ and $VPAC_2$; it also binds with high affinity to PAC_1, a receptor with low affinity for VIP.

$VPAC_1$

Previously known as the VIP, VIP_1, VIP/PACAP type II receptor, or PVR2 receptor, $VPAC_1$ was extensively characterized functionally and biochemically in the 1980s. It was cloned in 1992 from the rat lung by Ishihara *et al.* and the human counterpart was identified in 1993–1994. The human $VPAC_1$ contains 460 amino acids with a predicted MW of 51,929 (including a signal sequence of 30 amino acids); it exhibits 81% amino acid sequence identity with its rat counterpart. Paralleling the homology of their ligands, there is marked homology between $VPAC_1$ and the receptors for PACAP (PAC_1), secretin, GHRF, glucagon, parathyroid hormone (PTH), glucagon-like peptide I, and calcitonin. The human chromosomal location of $VPAC_1$ is 3p22. There is no known splice variant of this receptor.

As illustrated in Fig. 2, hydropathy analysis predicts seven transmembrane domains, characteristic of G-protein-coupled receptors, and a large N-terminal extracellular domain of 114 amino acids (excluding the signal sequence). Studies of chimeric secretin–VIP receptors support a key role for this N-terminal domain in VIP recognition. Similar to other G-protein-coupled receptors, the second and third intracellular loops may interact with G_s, the G-protein coupled to $VPAC_1$, and the intracellular C-terminus may mediate receptor desensitization. There are four potential sites for N-glycosylation, one site for cyclic AMP (cAMP)-dependent phosphorylation, and three sites for protein kinase

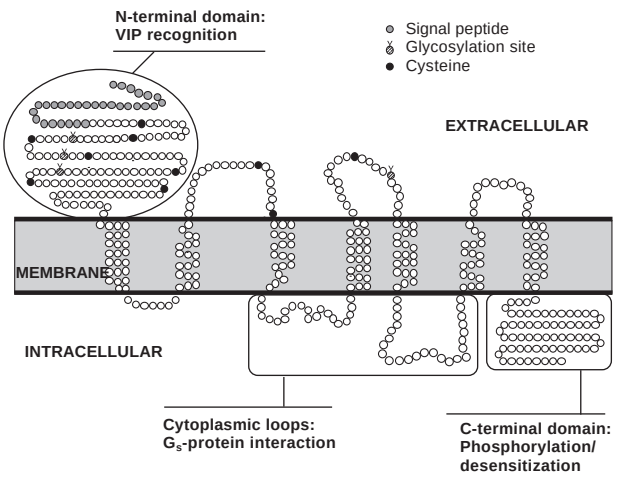


FIGURE 2 Predicted structure of $VPAC_1$. Each amino acid is represented as a circle and the amino acids of the signal peptide, cysteine residues, and potential glycosylation sites are indicated, respectively, by gray shading, black shading, and a Y symbol. The seven-transmembrane domains are characteristic of G-protein-coupled receptors. The large extracellular N-terminus is the domain for VIP recognition; the second and third cytoplasmic loops and the cytoplasmic C-terminal domain contain sites for G_s-protein interaction and phosphorylation/desensitization. Modified from Gaudin *et al.* (1995). *Biochem. Biophys. Res. Commun.* **211**, 901–908.

C phosphorylation. Of the 10 extracellular cysteines, the residues at positions 50, 63, 72, 86, 105, and 122 in the N-terminal domain are crucial for VIP binding. The Asp-3 of VIP penetrates into the transmembrane region of the receptor, in close proximity to the conserved Arg-188 and Lys-195 on the second transmembrane domain.

Functional studies of $VPAC_1$ demonstrate a VIP-binding affinity of 0.6 n*M* and the following order of potency: VIP = PACAP-27 > PACAP-38 > helodermin > GHRF = PHM > secretin. Selective $VPAC_1$ receptor agonists include the VIP/GRF hybrid [Lys-15, Arg-16, Leu-27]VIP(1–7)GRF(8–27)-NH_2, chicken [Arg-16]-secretin (which also activates the secretin receptor), and [Ala-11, -22, -28]VIP (>1000-fold discrimination between $VPAC_1$ and $VPAC_2$). [Acetyl-His-1, D-Phe-2, Lys-15, Arg-16]VIP(3–7)GRF(8–27)-NH_2 and (*N*-stearyl, norleucine-17)VIP hybrid are reported antagonists of $VPAC_1$.

Using Northern blotting, $VPAC_1$ is localized to lung > prostate > peripheral leukocyte, liver, brain, small intestine > colon, heart, spleen > placenta, kidney, thymus, testis (decreasing order of expression). By receptor autoradiography, $VPAC_1$ is expressed in many epithelial cells, such as hepatocytes, mucosal cells of the stomach and colon, acinar and/or duct cells of the breast, pancreas, and lung, and glandular cells of the thyroid, prostate, uterus, and adrenal medulla.

$VPAC_2$

$VPAC_2$, also known as the helodermin-preferring VIP receptor, VIP_2, PACAPR-3, or PVR3 receptor, was first cloned from the rat olfactory bulb and its human counterpart was subsequently identified. Its human chromosomal location is 7q36.3. Human $VPAC_2$ contains 438 amino acids (including a putative 23-amino-acid signal sequence) and exhibits an 86% amino acid identity with the rat $VPAC_2$. Homology with other receptors was much lower: 49% amino acid identity with human $VPAC_1$ and 52% identity with PAC_1. Similar to $VPAC_1$, the predicted structure of $VPAC_2$ includes seven transmembrane domains and a large extracellular N-terminal domain. There are 10 extracellular cysteines, conserved from $VPAC_1$, and three potential sites for N-glycosylation.

Binding and functional studies of $VPAC_2$ show the following order of potency: PACAP-38 ≥ helodermin ≥ VIP = PACAP-27. Unlike $VPAC_1$, $VPAC_2$ does not bind secretin, even at 1 μ*M*. Ro 25-1553 is a highly selective agonist for $VPAC_2$ and may be used for bronchodilation.

By receptor radioautography, $VPAC_2$ is found on smooth muscle cells (e.g., stomach, blood vessels) and on thymic lymphoid cells.

Related Receptors

A protein has been identified from the human jejunum exhibiting 100% homology with the 428 C-terminal amino acids of $VPAC_1$ but completely divergent at the N-terminal 67 amino acids. This protein does not bind VIP and its functional significance remains to be determined.

As discussed below, VIP may also interact with the single-transmembrane natriuretic peptide clearance receptor, coupled with G_{i1} or G_{i2}, to stimulate endothelial nitric oxide synthase (eNOS) in smooth muscle cells.

DISTRIBUTION AND METABOLISM

Distribution

VIP is widely distributed in the human body, primarily within the neurons of the central and peripheral nervous systems. Over the length of the gastrointestinal (GI) tract, it is present in large amounts in both its mucosal and muscular layers, with a trend toward more abundant peptide and higher mucosal/muscular

peptide ratio in the distal portions of this tract (with focal increases in the proximal duodenum and right colon). VIP is also present in the gallbladder and is abundant around sphincters (e.g., lower esophageal sphincter, pylorus, and sphincter of Oddi).

Within the gut wall, VIP is found in ~2.5% of the neurons of the myenteric, or Auerbach's, plexus and in ~45% of the neurons of the submucous, or Meissner's, plexus. As shown in Fig. 3, VIP-containing neurons in the myenteric ganglia project distally to other myenteric and submucous ganglia, to underlying circular muscles, and to prevertebral ganglia, suggesting a role for VIP in mediating the descending inhibition of a peristaltic wave. VIP-containing neurons in the submucous ganglia project to submucous arterioles and to the mucosa, suggesting additional functions for this peptide in regulating mucosal transport and local blood flow.

In the exocrine GI system, VIP is also present in the postganglionic cholinergic neurons that innervate glandular tissues such as submandibular salivary glands and pancreas. In these glands, the presence of VIP-containing nerve endings around small blood vessels, acini, and ducts again suggests a vasoregulatory and secretory function for VIP.

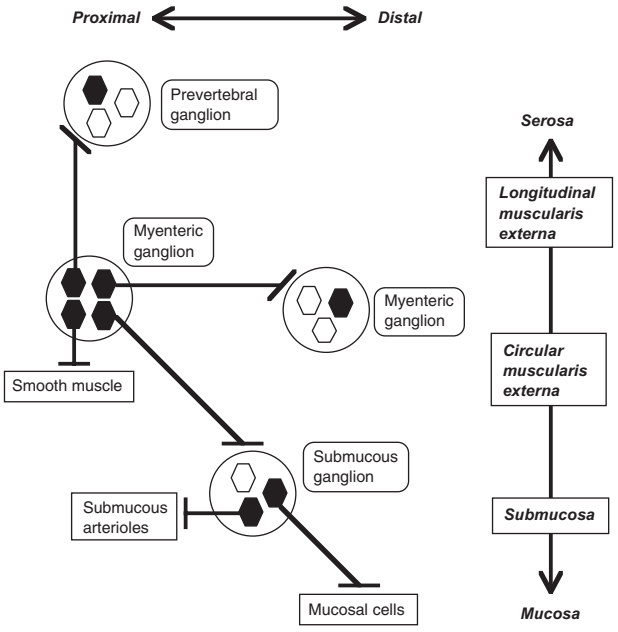


FIGURE 3 Organization of VIP-containing neurons in the GI tract. VIP-containing neurons (shown in black) in the myenteric and submucous ganglia of the GI tract and their projections to other ganglia and target cells (smooth muscles of the bowel wall, submucous arterioles, and secretory mucosal cells) are depicted. For orientation, the proximal (oral) and distal (anal) directions are shown at the top and the different layers of the bowel wall are shown on the right.

Vasoactive Intestinal Peptide Release and Metabolism

As a neurotransmitter, VIP is released from nerve endings (e.g., following neuronal depolarization) to act on adjacent cells. Locally, this VIP is cleared through internalization of the VIP–receptor complexes by the target cell, followed by intracellular degradation. Cell surface peptidases may also inactivate VIP, as reported for gastric smooth muscle membranes.

VIP may also evade local degradation to enter the vascular compartment. Indeed, VIP plasma concentrations (systemic basal levels: ~20–50 pg/ml; portal basal levels: 40–100 pg/ml) may increase following stimulation of VIP release (e.g., with oral lipids or intravenous calcium, oxytocin, or neostigmine). The entire GI tract is the main source for circulating VIP. VIP is rapidly cleared from the circulation, with a plasma half-life as short as 1–3 min. The major contribution of hepatic clearance in this process is suggested by the higher levels of VIP in the portal (versus systemic) circulation, elevated peptide levels in patients with hepatic failure, and hepatic extraction of labeled VIP injected into the portal circulation. In the systemic circulation, VIP is extracted by the lung.

BIOLOGIC EFFECTS OF VIP

Because VIP and its receptors are widely distributed in the GI tract, it has extensive effects on the digestive system. For over two decades, it has been established that VIP stimulates secretion from epithelial cells and relaxes smooth muscles; recently, it has also been recognized that VIP modulates inflammation and cellular growth and differentiation. The cellular mechanisms for these actions are illustrated in Fig. 4.

Stimulation of Secretion

The secretory effect of VIP has been established *in vitro* and *in vivo*, with cultured cells, animal models, and human subjects. VIP stimulates the secretion of fluid and electrolytes from the gallbladder, small intestine, and colon, of EGF and bicarbonate from duodenal Brunner's glands, and of bicarbonate and digestive enzymes from the exocrine pancreas. In promoting this secretory effect, VIP may act alone or in concert with other agents (e.g., VIP potentiates the secretory effect of cholecystokinin on pancreatic acini).

The mechanism underlying VIP-stimulated secretion has been elucidated with T84 colonocytes and is

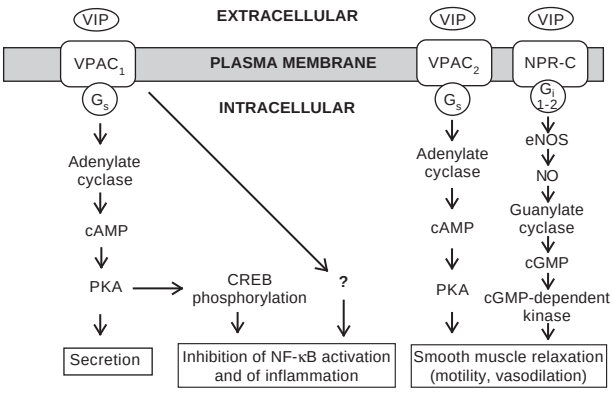


FIGURE 4 Biologic actions of VIP. The cellular mechanisms underlying the biologic effects of VIP on secretion, inflammation, and smooth muscle relaxation are depicted. $VPAC_1$ and $VPAC_2$, high-affinity VIP receptors; NPR-C, natriuretic peptide clearance receptor; G_s, G_{i1}, and G_{i2}, G-protein subtypes mediating the effect of VIP; eNOS, endothelial nitric oxide synthase; NO, nitric oxide; PKA, cAMP-dependent protein kinase A; CREB, cAMP regulatory element-binding protein. Modified from Murthy *et al*. (1998). *Am. J. Physiol*. **275**, C1409–C1416.

illustrative of VIP-induced secretion in other tissues. VIP interacts with $VPAC_1$, located on the basolateral membrane of these cells, to activate adenylate cyclase through G_s. The subsequent increase in intracellular cAMP will activate cAMP-dependent protein kinase (PKA), promote phosphorylation and activation of the relevant ion transport pathways (e.g., basolateral cAMP-dependent K^+ channel and apical cystic fibrosis transmembrane conductance regulator Cl^- channel), and induce the secretion of electrolytes (primarily Cl^-) and fluid. Agents acting via other signaling pathways may potentiate this response, either through cross talk between the transduction mechanisms or through activation of additional ion transport pathways (e.g., Ca^{2+}-activated K^+ channels).

Smooth Muscle Relaxation

In tissue and animal models, VIP induces the relaxation of smooth muscles from the stomach, small and large intestine, GI sphincters, gallbladder, and blood vessels. This effect is mediated through two pathways. By activating the G_s-coupled $VPAC_2$, VIP activates adenylate cyclase to promote intracellular cAMP production, stimulate PKA, and activate the proteins mediating smooth muscle relaxation. In addition, VIP may interact with the natriuretic peptide clearance receptor, coupled to G_{i1} and G_{i2}, to promote Ca^{2+} influx and $[Ca^{2+}]_i$ increase. The subsequent activation of the Ca^{2+}/calmodulin-dependent eNOS leads to the generation of nitric oxide (NO), NO-dependent activation of soluble guanylate cyclase, generation of cGMP, activation of cGMP-dependent kinase, and relaxation of smooth muscle cells.

In the gastric fundus, NOS and VIP are co-localized in the majority of myenteric neurons and NO also stimulates the release of VIP from nerve endings, further enhancing the interaction between VIP and NO. Conversely, VIP may also stimulate myenteric neurons to release substances that induce smooth muscle contraction (e.g., in guinea pig ileum).

Modulation of Inflammation

VIP suppresses the production of many pro-inflammatory agents by monocytes and macrophages. This effect, mediated through $VPAC_1$, occurs via inhibition of the transactivation of nuclear factor κB (NF-κB), a pleiotropic transcription factor that regulates the production of many pro-inflammatory agents [e.g., tumor necrosis factor α, interleukin-12 (IL-12), IL-1, IL-6, NO] and chemokines (e.g., IL-8, RANTES, monocyte chemoattractant protein-1, intracellular adhesion molecule-1). In unstimulated cells, NF-κB is present in the cytosol as a p50/p65 heterodimer, bound to the inhibitor IκB. On stimulation, IκB is phosphorylated and degraded, allowing NF-κB to translocate to the nucleus. In the nucleus, NF-κB promotes the transcription of inflammatory agents by binding to the corresponding DNA, in concert with other transcription factors, such as the CREB-binding protein (CBP, where CREB is cAMP regulatory element-binding protein) and TATA-box-binding protein (TBP).

VIP inhibits NF-κB transactivation at three levels. First, through a cAMP-independent pathway, VIP inhibits NF-κB nuclear translocation and subsequent DNA binding. Second, it induces the phosphorylation of CREB, allowing it to compete against p65 for CBP binding. Third, it inhibits the phosphorylation of TBP and its binding to p65 and the TATA-box. The effects of VIP on CBP and TBP are both mediated through cAMP. The significance of these actions on GI inflammation awaits further studies.

Effects on Cellular Growth and Differentiation

VIP also appears to modulate apoptosis and the growth and differentiation of neuronal and immune cells. It may also promote the proliferation of many neoplastic cells; $VPAC_1$ receptor antagonists have been used *in vitro* to inhibit tumor growth.

CLINICAL ASPECTS

Vasoactive Intestinal Peptide-Secreting Tumors (VIPoma)

In 1958, Verner and Morrison first described a syndrome of profuse, refractory, and watery diarrhea, hypokalemia, and achlorhydria in association with non-β islet cell pancreatic tumors. Known variously as pancreatic cholera, WDHA (for watery diarrhea, hypokalemia, and achlorhydria), and the VIPoma or Verner-Morrison syndrome, this entity is associated with elevated levels of VIP in the serum or corresponding tumor.

Pancreatic cholera is the third most common manifestation of pancreatic neuroendocrine tumors, after hypoglycemia from insulinomas and peptic symptoms from gastrinomas. VIPomas account for 2–7% of GI neuroendocrine tumors and occur with an annual incidence of 1 in 10^7 (mean patient age: 49 years; female/male preponderance: 3/1). Ninety percent of these tumors are encountered in the pancreas, where they usually occur as solitary tumors in the body or tail (75% of pancreatic cases). Nonpancreatic VIPomas occur mostly along the autonomic nervous system and in the adrenal medulla. Similar to other GI neuroendocrine tumors, approximately 60% of VIPomas are malignant. Most VIPomas also synthesize one or more additional neuropeptides (e.g., 39% of VIPomas secrete pancreatic polypeptide).

Whereas the liver and lung normally clear circulating VIP, excessive VIP secreted by tumors reaches distal organs to affect their function. Pancreatic cholera, occurring in half of the patients with VIPoma, is caused by the stimulation of secretion from intestinal crypt cells, manifested clinically as a watery, secretory, and fasting diarrhea associated with hypokalemia from fecal K^+ losses of up to 400 mEq/day. By inhibiting gastric secretion, VIPomas also cause achlorhydria or, more commonly, hypochlorhydria. Cross-reactivity between VIP and the hepatic glucagon receptor may account for the glucose intolerance and the hyperglycemia observed, respectively, in 50 and 18% of affected patients, whereas cross-reactivity with the PTH receptor may induce hypercalcemia in 25–75% of patients. Vasodilation promoted by VIP may cause an initial flushing and rash on the head and upper trunk in 20% of VIPoma patients; desensitization may subsequently reduce these symptoms.

Even though the VIPoma syndrome is a well-established entity, because of its rarity, it is still recognized only after an average duration of symptoms of 3 years. This diagnosis should be entertained for cases of secretory diarrhea productive of liters of isotonic watery stools that persist after 2–3 days of fasting. The diagnosis is supported by a fasting plasma VIP > 200 pg/ml.

VIPomas can be localized by conventional and endoscopic ultrasound, computed tomography scanning, angiography, and magnetic resonance. Because all neuroendocrine tumors (except insulinomas) express high-affinity somatostatin receptors, nuclear scanning with labeled octreotide, a stable somatostatin analogue, can effectively detect these tumors. Indeed, the experience with gastrinomas, representative of pancreatic neuroendocrine tumors, suggests that octreotide scanning is the single most sensitive method for imaging VIPomas, equal to all other imaging modalities combined (respective sensitivities of 58 and 70% for primary and metastatic tumors).

The most effective medical treatment for diarrhea is octreotide (100–150 μg subcutaneous every 8 h), which acts both on tumor cells to inhibit the synthesis and release of VIP and on intestinal crypt cells to inhibit secretion. Because of this dual effect, reduction of diarrhea (rather than normalization of VIP levels) may be a reasonable end-point. Adverse effects from octreotide include pain at the site of injection, imbalance of glucose metabolism, pancreatic exocrine insufficiency with steatorrhea, inhibition of gallbladder contraction, and gallstone formation.

Other agents advocated for the treatment of this syndrome include α_2-adrenergic agonists (e.g., clonidine and lidamidine), trifluoperazine, indomethacin, lithium, and metoclopramide. In addition, streptozotocin is a useful chemotherapeutic agent. An immunoglobulin G against VIP, which hydrolyzes VIP *in vitro*, may also have therapeutic potential.

For benign VIPomas, surgery after tumor localization will be curative. For malignant tumors, surgical debulking alone may be helpful in 40% of cases.

VIP Scanning

VIP receptors are present not only in many tissue types, but also on tumors derived from these tissues. $VPAC_1$ receptors are expressed by carcinomas of the breast, prostate, pancreas, lung, colon, stomach, liver, and bladder, as well as by lymphomas and meningiomas, whereas $VPAC_2$ receptors are expressed by leiomyomas. VIP receptors can therefore be used to detect these tumors. Indeed, in patients with various GI cancers (colorectal, pancreatic, gastric, carcinoid tumors, and insulinomas), most primary lesions and associated metastases are detected by scintigraphy with ^{123}I-VIP. The high quality of the positive scans reflects the abundance of VIP receptors on these tumors relative to normal tissue and blood cells, except for the lung

where background VIP uptake is high. The false-negative scans are probably caused by the absence of VIP receptors or their blockade by endogenous ligands. Transient hypotension is the only side effect observed with VIP scanning. Development of ^{99}Tc-labeled VIP may further enhance this scanning technique.

Vasoactive Intestinal Peptide in Cirrhosis

Because VIP is extracted by the liver and induces vasodilation, it has been postulated that, in cirrhosis, hepatic clearance of VIP is impaired, allowing VIP to produce some of the circulatory disturbances associated with this condition. In animal models and patients with cirrhosis, significant, but modest, increases in plasma VIP concentrations are observed. However, the circulatory effects of these increased VIP levels may be minimized by a blunted cardiovascular response. Indeed, although VIP infusion increased the portal tributary blood flow in normal rats, it did not affect any hemodynamic parameter in rats with cirrhosis caused by bile duct ligation.

SUMMARY AND CONCLUSION

Following its discovery in 1970, it was soon recognized that VIP stimulates epithelial cell secretion as well as smooth muscle cell relaxation and that VIP-producing tumors can cause a syndrome of watery diarrhea, hypokalemia, and achlorhydria. Additional roles for VIP in modulating inflammation, immune response, and cellular growth and differentiation are also being clarified. Furthermore, scintigraphy with radiolabeled VIP is being developed as a tool to detect GI tumors and clarifying the structure of VIP receptors and understanding their interaction with VIP will facilitate the design of beneficial agonists and antagonists. The relevance of VIP to GI physiology and pathophysiology is therefore continually expanding.

Acknowledgments

This work was partially supported by National Institutes of Health Grant RO1 DK 55885, VA Merit Review funding, and Cystic Fibrosis Foundation RDP R565. The author is indebted to Eric Nguyen for gathering source materials and Dr. Gabriel Makhlouf for clarifying the effects of VIP on smooth muscle cells.

See Also the Following Articles

Glucose-Dependent Insulinotropic Polypeptide (GIP) • Growth Hormone • Nitric Oxide • Pituitary Adenylate Cyclase Activating Peptide (PACAP) • Secretin • Vipoma

Further Reading

Bloom, S. R., Polack, J. M., and Pearse, A. G. E. (1973). Vasoactive intestinal peptide and watery diarrhea syndrome. *Lancet* II, 14–16.

Costa, M., and Furness, J. B. (1983). The origins, pathways and terminations of neurons with VIP-like immunoreactivity in the guinea-pig small intestine. *Neuroscience* **8**, 665–676.

Couvineau, A., Rouyer-Fessard, C., Darmoul, D., Maoret, J.-J., Carrero, I., Ogier-Denis, E., and Laburthe, M. (1994). Human intestinal VIP receptor: Cloning and functional expression of two cDNA encoding proteins with different N-terminal domains. *Biochem. Biophys. Res. Commun.* **200**, 769–776.

Delgado, M., and Ganea, D. (2001). Vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide inhibit nuclear factor-κB-dependent monocytic cell line THP-1. *J. Biol. Chem.* **276**, 369–380.

Harmar, A. J., Arimura, A., Gozes, I., Journot, L., Laburthe, M., Pisegna, J. R., Rawlings, S. R., Robberecht, P., Said, S. I., Sreedharan, S. P., and Washek, J. A. (1998). International Union of Pharmacology. XVIII. Nomenclature of receptors for vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide. *Pharmacol. Rev.* **50**, 265–270.

Itoh, N., Obata, K., Yanaihara, N., and Okamoto, H. (1983). Human preprovasoactive intestinal polypeptide contains a novel PHI-27-like peptide, PHM-27. *Nature* **304**, 547–549.

Lee, S. L., Huang, M., Ma, Z., and Rorstad, O. (1996). Vasoactive intestinal peptide in cirrhotic rats: Hemodynamic effects and mesenteric arterial receptor characteristics. *Hepatology* **23**, 1174–1180.

Murthy, K. S., Teng, B.-Q., Jin, J.-G., and Makhlouf, G. M. (1998). G protein-dependent activation of smooth muscle eNOS via natriuretic peptide clearance receptor. *Am. J. Physiol.* **275**, C1409–C1416.

Park, S. K., O'Dorisio, M. S., and O'Dorisio, T. M. (1996). Vasoactive intestinal polypeptide-secreting tumours: Biology and therapy. *Bailliere Clin. Gastroenterol* **10**, 673–696.

Said, S. I., and Mutt, V. (1970). Polypeptide with broad biological activity: Isolation from small intestine. *Science* **168**, 1217–1218.

Sreedharan, S. P., Huang, J.-X., Cheung, M.-C., and Goetzl, E. J. (1995). Structure, expression, and chromosomal localization of type I human vasoactive intestinal peptide receptor gene. *Proc. Natl. Acad. Sci. USA* **92**, 2939–2943.

Svoboda, M., Tastenoy, M., Van Rampelbergh, J., Goossens, J.-F., De Neef, P., Waelbroeck, M., and Robberecht, P. (1994). Molecular cloning and functional characterization of a human VIP receptor from SUP-T1 lymphoblasts. *Biochem. Biophys. Res. Commun.* **205**, 1617–1624.

Verner, J. V., and Morrison, A. B. (1974). Endocrine pancreatic islet disease with diarrhea *Arch. Intern. Med.* **133**, 492–500.

Verner, J. V. J., and Morrison, A. B. (1958). Islet cell tumor and a syndrome of refractory watery diarrhea and hypokalemia. *Am. J. Med.* **25**, 374–380.

Virgolini, I., Raderer, M., Kurtaran, A., Angelberger, P., Banyai, S., Yang, Q., Li, S., Banyai, M., Pidlich, J., Niederle, B., Scheithauer, W., and Valent, P. (1994). Vasoactive intestinal peptide-receptor imaging for the localization of intestinal adenocarcinomas and endocrine tumors. *N. Engl. J. Med.* **331**, 1116–1121.

Vipoma

MARK FELDMAN
University of Texas Southwestern Medical School and Presbyterian Hospital of Dallas

multiple endocrine neoplasia type 1 An autosomal dominantly inherited disorder associated with endocrine tumors of the parathyroid, pituitary, and endocrine pancreas, as well as in some patients with tumors of other endocrine glands and skin manifestations.

non-beta islet cell tumors Pancreatic endocrine tumors arising from cells other than the insulin-producing beta cells of the islets of Langerhans.

vipoma An endocrine tumor, usually of the pancreas, that produces vasoactive intestinal peptide.

Vipomas are rare tumors, with an annual incidence of 1 per 2 million or less. They are derived from neuroendocrine cells that produce and secrete vasoactive intestinal polypeptide (VIP). Most (80–90%) vipomas arise in the pancreas, are large (several centimeters), and are solitary. Rarely, they may arise from intestinal neuroendocrine tumors, bronchial carcinomas, pheochromocytomas, ganglioneuromas, or ganglioneuroblastomas. A small percentage of vipomas are associated with multiple endocrine neoplasia type 1 (MEN-1) syndrome, with prior, coexistent, or subsequent development of hyperparathyroidism, pituitary tumors, adrenal tumors, and/or thyroid tumors. The majority of vipomas are malignant.

Unlike most pancreatic tumors, vipomas do not usually cause symptoms despite their location in the pancreas. Thus, symptoms associated with other pancreatic tumors, such as pain in the epigastrium or back, nausea, vomiting, and jaundice, are conspicuously absent in these patients. Instead, symptoms in patients with vipomas derive from the actions of excessive amounts of VIP released into the portal circulation and ultimately into the systemic circulation.

PHYSIOLOGY

Vasoactive intestinal peptide (VIP) is a polypeptide with structural homologies to other peptides such as glucagon and secretin. Though originally described as a vasoactive intestinal-derived peptide that inhibits gastric acid secretion, VIP under physiologic conditions is produced mainly by enteric or central nervous system neurons and thus is a neuropeptide. The physiologic effects of VIP acting as a neuropeptide in the gut and the brain are not completely understood, although current evidence suggests that VIP acts in the gut as an inhibitory neurotransmitter that induces smooth muscle relaxation. For example, VIP in postganglionic esophageal neurons has been proposed to be the noncholinergic, nonadrenergic neurotransmitter that relaxes the lower esophageal sphincter, although nitric oxide has also been suggested to play this role. VIP is also a potent vasodilator.

CLINICAL FEATURES

Non-beta islet cell tumors of the pancreas may produce large amounts of peptides normally produced in the pancreas, such as glucagon and somatostatin, or peptides not normally produced in the pancreatic islets, such as gastrin or VIP. Vipomas are also called Verner-Morrison syndrome or the watery diarrhea hypokalemia hypochlorhydria syndrome, the latter of which describes some of the cardinal features of the disease. VIP, like cholera toxin and the heat-labile toxin of *Escherichia coli*, activates adenylate cyclase and increases concentrations of cyclic AMP within intestinal epithelial cells, and this increase has the effect of blocking NaCl absorption by these enterocytes. In addition, high concentrations of VIP, like the above-mentioned bacterial toxins, can stimulate intestinal chloride secretion. The consequence of reduced salt absorption and chloride secretion by the gut is a watery (nonbloody) diarrhea that contains large amounts of potassium and bicarbonate. The diarrhea in patients with vipomas can be prodigious and lead to life-threatening volume dehydration, hypotension or shock, hypokalemia, hyperchloremic metabolic acidosis, and hyponatremia. In fact, vipoma is sometimes referred to as pancreatic cholera syndrome because of its similarities to cholera. Classically, diarrhea in vipoma patients continues even when the patient fasts, although feeding can exacerbate the diarrhea due to the additional fluid and electrolytes entering the gut with a meal. Other symptoms (besides diarrhea) manifested by some patients with vipoma include weight loss, abdominal cramps, and flushing; the last symptom is a consequence of the vasoactive properties of VIP. Although high circulating levels of VIP also reduce gastric acid secretion in many of these patients, this

 Copyright 2004, Elsevier (USA). All rights reserved.

typically has no clinical consequences. Hypercalcemia and/or hyperglycemia may also be present in up to half of patients with vipomas.

DIAGNOSIS

Diagnosis of vipomas is difficult due to the rarity of the syndrome and many diseases that can produce chronic diarrhea. Furthermore, radioimmunoassay measurement of plasma VIP levels is performed only in certain specialized laboratories. One suggested approach is to measure the plasma VIP level in a reference laboratory only after an exhaustive search for other, more common causes of chronic diarrhea has been made (e.g., celiac sprue, Crohn's disease, laxative abuse). If plasma VIP is elevated (>190 pg/ml in most laboratories), a search for a tumor should be initiated since the test is highly specific. Assay sensitivity is fairly high (close to 90%), so repeating the test for VIP levels, if normal, has a finite but low utility. Probably the most useful test for localizing the tumor is endoscopic ultrasonography, but this test is invasive and expensive and requires considerable technical expertise. Radioactive scintigraphy using labeled octreotide is more widely available and is noninvasive. Abdominal computed tomography (CT) scan is also a good initial test because it is noninvasive and widely available. Most vipomas are large and therefore are evident on CT scan.

STAGING AND THERAPY

Once the diagnosis of vipoma has been established, the extent of the tumor should be addressed. If there is evidence of spread to the liver, which occurs at the time of diagnosis in approximately 40–80% of cases, palliative therapy of the patient's symptoms caused by the tumor is indicated. Palliative therapy to control symptoms primarily consists of long-acting somatostatin analogues that effectively control the diarrhea in most patients. Glucocorticoids may be effective if the patient fails to respond to a somatostatin analogue. There is some evidence that chemotherapy with a variety of agents including streptozotocin, adriamycin, 5-fluorouracil, interferon-α, and etoposide may benefit a few patients. Furthermore, liver transplantation has been utilized to treat metastatic disease to the liver refractory to systemic chemotherapy.

If there is no evidence of spread of the tumor beyond the pancreas on CT scan and/or endoscopic ultrasonography, or only to local lymph nodes within the projected resection specimen, the patient should be referred to an experienced pancreatic or gastrointestinal surgeon for an attempt at cure. Cure of vipoma is possible in approximately one-third of patients referred for surgery.

See Also the Following Articles

Multiple Endocrine Neoplasia (MEN) • Pancreatic Tumors, Other • Vasoactive Intestinal Peptide (VIP)

Further Reading

Cellier, C., Yaghi, C., Cuillerier, E., *et al.* (2000). Metastatic jejunal VIPoma: Beneficial effect of combination therapy with interferon-α and 5-fluorouracil. *Am. J. Gastroenterol.* **95**, 289–293.

Hengst, K., Nashan, B., Avenhaus, W., *et al.* (1998). Metastatic pancreatic VIPoma: Deteriorating clinical course and successful treatment by liver transplantation. *Z. Gastroenterol.* **36**, 239–245.

Jensen, R. T., and Norton, J. A. (2002). Pancreatic endocrine tumors. *In* "Gastrointestinal and Liver Disease" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th Ed., pp. 988–1016. W. B. Saunders, Philadelphia, PA.

Masel, S. L., Brennan, B. A., Turner, J. H., *et al.* (2000). Pancreatic vasoactive intestinal polypeptide-oma as a cause of secretory diarrhea. *J. Gastroenterol. Hepatol.* **15**, 457–460.

Yeh, C. N., Chen, M. F., and Chen, T. C. (2001). Surgical treatment of pancreatic vasoactive intestinal polypeptide-secreting tumor: A case report. *Hepatogastroenterology* **48**, 421–423.

Virtual Colonoscopy

JUDY YEE
University of California, San Francisco

adenomatous polyp Histologic type of colon polyp that has an increased risk of becoming malignant; may not be associated with symptoms, but may cause bleeding leading to anemia.

computer-aided detection Use of software to assess and locate areas of concern in scanned images, alerting a radiologist to evaluate an area that may represent a potential lesion. Computers could be trained to analyze images and mark areas of concern, acting as a second observer.

multidetector CT scanner Newest technological advance in computer tomography scanners, with multiple (4, 8, 16, or 32) detector rows, allowing flexibility in slice thickness and faster scanning ability. Thinner slices increase diagnostic ability and may be used as the source data for a broad range of three-dimensional applications.

Virtual colonoscopy, or computer tomography colonography, is an imaging technique that combines volumetrically acquired helical computer tomography scan data with advanced graphical software to create two- and three-dimensional views of the colon. This minimally invasive technique is a potential new tool for colorectal cancer screening and, compared to conventional colonoscopy, it may be more easily tolerated by patients.

INTRODUCTION

The majority of colorectal cancers arise from precursor adenomatous polyps. Detection and removal of these polyps while they are still benign can decrease the incidence of colorectal cancer. Current tools used for colorectal cancer screening include the fecal occult blood test (FOBT), flexible sigmoidoscopy, air contrast barium enema, and colonoscopy. Fecal occult blood testing is widely available and is the least expensive test but has been found to prevent the fewest cancers because of suboptimal accuracy. Colonoscopy is very sensitive for detecting polyps and cancers, but is the most invasive and expensive. Computed tomography (CT) colonography promises to change the way that colorectal cancer screening is performed. Performance studies have found that virtual colonoscopy has similar sensitivity to conventional colonoscopy for the detection of clinically significant polyps (those measuring 10 mm or larger).

CT TECHNIQUE

Bowel cleansing is performed prior to the CT study because colonic segments containing residual stool or fluid will be poorly evaluated. Patients are required to ingest either an electrolyte lavage solution or saline cathartic starting the day before the CT scan. Oral intake is also limited to clear liquids or a low-residue diet on the day before the procedure. Colonic distension is obtained by placement of a small rectal tube and retrograde instillation of air or carbon dioxide. Compared to air, carbon dioxide is resorbed more rapidly through the colonic wall and may therefore be more comfortable for patients. The use of intravenous glucagon as an antispasmodic agent is controversial and is no longer employed by some investigators, who have not found a clear benefit.

Patients are scanned in two positions, typically supine and prone. Segments of the colon that are poorly evaluated in one position due to inadequate distension or residual material may be viewed in the opposing position, which often improves distension or uncovers areas due to shifting of residual material. Multidetector CT scanners can scan the entire colon is less than 20 seconds in each position. Low-dose radiation CT protocols are used with thin slice thickness. The acquired CT data are then transferred to a computer workstation, where the radiologist interprets the study. Two- and three-dimensional views are used in an integrated approach to detect polyps and cancer (Fig. 1). The three-dimensional images of the inside of the colon may be viewed as a movie loop and have an appearance similar to what is seen on an actual colonoscopy. Most radiologists use enlarged axial images for primary interpretation and will correlate with the three-dimensional endoluminal views only for problem solving, which has been found to decrease interpretation times from 40 minutes down to about 15 minutes. Areas of the colon with poor cleansing or suboptimal distension will limit the diagnostic ability of the CT (Fig. 2).

 Copyright 2004, Elsevier (USA). All rights reserved.

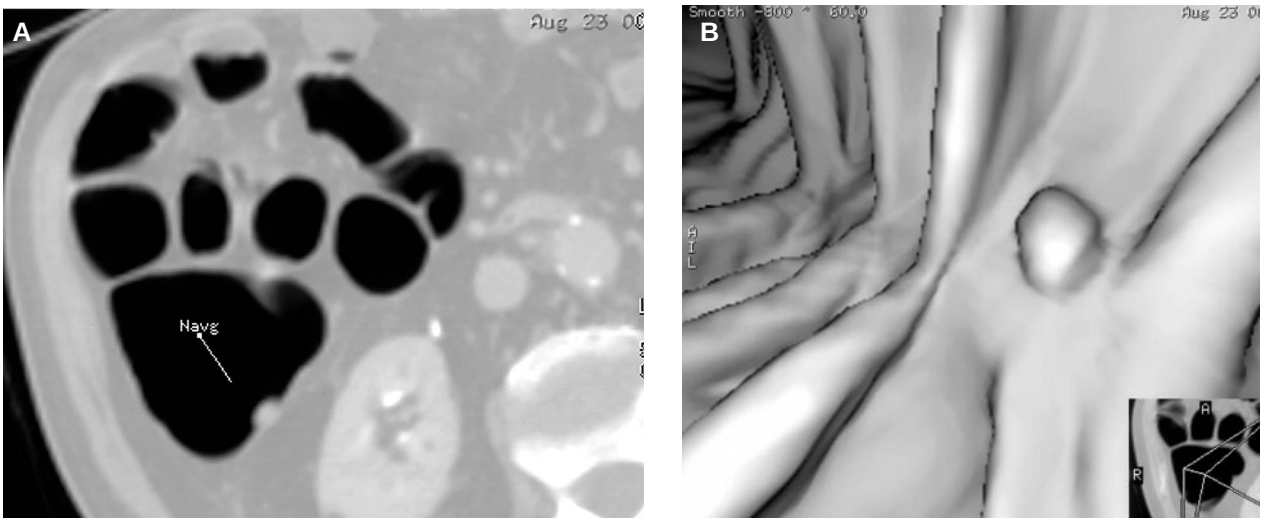


FIGURE 1 (A) Enlarged axial view of the ascending colon, showing a polyp located along the posterior wall. (B) Three dimensional endoluminal view demonstrates the appearance of the same ascending colon polyp.

PERFORMANCE DATA

Current studies to date have found that CT colonography has a sensitivity range of 83–100% and a specificity range of 93–100% for the detection of polyps 10 mm or larger when using a per patient matching scheme. When each polyp found on virtual colonoscopy is matched with each polyp found on colonoscopy, the per polyp sensitivity range is 75–100% for the detection of polyps 10 mm or larger. CT colonography has been found to have low sensitivity for small polyps (less than 5 mm) and for flat lesions. The majority of published studies have been performed using single-detector CT scanners and in high-risk or symptomatic patients. Confirmatory performance studies need to be performed using multidetector CT scanners and in asymptomatic or screening patients.

CT colonography can evaluate the proximal colon when fiber-optic colonoscopy is unsuccessful due to a distal obstructing lesion or colonic tortuosity. CT colonography is an excellent alternative test in patients who have contraindications for conventional colonoscopy.

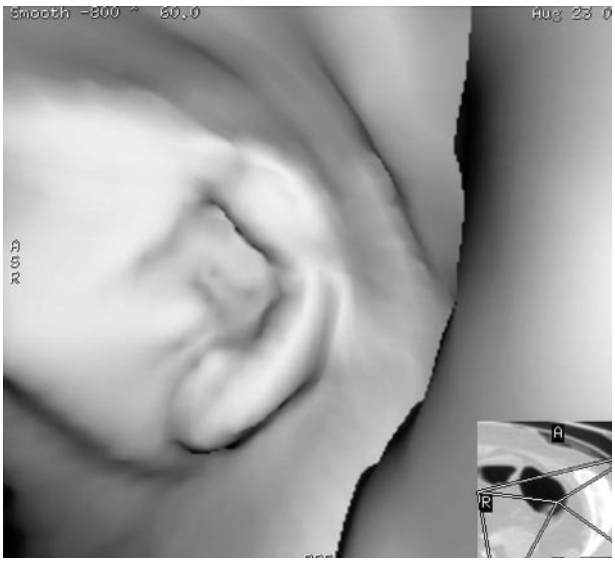


FIGURE 2 Appearance of an area of collapse on the three-dimensional endoluminal view. This imagery may simulate a carcinoma and must be correlated with the two-dimensional images.

ADVANTAGES OF CT COLONOGRAPHY

Colonoscopy has traditionally been considered the "gold standard" for evaluation of the colon. CT colonography is also a total colon examination technique (Fig. 3) and offers several advantages compared with traditional colonoscopy. CT colonography is a less invasive examination compared with colonoscopy and there is no need for intravenous sedation. The total procedure time for CT colonography is very short, taking less than 10 minutes for setup and complete scanning in two positions. In addition to being able to examine portions of the colon not seen when colonoscopy is unsuccessful, CT colonography can more accurately localize a lesion to a particular segment. CT colonography also has the ability to detect pathology

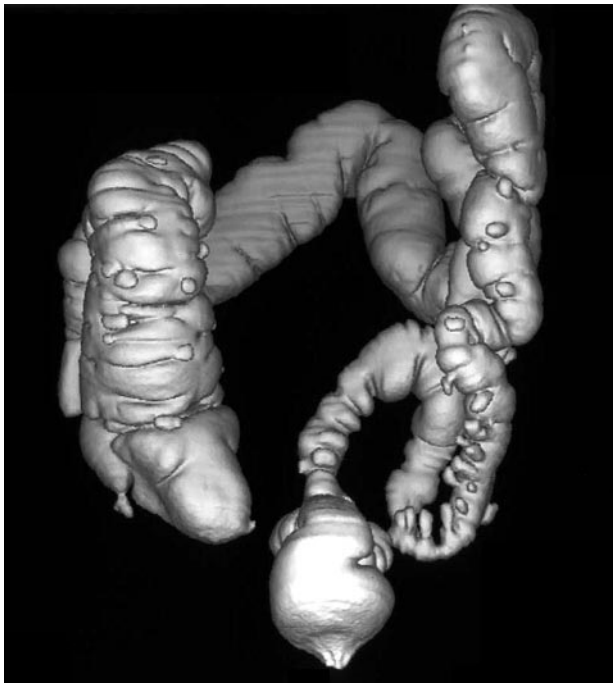


FIGURE 3 Three-dimensional external view of the entire colon in a patient with diffuse diverticulosis.

outside of the colon and can screen the remainder of the abdomen and pelvis without additional radiation.

LIMITATIONS OF CT COLONOGRAPHY

Patients are still required to restrict their diet and undergo a bowel cleansing regimen prior to the CT study, similar to colonoscopy, and this can be difficult to complete or may be perceived as an unpleasant experience. Patients may experience abdominal discomfort because of a distended colon during and immediately after the CT examination. The diagnostic ability of CT colonography examinations is dependent on adequate distension and cleansing of the colon. Therefore, detection rates are limited in the setting of suboptimal distension or poor colonic cleansing. Although CT colonography has excellent ability to detect large polyps and cancer, it has poor sensitivity for the detection of small polyps and flat lesions.

Radiologists need additional training to interpret CT colonography studies. Although the axial images are familiar to radiologists, the endoscopic view is a new visualization technique that requires training. A standardized cost needs to be established for CT colonography; the cost must be significantly lower than colonoscopy in order to make this technique available to the greatest number of patients for colorectal cancer screening.

FUTURE DIRECTIONS

Areas of research include the evaluation of alternative displays that allow the radiologist to view larger areas of the colonic surface at one time. A "virtual pathology" view splits the colon along its longitudinal axis, opening the colon so that it may be inspected like a surgical pathologic specimen. Maplike projections of the colon are also under investigation. The accuracy of these novel visualization methods needs to be determined. Computer-aided detection (CAD) of colorectal lesions is also being studied as a way to shorten interpretation times. Automated polyp detection computer software is under development. The applicable computer algorithms are based on the differences in curvature between polyps and folds or stool.

Fecal and fluid tagging protocols are being evaluated as a way to eliminate the need for cathartic bowel cleansing. Patients ingest positive oral contrast that tags residual material in the colon. This can then be combined with specific computer software that can electronically subtract the labeled material, leaving the colon effectively cleansed.

See Also the Following Articles

Colonic Ulcers • Colonoscopy • Colorectal Cancer Screening • Colorectal Adenocarcinoma • Colorectal Adenomas • Computed Tomography (CT) • Sigmoidoscopy

Further Reading

Beaulieu, C. F., Jeffrey, Jr., R. B., Karadi, C., Paik, D. S., and Napel, S. (1999). Display modes for CT colonography. Part II. *Radiology* **212**, 203–212.

Dachman, A. H., Kuniyoshi, J. K., Boyle, C. M., *et al.* (1998). CT colonography with three-dimensional problem solving for detection of colonic polyps. *AJR Am. J. Roentgenol.* **171**, 989–995.

Fenlon, H. M., Nunes, D. P., Schroy, P. C., Barish, M. A., Clarke, P. D., and Ferrucci, J. T. (1999). A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. *N. Engl. J. Med.* **341**, 1496–1503.

Ferruci, J. T. (2001). Colon cancer screening with virtual colonoscopy: Promise, polyps, politics. *AJR Am. J. Roentgenol.* **177**, 975–988.

Fletcher, J. G., Johnson, C. D., Welch, T. J., MacCarty, R. L., Ahlquist, D. A., Reed, J. E., Harmsen, W. S., and Wilson, L. A. (2000). Optimization of CT colonography technique: Prospective trial in 180 patients. *Radiology* **216**, 704–711.

Hara, A. K., Johnson, C. D., MacCarty, R. L., and Welch, T. J. (2000). Incidental extracolonic findings at CT colonography. *Radiology* **215**, 353–357.

Hara, A. K., Johnson, C. D., MacCarty, R. L., Welch, T. J., McCollough, C. H., and Harmsen, W. S. (2001). CT colonography: Single-versus multi-detector row imaging. *Radiology* **219**, 461–465.

Morrin, M. M., Kruskal, J. B., Farrell, R. J., Goldberg, S. N., McGee, J. B., and Raptopoulos, V. (1999). Endoluminal CT colonography after an incomplete endoscopic colonoscopy. *AJR Am. J. Roentgenol.* **172**, 913–918.

Summers, R. M., Beaulieu, C. F., Pusanik, L. M., Malley, J. D., Jeffrey, R. B., Glazer, D. I., and Napel, S. (2000). Automated polyp detector for CT colonography: Feasibility study. *Radiology* **216**, 284–290.

Yee, J., Akerkar, G. A., Hung, R. K., Steinauer-Gebauer, A. M., Wall, S. D., and McQuaid, K. R. (2001). Colorectal neoplasia: Performance characteristics of CT colonography for detection in 300 patients. *Radiology* **219**, 685–692.

Zalis, M. E., and Hahn, P. F. (2001). Digital subtraction bowel cleansing in CT colonography. *AJR Am. J. Roentgenol.* **176**, 646–648.

Vitamin A: Absorption, Metabolism, and Deficiency

MARC S. LEVIN

Washington University, St. Louis, Missouri and St. Louis VA Medical Center

provitamin A carotenoids A subset of carotenoids, which are a class of compounds synthesized by plants and microorganisms generally containing eight isoprenoid units, that can be oxidatively metabolized to produce retinal and retinoic acid and cleavage products known as apocarotenoids.

retinoids Naturally occurring compounds with vitamin A activity and synthetic analogues with or without vitamin A activity. These include analogues such as isotretinoin (13-*cis*-retinoic acid) that are used medically to treat acne.

retinyl esters Fatty acid conjugates of retinol that are the primary storage form of vitamin A in animal tissues.

vitamin A Fat-soluble substances that possess the biological properties of the prototypic vitamin A compound, all-*trans*-retinol, which is an unsaturated isoprenoid alcohol with five conjugated all-*trans* double bonds. Other important naturally occurring examples include retinal and retinyl esters. Dietary retinol and retinyl esters are referred to as preformed vitamin A.

Vitamin A is an essential nutrient that is required for normal growth, epithelial differentiation, fetal development, vertebrate morphogenesis, spermatogenesis, night vision, and a variety of other functions. Vitamin A also has an immunoregulatory role. Many of the actions of vitamin A are mediated by nuclear retinoic acid receptors that can bind to all-*trans*- or 9-*cis*-retinoic acid and act as transcriptional regulators to modulate gene expression. Although all-*trans*- and 9-*cis*-retinoic acid are vitamin A metabolites that mediate most of the biological actions of vitamin A, they are not sufficient for normal spermatogenesis or vision. There is great interest in the mechanisms of vitamin A action, the mechanisms of vitamin A absorption, and factors affecting its bioavailability because vitamin A deficiency remains a major global problem and because retinoids have tremendous potential as therapeutic and chemotherapeutic agents.

DIETARY SOURCES OF VITAMIN A

Animal Sources

Preformed vitamin A (principally retinyl palmitate and smaller quantities of other retinyl esters including stearate, myristate, and oleate) and very small quantities of provitamin A carotenoids are obtained from animal sources. Particularly rich source of preformed vitamin A include liver, whole milk, egg yolks, and fish.

Plant Sources

Dietary vitamin A is also obtained from plant sources as provitamin A carotenoids. Of the carotenoids, β-carotene is the most abundant. It is efficiently

 Copyright 2004, Elsevier (USA). All rights reserved.

Hara, A. K., Johnson, C. D., MacCarty, R. L., Welch, T. J., McCollough, C. H., and Harmsen, W. S. (2001). CT colonography: Single-versus multi-detector row imaging. *Radiology* **219**, 461–465.

Morrin, M. M., Kruskal, J. B., Farrell, R. J., Goldberg, S. N., McGee, J. B., and Raptopoulos, V. (1999). Endoluminal CT colonography after an incomplete endoscopic colonoscopy. *AJR Am. J. Roentgenol.* **172**, 913–918.

Summers, R. M., Beaulieu, C. F., Pusanik, L. M., Malley, J. D., Jeffrey, R. B., Glazer, D. I., and Napel, S. (2000). Automated polyp detector for CT colonography: Feasibility study. *Radiology* **216**, 284–290.

Yee, J., Akerkar, G. A., Hung, R. K., Steinauer-Gebauer, A. M., Wall, S. D., and McQuaid, K. R. (2001). Colorectal neoplasia: Performance characteristics of CT colonography for detection in 300 patients. *Radiology* **219**, 685–692.

Zalis, M. E., and Hahn, P. F. (2001). Digital subtraction bowel cleansing in CT colonography. *AJR Am. J. Roentgenol.* **176**, 646–648.

Vitamin A: Absorption, Metabolism, and Deficiency

MARC S. LEVIN

Washington University, St. Louis, Missouri and St. Louis VA Medical Center

provitamin A carotenoids A subset of carotenoids, which are a class of compounds synthesized by plants and microorganisms generally containing eight isoprenoid units, that can be oxidatively metabolized to produce retinal and retinoic acid and cleavage products known as apocarotenoids.

retinoids Naturally occurring compounds with vitamin A activity and synthetic analogues with or without vitamin A activity. These include analogues such as isotretinoin (13-*cis*-retinoic acid) that are used medically to treat acne.

retinyl esters Fatty acid conjugates of retinol that are the primary storage form of vitamin A in animal tissues.

vitamin A Fat-soluble substances that possess the biological properties of the prototypic vitamin A compound, all-*trans*-retinol, which is an unsaturated isoprenoid alcohol with five conjugated all-*trans* double bonds. Other important naturally occurring examples include retinal and retinyl esters. Dietary retinol and retinyl esters are referred to as preformed vitamin A.

Vitamin A is an essential nutrient that is required for normal growth, epithelial differentiation, fetal development, vertebrate morphogenesis, spermatogenesis, night vision, and a variety of other functions. Vitamin A also has an immunoregulatory role. Many of the actions of vitamin A are mediated by nuclear retinoic acid receptors that can bind to all-*trans*- or 9-*cis*-retinoic acid and act as transcriptional regulators to modulate gene expression. Although all-*trans*- and 9-*cis*-retinoic acid are vitamin A metabolites that mediate most of the biological actions of vitamin A, they are not sufficient for normal spermatogenesis or vision. There is great interest in the mechanisms of vitamin A action, the mechanisms of vitamin A absorption, and factors affecting its bioavailability because vitamin A deficiency remains a major global problem and because retinoids have tremendous potential as therapeutic and chemotherapeutic agents.

DIETARY SOURCES OF VITAMIN A

Animal Sources

Preformed vitamin A (principally retinyl palmitate and smaller quantities of other retinyl esters including stearate, myristate, and oleate) and very small quantities of provitamin A carotenoids are obtained from animal sources. Particularly rich source of preformed vitamin A include liver, whole milk, egg yolks, and fish.

Plant Sources

Dietary vitamin A is also obtained from plant sources as provitamin A carotenoids. Of the carotenoids, β-carotene is the most abundant. It is efficiently

 Copyright 2004, Elsevier (USA). All rights reserved.

absorbed and more readily converted to retinol than other carotenoids. Good sources of provitamin carotenoids include carrots, spinach, and other dark-colored fruits and vegetables. These foods also contain carotenoids that cannot be converted to vitamin A but are also thought to have beneficial health effects (e.g., lutein and lycopene).

Vitamin A Activity of Foods

Retinol activity equivalents (RAE), which correct for the bioavailability of provitamin A carotenoids, are used to compare the vitamin A activity of foods. One RAE is defined as 1 μg of retinol, 2 μg of β-carotene in oil (i.e., as a supplement), 12 μg of β-carotene in food, or 24 μg of other provitamin A carotenoids in foods. One international unit (IU) is equivalent to 0.3 μg of retinol. The recommended daily dietary intakes (RDI) for vitamin A are 900 RAE (3000 IU) and 700 (2333 IU) RAE per day for adult men and nonlactating women, respectively. There is no independent RDI for carotenoids. The Daily Value (DV) of 5000 IU found on food labels in the United States was based on older recommendations. The percentage DV on food labels provides the percentage of the recommended DV (based on a 2000-calorie diet) that is supplied by one serving.

VITAMIN A ABSORPTION AND METABOLISM

Digestion and Lumenal Events

Retinyl esters and carotenoids like other water-insoluble lipids are partially released from food by proteolysis and emulsified in the stomach before entering the small intestine. Within the intestinal lumen, retinyl esters are hydrolyzed to retinol and free fatty acids by pancreatic triglyceride lipase, intestinal brush border phospholipase B, and additional retinyl ester hydrolases. With the exception of esterified carotenoids (e.g., xanthophylls), carotenoids are absorbed without prior metabolic conversion within the intestinal lumen. Bile salts are required for the hydrolysis of retinyl esters and for the formation of mixed micelles that facilitate the solubilization of retinol and carotenoids. Following solubilization, retinol is efficiently transported across the brush border at physiological concentrations by a saturable, passive carrier-mediated mechanism. Carotenoids and pharmacological amounts of retinol are absorbed by nonsaturable, non-carrier-mediated, passive mechanisms. Most retinoids are absorbed in the proximal small intestine. The composition of the diet has been shown to affect the bioavailability of vitamin A and carotenoids. With adequate fat intake, 70–90% of preformed vitamin A is absorbed. Carotenoid absorption is particularly dependent on dietary composition and ranges from 5 to 50%. The bioavailability of β-carotene is markedly increased by high-fat diets. The efficiency of vitamin A absorption is diminished in patients with protein energy malnutrition and may be adversely affected by deficiencies in vitamin E and zinc.

Intracellular Metabolism

Following absorption, retinol, which is poorly soluble in the aqueous cytosol, is bound to the abundant cytosolic cellular retinol-binding protein type II (CRBP II). Most of the CRBP II-bound retinol is reesterified with palmitic acid and other long-chain fatty acids catalyzed by microsomal phosphatidylcholine-retinol *O*-acyltransferase (also known as lecithin:retinol acyltransferase), whereas unbound retinol can be esterified by retinol *O*-fatty-acyltransferase (acyl-CoA-retinol acyltransferase). CRBP II-bound retinol is also a substrate for retinol dehydrogenases that catalyze the reversible oxidation of retinol to retinal, which in turn can be irreversibly oxidized by retinal dehydrogenases to retinoic acid. The small amounts of retinoic acid that are absorbed or synthesized in the intestine may undergo isomerization or glucuronidation. Most provitamin A carotenoids undergo oxidative cleavage to produce apo-carotenoids and retinal, which can be converted to retinoic acid or retinol. β-Carotene can be cleaved centrally to produce two molecules of retinal. In some species, small quantities of carotenoids are exported without modification (e.g., humans and ferrets but not rats and pigs).

Extraintestinal Metabolism

Most of the retinyl esters, carotenoids, and apocarotenoids are packaged in chylomicron particles for secretion into intestinal lacteals. Retinoic acid and retinoic acid glucuronidates are absorbed via the portal circulation. Metabolism of chylomicrons by plasma lipoprotein lipase in the capillary endothelium produces remnants enriched in vitamin A as well as cholesterol ester and phospholipids. In addition to the liver, which is the major destination and storage site for vitamin A, chylomicron remnants are delivered to the lungs, kidneys, adipose tissue, muscles, spleen, and bone marrow. Within hepatocytes, retinol is reesterified with long-chain fatty acids for storage in sinusoidal stellate cells. Mobilization of hepatic retinyl ester stores is regulated to maintain serum retinol levels within a narrow range. Retinol generated by hydrolysis of retinyl esters is secreted from hepatocytes or stellate cells bound in a

complex to serum retinol-binding protein and transthyretin for delivery to target tissues. Within target tissues, cellular retinol-binding proteins homologous to intestinal CRBP II direct the metabolism of retinol along the same metabolic pathways utilized in the intestine. Carotenoids are cleaved to produce retinal, which is predominantly reduced to retinol, with smaller amounts oxidized to form retinoic acid. The discovery that carotenoid cleavage enzymes are located in extraintestinal tissues including the eye, kidney, testes, brain, and liver indicates that carotenoids can be direct sources of vitamin A and retinoic acid in these tissues. The oxidative inactivation of retinoic acid is mediated by cytochrome P450-dependent enzymes present in the liver and other tissues including skin, lung, and brain. The catabolism of retinoids produces polar metabolites that are excreted into the bile and urine. Although retinoyl and retinyl glucuronides and other retinoids excreted in the bile can be reabsorbed by the small intestine, the physiological significance of the enterohepatic circulation is not known.

VITAMIN A DEFICIENCY

Disorders that cause generalized fat malabsorption such as chronic pancreatitis and cystic fibrosis and those that result in decreased intestinal absorptive area such as Crohn's disease can result in vitamin A malabsorption and deficiency. Vitamin A deficiency resulting from dietary insufficiency is a major problem in developing nations. Chronic alcoholism is a more frequent cause in industrialized nations. Night blindness is the most sensitive indicator of deficiency in humans. Hypovitaminosis A is also associated with growth retardation, conjunctival xerosis, corneal damage, and impaired immune responses (e.g., mortality from measles is significantly increased in deficient children). Subclinical vitamin A deficiency (i.e., decreased vitamin A stores without overt signs of deficiency) is more prevalent than outright deficiency and may increase susceptibility to intestinal and respiratory infections and impair bone growth in children.

VITAMIN A TOXICITY

Symptoms of acute vitamin A toxicity resulting from the ingestion over a short period of time of large quantities of preformed vitamin A ($>660{,}000$ IU) include nausea and vomiting, vertigo, headache, and blurred vision. Chronic ingestion of vitamin A supplements or fortified foods exceeding three times the RDI leads to hypervitaminosis A. Symptoms include bone and muscle pain, alopecia, anorexia, visual impairment, and hyperlipidemia. Hepatomegaly is common and, particularly in the setting of alcoholism, hepatotoxicity can lead to venoocclusive disease or cirrhosis. Osteoporotic fractures are more common in postmenopausal women who chronically ingest a diet high in preformed vitamin A. Retinoic acid and synthetic retinoids used for medicinal purposes are highly teratogenic during the first trimester. Hypercarotenemia resulting from excessive intake of carotenoids causes yellow discoloration of the skin but is not associated with known toxicities.

See Also the Following Articles

Dietary Reference Intakes (DRI): Concepts and Implementation • Malabsorption • Malnutrition

Further Reading

Blomhoff, R. (ed.) (1994). "Vitamin A in Health and Disease." Dekker, New York.

Food and Nutrition Board, Institute of Medicine (2002). Vitamin A. *In* "Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc," pp. 82–161. National Academy Press, Washington, DC. Available at http://books.nap.edu/books/0309072794/html/82.html.

Harrison, E. H., and Hussain, M. M. (2001). Mechanisms involved in the intestinal digestion and absorption of dietary vitamin A. *J. Nutr.* **131**, 1405–1408.

Levin, M. S. (1994). Intestinal absorption and metabolism of vitamin A. *In* "Physiology of the Gastrointestinal Tract" (L. R. Johnson, ed.), 3rd Ed., Vol. 2, pp. 1957–1978. Raven Press, New York.

Sporn, M. B., Roberts, A. B., and Goodman, D. W. (eds.) (1994). "The Retinoids: Biology, Chemistry and Medicine," 2nd Ed. Raven Press, New York.

Vitamin B_{12}: Absorption, Metabolism, and Deficiency

JEAN-LOUIS GUÉANT AND FARÈS NAMOUR
University of Nancy, France

cobalamin analogues Corrinoids free from any vitamin activity.
haptocorrin Corrinoid binding protein, also named R binder, present in the digestive tract and the blood; the functions are not clearly established and may include the clearance of cobalamin analogues from blood.
intrinsic factor Glycoprotein secreted by the stomach; binds specifically to cobalamin prior to the receptor-mediated uptake of the intrinsic factor–cobalamin complex in the ileum.
transcobalamin Specific blood transport protein that delivers cobalamin to cells by a receptor-mediated endocytosis.
vitamin B_{12} Micronutrient that belongs to the family of corrinoids; present in animal-derived alimentary tracts and synthesized only by microorganisms. B_{12} corresponds to several vitamers named cobalamins, including deoxyadenosylcobalamin, the cofactor of methylmalonyl-CoA mutase, and methylcobalamin, the cofactor of methionine synthase.

Vitamin B_{12} (a cobalamin) is synthesized by a wide variety of microorganisms, including telluric bacteria and bacteria from the rumen and the intestine. Foods of animal origin (meat, liver, kidney, eggs, milk, fish, and shellfish) contain vitamin B_{12}, which is the only organic molecule to incorporate an atom of cobalt and to contain a ribonucleotide with a 5,6-dimethylbenzimidazole base. As cofactors of methionine synthase and methylmalonyl-coenzyme A mutase, cobalamins play a key role in homocysteine metabolism, energy metabolism, and DNA replication. Severe cobalamin deficiency causes macrocytic and/or megaloblastic anemia with leukoneutropenia and thrombocytopenia, digestive mucosa atrophy, and neurological symptoms such as psychiatric, cognitive, and proprioceptive disorders, related in part to demyelination.

PROTEINS INVOLVED IN VITAMIN B_{12} ASSIMILATION AND METABOLISM

Proteins involved in assimilation and activity of the cobalamin vitamin B_{12} (Fig. 1) include intrinsic factor, intrinsic factor receptor, transcobalamin, transcobalamin receptors, and haptocorrin.

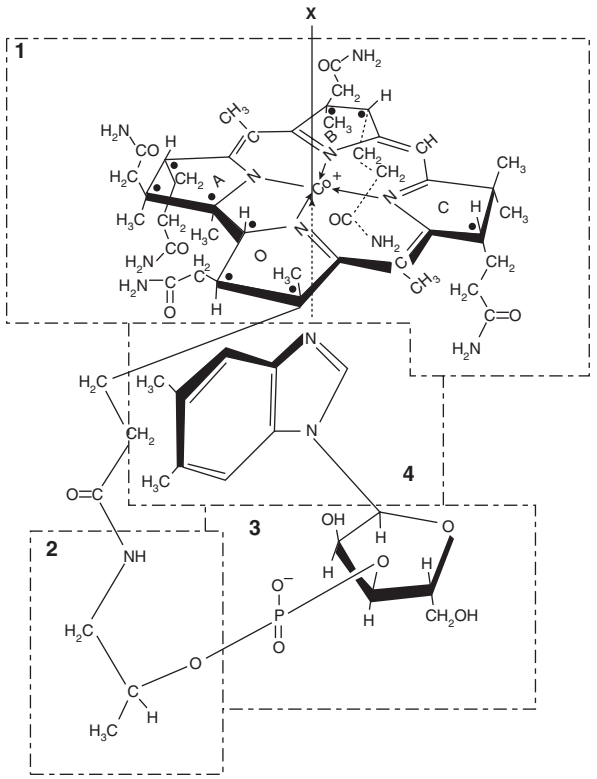


FIGURE 1 Cobalamin structure. (1) Tetrapyrrole group; (2) amino-1-propanol-2; (3) ribose-3′-phosphate group; (4) 5,6-dimethylbenzimidazole group; X represents a hydroxyl, methyl, 5′-desoxyadenosyl cyanide, or sulfuryl group.

Intrinsic Factor

Intrinsic factor (IF) is synthesized in humans by the parietal cells of the fundus (fornix; upper area of the stomach) and the gastric body (area of the stomach above the angulus). IF secretion is considerably greater than human daily requirements for IF and occurs by a process closely dependent on membrane translocation. The human gene encoding IF is situated in chromosome 11. The N-lactosaminic and O-glycosidic glycans of the carbohydrate core of IF constitute

 Copyright 2004, Elsevier (USA). All rights reserved.

9.2–15% of the molecule. Removing the carboxyl terminus of the molecule results in all cobalamin (Cbl) binding activity being lost; the amino terminus is important in the conformational changes required for receptor binding. It is possible that there exists an IF "pocket" that could accept the Cbl nucleotide and at least one pyrrole side chain.

Intrinsic Factor–Cobalamin Receptor

In the human intestine, the greatest amount of the intrinsic factor–cobalamin receptor (IFCR) seems to occur in the final 60 cm of the ileum. IFCR is also expressed in the kidney, in the adult and fetal intestine and colon, in the visceral yolk sac and the placenta, and in HT29 and Caco-2 adenocarcinoma cell lines. Calcium is involved in IFCR oligomerization and IF–Cbl binding is inhibited by wheat germ agglutinin. A 460-kDa multiligand protein named cubilin binds to IF–Cbl and also to apolipoprotein A-I (apoA-I), albumin, and immunoglobulin light chains. It has been concluded that cubilin and IFCR are identical, despite different physicochemical properties, which may be explained by a specific proteolytic cleavage in the intestine. Cubilin is a peripheral membrane protein without a clear membrane-spanning domain, suggesting the requirement for a third protein in the endocytosis mechanism.

Transcobalamin

Transcobalamin (TC) is a 43-kDa nonglycosylated protein synthesized in most tissues and encoded by a gene located on chromosome 22. It binds to Cbl with a high affinity and delivers blood Cbl to cells by receptor-mediated endocytosis. A low concentration of holo-TC in serum is an indicator of inadequate dietary intake or impaired assimilation. The synthesis of TC increases in inflammatory as well as in neoplastic disorders. A single nucleotide polymorphism in codon 259 affects the blood concentration of the two corresponding isoproteins and behaves as a genetic determinant of homocysteine. Alternative cleavage of the signal peptide also generates two TC isoforms.

Transcobalamin Receptors

The receptor monomer is present as a dimer in the plasma membrane and is up-regulated by cell proliferation. Uptake of TC–Cbl has been demonstrated in the liver, kidney, heart, spleen, intestine, and lung. A distinct receptor, megalin, is a 600-kDa multiligand protein expressed in the proximal tubules of the kidney and also in other absorptive epithelia. Receptor-mediated endocytosis of TC–Cbl by megalin may account for the accumulation of Cbl in the kidney.

Haptocorrin

The glycoprotein haptocorrin (HC) carries Cbl and other corrinoids. In the digestive tract, it is secreted in saliva, bile, and pancreatic fluid. It is also expressed in epithelial and glandular cells and granulocytes. Haptocorrin accounts for more than 50% of Cbl bound in gastric juice and 80% in serum. The haptocorrin carbohydrate core represents 30–40% of the molecule and protects part of the molecule from intraluminal proteolysis. The role of haptocorrin in Cbl metabolism is unclear. It may participate in regulating bacterial growth in secretion fluids and in clearing blood Cbl analogues in bile. Asialo-haptocorrin is involved in the liver and intestinal uptake of corrinoids via the receptor of asialoglycoproteins. Although lysosomes are the usual end point for the endocytic pathway for ligands bound to this receptor, HC escapes unaltered into bile by a transcytotic mechanism.

ASSIMILATION AND METABOLISM OF VITAMIN B_{12}

Intraluminal Stage

Cobalamin bound to dietary food products is released via sequential exposure to heat from cooking and to hydrochloric acid-dependent peptic digestion. The Schilling test, performed with orally ingested radiolabeled Cbl (incorporated in a dietary item such as trout or chicken), can be used to detect patients with chronic gastritis, who are unable to release Cbl from its food binders. In gastric juice, ingested Cbl is exposed to both IF and HC, the latter derived mainly from saliva. However, Cbl binds preferentially to HC, due to a higher affinity for this binding protein. Transfer of Cbl from HC to IF will occur in the intestinal lumen, where there is partial digestion of HC and neutralization of chyme. The intraluminal degradation of biliary HC is therefore a key step of an enterohepatic cycle that permits the selective reabsorption of biliary Cbl in the ileum and the elimination of Cbl analogues in the feces, in a form coupled with partially degraded HC.

Intestinal Parietal Transport

IF–Cbl uptake occurs in humans by a receptor-mediated endocytosis in the distal ileum. The receptor binding requires a neutral pH and the presence of calcium. Most of the studies on intestinal parietal transport (Fig. 2) have been performed with Caco-2 and HT29

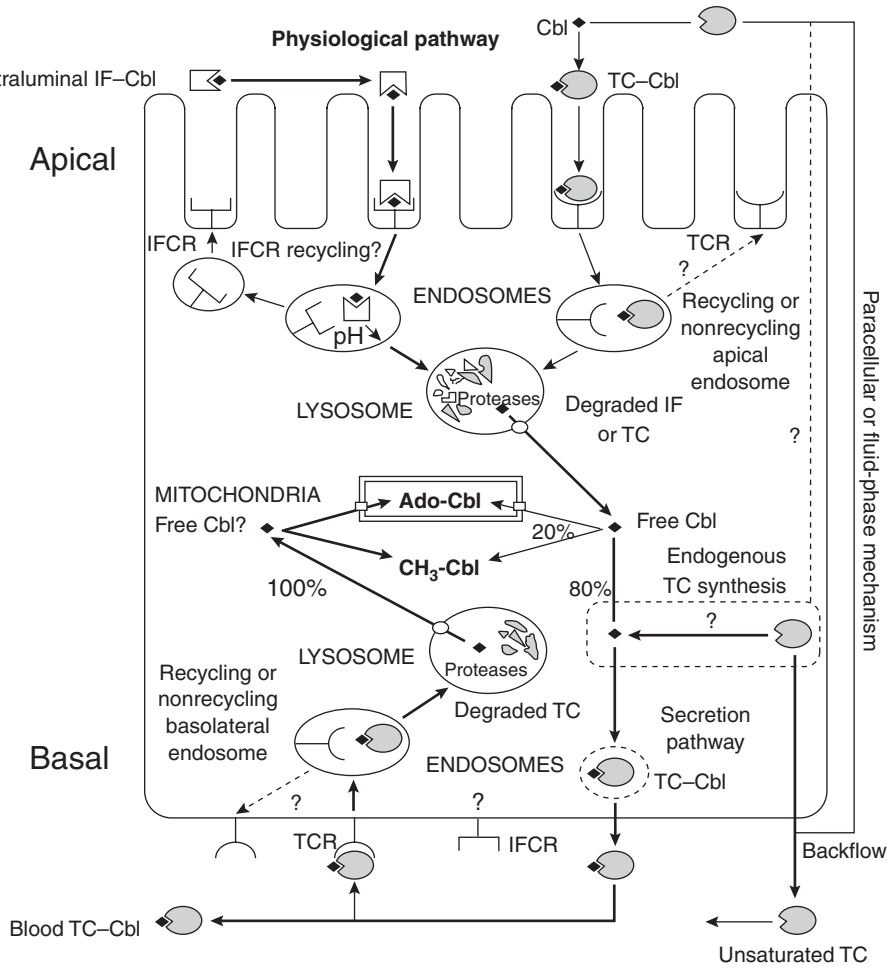


FIGURE 2 Intestinal parietal transport of vitamin B_{12}. The physiological pathway involves the apical intrinsic factor–cobalamin receptor (IFCR)-mediated endocytosis of intrinsic factor–cobalamin (IF–Cbl), release of cobalamin and transfer to endogenous transcobalamin (TC), and basolateral sorting of TC–cobalamin (TC–Cbl). Blood cobalamin can be internalized via a transcobalamin receptor (TCR)-mediated endocytosis of TC–Cbl, on the basolateral side and also on the apical side of Caco-2 cells.

cells. The IF–Cbl complex is dissociated from its apical receptor in early endosomes, under the effect of acid pH. IF is degraded by leupeptin-sensitive proteases such as cathepsin L and by glycosidases. Lysosomes play a key role in free Cbl release to the cytoplasm. Part of the apically transported dietary Cbl can be directly transformed into coenzymes by the enterocyte for its own metabolism. The Cbl sorting to the basolateral side requires the transfer of Cbl to TC in a still unknown cell compartment. The pathways of Cbl receptor-mediated endocytosis are similar in the apical and basolateral sides, except that IF and TC receptors are expressed mainly in the apical and the basolateral sides, respectively. In addition to the TC–Cbl receptor, megalin has been shown to bind to TC–Cbl and mediate its endocytosis. Colocalization of cubilin with megalin in both intestinal and renal epithelia suggests that megalin may be involved in the cell trafficking of IF–Cbl–cubilin. Because of its apical expression in the kidney proximal tubules, renal megalin may function in the tubular reabsorption of TC–Cbl, thus preventing Cbl loss in the urine.

Intracellular Metabolism

After endocytosis of blood TC–Cbl, the TC receptor is dissociated from the TC–Cbl complex in endosomes and TC is degraded and Cbl is released in lysosomes. In the cytoplasm, cobalamin undergoes reduction from cob(III)alamin to cob(II)alamin, which can then be

methylated to form CH_3-Cbl and assist in the methionine synthase conversion of homocysteine to methionine. Alternatively, cob(III)alamin or cob(II)alamin can enter the mitochondria, be further reduced to cob(I) alamin, adenosylated to Ado-Cbl, and participate in the conversion of methylmalonyl-CoA to succinyl-CoA by methylmalonyl-CoA mutase.

ETIOLOGIES OF VITAMIN B_{12} DEFICIENCY

Dietary B_{12} Deficiency

The required dietary allowance of vitamin B_{12} has been recently estimated to 2.4 μg/day in adults. An inadequate dietary intake of cobalamin can be observed in vegans and in breast-feeding infants of either vegan mothers or mothers with previously undiagnosed pernicious anemia. However, a vegetarian diet causing B_{12} deficiency is uncommon in western countries, because many foods are fortified with vitamin B_{12}.

Addison's Anemia and Chronic Atrophic Gastritis

Addison's anemia is a type A gastritis; it is also referred to as pernicious anemia or as Biermer's anemia in European countries. It usually occurs in patients aged 60 years or older, may be genetically determined, and is slightly more common in women than in men. Addison's anemia is frequently associated with other diseases, such as Hashimoto's thyroiditis, adrenocortical insufficiency, hemolytic anemia, primary biliary cirrhosis, diabetes, vitiligo, and hypogammaglobulinemia. In addition to megaloblastic anemia, biological data include a low serum Cbl concentration, an abnormal Schilling test that can be corrected in the presence of IF, a drastic reduction in gastric acid and IF secretion, and the presence of serum anti-IF antibodies (which is pathognomonic of the illness) and antiparietal cell autoantibodies in 60% and 80–90% of cases, respectively. There are two types of anti-IF autoantibodies, those that prevent Cbl–IF from binding to its receptor and those that block Cbl from IF binding. The autoantibodies recognize an area of IF that includes the sequence encompassing residues 251–265. A particular clinical form of pernicious anemia occurs in young patients, usually between the ages of 10 and 20 years, and is very often associated with autoimmune outbreaks and a family history.

Juvenile Pernicious Anemia

Juvenile pernicious anemia is a quite distinct entity. It occurs during the early years of life and may affect several members of a sibling group. Acidic secretion is normal. IF is absent from the gastric juice and there are no serum anti-IF autoantibodies. The presence of degradable IF in neutralized gastric juice has been described in two twins. In another case, IF could not bind to the ileal receptor.

Protein-Bound Cobalamin Malabsorption

Protein-bound cobalamin malabsorption is a type of deficiency that is sometimes difficult to distinguish from genuine Addison's anemia. It is caused partly by a failure to release dietary Cbl from its protein medium, even if IF secretion is sufficient to enable its absorption. Achlorhydria sometimes causes bacterial overgrowth, which accentuates Cbl malabsorption.

Gastrectomy

Total gastrectomy is inevitably accompanied by Cbl deficiency, because IF secretion is lacking. Cbl deficiency manifests when the body's reserves are exhausted, i.e., about 3–7 years after gastrectomy. Malabsorption is found in 19–30% of cases of partial gastrectomy. The pathogenesis involves a deficiency of IF and proton secretion, with malabsorption of protein-bound Cbl and blind-loop syndrome.

Exocrine Pancreatic Insufficiency

Malabsorption of crystalline Cbl may be partly the result of a lack of haptocorrin degradation due to inadequate secretion of pancreatic proteases. However, exocrine pancreatic insufficiency is rarely associated with a Cbl deficiency.

Obstructive Jaundice

The Schilling test reveals malabsorption of crystalline Cbl in about 40% of cases of obstructive jaundice and in about 50% of patients with external bile drainage. Interruption of bile drainage provides secondary correction. Bile appears to have an effect on binding of the Cbl–IF complex with the ileal receptor.

Bacterial Overgrowth

An imbalance of the intestinal flora, known as blind-loop syndrome, occurs in various pathological circumstances, including Finsterer- or Polya-type gastrectomies, segmentary intestinal resection with end-to-side anastomosis, ileocolic resection, inflammatory intestinal diseases, diverticulosis, or gastric achlorhydria. Bacteria possess binding sites that can trap Cbl after

bacterial proteases and glycosidases act to dissociate the IF–Cbl complex.

Tropical Sprue

The mechanism involved in B_{12} deficiency in tropical sprue may involve the intraluminal and parietal phases of Cbl assimilation. Cbl malabsortpion can be reversed by antibiotic treatment over a period of time.

Parasitic Infestations

The main parasites involved in B_{12} deficiency are *Diphyllobothrium latum* (causing diphyllobothriasis, or fish tapeworm) and *Giardia lamblia* (lambliasis); fish tapeworm infestation, however, is now only anecdotal. Infection is accompanied by a decrease in IF secretion and atrophic gastritis. The parasite traps Cbl bound with IF as well as free Cbl. In cases of lambliasis, malabsorption is primarily the result of Cbl being trapped by the parasite, but also the result of mucosal alteration.

Zollinger–Ellison Syndrome

In Zollinger–Ellison syndrome, hypergastrinemia is accompanied by hyperacidity and irreversible inactivation of pancreatic lipase in the duodenum, producing luminal lipase deficiency.

Malabsorption of Iatrogenic Origin

Long-term treatment with proton pomp inhibitors such as omeprazol may occasionally produce a Cbl deficiency related to inadequate assimilation of protein-bound Cbl. Cholestyramine is a resin that can chelate IF by ion exchange. Colchicine, alcohol, and antibiotics can act as inhibitors of IF–Cbl endocytosis. Occasionally, patients with undiagnosed vitamin B_{12} deficiency who have received nitrous oxide anesthesia may experience neurologic abnormalities in the weeks after the exposure.

Congenital Defects of B_{12} Assimilation

A small number of cases with congenital deficiency of IF have been reported. Mutations appear to be both structural and regulatory. One patient had a combined deficiency of IF and HC, suggesting a genetic linkage between regions located on chromosome 11 that encode the two proteins. Three siblings in one family had an abnormal IF that was markedly susceptible to gastric acid and to proteolytic enzymes. The selective intestinal malabsorption of Cbl, also known as the Gräsbeck–Imerslund syndrome or megaloblastic anemia 1 (MGA1), is a rare autosomal recessive disease, often associated with a proteinuria independent of Cbl deficiency. The pathogenesis may correspond to a defect of expression, stability, and/or trafficking of IFCR. Abnormal binding of IF–Cbl to the receptor has been found in ileum biopsies and urine assays of patients. Two cubilin gene mutations, FM1 and FM2, have been identified in some (not all) MGA1 patients.

Inherited Defects of B_{12} Blood Transport

Inherited TC deficiency manifests in the first few month of life. Most of the 30 known patients lack immunologically detectable TC. Two deletion mutations have been detected in the TC gene of one patient. Because of the role of TC in blood transport and ileal transcytosis of Cbl, its absence impairs Cbl absorption as well as cell delivery.

Inherited Defects of B_{12} Cellular Metabolism

Patients with deficiencies in intracellular cobalamin metabolism and utilization have been classified by fibroblast complementation analysis into eight groups. Patients who have defects in methionine synthase (MS) (*cblG*) or in the regeneration of methylcobalamin (MeCbl) (*cblE*) have homocystinuria and hyperhomocysteinemia. Those who have defects in the mutase (mut^0, mut^-) or in the formation of deoxyadenosylcobalamin (AdoCbl) (*cblA*, *cblB*) have methylmalonic aciduria and methylmalonic acidemia. Patients with blocks in the early steps of cobalamin cellular transport cannot synthesize MeCbl or AdoCbl (*cblC*, *cblD*, *cblF*), which results in a combined increase of homocysteine and methylmalonic acid.

TREATMENT OF VITAMIN B_{12} DEFICIENCY

Intramuscular injections of vitamin B_{12} (usually 1000 μg/month) have been the classic treatment for vitamin B_{12} deficiency. High-dose oral vitamin B_{12} pills and injections of B_{12} have the same efficiency for curing megaloblastic anemia and neurologic disorders. Patients can be managed with oral doses of 300–2000 μg/day.

See Also the Following Articles

Cobalamin Deficiency • Dietary Reference Intakes (DRI): Concepts and Implementation • Intrinsic Factor • Malnutrition • Pernicious Anemia

Further Reading

Alpers, D. H. (1999). Intrinsic factor, haptocorrin and their receptor. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 411–440. John Wiley & Sons, New York.

Christensen, E. I., and Birn, H. (2002). Megalin and cubilin: Multifunctional endocytic receptors. *Nat. Rev. Mol. Cell Biol.* **3**, 256–266.

Guéant, J. L. (1990). Malabsorption of cobalamins. *In* "Cobalamin and Related Binding Proteins in Clinical Nutrition" (J. L. Guéant and J. P. Nicolas, eds.), pp. 55–68. Elsevier, Amsterdam.

Pons, L., Guy, M., Lambert, D., Hatier, R., and Guéant, J. L (2000). Transcytosis and coenzymatic conversion of (^{57}Co) cobalamin bound to either endogenous transcobalamin II or exogenous intrinsic factor in Caco-2 cells. *Cell Physiol. Biochem.* **10**, 135–148.

Rosenblatt, D. S. (1999). Inborn errors of cobalamin metabolism. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 367–384. John Wiley & Sons, New York.

Rothenberg, S. P., Quadros, E. V., and Regec, A. (1999). Transcobalamin II. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 441–473. John Wiley & Sons, New York.

Seetharam, B. (1999). Receptor-mediated endocytosis of cobalamin (vitamin B12). *Annu. Rev. Nutr.* **19**, 173–195.

Vitamin K: Absorption, Metabolism, and Deficiency

CHALLA AJIT AND SANTIAGO J. MUNOZ
Albert Einstein Medical Center, Philadelphia

menaquinone Vitamin K_2.
phylloquinone Vitamin K_1.

Vitamin K was discovered more than 50 years ago following investigation of a hemorrhagic disease in cattle and chicks that was corrected by vitamin K dietary supplements. The vitamin's name stems from the German word *Koagulationsvitamin*. Vitamin K and its derivatives have a 2-methyl-1,4-naphthoquinone nucleus with a lipophilic side chain at position 3. This lipophilic chain gives vitamin K its unique properties. The most abundant form of vitamin K is phylloquinone (vitamin K_1), which is found in green leafy vegetables and certain plant oils. Another form, menaquinone (vitamin K_2), has several subtypes; some are synthesized by the intestinal microflora (subtypes MK-7, MK-8, MK-10, and MK-11) and other nutritionally significant variants (MK-4) occur only in meats (especially liver). In contrast to vitamin K_1, the contribution of vitamin K_2 to human nutrition is poorly understood.

ABSORPTION

Most of the absorption of vitamin K occurs in the distal small intestine. Dietary phylloquinone is protein bound and requires pancreatic enzymes for liberation into the lumen of the small intestine. The highly lipophilic cleaved products are solubilized into micelles by bile and are absorbed into the lymphatic circulation after being incorporated into chylomicrons. Ultimately, these chylomicrons enter the portal circulation and reach the liver by a carrier-dependent process. When the intraluminal levels of bile salts fall below the critical micellar concentration required for absorption, deficiency of vitamin K and other fat-soluble vitamins is likely to develop. Therefore, vitamin K absorption depends on the functional integrity of the liver, small intestine, and pancreas.

METABOLISM

In humans, vitamin K functions as a cofactor for the endoplasmic enzyme γ-glutamylcarboxylase. This enzyme is involved in a unique posttranslational carboxylation reaction, in which glutamate residues on various proteins are converted into γ-carboxyglutamate (Gla). This reaction occurs during the last stages of protein synthesis. The resulting Gla residues are characteristic

Copyright 2004, Elsevier (USA). All rights reserved.

Further Reading

Alpers, D. H. (1999). Intrinsic factor, haptocorrin and their receptor. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 411–440. John Wiley & Sons, New York.

Christensen, E. I., and Birn, H. (2002). Megalin and cubilin: Multifunctional endocytic receptors. *Nat. Rev. Mol. Cell Biol.* **3**, 256–266.

Guéant, J. L. (1990). Malabsorption of cobalamins. *In* "Cobalamin and Related Binding Proteins in Clinical Nutrition" (J. L. Guéant and J. P. Nicolas, eds.), pp. 55–68. Elsevier, Amsterdam.

Pons, L., Guy, M., Lambert, D., Hatier, R., and Guéant, J. L (2000). Transcytosis and coenzymatic conversion of (^{57}Co) cobalamin bound to either endogenous transcobalamin II or exogenous intrinsic factor in Caco-2 cells. *Cell Physiol. Biochem.* **10**, 135–148.

Rosenblatt, D. S. (1999). Inborn errors of cobalamin metabolism. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 367–384. John Wiley & Sons, New York.

Rothenberg, S. P., Quadros, E. V., and Regec, A. (1999). Transcobalamin II. *In* "Chemistry and Biochemistry of B12" (R. Banerjee, ed.), pp. 441–473. John Wiley & Sons, New York.

Seetharam, B. (1999). Receptor-mediated endocytosis of cobalamin (vitamin B12). *Annu. Rev. Nutr.* **19**, 173–195.

Vitamin K: Absorption, Metabolism, and Deficiency

CHALLA AJIT AND SANTIAGO J. MUNOZ
Albert Einstein Medical Center, Philadelphia

menaquinone Vitamin K_2.
phylloquinone Vitamin K_1.

Vitamin K was discovered more than 50 years ago following investigation of a hemorrhagic disease in cattle and chicks that was corrected by vitamin K dietary supplements. The vitamin's name stems from the German word *Koagulationsvitamin*. Vitamin K and its derivatives have a 2-methyl-1,4-naphthoquinone nucleus with a lipophilic side chain at position 3. This lipophilic chain gives vitamin K its unique properties. The most abundant form of vitamin K is phylloquinone (vitamin K_1), which is found in green leafy vegetables and certain plant oils. Another form, menaquinone (vitamin K_2), has several subtypes; some are synthesized by the intestinal microflora (subtypes MK-7, MK-8, MK-10, and MK-11) and other nutritionally significant variants (MK-4) occur only in meats (especially liver). In contrast to vitamin K_1, the contribution of vitamin K_2 to human nutrition is poorly understood.

ABSORPTION

Most of the absorption of vitamin K occurs in the distal small intestine. Dietary phylloquinone is protein bound and requires pancreatic enzymes for liberation into the lumen of the small intestine. The highly lipophilic cleaved products are solubilized into micelles by bile and are absorbed into the lymphatic circulation after being incorporated into chylomicrons. Ultimately, these chylomicrons enter the portal circulation and reach the liver by a carrier-dependent process. When the intraluminal levels of bile salts fall below the critical micellar concentration required for absorption, deficiency of vitamin K and other fat-soluble vitamins is likely to develop. Therefore, vitamin K absorption depends on the functional integrity of the liver, small intestine, and pancreas.

METABOLISM

In humans, vitamin K functions as a cofactor for the endoplasmic enzyme γ-glutamylcarboxylase. This enzyme is involved in a unique posttranslational carboxylation reaction, in which glutamate residues on various proteins are converted into γ-carboxyglutamate (Gla). This reaction occurs during the last stages of protein synthesis. The resulting Gla residues are characteristic

Copyright 2004, Elsevier (USA). All rights reserved.

of a limited number of proteins found in liver, bone, and blood vessels.

Vitamin K occurs naturally in the quinone oxidized state and is reduced to the hydroquinone form (vitamin KH_2) by the enzyme vitamin K epoxide reductase (VKOR). In turn, vitamin KH_2 becomes the active cofactor for the vitamin K-dependent enzyme γ-glutamylcarboxylase. Vitamin KH_2 is oxidized to vitamin K 2,3-epoxide (KO). Indeed, this reaction provides the energy driving the carboxylation reaction by the vitamin K-dependent γ-glutamylcarboxylase. Under normal conditions, for each molecule of Gla generated, one molecule of vitamin K epoxide is also formed. There is a strict one-to-one stoichiometric relation between the conversion of vitamin KH_2 to KO and the number of protein bound Gla residues formed.

After protein degradation, Gla residues are excreted in urine. It is clear that the number of carboxylation reactions far exceeds the number of vitamin K molecules available. The short-lived, highly reactive epoxide (KO) is potentially toxic and is recycled back to vitamin KH_2 by the enzyme VKOR. Warfarin inhibits VKOR, which leads to insufficient generation of vitamin KH_2 by interrupting the recycling of vitamin K_2 (see Fig. 1). This results in a vitamin K-depleted state, leading to clinical manifestations and raising the nutritional requirement. In such situations, a second enzyme, NAD(P)H dehydrogenase [also known as detoxifying (DT) diaphorase], reduces quinone to vitamin KH_2 but not the epoxide (KO). Interruption of this cycle leads to production of decarboxyglutamate. Measurement of proteins containing decarboxyglutamate gives an early indication of vitamin K deficiency.

Vitamin K plays a major role in the synthesis of clotting factors. The Gla-containing coagulation proteins promoting clotting include prothrombin and factors VII, IX, and X. Vitamin K is also involved in the biosynthesis of the antithrombotic proteins C and S. With the exception of protein S (produced in many tissues), all these proteins are exclusively synthesized in the liver.

Abnormal bleeding is a cardinal sign of vitamin K deficiency. The function of Gla residues on blood coagulation proteins is to facilitate adhesion to the negatively charged phospholipids on the surface of activated platelets. Gla residues, along with calcium, are internalized to the core, leading to exposure of the phospholipid binding domains.

Other important Gla-containing proteins, such as protein Z, osteocalcin, and matrix growth protein (MGP), are involved in equally important metabolic functions. Protein Z helps thrombus adherence to sites of injury. Osteocalcin is a product unique to bone tissue produced by osteoblasts, the deficiency of which may be related to development of osteoporosis. MGP play a crucial role in bone formation and in addition to protein S contributes to maintaining vascular stability. All Gla-containing proteins are synthesized in the endoplasmic reticulum in a propeptide form. The propeptide contains the γ-carboxylation recognition site characteristic of the vitamin K-dependent proteins, except in MGP, where this site resides in the mature protein.

DEFICIENCY

The dietary requirement for vitamin K is approximately 100−200 μg/day and the current recommended daily allowance is 65−80 μg/day, in addition to the estimate for what is produced by colonic synthesis. Vitamin K deficiency in an otherwise healthy adult is rare. Acquired deficiency can occur in patients on long-term antibiotics and parenteral nutrition or with gastrointestinal diseases characterized by maldigestion, malabsorption, hepatobiliary disease (diminished hepatocellular function or cholestasis), and other disease states. Salicylates, anticonvulsants, and large doses of vitamin E (>1200 IU/day) have been associated with

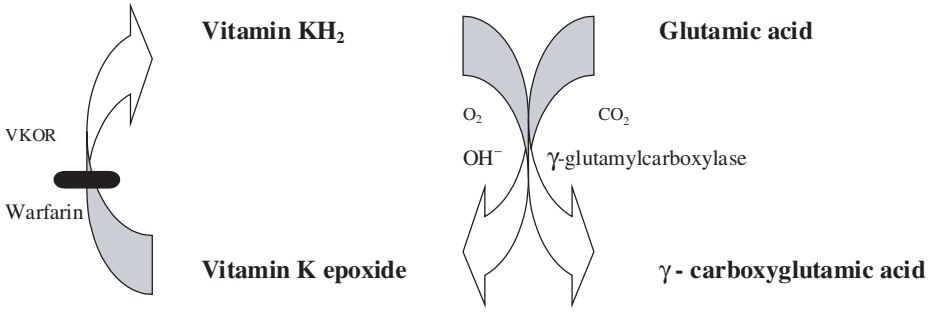


FIGURE 1 Biosynthetic pathway for vitamin K-dependent production of γ-carboxyglutamic acid. Warfarin inhibits the enzyme vitamin K epoxide reductase (VKOR).

vitamin K deficiency in some patients. A state of therapeutic deficiency is induced by administration of coumadin-based anticoagulants. An important acquired vitamin deficiency state is hemorrhagic disease of the newborn. The normal newborn has a moderate deficiency of vitamin K-dependent coagulation factors. Adult levels are achieved at about 3 months of age. Hemorrhagic disease of the newborn is now uncommon due to routine administration of vitamin K at birth.

Vitamin K status can be determined by various methods. Until recently, direct measurement of vitamin K was difficult. Plasma phylloquinone levels can now be measured directly by high-pressure liquid chromatography. Plasma levels of vitamin K primarily reflect dietary intake over the previous 24 hours. In addition to direct measurement of plasma phylloquinone level, vitamin K status can be measured by indirect methods. A widely used indicator is the determination of the prothrombin time (PT), which becomes prolonged in vitamin K deficiency. PT is nonspecific for vitamin K deficiency because it also becomes prolonged in hepatocellular dysfunction and certain hematological disorders (coagulopathies). The specificity of PT can be increased by the so-called vitamin K test. In this test, reversal of prolonged PT by parenteral vitamin K administration confirms vitamin K deficiency caused by inadequate dietary intake or intestinal malabsorption. Improvement of the PT after parenteral vitamin K is less consistent with coagulopathy of parenchymal hepatocellular disease. Improvement of PT is noted within 8 hours, with a maximum improvement at 48 hours. Absence of improvement or a minimal change in PT suggests investigating hepatocellular causes of prolonged PT.

In emergencies, vitamin K can be administered by slow intravenous injection, but there is risk of arterial hypotension and rarely anaphylaxis. Intramuscular injection is not generally recommended. The PT test can respond rapidly to changes in levels of factor VII, which has a half-life of only 4–7 hours. However, PT is considered relatively insensitive for detecting vitamin K deficiency because it becomes prolonged when factor VII levels fall below 40%. PT is even less sensitive to prothrombin concentration, becoming prolonged only when prothrombin levels fall below 30% of normal. Another indirect method to assess vitamin K status involves measurement of decarboxylated osteocalcin. This substance would appear a relatively sensitive indicator of vitamin K deficiency, because its affinity for γ-glutamylcarboxylase is lower than that of other vitamin K-dependent proteins. Therefore, detection of decarboxylated osteocalcin may provide an early indication of mild vitamin K deficiency.

See Also the Following Articles

Dietary Reference Intakes (DRI): Concepts and Implementation • Malnutrition • Parenteral Nutrition • Small Intestine, Absorption and Secretion

Further Reading

Booth, S. L., and Saltzman, E. (2001). *In* "Rakel: Conn's Current Therapy," 53rd Ed., pp. 609–611. W. B. Saunders, Philadelphia.

Food and Nutrition Board, Institute of Medicine (2001). Panel on micronutrients. Standing Committee on the Scientific Evaluation of Dietary Reference Intakes. Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. National Academy Press, Washington, D.C.

Furie, B., Bouchard, B. A., and Furie, B. C., (1999). Vitamin K-dependent biosynthesis of gamma carboxyglutamic acid. *Blood* **93**(6), 1798–1808.

Grosset, A. B. M., and Rodgers, G. M. (1998). *In* "Wintrobe's Clinical Hematology," 10th Ed., Vol. 2, pp. 1733–1767. Williams and Wilkins, Philadelphia.

Rutherford, C. J., and Frenkel, E. P. (1999). *In* "Diseases of Liver" (E. R. Schiff, M. F. Sorrell, and W. C. Maddrey, eds.), 8th Ed., Vol. 1, pp. 583–596. Lippincott-Raven, Philadelphia.

Veermer, C., and Schurgers, L. J. (2000). A comprehensive review of vitamin K and vitamin K antagonists. *Hematol. Oncol. Clin. North Am.* **14**(2), 339–353.

Volvulus

RICHARD H. TURNAGE, LESTER W. JOHNSON, AND QUYEN CHU

Louisiana State University Health Sciences Center and Overton Brooks Veterans Affairs Medical Center, Shreveport

closed-loop obstruction Type of mechanical obstruction in which both the proximal and distal segments of bowel are obstructed.

volvulus Abnormal twisting of a segment of bowel, resulting in a mechanical obstruction of the proximal and distal parts of the involved segment.

Volvulus is an important cause of acute intestinal obstruction in infants and of gastric and intestinal obstruction in adults. Volvulus, defined as an abnormal twisting of a segment of bowel, results in a mechanical obstruction of the proximal and distal segments of the involved bowel. The resultant closed-loop obstruction has a particularly high risk of strangulation, necrosis, and perforation. The anatomic requirements underlying the development of volvulus include a redundant segment of bowel that is freely moveable within the peritoneal cavity, with close approximation to the points of fixation of the bowel.

GASTRIC VOLVULUS

Anatomy and Etiology

The stomach is fixed inferiorly by the duodenum, and its normal orientation is maintained by the gastrosplenic, gastroduodenal, gastrophrenic, and gastrohepatic ligaments. About two-thirds of cases of gastric volvulus are associated with a congenital or acquired diaphragmatic hernia in which the stomach (or part of the stomach) is positioned within the chest. One-third of cases of gastric volvulus occur below the diaphragm and are due to laxity of the ligamentous attachments of the stomach to the duodenum, liver, spleen, and diaphragm. Sliding paraesophageal hernias are not associated with gastric volvulus because the stomach maintains its normal configuration as part of it ascends into the posterior mediastinum.

Gastric volvulus may be classified anatomically as organoaxial or mesenteroaxial, depending on the axis on which the volvulus occurs. Most cases of gastric volvulus are organoaxial in nature, with the stomach twisting on its long axis. This axis most commonly passes through the gastroesophageal and the gastropyloric junctions, with the antrum rotating anteriorly and superiorly and the fundus rotating posteriorly and inferiorly, causing torsion of the greater curvature at a point along its length. This type of gastric volvulus is most often associated with diaphragmatic hernias and causes gastric ischemia and infarction in 5–28% of cases. Mesenteroaxial gastric volvulus occurs when the stomach folds on an axis extending from the lesser to the greater curvature such that the antrum twists anteriorly and superiorly. This type of volvulus occurs in about one-third of cases and is more likely to be incomplete, intermittent, and associated with chronic symptoms.

Clinical Manifestations

Patients with acute gastric volvulus experience sudden, severe upper abdominal or chest pain. In those cases in which the stomach is located within the abdomen, there is upper abdominal distension. Patients with herniation of the stomach through a diaphragmatic defect present with severe chest pain radiating to the arms and neck with accompanying dyspnea, misleading the unsuspecting clinician into believing the patient may be suffering acute myocardial infarction. Persistent retching productive of little or no vomitus is a common manifestation. The triad of abdominal or chest pain, violent retching, and the inability to pass a nasogastric tube should lead to a strong clinical suspicion of acute gastric volvulus. Patients with chronic gastric volvulus may have mild, nonspecific, or intermittent symptoms such as epigastric discomfort, heartburn, abdominal fullness, or bloating. A large, unusual gas-filled viscus in the chest or upper abdomen on plain abdominal radiograph suggests the diagnosis, which may be confirmed by barium-swallow radiographic study or computed tomography.

Treatment

Surgical therapy for patients presenting with acute gastric volvulus, after initial resuscitation, involves laparotomy with operative reduction, resection of necrotic tissue, and repair of precipitating conditions such as a

 Copyright 2004, Elsevier (USA). All rights reserved.

hiatal hernia. On occasion, a nasogastric tube may be placed to allow resuscitation and an elective repair, but more frequently a tube cannot be placed and immediate operation is mandatory. Patients presenting with chronic recurrent volvulus should have a careful preoperative identification of precipitating conditions followed by operative repair. In those cases in which primary volvulus occurs without obvious cause, gastropexy may be performed by tacking the anterior wall of the stomach to the parietal peritoneum of the abdominal wall.

MIDGUT VOLVULUS

The most common form of small bowel volvulus is midgut volvulus.

Etiology

Midgut volvulus is an important cause of intestinal obstruction in neonates and infants and is caused by incomplete (or absent) rotation of the embryonic intestine during the fifth to twelfth week of gestation. From 50 to 75% of cases are discovered in the first month of life and 90% of cases are detected before 1 year of age.

The most common abnormality associated with midgut volvulus is nonrotation, in which there is inadequate counterclockwise rotation of the midgut loop around the superior mesenteric artery (SMA). This causes the duodenojejunal junction to be located to the right of the midline, as is the remainder of the small intestine. The colon resides in the left abdomen with the cecum near the midline. A narrow mesenteric pedicle predisposes to volvulus, and the peritoneal attachments (Ladd's bands), which extend anterior and lateral to the duodenum to fix the cecum to the posterior body wall, may obstruct the duodenum.

Clinical Presentation

The discovery of malrotation during infancy occurs when duodenal obstruction from Ladd's bands or midgut volvulus develops. Infants with duodenal obstruction from Ladd's bands present with proximal duodenal obstruction and bilious vomiting with little abdominal distension. Midgut volvulus may also present with obstructive symptoms; however, the development of intestinal ischemia due to strangulation may cause transmural necrosis with acidemia, thrombocytopenia, and sepsis.

Diagnosis

The high risk of intestinal ischemia and necrosis from midgut volvulus and the associated very high mortality rate mandate aggressive diagnosis and management of neonatal intestinal obstruction. A plain abdominal radiograph in babies with midgut volvulus will demonstrate a distended stomach and proximal duodenal bulb with a paucity of small bowel gas. An upper gastrointestinal contrast study will be diagnostic by demonstrating malpositioning of the duodenojejunal junction to the right of the midline, with the small intestine on the right and the cecum and ascending colon to the left. The contrast study may also demonstrate a characteristic corkscrew or coiled appearance in the third or fourth portions of the duodenum. A barium enema demonstrating the cecum in the right lower quadrant does not exclude the possibility of malrotation and midgut volvulus.

Treatment

The treatment of intestinal malrotation, whether manifested by duodenal obstruction from Ladd's bands or midgut volvulus, is operative. In the latter case, the diagnosis should be followed by laparotomy because a delay of even hours may mean the difference between viable or gangrenous bowel. Operative repair of malrotation is achieved by the Ladd procedure, which consists of relief of the volvulus and division of the peritoneal bands tethering the cecum, small bowel mesentery, mesocolon, and duodenum around the base of the superior mesenteric artery. This allows the mesenteric leaves to open widely and is associated with a very low incidence of recurrent volvulus. Meticulous and complete mobilization of the entire duodenum with division of all anterior, lateral, and posterior peritoneal attachments relieves the extrinsic compression and obstruction of the distal duodenum.

COLONIC VOLVULUS

Volvulus causes about 10–15% of all colonic obstructions in the United States and other Western countries and about 1–4% of all cases of intestinal obstruction. In Eastern Europe and parts of Africa and Asia, colonic volvulus accounts for 20–50% of all intestinal obstructions. It is the second most common cause of colonic obstruction, the most common being adenocarcinoma. The sigmoid colon and cecum are the most frequent sites of colonic volvulus, accounting for about 75 and 22% of all cases, respectively. Rare sites for colonic volvulus include the transverse colon (2%) and splenic flexure (<1%).

Etiology and Pathophysiology

In each case, a freely mobile segment of intra-abdominal colon twists or folds on fixed afferent and efferent limbs of the bowel, causing a closed-loop obstruction. As fluid and gas accumulate within the involved segment of bowel, intraluminal pressure rises, and when it exceeds capillary pressure, the colonic wall becomes ischemic. Closed-loop obstructions, such as that occurring with colonic volvulus, are associated with a high rate of ischemia, infarction, and perforation.

Clinical Presentation

The most common clinical manifestations of colonic volvulus are acute abdominal distension and lower abdominal pain. When present, the pain varies from a vague discomfort accompanying abdominal distension to the excruciating pain of peritonitis. Severe unremitting pain suggests gangrenous bowel and peritonitis. The duration of symptoms in patients with sigmoid volvulus is often less than that in patients with malignant or benign strictures. Abdominal tenderness occurs in less than one-third of patients with colonic volvulus. Patients with sigmoid volvulus are often in the seventh or eighth decade of life and often have various comorbid illnesses. A history of chronic constipation and laxative use is also a frequent finding in patients with cecal or sigmoid volvulus. Patients with cecal volvulus tend to be younger than patients with sigmoid volvulus and have a history of prior abdominal operations or distal obstruction. The most common physical finding is massive abdominal distension, which is always present. Significant abdominal tenderness with evidence of peritonitis suggests impending or actual colonic necrosis and perforation and mandates emergent operative intervention.

Diagnostic Evaluation

The initial diagnostic approach to patients suspected of having colonic obstruction includes plain abdominal radiographs in the supine and upright positions. The abdominal radiograph of patients with sigmoid volvulus will demonstrate a markedly dilated sigmoid colon and proximal bowel with minimal gas in the rectum. The standard radiographic appearance of sigmoid volvulus is a distended ahaustral sigmoid loop, i.e., "bent inner-tube" appearance; the apex of which is directed toward the right shoulder. The classic radiographic features of cecal volvulus include (1) a massively dilated cecum located in the epigastrium or left upper quadrant, (2) a kidney bean shape of the distended cecum, (3) distended loops of small bowel, suggesting small bowel obstruction, and (4) a single, long air–fluid level present on upright or decubitus films. In these instances, the massively distended cecum extends across the abdominal midline and is "directed" toward the left upper quadrant or left midabdomen. These "classic" radiographic findings are seen in 40–60% of cases.

The remaining diagnostic studies in patients with colonic obstruction are predicated on the presence or absence of peritonitis and the degree of obstruction (i.e., partial or complete). Patients with peritonitis should undergo resuscitation and urgent laparotomy without further diagnostic procedures, whereas patients without evidence of bowel wall ischemia and an abdominal radiograph suggestive of a distal complete obstruction should undergo proctosigmoidoscopy. This procedure will demonstrate the site and nature of distal strictures and in the case of sigmoid volvulus may allow decompression.

If the obstruction is proximal to the area visualized by proctosigmoidoscopy, a water-soluble contrast enema will confirm the diagnosis of colonic obstruction and delineate the site of obstruction. In patients with sigmoid or cecal volvulus and an equivocal plain abdominal radiograph, a water-soluble contrast enema may be helpful by demonstrating a point of torsion (e.g., a mucosal spiral pattern, or "bird's beak" sign). The use of water-soluble contrast media obviates the risk of barium impaction at the site of obstruction and barium peritonitis in the case of unrecognized perforation.

Although colonoscopy may be useful in patients with partial colonic obstruction, it has little role in the initial evaluation of patients suspected of having complete obstruction. The insufflation of air or carbon dioxide through the endoscope into the obstructed bowel may exacerbate colonic distension and precipitate perforation.

Treatment and Outcome

Resuscitation of patients with colonic obstruction includes restoration of intravascular volume, correction of electrolyte abnormalities, and nasogastric aspiration. The urgency with which the obstruction must be decompressed is dependent on the degree of obstruction (partial or complete) and the clinical presentation of the patient (evidence of strangulation or not). As alluded to earlier, the initial management of patients with sigmoid volvulus without evidence of peritonitis is proctoscopic decompression of the obstruction, often assisted by placing a rectal tube into the obstructed bowel. In a compilation of 19 American series including 595 patients with sigmoid volvulus, Ballantyne reported that

proctoscopy, either alone or combined with a rectal tube, successfully reduced the volvulus in 70–80% of attempts. The placement of a rectal tube for 48 hours may minimize the possibility of early recurrence. Successful reduction of sigmoid volvulus has also been reported with flexible sigmoidoscopy or colonoscopy; however, the procedure must be performed with limited manipulation and minimal insufflation of air or carbon dioxide.

The risk of recurrence following nonoperative reduction of a sigmoid volvulus is 40–50%. Thus, following proctoscopic decompression, the patient should undergo mechanical cleansing of the bowel, followed by elective sigmoid resection. This multistep approach allows time for optimization of the patient's condition prior to laparotomy, resection, and the performance of a primary anastomosis. Recurrence rates following this approach are less than 3%. Patients requiring emergent laparotomy for strangulated sigmoid volvulus require sigmoid resection with an end colostomy and Hartmann pouch.

The role of initial nonoperative management of patients with cecal volvulus is less well defined than that of sigmoid volvulus. Although colonoscopy has been successfully employed to reduce the volvulus, the risk of perforation of the thinned, often ischemic, cecum is substantial, as is the danger of missing a segment of necrotic bowel with delay of definitive resection. Current options for the operative management of cecal volvulus include cecopexy, cecostomy, and resection. Detorsion alone or when combined with appendectomy is associated with a high recurrence rate. In some cases, the performance of a cecopexy, in which the right colon is anchored to the peritoneum of the right paracolic gutter with or without a cecostomy, is favored. Right colectomy with primary ileotransverse colostomy effectively prevents recurrent volvulus and is the procedure of choice of a number of surgeons, including the authors. Ballantyne reported 27 patients with cecal volvulus who were treated with resection and primary anastomosis with no operative mortality and no recurrence of volvulus in 5 years of followup.

Overall, the mortality rate for patients with colonic volvulus is about 8%, with the major predictive factor for mortality being the presence of gangrenous bowel. The incidence of gangrenous bowel in patients with either cecal or sigmoid volvulus is 15–20%. In a review of 18 American studies including 299 patients with sigmoid volvulus, the mortality rate for patients with gangrenous colons was 80%, whereas only 10% of patients without colonic necrosis died.

See Also the Following Articles

Colectomy • Colonic Obstruction • Duodenal Obstruction • Gastric Volvulus • Gastrostomy • Malrotation

Further Reading

Agrez, M., and Cameron, D. (1981). Radiology of sigmoid volvulus. *Dis. Colon Rectum* **24**, 510.

Anderson, J. R., and Welch, G. H. (1986). Acute volvulus of the right colon: An analysis of 69 patients. *World J. Surg.* **10**, 336.

Ballantyne, G. H. (1982). Review of sigmoid volvulus: History and results of treatment. *Dis. Colon Rectum* **25**, 494.

Ballantyne, G. H. (1982). Review of sigmoid volvulus: Clinical patterns and pathogenesis. *Dis. Colon Rectum* **25**, 823.

Ballantyne, G. H., Brandner, M. D., Beart, R. W., Jr., *et al.* (1985). Volvulus of the colon: Incidence and mortality. *Ann. Surg.* **202**, 83.

Brothers, T. E., Strodel, W. E., and Eckhauser, F. E. (1987). Endoscopy in colonic volvulus. *Ann. Surg.* **206**, 1.

Burke, J. B., and Ballantyne, G. H. (1984). Cecal volvulus: Low mortality at a city hospital. *Dis. Colon Rectum* **27**, 737.

Carter, R., Brewer, L. A., and Hinshaw, D. B. (1980). Acute gastric volvulus. *Am. J. Surg.* **140**, 99.

Grossman, E. M., Longo, W. E., Stratton, M. D., *et al.* (2000). Sigmoid volvulus in Department of Veterans Affairs Medical Centers. *Dis. Colon Rectum* **43**, 414.

Haskin, P. H., Teplick, S. K., Teplick, J. G., *et al.* (1981). Volvulus of the cecum and right colon. *J. Am. Med. Assoc.* **245**, 2433.

Turnage, R. H., and Bergen, P. (2002). Intestinal obstruction and ileus. *In* "Sleisinger's Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management," (M. Feldman, L. S. Friedman, and M. H. Sleisinger, eds.), 7th Ed., pp. 2113–2128. W. B. Saunders, Philadelphia.

Water-Soluble Vitamins: Absorption, Metabolism, and Deficiency

HAMID M. SAID
University of California, Irvine and VA Medical Center, Long Beach

adaptive regulation A process by which a transport event is up- or down-regulated by conditions in the surrounding environment, such as the level of a particular substrate in the diet or in the extracellular fluid.
carrier-mediated membrane transport Mechanism for transporting a substrate that involves a defined protein system located at the plasma membrane of the cell.
ontogenic regulation A process describing developmental changes, for example, in a transport event.
transport by passive diffusion A process by which a solute moves from one side of the membrane to the other without the need for a specialized transporting system.
vitamin suboptimal level Level of the vitamin that falls below the normal range.

The water-soluble vitamins are a structurally unrelated group of organic compounds that share the same features of being essential for normal cellular functions, growth, and development and exist in minute quantities in the diet. Humans and other mammals obtain these micronutrients from exogenous sources (e.g., diet) via intestinal absorption. Thus, the intestine plays a central role in the regulation of normal body homeostasis of these nutrients. As a result, deficiency and suboptimal levels of these nutrients occur in a variety of disease conditions and states that affect the intestine. This includes both inherited defects in the intestinal transport systems involved and secondary causes such as intestinal diseases (e.g., inflammatory bowel disease, celiac disease), excessive alcohol intake, drug interactions, intestinal resection, and aging. This article will provide a brief description of the metabolic role of these nutrients and the current understanding of their intestinal absorption mechanisms and regulation.

VITAMIN B1

Metabolic Role and Deficiency

Vitamin B1 (thiamine), in its pyrophosphate form, plays a critical role in normal carbohydrate metabolism, participating in the decarboxylation of pyruvic and α-ketoglutamic acids and in the utilization of pentose in the hexose monophosphate shunt. Thiamine deficiency in humans leads to a variety of clinical abnormalities including neurological (neuropathy and/or Wernicke-Korsakoff syndrome) and cardiovascular disorders (e.g., peripheral vasodilation, biventricular myocardial failure, edema, and potentially acute fulminant cardiovascular collapse). Thiamine deficiency represents a significant nutritional problem in both developed and underdeveloped countries. In developed countries, thiamine deficiency has been reported in a high percentage of alcoholics, in diabetics, in patients with renal diseases, acquired immunodeficiency syndrome, cancer, celiac disease, and congestive heart failure, and in those on long-term use of the diuretic furosemide. Thiamine deficiency also occurs in thiamine-responsive megaloblastic anemia due to a genetic defect in the thiamine transport system SLC19A2. Furthermore, deficiency of thiamine has been reported in the elderly despite an average daily intake of the vitamin that exceeds the recommended requirement.

Intestinal Thiamine Absorption

Humans and other mammals cannot synthesize thiamine and thus must obtain the vitamin from exogenous sources via intestinal absorption. The intestine is exposed to two sources of thiamine: a dietary source and a bacterial source (the latter is via synthesis of the vitamin by the normal microflora of the large intestine). Previous studies using a variety of human and animal intestinal preparations have established the involvement of a specialized, carrier-mediated mechanism for thiamine uptake in the small intestine. This includes studies with intestinal biopsies, surgical specimens, purified jejunal brush border membrane vesicles isolated from human organ donors, and cultured human-derived intestinal epithelial cell lines. The identified uptake system was found to transport the vitamin via a proton gradient-dependent exchange mechanism. A similar carrier-mediated thiamine uptake mechanism was also found at the apical membrane of human colonocytes and is believed to be responsible for the uptake

 Copyright 2004, Elsevier (USA). All rights reserved.

of thiamine that is produced by the normal microflora of the large intestine. Following transport into the cell, thiamine is phosphorylated by the action of the cytoplasmic thiamine pyrophosphokinase, a process probably aimed at increasing the retention of the vitamin. It then exits the cell across the intestinal basolateral membrane as free (unphosphorylated) thiamine. The mechanism of transport of thiamine across the human basolateral membrane also occurs via a specialized carrier-mediated mechanism. One of the factors that have been shown to interfere with the normal intestinal thiamine uptake process *in vivo* is excessive alcohol intake. Also, the intestinal thiamine uptake process has been shown to be sensitive to the inhibitory effect of the diuretic amiloride.

Molecular Identity of the Intestinal Thiamine Transport System(s)

Insight into the molecular identity of the intestinal thiamine transport process has begun to emerge in recent years with the cloning of two human thiamine transporters, SLC19A2 and SLC19A3. Both of these transporters were found to be expressed in the human intestine. In addition, the *SLC19A2* promoter appears to be very active following transfection into intestinal epithelial Caco-2 cells. Although these studies point to the possible involvement of these transporters in intestinal thiamine absorption, further studies are required to establish their relative contribution.

Regulation of Intestinal Thiamine Uptake

Studies with human-derived cultured intestinal epithelial cells have shown that the thiamine uptake process is under the regulation of an intracellular Ca^{2+}–calmodulin-mediated pathway. Also, thiamine deficiency has been reported to up-regulate thiamine uptake by human intestinal biopsy specimens.

VITAMIN B2

Metabolic Role and Deficiency

Vitamin B2 (riboflavin), in its coenzyme forms, riboflavin-5′-phosphate (FMN) and flavin adenine dinucleotide (FAD), is involved in many critical metabolic reactions including amino acid carbohydrate and lipid metabolism and in the conversion of folic acid and pyridoxine into their coenzyme forms. Deficiency of riboflavin in humans leads to a variety of clinical abnormalities including degenerative changes in the nervous system, endocrine dysfunction, anemia, and skin disorders. Riboflavin deficiency and suboptimal levels have been reported in different conditions including alcoholism, patients with diabetes mellitus, inflammatory bowel disease (IBD), or human immunodeficiency virus infection and those receiving chemotherapy.

Intestinal Riboflavin Absorption

Humans and other mammals have lost the ability to synthesize riboflavin and thus must obtain riboflavin from exogenous sources via intestinal absorption. The intestine is exposed to riboflavin from two sources: the diet and the bacterially synthesized vitamin in the large intestine. Dietary riboflavin exists mainly in the form of FMN and FAD. These forms are hydrolyzed to free riboflavin in the intestinal lumen prior to absorption, a process performed by the action of intestinal phosphatases. The mechanism of absorption of dietary riboflavin has been extensively studied using a variety of human and animal intestinal preparations. These studies have shown the proximal part of the small intestine to be the preferential site of riboflavin absorption. Also, a specialized, carrier-mediated mechanism was found to be involved in riboflavin uptake. The amount of bacterially synthesized riboflavin in the large intestine varies depending on the diet, being higher following consumption of a vegetable-based diet than of a meat-based diet. In addition, the large intestine has been shown to be capable of absorbing a significant amount of lumenally introduced riboflavin. The mechanism of riboflavin absorption in the colon has been studied with cultured human colonic epithelial cells and has been shown to occur via a carrier-mediated system that is similar to that observed in the small intestine. Factors that have been shown to interfere with the normal intestinal riboflavin uptake process include alcohol and certain tricyclic drugs, such as chlorpromazine. Nothing is currently known about the molecular identity of the intestinal riboflavin uptake system.

Regulation of Intestinal Riboflavin Uptake

The intestinal riboflavin uptake process has been shown to be adaptively regulated by extracellular substrate levels and by specific intracellular protein kinase-mediated regulatory pathways. Oversupplementation with pharmacological amounts of riboflavin has been shown to lead to a significant and specific down-regulation in riboflavin uptake. On the other hand, riboflavin deficiency causes a significant and specific up-regulation in intestinal riboflavin uptake. The latter effect appears to be mediated via an increase in the

number (and/or activity) of the riboflavin uptake carriers with no changes in their affinity. The intestinal riboflavin uptake process was also found to be under the regulation of an intracellular PKA-mediated pathway. This pathway appears to exert its effect via a decrease in riboflavin carrier activity/number. Furthermore, the intestinal riboflavin uptake process was found to be ontogenically regulated during early stages of life.

VITAMIN B3

Metabolic Role and Deficiency

The main function of vitamin B3 (also known as niacin or nicotinic acid) is to act as a precursor for the synthesis of two important coenzymes: NAD and NADP. Both of these coenzymes are involved in maintaining the redox state of the cell. Most NAD- and NADP-linked enzymes are involved in catabolic reactions, such as glycolysis and the pentose phosphate shunt. In addition, niacin appears to have a beneficial effect as a lipid-lowering agent. Severe niacin deficiency occur in humans and leads to pellagra, a disease characterized by skin lesions, weight loss, diarrhea, inflammation of mucous membranes, vertigo, and mental confusion. Niacin deficiency occurs in alcoholics and in patients that carry mutations in the tryptophan transport gene (Hartnup's disease). Niacin deficiency occurs in the latter population because tryptophan is a precursor for nicotinic acid production in the body.

Intestinal Nicotinic Acid Absorption

The body obtains niacin from two sources: endogenously via conversion of tryptophan to niacin, a process that occurs mainly in the liver and kidneys, and exogenously from the diet via intestinal absorption. Limited studies regarding the cellular and molecular aspects of intestinal niacin absorption are available. Uptake studies have shown the involvement of a carrier-mediated, pH-dependent uptake system for nicotinic acid in mammalian intestine. The reported pH dependence of intestinal nicotinic acid uptake process is in contrast to the reported sodium dependency of the vitamin uptake process in renal epithelial cells. Some of the internalized nicotinic acid undergoes metabolism to intermediates of the Preiss-Handler pathway for NAD biosynthesis. Nothing is currently known about the cellular regulation of the intestinal nicotinic uptake process.

VITAMIN B6

Metabolic Role and Deficiency

Vitamin B6 consists of a group of structurally related micronutrients (pyridoxine, pyridoxal, and pyridoxamine) that exist in both unphosphorylated and phosphorylated forms. Pyridoxal phosphate is the most biologically active derivative and functions as a cofactor in several enzymatic reactions including those involved in amino acid metabolism. Vitamin B6 deficiency leads to a variety of clinical abnormalities including peripheral neuritis, depression, confusion, and dermatitis. Deficiency and/or suboptimal levels of vitamin B6 occur in alcoholics, patients with renal disease, diabetics, those on long-term use of certain medications (e.g., isoniazid and other hydrazines), and those with B6-responsive anemia.

Intestinal Vitamin B6 Absorption

Humans and other mammals cannot synthesize vitamin B6 and thus must obtain the vitamin via intestinal absorption. Dietary phosphorylated forms of vitamin B6 are hydrolyzed in the intestinal lumen prior to absorption. Absorption of the unphosphorylated vitamin B6 has been reported to occur via a nonsaturable simple diffusion process. Following absorption, the various components of vitamin B6 accumulate within the absorptive cell, primarily in the phosphorylated form. The phosphate esters then are dephosphorylated over time, with final transfer of the products to the circulation across the basolateral membrane. Although vitamin B6 status was found to be regulated homeostatically, this does not appear to be due to alterations in intestinal absorption of the vitamin.

VITAMIN B12

Vitamin B12 will be discussed only briefly here.

Metabolic Role and Deficiency

Vitamin B12 (also known as cobalamin, Cbl) is important for maintaining the normal differentiation, proliferation, and metabolic status of all cells. It acts as a coenzyme for two key enzymatic reactions: the conversion of homocysteine to methionine and the conversion of methylmalonyl coenzyme A (coA) to succinyl coA. Deficiency of Cbl leads to intracellular accumulation and eventual secretion into the plasma of the metabolites of these two reactions, homocysteine and methylmalonic acid. Elevated plasma levels of these two metabolites are a strong indication of an intracellular

deficiency of Cbl. Intracellular Cbl deficiency can arise due to multiple causes that include both acquired and inherited disorders. Strict vegetarians and vegans develop Cbl deficiency due to a lack of intake, and patients with tapeworm infestation or bacterial overgrowth develop Cbl deficiency due to competition for dietary Cbl. These patients can be treated successfully with adequate intake of Cbl when the underlying cause is eliminated with antibiotics. However, patients with gastric surgery (partial or complete) or surgery of the ileum (chronic inflammation) will develop Cbl deficiency as they will not be able to absorb Cbl due to decreased levels or total loss of intrinsic factor or the ileal receptor, respectively. In these patients, along with children who have inherited disorders involving intrinsic factor, its ileal receptor, or the plasma transporter transcobalamin II or a number of defects that involve its intracellular utilization, deficiency of Cbl is permanent and must be treated on a regular basis with intramuscular injections of pharmacological doses of Cbl. Another factor that interferes with intestinal Cbl absorption in humans is excessive ethanol consumption.

Intestinal Cobalamin Absorption

Humans and other mammals must obtain Cbl from the diet because they cannot synthesize the vitamin. After digestion, free Cbl binds to a hydrophobic protein ligand, the gastric intrinsic factor (IF), prior to absorption. Cbl bound to the IF is then endocytosed via a distinct cell surface receptor, the intrinsic factor–cobalamin receptor (also called cubilin), located at the brush border membrane of distal ileum epithelia. In humans, IF is localized mainly to the parietal cells, but it is also detected at the margins of the anatomical regions in clusters of chief cells, enteroendocrine cells, and endothelial cells. IF is a glycoprotein with one Cbl-binding site. It is believed that the carbohydrate moiety of IF affords protection from proteolytic degradation. This is not too surprising considering the ability of IF to remain structurally and functionally intact even after its exposure to proteases of the stomach and intestinal lumen. Following endocytosis of the IF–Cbl complex, the IF portion is degraded. This is followed by intracellular formation of a transcobalamin II–Cbl complex and release into circulation across the basolateral membrane.

VITAMIN C

Metabolic Role and Deficiency

Vitamin C (ascorbic acid) is required by all cells; it functions as a powerful antioxidant, assists in maintaining metal ions in their reduced forms, and plays a role in a variety of hydroxylation reactions. Vitamin C deficiency leads to scurvy, a disease that is primarily due to an abnormality in collagen formation that leads to, among other things, bleeding from the gums and joint, muscle, and dystrophic hair deformities. The incidence of vitamin C deficiency or scurvy is rare in developed countries and usually occurs in certain populations at risk, such as individuals with alcoholism and the elderly.

Intestinal Vitamin C Absorption

Primates and guinea pigs are the only mammals that cannot synthesize ascorbic acid from glucose endogenously and thus have a dietary requirement for the vitamin and obtain it via intestinal absorption. Dietary vitamin C occurs in two forms: the reduced form (i.e., ascorbic acid, AA) and the oxidized form (i.e., dehydroascorbic acid, DHAA). Intestinal absorption of AA has been the subject of extensive investigations using a variety of human and animal intestinal preparations. Uptake was found to occur without significant metabolic alterations in the transported substrate and to involve a specialized, Na^+-dependent, carrier-mediated system that is localized at the apical membrane of the absorptive enterocytes. Exit of AA from the enterocyte across the intestinal basolateral membrane occurs via a Na^+-independent carrier-mediated system.

As to DHAA, the amount of this form of the vitamin in the human diet is not fully documented but it is likely to increase during prolonged storage of food. The intestinal absorptive cells are capable of taking up DHAA and metabolizing it to the reduced ascorbic acid. The mechanism of uptake of DHAA across the intestinal brush border membrane appears to be via a Na^+-independent transport process. This process is believed to involve the mammalian facilitative hexose transporters and appears to represent a physiologically significant pathway for uptake and accumulation of vitamin C by many different cells.

Molecular Identity of Ascorbic Acid Transport System(s)

Two Na^+-dependent L-ascorbic acid transporters, the so-called sodium-dependent vitamin C transporters 1 and 2 (SVCT1 and SVCT2, respectively), have been cloned and shown to be expressed in intestinal absorptive cells. The level of expression of SVCT1 is markedly higher than that of SVCT2, and SVCT 1 appears to play a more important role in the overall intestinal AA absorption process. Both transporters have higher selectivity

for L-ascorbic acid than for D-isoascorbic acid and dehydroascorbic acid; also, they both transport L-ascorbic acid via an electrogenic Na^+-dependent process (stoichiometric ratio of 2 : 1 for Na^+ to ascorbic acid). SVCT1 was found to share 65% identity with SMVT2 at the amino acid level, but no identity was found between these two transporters and other mammalian membrane transporters.

Regulation of Intestinal AA Uptake

Animal studies have shown that the process of intestinal AA uptake is down-regulated following oversupplementation of the vitamin. No information, however, exists as to the molecular mechanism(s) involved in this adaptive regulation.

BIOTIN

Metabolic Role and Deficiency

Biotin acts as a coenzyme for four carboxylases that are essential for gluconeogenesis, the metabolism of several branched-chain amino acids, and fatty acid synthesis. Deficiency of biotin leads to serious clinical abnormalities, which include neurological disorders, growth retardation, and dermal abnormalities. Also, at least in animals, biotin deficiency during pregnancy leads to embryonic growth retardation, congenital malformation, and death. The incidence of biotin deficiency and suboptimal conditions has been reported with increased frequency in recent years. Biotin deficiency has been reported in patients with inborn errors of biotin metabolism, in patients on long-term therapy with anticonvulsant agents, and in patients on long-term parenteral nutrition. Suboptimal biotin levels have also been reported during pregnancy, in substantial numbers of alcoholics, in patients with inflammatory bowel diseases, and in infants with seborrheic dermatitis and Leiner's disease.

Intestinal Biotin Absorption

Mammalian cells cannot synthesize biotin and thus must obtain the vitamin from exogenous sources via absorption in the intestine. The intestine is exposed to two sources of biotin: a dietary source and a bacterial source (the latter being via synthesis of the vitamin by the normal microflora of the large intestine). Dietary biotin exists in free and protein-bound forms. The latter form is digested first by gastrointestinal proteases and peptidases to generate biocytin (*N*-biotinyl-L-lysine) and then is converted to free biotin by the action of the enzyme biotinidase. The mechanism of absorption of free biotin in the small intestine has been studied using a variety of human and animal intestinal preparations. These studies have shown the involvement of a specialized, Na^+-dependent carrier-mediated mechanism for biotin uptake across the intestinal brush border membrane. At the basolateral membrane, biotin was found to exit the cell via a Na^+-independent carrier system. In adult humans, biotin uptake is significantly higher in the duodenum and jejunum than in the ileum. Recently, it has been recognized that the intestinal biotin transport system also transports the structurally and functionally unrelated water-soluble vitamin pantothenic acid and the metabolically important substrate lipoate. Thus, a new name was given to the transport system: the sodium-dependent multivitamin transporter (SMVT). The nutritional consequences of competition between these vitamins for a common transporter are not currently known. As to transport of the bacterially synthesized biotin in the large intestine, an efficient Na^+-dependent, specialized, carrier-mediated system (similar to that of the small intestine) has been identified in human colonocytes. Among the factors that have been shown to negatively interfere with the normal intestinal biotin transport process are the long-term use of certain anti-epileptic medications (Primidone and Tegretol) and alcohol intake.

Molecular Identity of the Intestinal Biotin Transport System(s)

The molecular identity of the intestinal biotin transport system, SMVT, was delineated following cloning of its cDNA from a number of species including humans and rats. At least in rats, significant heterogeneity was found in the 5′-untranslated region of the SMVT cDNA, with four distinct variants (I, II, III, and IV) being identified. Variant II was found to be the predominant form expressed in the small and large intestine. Also, villus cells were found to express significantly more SMVT message than crypt cells. Furthermore, the 5′-regulatory regions of the human and rat SMVT genes have been cloned and characterized and multiple promoters that drive transcription of these genes have been identified.

Regulation of Intestinal Biotin Uptake

The intestinal absorption of biotin was found to be regulated both by extracellular substrate level and by specific intracellular protein kinase-mediated pathways. Biotin deficiency was found to lead to a specific and significant up-regulation in biotin intestinal uptake. On the other hand, oversupplementation with pharmacological doses of biotin was found to lead to a specific and significant down-regulation of

the intestinal biotin uptake. These effects were found to be mediated mainly via changes in the number of the biotin transport carriers. Also, an intracellular protein kinase C and a Ca^{2+}/calmodulin-mediated pathway appear to play a role in regulating the intestinal biotin uptake process. Furthermore, the intestinal biotin uptake process was shown to undergo ontogenic regulation during early postpartum development.

FOLIC ACID

Metabolic Role and Deficiency

The coenzyme derivatives of the vitamin folic acid (folate) are necessary for the synthesis of purine and pyrimidine precursors of nucleic acids, for the metabolism of certain amino acids, and for the initiation of protein synthesis in mitochondria. Folate deficiency leads to a variety of clinical abnormalities including megaloblastic anemia and growth retardation. Recent studies have also suggested that low folate levels may play a role in the etiology of coronary artery disease (via folate involvement in the metabolism of homocysteine) and colonic, cervical, and lung neoplasia. In contrast to the negative consequences of folate deficiency and suboptimal levels, optimization of folate body homeostasis has been shown to be effective in preventing certain diseases (e.g., prevention of neural-tube defects) and has also been reported to decrease the risk of developing certain types of cancer, most notably in the colon. Folate deficiency is a highly prevalent vitamin deficiency throughout the world. In the Western Hemisphere, a significant number of folate deficiency cases result from impairment in intestinal absorption of the vitamin. This impairment is common in a variety of intestinal diseases (e.g., IBD) and has also been associated with the use of certain pharmacological agents (e.g., sulfasalazine, phenytoin). Furthermore, impairment in intestinal folate absorption due to congenital defects has also been reported.

Intestinal Folate Absorption

Humans and other mammals cannot synthesize folate and thus must obtain the vitamin via intestinal absorption. The intestine is exposed to folate from two sources: a dietary source, where the vitamin is absorbed in the small intestine, and a bacterial source, where the vitamin is synthesized by normal microflora in the large intestine. The main form of dietary folate is the folate polyglutamate form. This form must be converted to the folate monoglutamate form prior to absorption, a process achieved via the special action of the intestinal folate hydrolase. The absorption of the released folate monoglutamates has been the subject of extensive investigations. These investigations have established that (1) the proximal part of the small intestine is the main site of absorption of dietary folate with significantly less uptake in the ileum; (2) a specialized, carrier-mediated system is involved in folate uptake across the apical membrane of the absorptive cells that transports the vitamin by an electroneutral process; (3) the uptake system is highly dependent on an acidic extracellular pH with significantly less uptake at neutral and alkaline pH; and (4) the uptake system transports reduced (e.g., 5-methyltetrahydrofolate, 5-MTHF), oxidized (e.g., folic acid), and substituted (e.g., the folate analogue methotrexate, MTX) folate derivatives with similar affinities. The last two features of the intestinal folate uptake process are unique to the intestine and are different from the widely characterized folate uptake process of nonpolarized cells, such as that of the mouse leukemia L1210 cells. The intestinal epithelial cells are also unique compared to other absorbing (transporting) epithelial cells such as those of the kidney and the placenta, in that they do not utilize the other folate uptake mechanism via the folate receptor (also known as the membrane folate-binding protein) under normal conditions. The factors that negatively affect intestinal folate absorption are alcohol consumption and the use of certain pharmacological agents, such as sulfasalazine and phenytoin.

As to the bacterially synthesized folate in the large intestine, studies using apical membrane vesicles isolated from the colon of organ donors and cultured human colonic epithelial NCM 460 cells have shown the uptake to be via a specific, carrier-mediated, pH-dependent, uptake mechanism that is sensitive to the effect of the anion exchange inhibitor DIDS. The system appears to be similar (or identical) to that in the small intestine.

Molecular Identity of the Intestinal Folate Transport System

The molecular identity of the intestinal folate uptake system has also been delineated by cloning its cDNA from human and mouse intestine. The open reading frames of the human and mouse intestinal cDNA clones were found to be very similar (or identical in the case of the mouse) to the open reading frame of the cDNA of the reduced folate carrier cloned from nonpolarized cells of the respective species. This is despite the fact that the characteristics of the folate uptake process in these distinct cell types are different. Subsequent studies have shed some light onto how the reduced folate carrier (RFC) exhibits such different characteristics in distinct cells and suggested that the cell environment may play a role in causing the difference.

Regulation of Intestinal Folate Uptake

The intestinal folate absorption process appears to be under the regulation of extracellular and intracellular mechanisms. Dietary folate deficiency has been shown to lead to a significant and specific up-regulation in folate uptake. This up-regulation appears to involve transcriptional regulatory mechanisms. Intracellular protein tyrosine kinase- and protein kinase A-mediated pathways have also been reported to play a role in regulating the intestinal folate uptake process. Furthermore, the intestinal folate uptake process was found to undergo ontogenic regulation during the early stages of life.

PANTOTHENIC ACID

Pantothenic acid is needed for the synthesis of coenzyme A and acyl carrier protein in mammalian cells and thus is important in the metabolism of carbohydrate, fat, and, to a lesser extent, protein. Pantothenic acid deficiency in humans is rare, but its induction leads to depression and fatigue. As with most other water-soluble vitamins, mammals lack the ability to synthesize pantothenic acid and thus must obtain the vitamin from exogenous sources via intestinal absorption. The intestine is exposed to two sources of pantothenic acid: dietary and bacterial. Dietary pantothenic acid exists mainly in the form of coA, which is hydrolyzed to free pantothenic acid in the intestinal lumen prior to absorption. The mechanism of absorption of pantothenic acid in the small intestine involves the same carrier-mediated Na^+-dependent system that transports biotin (i.e., SMVT; see Biotin above). Similarly, colonic absorption of pantothenic acid was found to involve the same biotin Na^+-dependent, carrier-mediated system. It is worth mentioning that the interaction between biotin and pantothenic acid transport has also been seen in other tissues and organs, such as the blood–brain barrier, the heart, and the placenta. However, the nutritional consequences of such interactions have not been investigated.

SUMMARY

The water-soluble vitamins are essential for normal human health and well-being. Deficiency of these micronutrients leads to a variety of clinical abnormalities, some of which have been recognized since ancient times. For humans, water-soluble vitamins are obtained from exogenous sources via intestinal absorption. Studies using a variety of human and animal intestinal preparations including intact tissues, cellular preparations, and membrane vesicles have established that for most of these micronutrients, efficient and specialized carrier-mediated transport systems exist. Some of these systems are driven by Na^+ gradients, others are driven by H^+ and/or OH^- gradients, and yet others are not affected by an ion gradient. In addition, the molecular identity of some of these transport systems has been delineated by cloning and their promoters have been characterized. Furthermore, it is now known that a variety of intestinal diseases and conditions negatively interfere with the normal absorption process of these nutrients and thus warrant closer attention by the treating physician to avoid deficiency and prevent suboptimal levels from occurring.

Acknowledgments

This work was supported by grants from the Department of Veterans Affairs and the National Institutes of Health (DK56061 and DK58057).

See Also the Following Articles

Cobalamin Deficiency • Dietary Reference Intakes (DRI): Concepts and Implementation • Folate Deficiency • Malnutrition • Small Intestine, Absorption and Secretion • Vitamin B12: Absorption, Metabolism, and Deficiency

Further Reading

Cooperman, J. M., and Lopez, R. (1984). Riboflavin. *In* "Handbook of Vitamins: Nutritional, Biochemical and Clinical Aspects" (L. J. Machlin, ed.), pp. 299–327. Dekkes, New York.

Said, H. M., Rose, R., and Seetharam, B. (2000). Intestinal absorption of water-soluble vitamins: Cellular and molecular aspects. *In* "Current Topics in Membranes: Gastrointestinal Transport Molecular Physiology" (K. E. Barrett and M. Donowitz, ed.), pp. 35–76. Academic Press, San Diego, CA.

Seetharam, B. (1994). Gastric intrinsic factor and cobalamin absorption. *In* "Physiology of the Gastrointestinal Tract" (L. Johnson, ed.), chap. 62, pp. 1997–2026. Raven Press, New York.

Wrong, O. M., Edmonds, C. J., and Chadwich, V. S. (1981). Vitamins. *In* "The Large Intestine: Its Role in Mammalian Nutrition and Homeostasis," pp. 157–166. Wiley, New York.

Webs

STEPHEN A. OLJESKI AND ROY MCCAULEY
Tufts–New England Medical Center

antrum Distal portion of the stomach between the body and the pylorus.
atresia Nondevelopment of an expected structure.
double-bubble sign Radiographic appearance signifying duodenal obstruction. Consists of large air collections in the stomach and in the duodenum, without visualization of gas more distally in the gastrointestinal tract.
duodenum The first portion of the small intestine, located between the stomach and the jejunum.
stenosis Narrowing.
web A thin sheet of tissue or large mucosal fold protruding into the lumen of a hollow organ, often causing obstruction.
windsock A type of partially obstructing duodenal web that balloons into the more distal duodenum like an internal diverticulum.

Gastrointestinal webs are an uncommon form of congenital or acquired stenosis found in a variety of locations within the gastrointestinal tract. The webs discussed herein include esophageal, antral, and duodenal webs. They may be congenital or acquired and they present in both adults and children.

INTRODUCTION

As outlined above, a gastrointestinal (GI) web is an uncommon finding and can be identified in a number of locations throughout the alimentary tract. This article is concerned with three of the more common GI webs: esophageal, antral, and duodenal. The common clinical presentations, diagnoses, imaging findings, and current therapies will be outlined briefly, though it is important to note that theories of pathogenesis and treatment continue to evolve. Pertinent radiographs will also be included and links with other disorders will be examined.

ESOPHAGEAL WEBS

Esophageal webs (often equated with rings) may be found as either congenital or acquired lesions. Many can be linked with specific disease states, whereas others are found as isolated lesions. The presentation, diagnosis, and therapy depend on both the putative mechanism and the age of the patient. In general, webs rarely protrude more than 5 mm into the lumen, are eccentrically located, and average 2–4 mm in thickness (Fig. 1). Most are asymptomatic. They can be found anywhere along the esophagus, though most are anterior and postcricoid. Webs in the upper esophagus are covered with squamous mucosa, whereas those in the lower esophagus are often covered with columnar gastric mucosa. All are believed to contain mucosa and submucosa with scanty muscular fibers.

Congenital Esophageal Webs

Congenital esophageal webs are relatively uncommon and can be located at any level. If the web is complete and totally excludes the lumen, it will present as a form of esophageal atresia. These are thought to be secondary to failure of recanalization of the primitive

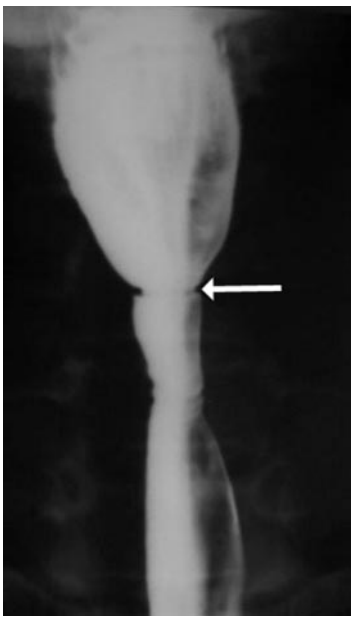


FIGURE 1 Arrow indicates the location of the esophageal web.

 Copyright 2004, Elsevier (USA). All rights reserved.

foregut. Single webs are more common than multiple webs. The prevalence is unknown, but is estimated to be 1 in 25,000–50,000 live births, with females more commonly affected than males. There is no ethnic predilection. Although isolated webs have been reported, they are more likely to occur in conjunction with tracheoesophageal fistulas, Down's syndrome, VACTERAL malformations, and prematurity. Webs severely compromising the lumen are often discovered at birth by inability to pass a suction catheter. Others may not present until later in childhood or adulthood (usually before the age of 40), when inability to feed, vomiting, and aspiration are noted. A web that severely narrows the lumen is not compatible with life and early endoscopic dilation or transthoracic surgical removal is necessary following stabilization of the patient's clinical condition. Diagnosis is most often made by endoscopy. Upper GI studies can also be diagnostic, though there is a real risk of aspiration of the contrast medium.

A special case of congenital webs is the Schatzki ring. Also composed of both mucosa and submucosa, it is typically located at the gastroesophageal junction and is believed to mark the proximal margin of a hiatal hernia. Although some believe that the ring offers protection against gastroesophageal reflux, this is not generally agreed upon. Meat impaction and progressive dysphagia are common, though many adults are asymptomatic. In older patients, gastroesophageal reflux disease (GERD) should be ruled out.

Acquired Esophageal Webs

Acquired webs, found in both children and adults, are secondary to a variety of causes. Most are postcricoid. The true incidence of acquired webs is unknown, as 95% are asymptomatic. Webs are found in approximately 6–14% of barium examinations of the upper gastrointestinal tract and diagnosis can be made by both upper GI barium study and endoscopy. The common clinical complaint is intermittent dysphagia for solid foods, alternating with long periods of normal swallowing. Dysphagia is most likely if the web narrows the lumen to less than 13 mm in diameter. Acquired esophageal webs can be found associated with chronic GERD, Plummer-Vinson syndrome, hiatal hernias, circatrious pemphigoid, chronic infection, celiac sprue, connective tissue disorders, or ingestion of corrosive agents. Treatment of the underlying condition, such as correction of the anemia in Plummer-Vinson syndrome or anti-reflux medications in GERD, will often correct the web. If this is not the case, endoscopic dilation or surgical disruption can be performed. Progressive dysphagia or increasing pain should prompt a search for underlying malignancy.

ANTRAL WEBS

Congenital gastric outlet obstruction resulting from antral webs is quite uncommon and represents less than 1% of stenoses of the alimentary tract. The cause of the defect is unknown, but failure of recanalization of the primitive foregut is again implicated. A small number of nonobstructing webs can be seen in association with conditions such as hypertrophic pyloric stenosis. Sex distribution is approximately equal. Most cases of complete antral webs present on the first day of life with nonbilious vomiting, feeding difficulties, and abdominal distension with no distal gas. Gastric perforation is a real danger. However, the degree of obstruction dictates the severity of symptoms. Older children and adults with less severe narrowing may present with nausea, vomiting, abdominal pain, and weight loss. Diagnosis of an antral web (Fig. 2) is suggested by the findings of a thin persistent septum near the pyloric channel on barium examination. Disruption of peristalsis to the level of the antrum is also suggestive of the diagnosis. Endoscopy can confirm the diagnosis. Treatment of this rare lesion entails correction of metabolic disturbances, as well as nasogastric tube placement. Surgical or endoscopic repair, consisting of simple excision or pyloroplasty, can then be undertaken.

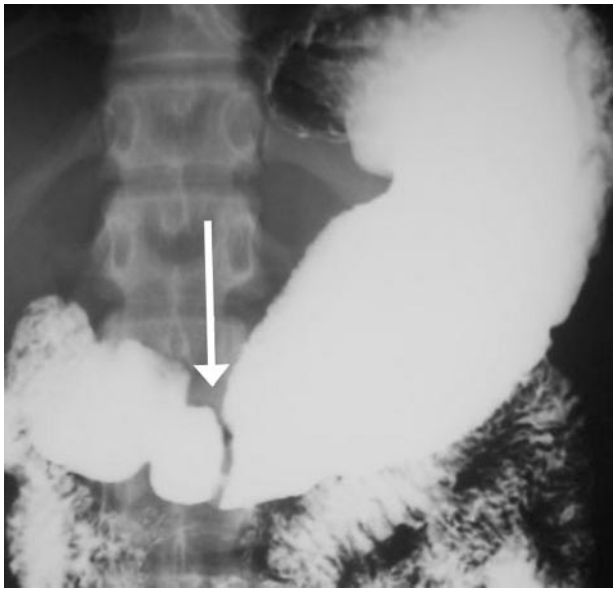


FIGURE 2 Circumferential narrowing of the gastric antrum by a web (arrow).

DUODENAL WEBS

As is the case with the esophageal webs discussed above, duodenal webs may form as a result of primary failure of recanalization of the primitive foregut in the fourth to fifth week of development, though this is debated. Nearly all duodenal webs are congenital in origin and nearly all are discovered within a few days of birth, although diagnosis may be delayed until adulthood if the degree of obstruction allows the passage of food. Single webs are more common than multiple webs and are most often found in the first and second portions of the duodenum. Complete duodenal webs are the most common forms of duodenal atresia, which occurs in approximately 1 in 1000 births. The webs take the form of an intact membrane seen obstructing the lumen. Incomplete webs may present as the "windsock deformity," in which an incomplete and distensible membrane spanning the lumen stretches and balloons downstream. It is often associated with abnormalities of the biliary system.

As is alluded to above, the diagnosis of duodenal webs is usually made on the first day of life. Plain radiographs will demonstrate the "double-bubble sign" (Fig. 3) of air in the distended stomach and proximal duodenum without visualization of more distal gas. Contrast studies are rarely indicated and may be precluded secondary to the risk of aspiration of the contrast medium. Annular pancreas and malrotation of the bowel with Ladd's bands must be considered in the differential diagnosis. Increasingly, prenatal ultrasound can suggest the diagnosis.

A number of associated conditions are seen in patients with duodenal webs, including polyhydramnios (5–20%), Down's syndrome (20–30%), VACTERAL anomalies, jaundice, esophageal atresia, and congenital heart disease. The initial treatment of the affected infant includes nasogastric tube decompression and correction of electrolyte balance. Associated abnormalities such as those noted above should be excluded before repair, as 33% of these infants have life-threatening abnormalities. Repair by duodenodudenostomy or duodenoplasty can then be undertaken.

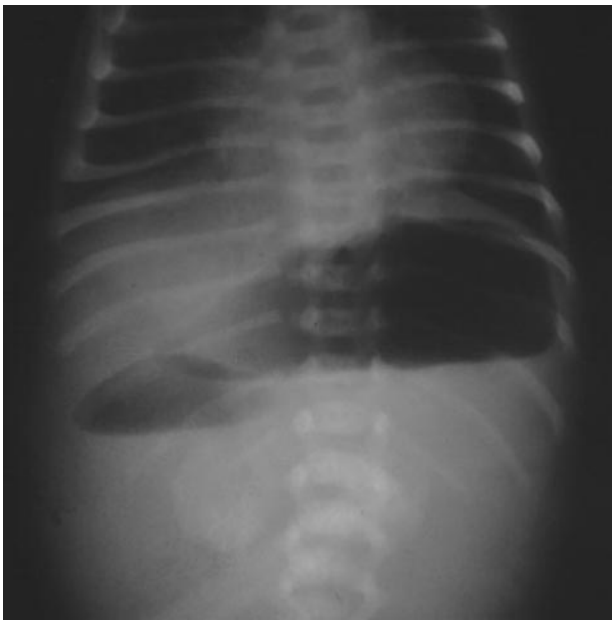


FIGURE 3 The "double-bubble" sign indicative of duodenal obstruction.

See Also the Following Articles

Esophagus, Development • Gastroesophageal Reflux Disease (GERD) and Congenital Esophageal Obstructive Lesions, Pediatric • Neonatal Tracheoesophageal Anomalies • Pyloric Stenosis

Further Reading

Anderson, L. S., *et al.* (1973). Cartilaginous esophagus: A cause of esophageal stenosis in infants and children. *Radiology* **108**, 66.

Buckley, K., *et al.* (1998). Schatzki's ring in children and young adults: Clinical and radiologic findings. *Pediatr. Radiol* **28**, 884–886.

Chen, Y. M., *et al.* (1987). Natural progression of the lower esophageal ring. *Gastrointest. Radiol.* **12**, 93–98.

Ekberg, O., and Nylander, G. (1983). Webs and web-like formations in the pharynx and cervical esophagus. *Diag. Imaging* **52**, 10–18.

Hirano, I., *et al.* (2000). Clinical and manometric features of the lower esophageal ring. *Am. J. Gastroenterol.* **95**, 43–49.

Johnson, M. C. (2001). The esophagus. *Primary Care* **28**, 1–20.

Ladd, A. P., and Madura, J. A. (2001). Congenital duodenal anomalies in the adult. *Arch. Surg.* **136**, 576–584.

Wong, P. (2002). Esophageal webs and rings. Available at http://www.emedicine.com.

Whipple's Disease

RENE MAHNEL AND THOMAS MARTH
Deutsche Klinik für Diagnostik, Wiesbaden, Germany

Tropheryma whipplei The bacterium causing Whipple's disease.

Whipple's disease A rare chronic disease. The main symptoms are weight loss, arthralgia, diarrhea, and abdominal pain. The current medical treatment still depends on empiric antibiotic regimens with variable rates of success.

Whipple's disease was first described in 1907 as intestinal lipodystrophy. It is a rare infectious disease caused by the actinomycete *Tropheryma whipplei*. Whipple's disease occurs mostly in middle-aged caucasian men. As of 1987, approximately 700 cases have been reported in the literature.

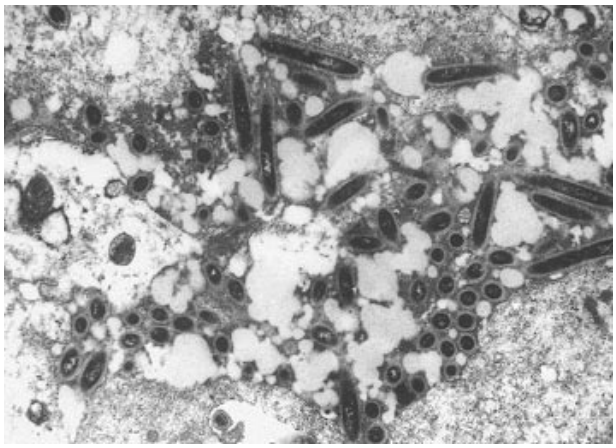


FIGURE 1 Electron microscopic view of *T. whipplei* (×20,000), showing abundant rod-shaped bacteria with an approximate size of 0.5−1.5 × 2.5 μm.

ETIOLOGY, CLINICAL FEATURES, AND DIAGNOSIS

T. whipplei is probably a ubiquitously occurring bacterial agent as its presence in soil, in waste water, and in the saliva of healthy persons has been ascertained by polymerase chain reaction (PCR). It is an actinomycete that was phylogenetically characterized in 1991 and now has been cultured and formally named. The bacterium has a typical morphology with a trilamellar cell wall as observed by electron microscopy (Fig. 1). *T. whipplei* can be found intra- and extracellularly and infested tissue pathognomonically reveals a strong positivity for polysaccharides, mucoproteins, and glucoproteins in the periodic acid Schiff (PAS) reaction. *T. whipplei* is poorly gram-stained in tissue but is gram-negative in cell culture. Histologically, the frequently affected small intestine often appears thickened and edematous with a marked infiltration by large macrophages and lipid deposits, which are the result of villous lymphatic blockade.

There is evidence that host factors contribute to the clinical manifestations of Whipple's disease. Cellular immune function, i.e., macrophages and T lymphocytes, seem to be primarily affected. In addition to a reduced proliferation rate of T lymphocytes after stimulation, a reduced CD4/CD8 ratio can be found. Other alterations are an increased memory T-cell subpopulation with a simultaneous decrease in naive T cells. A diminished expression of CD11b and a significantly reduced production of T_H1 cytokines (interleukin-12 and interferon-γ) after stimulation of monocytes *in vitro* are also seen. The pathogenetic significance of the latter immune defects is underscored by the fact that they are not only present in florid disease but persist even in patients with long-term remission.

The main clinical features are weight loss due to abdominal manifestations, fever, or inappetence, as well as nondestructive, migratory, seronegative arthropathies often preceding other symptoms by many years. Abdominal symptoms such as diarrhea and abdominal pain may be accompanied by the full spectrum of a malabsorption syndrome. Therefore, Whipple's disease was considered mainly a gastrointestinal disease for a long time. However, as the bacterium spreads systemically, numerous other organs and systems, such as the heart, lungs, and central nervous system, may be affected (Table I).

The diagnosis is usually made by duodenal biopsy. Diagnosis may also be established, for example, from lymph node biopsies, endocardial tissue, synovia, and

 Copyright 2004, Elsevier (USA). All rights reserved.

TABLE I Clinical Features of Whipple's Disease

	Approximate incidence (%)
Major clinical features	
Weight loss	95
Arthropathy	85
Diarrhea	75
Frequent signs and symptoms	
Systemic	
Fever	50
Lymphadenopathy	50
Hyperpigmentation	45
Gastrointestinal	
Abdominal pain	65
Occult bleeding	25
Abdominal mass	20
Splenomegaly	15
Hepatomegaly	15
Ascites	10
Other	
Hypotension	40
Peripheral edema	35
Cardiac murmurs	35
Myalgia	25
Chronic cough	20
Other less frequent clinical signs	
Pleuritis	
Pleural effusion	
Endocarditis	
Muscle wasting	
Glossitis	
Peripheral neuropathy	
Eye involvement (e.g., visual loss, uveitis, retinitis)	
CNS involvement (e.g., dementia, ophthalmoplegia, myoclonus, ataxia, nystagmus)	

cerebrospinal fluid (CSF) and these tests should also be confirmed by duodenal biopsy. The presence of bacterial DNA can be verified using PCR. Bacterial DNA often can be found in the CSF at the time of diagnosis.

THERAPY

Due to the small number of cases, antibiotic treatment is based on empiric observations and there is still no evidence-based therapy regimen. In the initial therapy, ceftriaxone, which has good CNS penetration, is frequently used at a dose of 2 g/day intravenously for 2 weeks. This should be followed by a long-term oral antibiotic regimen. Most frequently, trimethoprim/sulfamethoxazole (800 mg/160 mg), which has a good penetration of the blood–brain barrier, is used for a >1-year period. In many cases, this therapy is able to eliminate *T. whipplei*, which should be checked histologically after 6 and 12 months. However, in cases with refractory or relapsing disease, or in the case of a CNS manifestation, the prognosis may be poor. Up to 40% of all patients may have a clinical relapse with a higher incidence of neurological symptoms.

See Also the Following Articles

Diarrhea • Malabsorption

Further Reading

Dobbins, W. O., III (1987). "Whipple's Disease." Charles C. Thomas, Springfield, IL.

Dutly, F., and Altwegg, M. (2001). Whipple's disease and "*Tropheryma whipplei*." *Clin. Microsc. Rev.* **14**, 561–583.

Fenollar, F., and Raoult, D. (2001). Molecular techniques in Whipple's disease. *Expert Rev. Mol. Diagn.* **1**, 299–309.

Marth, T. (1999). Whipple's disease. *In* "Principles and Practice of Infectious Disease" (G. L. Mandell, J. E. Bennett, and R. Dolin, eds.), 5th Ed., pp. 1170–1174. Churchill Livingstone, Philadelphia, PA.

Marth, T., and Raoult, D. (2003). Whipple's disease. *Lancet* **36**, 239–246.

Wilson's Disease

MICHAEL L. SCHILSKY
Recanati/Miller Transplantation Institute, Mount Sinai Medical Center, New York

ATP7B Human copper-transporting ATPase that is the product of the Wilson's disease gene (*ATP7B*). The gene is expressed mainly in hepatocytes; the encoded protein maintains copper homeostasis by exporting copper from the hepatocyte into bile.

ceruloplasmin Copper-containing serum glycoprotein produced mainly in the liver; binds six copper atoms per molecule of protein.

Kayser–Fleischer ring Corneal deposits of copper present in Descemet's membrane; visible on slit lamp examination, and, when more pronounced, by direct visualization.

In Wilson's disease, copper accumulates to toxic levels in the liver, brain, and other sites in the body. When discovered in a timely fashion, many of the toxic effects of copper accumulation can be prevented or reversed by medical therapy, which must be continued for life. Failure to initiate treatment, or interruption of treatment, can lead to liver injury and liver failure and death, or to neurological and psychiatric signs and symptoms. In some patients with fulminant liver failure due to Wilson's disease, or in patients with advanced liver disease that fails to respond to medical treatment, liver transplantation can be lifesaving and curative.

INTRODUCTION

Wilson's disease is present in all populations, with an incidence of ~1/30,000 individuals. The gene for Wilson's disease, *ATP7B*, encodes a copper-transporting ATPase that resides mainly within hepatocytes.

PATHOPHSYIOLOGY

Wilson's disease (WD) is an autosomal recessive disorder in which copper accumulates to toxic levels in the liver and subsequently in the neurologic system and other tissues. The gene for WD, *ATP7B*, is located on chromosome 13 and encodes a copper-transporting ATPase expressed mainly in hepatocytes. Individuals homozygous for a single *ATP7B* mutation, or more commonly with two different mutations of this gene (compound heterozygotes), have reduced or failed copper transport activity of the ATP7B protein. Loss of this physiologic copper transport function reduces biliary copper excretion; this is responsible for hepatic copper accumulation and leads to a reduced copper incorporation into ceruloplasmin and thereby to decreased levels of the ATP7B protein in the circulation of most patients, due to the instability of the protein without copper. Tissue accumulation of copper results in cellular injury by oxidative damage to membranes, organelles such as mitochondria, and DNA, and to altered protein synthesis, among other effects. Liver injury manifests as mild steatosis and inflammation early on, and to chronic injury, fibrosis, cirrhosis, and liver failure. Excess copper appears in the circulation following hepatic copper accumulation and eventually toxic accumulation in the central nervous system with resultant neurologic or psychiatric manifestations. Other organs that may be affected by toxic copper accumulation include the kidneys, heart, and bone.

DIAGNOSIS

The clinical presentation of Wilson's disease varies widely, but mainly presents as hepatic or neurologic disease. Liver disease typically presents within the first two decades of life, and later on, typically in the third decade, neurologic or psychiatric signs and symptoms may be evident. Patients with liver disease may be asymptomatic or they may experience symptoms of chronic liver disease such as fatigue or jaundice, or may manifest signs or symptoms of portal hypertension such as ascites and varices without or with bleeding. Some patients may have chronic hepatitis with features indistinguishable from autoimmune hepatitis. In about 5% of cases, the sudden onset of jaundice or ascites with associated hemolysis heralds the onset of acute fulminant hepatic failure.

Neurologic manifestations of WD are mainly due to injury to the basal ganglia, but many different parts of the brain may be affected. Symptoms may include diminished motor coordination, tremor with deteriorating handwriting and micrographia, drooling, dysarthria, dystonia, and spasticity. Transfer dysphagia with

 Copyright 2004, Elsevier (USA). All rights reserved.

the risk of aspiration may occur in some patients. Psychiatric ailments range from behavioral changes and anxiety disorders to depression or even psychosis.

WD must be considered in any patient below the age of 45 years with unexplained liver disease or cirrhosis, and in patients with neurologic and psychiatric symptoms and evidence of liver disease. Individuals with hepatic histology suspicious for WD, pediatric patients with autoimmune features not responsive to steroids and lacking typical markers, patients with brain imaging demonstrating compatible findings, and patients in whom Kayser–Fleischer rings are identified should be evaluated for WD.

The diagnosis of WD is established by a combination of clinical signs and biochemical studies or by molecular methods (see Table I). Clinical findings that include Kayser–Fleischer (KF) rings and corneal deposits of copper in Descemet's membrane are pathognomonic and are present in almost all patients with WD with neuropsychiatric manifestations; these symptoms are best detected by slit lamp examination. However, KF rings may be present in only ~50% of patients presenting with liver disease, and may rarely be found in patients with chronic cholestasis such as primary biliary cirrhosis or primary sclerosing cholangitis. Sunflower cataracts are another rare ocular finding associated with WD, and lunulae cerulae (blue lunules) are another dermatologic abnormality present very rarely in patients with WD. Other nondiagnostic findings include evidence of stigmata of chronic liver disease and cirrhosis and Parkinsonian features on neurological examination.

Reduced levels of serum ceruloplasmin (typically <20 mg/dl) are present in ~95% of patients with WD, but also in 20% of heterozygous carriers without manifestations of WD. Ceruloplasmin levels in the circulation may also be reduced in patients with severe hepatic insufficiency, in protein-losing enteropathy or

TABLE I Diagnostic Testing for Wilson's Disease

Test	Results
Ophthalmologic slit lamp examination	Kayser–Fleischer rings and sunflower cataracts are absent in normal individuals; Kayser–Fleischer rings absent early on in patients with WD; present in only 50% of patients with hepatic presentation; found in 98% of patients with neurologic signs or symptoms; sunflower cataracts may also be seen on slit lamp exam in patients with WD; Kayser–Fleischer rings and sunflower cataracts abate with treatment for WD
Serum ceruloplasmin	Normal 20–40 mg/dl, elevated with acute phase and in pregnancy and with use of estrogens; less than 20 mg/dl in 95% of patients and 20% of heterozygous carriers; physiologically decreased in newborns, undetectable in rare patients with aceruloplasminemia; below 20 mg/dl with severe hepatic insufficiency and in severe protein-loss states
Serum copper	Normally ~100 μg/dl, decreased in most patients, typically <80 μg/dl; proportion not bound to ceruloplasmin >10% in patients and total serum copper and markedly elevated above 200 μg/dl in fulminant hepatitis due to Wilson's disease
24-Hour urinary copper	Normal <50 μg/24 h, greater than 100 μg/24 h in most symptomatic patients and following chelation treatment; postpenicillamine (500 mg given twice, 12 h apart) >1600 μg in untreated patients
Hepatic copper concentration	Normal <40 μg/g dry wt liver; greater than 250 μg/g dry wt liver in patients; may be increased in other cholestatic disorders, idiopathic copper toxicosis
Hepatic histology	Abnormal findings in patients include steatosis, glycogen nuclei, fibrosis, chronic active hepatitis and cirrhosis; marked degeneration of hepatocytes, pleiocytosis, and nuclear irregularities in fulminant hepatitis
Electron microscopy for hepatic ultrastructure	Abnormal in untreated patients; dilatation of mitochondrial cristae and crystalline deposits (present when steatosis is seen on light microscopy); dense lysosomes later on
Histochemistry for copper	Absent normally; in some patients, positive staining is present in some but not all liver nodules; absent staining does not exclude Wilson's disease
Radiological imaging of the brain: MR or CT imaging	Normal in the absence of Wilson's disease and in many patients early on; findings in patients with neurologic or psychiatric symptoms include atrophy and alterations in basal ganglia, subcortical white matter, midbrain, and pons; abnormalities can be present in some asymptomatic patients
Molecular diagnostic studies: haplotype analysis and mutation analysis	Haplotype: same as proband in patients, different in carriers and unaffected siblings; mutations: patients have disease-specific mutations on each allele; most useful in populations with dominant mutations

nephropathy, and in the rare genetic disorders Menke's disease and aceruloplasminemia. Serum copper not bound to ceruloplasmin is increased in untreated patients, but total serum copper is typically reduced because ~90% is bound to ceruloplasmin, which is reduced in the vast majority of patients. Urinary copper excretion is increased to >100 μg for 24 hours in most untreated patients with symptomatic disease. Low serum uric acid levels are found in many patients, and aminoaciduria and electrolyte changes due to renal tubular acidosis may be present.

Liver biopsy still plays an important role in the diagnosis of WD. Copper quantitation from the biopsy specimen is a critical test and reveals a level above 250 μg/g dry weight in almost all untreated patients. Characteristic histologic features of the earlier stages of the disorder include steatosis, microvesicular and macrovesicular fatty changes, with glycogenated nuclei in some hepatocytes. Varying degrees of fibrosis and inflammatory infiltrate can be found. Cirrhosis is present in most patients by the time they present with hepatic insufficiency or with neurological disease. Histochemical staining of liver biopsy specimens for copper typically reveals positive areas in some nodules, with absence in others. The absence of histochemically identifiable copper does not exclude WD. Ultrastructural analysis of liver specimens at the time steatosis is present reveals specific mitochondrial abnormalities that include dilated cristae and crystalline inclusions. For patients presenting with neurological or psychiatric disease, common abnormalities present on brain imaging that should suggest a diagnosis of WD include increased density of the basal ganglia on computer tomography (CT) or hyperintensity of this region on T2 magnetic resonance (MR) imaging. Abnormal findings are not limited to the basal ganglia, and other abnormalities have been described, including regions of the pons, white matter, and cerebellum.

The diagnosis of WD is established by the presence of KF rings and a decreased level of serum ceruloplasmin, of KF rings and neurologic or psychiatric symptoms, and of elevated hepatic copper levels with appropriate histology. Haplotype or polymorphism analysis can be used to identify affected siblings or offspring; however, the diagnosis must be firmly identified in the proband. Direct identification of *ATP7B* mutations is possible but is limited by the size of the gene and the large numbers of mutations associated with the WD phenotype, and the limited availability of testing.

Patients with fulminant hepatic failure due to WD have markedly elevated serum copper concentration and urinary copper excretion, and most have a lowered serum ceruloplasmin. Kayser–Fleischer rings may be present in about 50% of patients. These patients invariably have cirrhosis underlying the acute liver injury, which likely accounts for the relatively low levels of transaminase elevations despite the massive hepatic injury. Unique to acute fulminant liver failure due to WD is a relatively low alkaline phosphatase level, with the ratio of alkaline phosphatase (units/liter) to bilirubin (milligrams/deciliter) of ≤2.0. Hepatic copper is almost always elevated above 250 μg/g dry weight in these individuals despite the massive hepatic necrosis and presence of cirrhosis.

TREATMENT

Treatment options for WD include lifelong medical therapy with oral chelating agents or zinc (see Table II), and liver transplantation. Dietary restrictions of copper intake are recommended along with medical therapy. The chelating agents penicillamine and trientine promote renal copper excretion and are recommended as first-line therapy for symptomatic patients with hepatic or neurologic disease, and, at lower dosages, these agents may be used for maintenance therapy. Zinc acts to block copper absorption from the gut at the level of the enterocyte, and is mainly used for maintenance therapy or initial therapy of asymptomatic patients. Liver transplantation restores a normal phenotype with respect to copper metabolism, and no further therapy specific for Wilson's disease is required.

Initial treatment with penicillamine consists of dosages that are increased to 1–2 g/day over a few weeks, and 750–1000 mg/day is used for maintenance therapy. Some patients with WD with neurologic symptoms initially treated with this drug have worsening neurologic disease that may become irreversible. Early side effects include hypersensitivity reactions, and nephrosis or a lupuslike syndrome with associated hematuria, proteinuria, and positive antinuclear antibodies will occur later on. Late reactions include dermatologic side effects such as progeric skin changes, elastosis perforans serpingosa, pemphigous or pemphigoid lesions, lichen planus, and apthous stomatitis, and rarely systemic anaphylaxis, myasthenia gravis, polymyositis, dysgeusia and loss of taste, IgA depression, and serous retinitis. An adverse drug reaction to penicillamine should prompt its discontinuation and a change to an alternative treatment. Monitoring of therapy includes clinical examination and biochemical determination of non-ceruloplasmin-bound copper (derived from subtracting the amount of ceruloplasmin-bound copper from the total serum copper),

TABLE II Treatment Options for Wilson's Disease

Medication	Mode of action	Dosages (total/24 h)	Adverse effects
d-Penicillamine (Cuprimine)	Chelating agent, induces cupriuresis	750–2000 mg in two to three divided doses apart from meals; maintenance 750–1000 mg; requires supplemental pyridoxine	Worsening with initial treatment of neurlogic symptoms in 10–50% of cases; initial hypersensitivity reactions; marrow suppression; lupuslike syndrome; nephrosis; dermatologic toxicity; rare Goodpastures syndrome
Trientine (Syprine)	Chelating agent, induces cupriuresis	750–1500 mg in two to three divided doses apart from meals; maintenance, 750–1000 mg	Lupuslike syndrome; nephrosis; marrow suppression
Zinc (Galzin)	Blocks intestinal uptake of copper by enterocytes	100–150 mg in two to three divided doses apart from meals	Gastric irritation

blood counts, liver functions testing, and urinalysis to look for the appearance of cellular elements or protein. Urinary copper excretion over 24 hours is usually greater than 1 mg/day during initial treatment and afterward is typically reduced to between 250 and 500 μg/day, but is dosage dependent.

Trientine, developed for treating penicillamine-intolerant patients, may be used as first-line therapy when administered in two to four divided dosages totaling up to 1500 mg/day, and as maintenance therapy with a reduced dosage of 750–1000 mg/day. Trientine has a better safety profile compared to penicillamine, with only extremely rare reports of hypersensitivity reactions. Known toxicity includes bone marrow suppression, sideroblastic anemia (the result of an induced copper deficiency), and lupuslike reactions. Monitoring of therapy is the same as for penicillamine.

Zinc, used mainly as a maintenance therapy and for the initial therapy of asymptomatic patients, is administered in three daily dosages of 50 mg (of elemental zinc) for adults and 25 mg for children. This dosage effectively achieves a negative copper balance in most individuals. The most important side effect of zinc affecting a minority of patients is gastric upset, and this may be related to the particular zinc salt used. Monitoring of zinc therapy includes determination of non-ceruloplasmin-bound copper and 24-hour urinary copper excretion, the latter being below 100 μg/day in effectively treated patients.

Tetrathiomolybdate is a potent copper chelator that remains currently under investigation for the initial treatment of neurologic WD in the United States. Reversible side effects of tetrathiomolybdate therapy include anemia and neutropenia. Further recommendations regarding this drug await the completion of ongoing clinical trials.

Patients with fulminant hepatic failure due to WD and those in whom medical therapy is ineffective or is interrupted by resulting liver failure require liver transplantation for survival. Following the onset of fulminant liver failure in WD, there is a large release of copper into the circulation that contributes renal tubular injury and reduced kidney function. Plasma exchange, plasmapheresis, and other forms of dialysis, including albumin dialysis, are useful to help rapidly lower copper levels in the circulation and to limit hemolysis and further renal damage, but have not eliminated the need for transplantation of these individuals.

Treatment for WD must continue throughout pregnancy. There are recorded successful pregnancies with administration of penicillamine, trientine, and zinc. No dosage reduction is needed for zinc therapy, whereas the smallest effective dosage of penicillamine or trientine should be administered during pregnancy and until wound healing occurs. Similarly for patients undergoing elective surgical procedures, dosages of chelating agents should be minimized until wound healing is achieved.

PROGNOSIS

Medical treatment of asymptomatic patients prevents the development of liver or neurological disease. The long-term survival of patients with WD with medical therapy is excellent, even when chronic liver disease and cirrhosis are present at the outset. Patients with cirrhosis still may manifest signs or symptoms due to portal hypertension, including esophageal or gastric varices or ascites. In some patients, neurologic symptoms may

improve with therapy, whereas other patients may worsen during the initial phase of treatment or may develop neurological disease no longer responsive to therapy aimed at WD.

Liver transplant also offers excellent survival, approaching 90% for patients with acute liver failure due to WD, well above the survival of $\sim$60% for all others transplanted for fulminant liver failure. Patients with WD surviving beyond 1 year posttransplant typically have excellent long-term survival.

See Also the Following Articles

Ascites • Cirrhosis • Liver Biopsy • Liver Disease, Pregnancy and • Liver Transplantation • Portal Hypertension and Esophageal Varices • Trace Minerals: Metabolism and Deficiency (Zinc, Copper, Selenium, Manganese)

Further Reading

Brewer, G. J., Dick, R. D., Johnson, V. D., Brunberg, J. A., Kluin, K. J., and Fink, J. K. (1998). Treatment of Wilson's disease with zinc. XV. Long-term follow-up studies. *J. Lab. Clin. Med.* **132**, 264–278.

Brewer, G. J., Dick, R. D., Johnson, V. D., Fink, J. K., Kluin, K. J., and Daniels, S. (2001). Treatment of Wilson's disease with zinc. XVI. Treatment during the pediatric years. *J. Lab. Clin. Med.* **137**, 191–198.

Brewer, G. J., Johnson, V. D., Dick, R. D., Hedera, P., Fink, J. K., and Kluin, K. J. (2000). Treatment of Wilson's disease with zinc. XVII. Treatment during pregnancy. *Hepatology* **31**, 364–370.

Emre, S., Atillasoy, E. O., Ozdemir, S., Schilsky, M. L., Rathna Varma, C. V. R., Thung, S. N., *et al.* (2001). Orthotopic liver transplantation for Wilson's disease: A single center experience. *Liver Transpl.* **72**, 1232–1236.

Martins da Costa, C., Baldwin, D., Portmann, B., Lolin, Y., Mowat, A. P., and Mieli-Vergani, G. (1992). Value of urinary copper excretion after penicillamine challenge in the diagnosis of Wilson's disease. *Hepatology* **15**, 609–615.

Scheinberg, I. H., and Sternlieb, I. (1984). "Wilson's Disease." W. B. Saunders Co., Philadelphia.

Schilsky, M. L., Scheinberg, I. H., and Sternlieb, I. (1994). Hepatic transplantation for Wilson's disease: Indication and outcome. *Hepatology* **19**, 583–587.

Steindl, P., Ferenci, P., Dienes, H. P., Grimm, G., Pabinger, I., Madl, C., *et al.* (1997). Wilson's disease in patients presenting with liver disease: A diagnostic challenge. *Gastroenterology* **113**, 212–218.

Sternlieb, I. (2000). Wilson's disease and pregnancy. *Hepatology* **31**, 531–532.

Xerostomia

JONATHAN A. SHIP
New York University

dysgeusia Difficulty in tasting.
dysphagia Difficulty in swallowing.
halitosis Subjective complaint of bad breath.
salivary hypofunction Objective diminution of salivary output.
xerostomia Subjective complaint of a dry mouth.

Xerostomia is the subjective complaint of a dry mouth due to salivary hypofunction. Saliva provides host protection, assists in the initiation of food and fluid intake, and enables communication through speech. Without adequate salivary output, oral and pharyngeal health declines as does a person's quality of life. Xerostomia becomes more common with aging and is caused by numerous intraoral and systemic factors.

ROLE OF SALIVA IN ORAL AND SYSTEMIC HEALTH

The adequate secretion of saliva is critical for preserving the health of the oral cavity and gastrointestinal system. It provides a physical barrier against irritants, lubricates oral mucosal tissues, aids in the digestion of fats and carbohydrates, and assists with mastication, deglutition, taste, and retention of dentures. Saliva maintains a stable pH in the mouth and esophagus and helps prevent tooth decay and remineralize the dental hard tissues. Saliva contains antibacterial, antiviral, and antifungal factors that help maintain the oral commensal flora.

ORAL AND SYSTEMIC EFFECTS OF SALIVARY HYPOFUNCTION

Salivary function is age-stable in healthy adults, yet xerostomia is more common in the elderly due to oral and systemic factors that cause salivary hypofunction. Inadequate output causes persistent complaints of xerostomia and halitosis and leads to multiple oral problems: new and recurrent dental caries, mucositis, pain, and increased susceptibility to microbial infections. The most prevalent infection is caused by *Candida albicans*, observed as angular cheilitis of the lips, erythematous candidiasis (denture stomatitis), and pseudomembraneous candidiasis. Salivary hypofunction also causes problems with chewing and wearing dentures, dysphagia, and dysgeusia, resulting in altered food and fluid selection, compromising nutrition.

INTRAORAL SOURCES OF SALIVARY PATHOLOGY

Bacterial infections (e.g., parotitis) occur secondary to salivary hypofunction. Colonization and infection (e.g., *Staphylococcus aureus*, *Staphylococcus pyogenes*, *Streptococcus pneumonia*, *Escherichia coli*) occur following obstruction of a duct, causing swelling, purulence, and pain. Viral infections preferentially involve parotid glands: paramyxovirus (mumps) and cytomegalovirus infections affect young and immunocompromised adults. Obstructions of excretory ducts are caused by trauma in the lips (mucocele of a minor gland) and floor of the mouth (ranula of the sublingual or submandibular gland) or by calcification of mucous plugs in major glands (sialoliths, stones).

Most salivary tumors are benign, the most common being pleomorphic adenoma of the parotid gland, which presents as an asymptomatic mass in the tail of the gland. Malignant tumor incidence increases with age; they are most prevalent in the palatal minor glands and the submandibular and sublingual glands. Mucoepidermoid carcinoma is the most common malignant salivary gland tumor, followed by adenoid cystic carcinoma, acinic cell carcinoma, and adenocarcinoma. Presentation includes a swelling with nerve paralysis, pain, or facial paresis.

EXOGENOUS AND SYSTEMIC SOURCES OF SALIVARY PATHOLOGY

Medications are the most likely cause of salivary hypofunction, particularly those with anticholinergic effects (e.g., antidepressants, sedatives, tranquilizers, antihistamines, cytotoxic agents, antihypertensives, anti-Parkinson's disease drugs, and anti-spasmodic/anti-seizure drugs). Chemotherapy can cause acute

 Copyright 2004, Elsevier (USA). All rights reserved.

yet reversible changes in salivary function and radioactive iodine (^{131}I) causes parotid but not submandibular dysfunction. Head and neck radiotherapy, used for cancer, is the prime exogenous source of salivary hypofunction. Radiotherapy-induced xerostomia is usually irreversible and permanent, causing a lifetime of oral–pharyngeal disorders.

Sjögren's syndrome, an autoimmune exocrinopathy, is the most common disorder causing salivary hypofunction. It is characterized by progressive and irreversible xerostomia and xerophthalmia (dry eyes). Other autoimmune diseases with salivary hypofunction include rheumatoid arthritis, scleroderma, and systemic lupus erythematosus. Human immunodeficiency virus infection, diabetes, Alzheimer's disease, Parkinson's disease, strokes, and cystic fibrosis are also associated with salivary hypofunction.

TREATMENT OF SALIVARY HYPOFUNCTION

Appropriate diagnosis is required before initiation of therapy. To prevent the development of xerostomia-induced oral sequelae, frequent dental evaluations are required, as is oral hygiene after each meal and before bedtime with fluoridated dentifrices, gels, and rinses to avoid dental caries. Regular hydration with water, not sugared beverages, is helpful. Sugar-free chewing gum, candies, and mints can stimulate remaining salivary secretions, whereas artificial saliva and lubricants may assist when there is no remaining salivary tissue. Counseling regarding a well-balanced nutritionally adequate diet should be provided. Dysphagia can be improved with oral moisturizers, lubricants, and careful use of fluids during eating. Candidiasis responds to topical antifungal agents (e.g., clotrimazole, nystatin) and appropriate denture hygiene, and parotitis is treated with broad-spectrum antibiotics (e.g., Augmentin).

Salivary lesions and tumors must undergo histopathological evaluation and imaging (sialograms, radiographs, radioactive isotype scintiscans, magnetic resonance imaging, and computed tomography scans). Most tumors are treated with excision and possibly radiotherapy. Drug-induced xerostomia requires assessment of all medications, with attempts to eliminate, diminish, or substitute medications causing hypofunction. Secretagogues are available (pilocarpine 5 mg tid and qhs; cevimeline 30 mg tid) to increase secretions and diminish xerostomia in patients with remaining exocrine tissue.

See Also the Following Articles

Candidiasis • Dysphagia • Halitosis • Salivary Glands, Physiology • Sjögrens Syndrome

Further Reading

Atkinson, J. C., and Wu, A. (1994). Salivary gland dysfunction: Causes, symptoms, treatment. *J. Am. Dent. Assoc.* **125**, 409–416.

Fox, P. C. (1997). Management of dry mouth. *Dent. Clin. N. Am.* **41**, 863–876.

Fox, R. I., Stern, M., and Michelson, P. (2000). Update in Sjögren syndrome. *Curr. Opin. Rheumatol.* **12**, 391–398.

Narhi, T. O., Meurman, J. H., and Ainamo, A. (1999). Xerostomia and hyposalivation: Causes, consequences and treatment in the elderly. *Drugs Aging* **15**, 103–116.

Silverman, S., Jr. (1999). Oral cancer: Complications of therapy. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endo.* **88**, 122–126.

Sreebny, L. M., and Schwartz, S. S. (1997). A reference guide to drugs and dry mouth. 2nd Ed. *Gerodontology* **14**, 33–47.

Yersinia

EDWARD J. BOTTONE
Mt. Sinai School of Medicine, New York

adenitis Glandular inflammation.
bacteriophage typing Bacterial typing that differentiates between bacteria or strains of bacteria by their susceptibility to one or more bacteriophages.
erythema nodosum Disorder characterized by tender red bumps, usually found on the front of the lower leg.
ileitis Inflammation of the ileum.
Reiter's syndrome Chronic form of inflammatory arthritis characterized by the inflammation of the joints, the eyes, and the genital, urinary, or gastrointestinal system.

The genus *Yersinia* is composed of 10 species of gram-negative bacilli, of which three, *Yersinia pestis* (plague bacillus), *Yersinia enterocolitica*, and *Yersinia pseudotuberculosis*, are well-recognized human pathogens. Whereas isolates of *Y. pestis* and *Y. pseudotuberculosis* are inherently pathogenic, virulence among *Y. enterocolitica* isolates differs by biogroup and serogroup.

MICROBIOLOGY

Many of the identifying characteristics and virulence factors of *Yersinia enterocolitica* are influenced by growth temperature. For instance, *Y. enterocolitica* is nonmotile at 37°C but is motile at 25°C; plasmid-encoded virulence factors are expressed at 37°C and are repressed or absent at 25°C. Therefore, full characterization of an isolate as *Y. enterocolitica* should include tests determined at two incubation temperatures. *Yersinia enterocolitica* grows on routine bacteriologic media used for the isolation of other enteric bacteria. On these media, however, because *Y. enterocolitica* produces only pinpoint colonies after 24 hours of incubation, colonies may be overlooked in stool cultures containing a multiplicity of bacterial species. The use of cefsulodin–irgasan–novobiocin (CIN) agar enhances recovery; *Y. enterocolitica* produces red colonies on this medium.

Culture of stools of patients with acute gastroenteritis due to pathogenic *Y. enterocolitica* serogroups (O3, O5,27, O8, O9) produces numerous colonies on routine isolation media. However, in those instances when *Y. enterocolitica* is sought during convalescence from acute enteritis, or for surveillance during an outbreak, cold (4°C) enrichment of stool specimens in phosphate-buffered saline for up to 4 weeks with periodic subculture may enhance recovery. *Yersinia enterocolitica* isolates recovered after prolonged cold enrichment should be biogrouped and serogrouped before clinical significance is ascribed to the isolate. In many instances, such isolates have an environmental origin (biogroup 1A) and lack virulence attributes.

Yersinia enterocolitica biogrouping schemas delineate six biogroups that correlate with ecologic distribution, serogroup designation, and human pathogenic potential. In the United States, biogroup 4, serogroup O3 is the most frequently isolated bio/serogroup, exceeding serogroup O8. Using antisomatic O antisera, *Y. enterocolitica* isolates may be serologically grouped into approximately 60 serogroups. Bacteriophage typing of an isolate also correlates biogroups and serogroups to secondary autoimmune sequelae and epidemiologic distribution.

EPIDEMIOLOGY

Yersinia enterocolitica is widely distributed in terrestrial and freshwater ecosystems and has been recovered from the intestinal tract of numerous mammalian species. *Yersinia enterocolitica* infections have been documented almost globally. Whereas a cooler seasonal prevalence has been recognized in European countries, such a correlation has not been found in the United States or other countries. The frequency of infection seems to correlate with the porcine reservoir of pathogenic serogroups (O3, O5,27, O8, O9) and consumption of undercooked or raw pork products, or their preparation.

Although most sporadic infections with *Y. enterocolitica* cannot be traced to a specific exposure, ingestion of contaminated milk or other foods, contact with sick animals or perhaps index cases, and transfusion of contaminated blood are also routes of transmission and acquisition.

 Copyright 2004, Elsevier (USA). All rights reserved.

PATHOGENESIS

Gastrointestinal disease is the most common clinical manifestation of *Y. enterocolitica* infection. Once ingested, *Y. enterocolitica*, aided by its motility, traverses the intestinal lumen and binds to and penetrates M cells, to invade underlying Peyer's patches of the distal ileum; this results in the formation of microabscesses, ulceration of the overlying epithelium, and an inflammatory reaction. Spread to the mesenteric lymph nodes may lead to abscesses in the medullary region and pain in the lower quadrant, mimicking appendicitis. Extensive ulceration of the intestinal tract and death have occurred in the course of *Y. enterocolitica* serogroup O8 infection. Acute enteritis with inflammatory cells and occasionally bloody, watery stools characterize infection in children. Concomitant *Y. enterocolitica* bacteremia may be present in infants with enteritis. In young adults, acute terminal ileitis and mesenteric adenitis are the more common presentations.

Yersinia enterocolitica extraintestinal disease depends on host immune status and the pathogenic potential of the infecting strain. A striking characteristic of *Y. enterocolitica* bacteremia is the increased incidence in people with iron overload diseases or in those being administered iron chelating agents. Secondary manifestations of bacteremia may involve every organ of the body, including the endovasculature and the central nervous system, and may even lead to cutaneous lesions.

Bacteremia with *Y. enterocolitica* may also occur by transfusion of contaminated blood. After the first report in 1975, 36 cases of *Y. enterocolitica* transfusion associated-bacteremia were documented up to 1996. These cases were predominantly associated with transfusion of packed red blood cells, including autologous transfusions. Shock, with and without disseminated intravascular coagulation, occurred in all 36 recipients, and death occurred in 19 patients. The source of *Y. enterocolitica* contamination was asymptomatic bacteremia in the donor.

Secondary immunologically mediated nonsuppurative sequelae of *Y. enterocolitica* infection include polyarticular arthritis and erythema nodosum (most common) and Reiter's syndrome, reported mainly among northern Europeans. Most patients manifesting postyersinial reactive arthritis are HLA-B27 positive.

VIRULENCE FACTORS

Human pathogenic strains of *Y. enterocolitica* possess plasmid- and chromosomal-encoded virulence determinants. Those specified by a 70- to 75-kb plasmid are expressed mainly at 37°C and confer resistance to phagocytosis and intracellular killing by macrophages. Chromosomally encoded determinants include an outer membrane protein named invasin, maximally produced at 25°C, which binds to host cells through β1 integrins, aiding initial colonization. Adaptation to 37°C triggers the plasmid-encoded virulence attributes and another chromosomally encoded outer membrane attachment/invasion locus (Ail) protein. Presence of the gene for Ail also allows *Yersinia* to resist serum bactericidal activity. *Yersinia enterocolitica* produces a heat-stable enterotoxin but its role in diarrheal disease is disputed because it is maximally produced below 30°C in late log-phase broth cultures.

TREATMENT

Yersinia enterocolitica is susceptible *in vitro* to trimethoprim-sulfamethoxazole (TMP-SMX), third-generation cephalosporins, aminoglycosides, imipenem, aztreonam, and quinolones. There is a trend of increasing resistance to TMP-SMX and chloramphenicol in Spain. Because the mean duration of symptoms in children is 14 days, controversy exists over the need to treat uncomplicated *Y. enterocolitica* enteritis. Extraintestinal infection, such as bacteremia in infants with enteritis, in immunocompromised hosts, or in people with iron overload merits antiyersinial therapy.

See Also the Following Articles

Foodborne Diseases • Food Poisoning • Food Safety • *Salmonella*

Further Reading

Bliska, J. B., and Falkow, S. (1992). Bacterial resistance to complement killing mediated by the Ail protein of *Yersinia enterocolitica*. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3561–3565.

Bottone, E. J. (1997). *Yersinia enterocolitica*: The charisma continues. *Clin. Microbiol. Rev.* **10**, 257–276.

Bottone, E. J. (1977). *Yersinia enterocolitica*: A panoramic view of a charismatic microorganism. *Crit. Rev. Microbiol.* **5**, 211–241.

Cover, T. L., and Aber, R. C. (1989). *Yersinia enterocolitica*. *N. Engl. J. Med.* **321**, 16–24.

Lee, L. A., Gerber, A. R., Lonsway, D. R., *et al.* (1990). *Yersinia enterocolitica* infections in infants and children associated with the household preparation of chitterlings. *N. Engl. J. Med.* **322**, 984–987.

Tauxe, R. V., Vandepitte, J., Wauters, G., *et al.* (1987). *Yersinia enterocolitica* infections and pork: The missing link. *Lancet* **1**, 1129–1132.

Van Noyen, R., Vandepitte, J., and Wauters, G. (1980). Nonvalue of cold enrichment of stools for isolation of *Yersinia enterocolitica* serotype 3 and 9 from patients. *J. Clin. Microbiol.* **11**, 127–131.

Zenker's Diverticulum

AMINE HILA AND DONALD O. CASTELL
Medical University of South Carolina, Charleston

Zenker's diverticulum Outpouching in the posterior pharyngeal wall immediately above the upper esophageal sphincter.

A diverticulum is a sac that protrudes from the wall of the gastrointestinal tract. A true diverticulum contains all layers of the gastrointestinal wall. Most diverticula are false diverticula, consisting of mucosa, submucosa, and a few muscle fibers.

INTRODUCTION

Zenker's diverticulum, also called pharyngoesophageal diverticulum, is described as an esophageal pouch. In fact, this type of diverticulum occurs in the pharynx, above the upper esophageal sphincter (UES), and should be considered a hypopharyngeal diverticulum. Pharyngoesophageal diverticulum was first described by Ludlow in 1769. In 1877, Zenker and von Ziemssen, in a compilation of five cases of their own together with 22 previously reported cases, clarified the then uncertain nature of a disease defined earlier by Ludlow as a "preternatural dilatation of, and bag formed in, the pharynx." Subsequently, this condition was given Zenker's name.

Esophageal diverticula are rare, occurring in less than 1% of upper gastrointestinal X rays and accounting for less than 5% of cases of dysphagia. The prevalence of Zenker's diverticulum has been reported to range between 0.01 and 0.11%. The prevalence is higher in the elderly, predominantly in women, in whom the occurrence of Zenker's diverticulum increases to about 50% after the age of 70. Zenker's diverticulum grows slowly but continuously. Even if it may be initially totally asymptomatic, with growth it can manifest at a later stage.

ANATOMY

The pharyngeal constrictor muscles form a funnel, and the mouth of the esophagus is like a transverse slit at the bottom of this funnel. At the esophageal inlet, the fibers of the cricopharyngeal muscle run transversely, thus forming the UES. Above the cricopharyngeal muscle, the walls of the hypopharynx contain the oblique fibers of the inferior constrictor muscles. Between the transverse fibers of the cricopharyngeal below and the oblique fibers of the inferior constrictors above, a triangular area contains fewer muscle fibers and constitutes a region of relative weakness; this is called the triangle of Killian, or Killian's dehiscence. The mucosa of the hypopharynx is allowed to bulge posteriorly at Killian's triangle, and with time, a pouch may develop, forming a Zenker's diverticulum.

ETIOLOGY AND PATHOPHYSIOLOGY

There is general agreement on the location of Zenker's diverticulum, but several pathogenic mechanisms have been offered, yet none of them has been definitely proved. Age has been considered an important factor. A decrease in tissue elasticity leading to an increased weakness of the triangle of Killian may explain why Zenker's diverticulum is rarely seen before 40 years of age.

The most widely accepted mechanism for the development of a Zenker's diverticulum is a functional disturbance during swallowing in the pressure relationships between the hypopharynx and the UES. Popular hypotheses have included increased resting pressure of the sphincter, lack of complete UES relaxation, and incoordination between hypopharyngeal contraction and sphincter relaxation. At present, poor UES compliance due to a restrictive myopathy rather than cricopharyngeal incoordination appears to account for the genesis of Zenker's diverticulum. The consistent finding in recent studies is increased intrabolus pressure preceding the pharyngeal contraction during swallowing.

DIAGNOSIS

Symptoms of Zenker's diverticulum depend on the stage of the disease. The frequency and severity of symptoms increase as the disease advances. Clinically, Zenker's diverticulum presents, in order of frequency, with

 Copyright 2004, Elsevier (USA). All rights reserved.

oropharyngeal dysphagia (in up to 98% of cases), regurgitation of undigested food, aspiration, noisy deglutition, halitosis, and changes in the voice. A sensation of sticking in the throat or of vague irritation can be present in the early stages. With time, patients often learn special maneuvers to empty the pouch by pressing on the neck or coughing and clearing the throat. Rarely does the pouch become so large as to obstruct the esophagus.

The most potentially dangerous complications are aspiration, pneumonia (occurring in about 30% of cases), and perforation. Massive bleeding in the diverticulum has been reported. Obstruction and fistula between the diverticulum and the trachea are rare. Carcinoma is an unlikely complication, arising in about 0.4% of cases.

Definitive diagnosis of Zenker's diverticulum is usually accomplished by contrast radiographic examination of the pharynx. Barium swallow with particular attention to the oropharyngeal area is the most helpful procedure, particularly when videoradiography is used. Small diverticula are less likely to be missed by careful evaluation of lateral and oblique views. The diverticulum is seen to protrude in the posterior midline, generally to the left cervical region (10% to the right), and barium tends to fill the pouch before progressing into the esophagus. In patients with symptomatic Zenker's diverticulum, the pouches usually are more than 2 cm in diameter. Figure 1 illustrates a typical Zenker's diverticulum seen on barium swallow.

Endoscopy is technically difficult and rarely indicated except when cancer is suspected following radiological studies. If endoscopy is indicated for other reasons, it is important to be cautious not to enter the diverticulum and risk perforation.

Manometric evaluation of patients with Zenker's diverticulum is not required from the clinical standpoint, although it is the only technique that will document the increased hypopharyngeal intrabolus pressure. It should be primarily reserved for clinical research. There is no conclusive evidence linking Zenker's diverticulum with gastroesophageal reflux disease. However, patients diagnosed with Zenker's diverticulum should be thoroughly evaluated if they have symptoms such as heartburn or chest pain. Esophageal manometry may confirm the presence of concomitant motility abnormalities contributing to dysphagia.

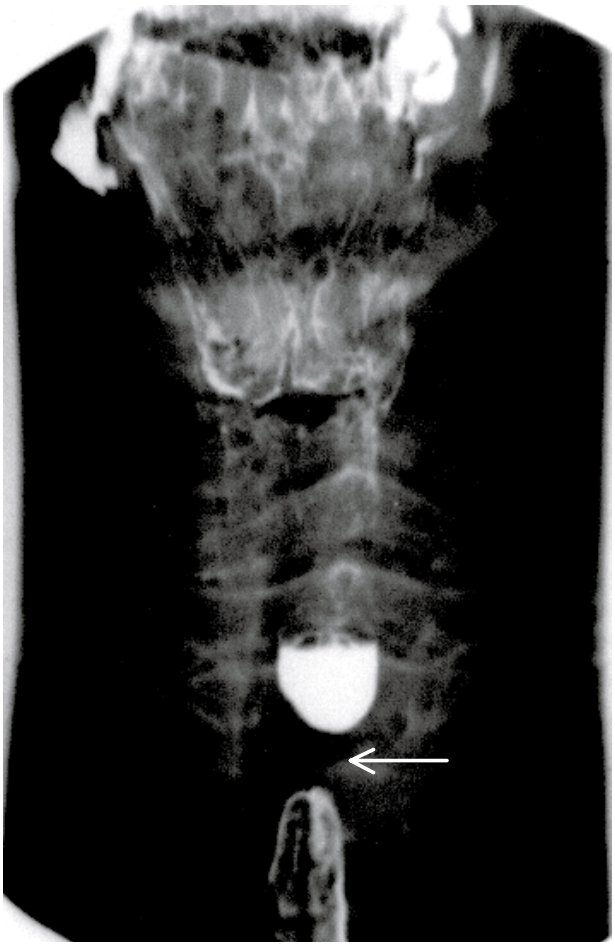


FIGURE 1 Typical Zenker's diverticulum as seen on an anteroposterior view during a barium swallow. The arrow indicates the location of the cricopharyngeal muscle.

TREATMENT

The therapeutic choice in patients with Zenker's diverticulum is dictated by symptoms. Small diverticula discovered by chance in asymptomatic patients do not require any intervention and may be followed by periodic barium swallows. When needed, the only effective treatment of Zenker's diverticulum is surgical. Surgical techniques include, separately or in combination, diverticuloplexy, diverticulectomy, and cricopharyngeal myotomy. Open surgical treatment of Zenker's diverticulum may result in complications such as fistulas, infection, vocal cord paralysis, and aspiration. These complications are quite rare. Endoscopic treatment is emerging as a viable alternative to open surgery. It includes different techniques with or without laser or stapling. There are no comparative studies that allow us to conclude which treatment technique to recommend. At this time, it seems that open surgical procedures are more in favor in the United States, but endoscopic techniques are gaining more advocates as they are being refined.

See Also the Following Articles

Aging • Dysphagia • Esophageal Surgery • Halitosis • Rumination Syndrome

Further Reading

Achkar, E. (1999). Esophageal diverticula. *In* "The Esophagus" (D. O. Castell, J. E. Richter, eds.), 3rd Ed. Lippincott Williams & Wilkins Philadelphia.

Fulp, S. R., and Castell, D. O. (1992). Manometric aspects of Zenker's diverticulum. *Hepato-Gastroenterol.* **39**, 123–126.

Mendez, L., Friedman, L. S., and Castell, D. O. (1991). Swallowing disorders in the elderly. *Clin. Geriat. Med.* **7**(2), 215–230.

Watemberg, S., Landau, O., and Avrahami, R. (1996). Zenker's diverticulum: Reappraisal. *Am. J. Gastroenterol.* **91**(8), 1494–1498.